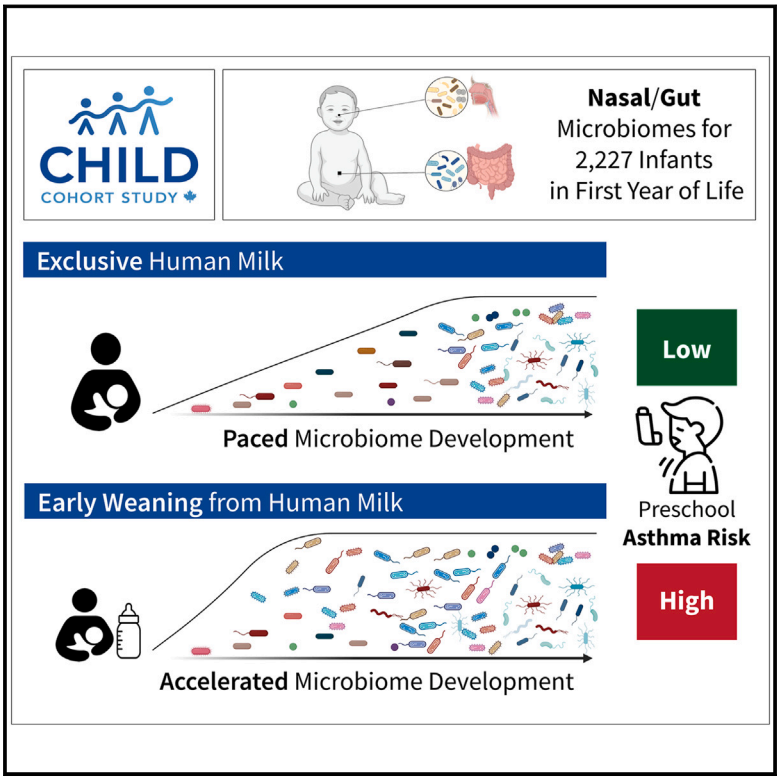
Cell

Microbial colonization programs are structured by breastfeeding and guide healthy respiratory development

Graphical abstract



Highlights

- Early breastfeeding cessation accelerates premature gain of microbial species/functions
- Breastfeeding over 3 months aids gradual microbiome maturation, protecting from asthma
- Microbiome colonization patterns and human milk composition accurately predict asthma
- Timely *R. gnavus* acquisition and tryptophan metabolism link human milk to asthma defense

Authors

Liat Shenhav, Kelsey Fehr, Myrtha E. Reyna, ..., Padmaja Subbarao, Michael G. Surette, Meghan B. Azad

Correspondence

liat.shenhav@nyulangone.org (L.S.), padmaja.subbarao@sickkids.ca (P.S.), surette@mcmaster.ca (M.G.S.), meghan.azad@umanitoba.ca (M.B.A.)

In brief

Human milk serves as a "pacemaker," indirectly protecting against asthma by regulating nasal and gut microbiome development during the first year of life.

Shenhav et al., 2024, Cell *187*, 5431–5452
September 19, 2024
https://doi.org/10.1016/j.cell.2024.07.022

CellPress

Cell

CellPress
OPEN ACCESS

Article

Microbial colonization programs are structured by breastfeeding and guide healthy respiratory development

Liat Shenhav,[1,2,3,19,20,21,*] Kelsey Fehr,[4,5,19] Myrtha E. Reyna,[6] Charisse Petersen,[7] Darlene L.Y. Dai,[7] Ruixue Dai,[6] Vanessa Breton,[6] Laura Rossi,[8] Marek Smieja,[8] Elinor Simons,[4,5] Michael A. Silverman,[9,10,11] Maayan Levy,[10,11] Lars Bode,[12,13,14] Catherine J. Field,[15] Jean S. Marshall,[16] Theo J. Moraes,[6] Piush J. Mandhane,[17] Stuart E. Turvey,[7] Padmaja Subbarao,[6,18,20,*] Michael G. Surette,[8,20,*] and Meghan B. Azad[4,5,20,*]

[1]Institute for Systems Genetics, New York Grossman School of Medicine, New York University, New York, NY, USA
[2]Department of Microbiology, New York Grossman School of Medicine, New York University, New York, NY, USA
[3]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA
[4]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, MB, Canada
[5]Manitoba Interdisciplinary Lactation Centre (MILC), Children's Hospital Research Institute of Manitoba, Winnipeg, MB, Canada
[6]Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada
[7]Department of Pediatrics, BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada
[8]Department of Medicine, McMaster University, Hamilton, ON, Canada
[9]Division of Infectious Disease, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
[10]Institute for Immunology and Immune Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[11]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[12]Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA
[13]Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence (MOMI CORE), University of California, San Diego, La Jolla, CA, USA
[14]Human Milk Institute (HMI), University of California, San Diego, La Jolla, CA, USA
[15]Department of Agriculture, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada
[16]Department of Department of Microbiology & Immunology, Dalhousie University, Halifax, NS, Canada
[17]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
[18]Department of Physiology, University of Toronto, Toronto, ON, Canada
[19]These authors contributed equally
[20]Senior author
[21]Lead contact
*Correspondence: liat.shenhav@nyulangone.org (L.S.), padmaja.subbarao@sickkids.ca (P.S.), surette@mcmaster.ca (M.G.S.), meghan.azad@umanitoba.ca (M.B.A.)
https://doi.org/10.1016/j.cell.2024.07.022

SUMMARY

Breastfeeding and microbial colonization during infancy occur within a critical time window for development, and both are thought to influence the risk of respiratory illness. However, the mechanisms underlying the protective effects of breastfeeding and the regulation of microbial colonization are poorly understood. Here, we profiled the nasal and gut microbiomes, breastfeeding characteristics, and maternal milk composition of 2,227 children from the CHILD Cohort Study. We identified robust colonization patterns that, together with milk components, predict preschool asthma and mediate the protective effects of breastfeeding. We found that early cessation of breastfeeding (before 3 months) leads to the premature acquisition of microbial species and functions, including *Ruminococcus gnavus* and tryptophan biosynthesis, which were previously linked to immune modulation and asthma. Conversely, longer exclusive breastfeeding supports a paced microbial development, protecting against asthma. These findings underscore the importance of extended breastfeeding for respiratory health and highlight potential microbial targets for intervention.

INTRODUCTION

Early-life gut and nasal microbiome and healthy development

The ecological process of microbial colonization during early life is crucial. Experiments in germ-free animals have shown that microbial colonization induces anatomical development, increases epithelial cell turnover rates, and kick-starts the maturation of the gut-associated lymphoid tissue.[1,2] Similarly, in the first months of life, infants rely on the acquisition of new microbial species and functions in order to complete immunological and physiological development. The establishment of the gut

microbiome adheres to consistent patterns,[3–11] with initial vertical transmission at birth.[12,13] Subsequently, the infant gut microbiome undergoes distinct shaping influenced by various physiological, dietary, and environmental factors.[14] Similarly, the dynamic respiratory tract microbiome undergoes marked changes in early life and is associated with various host and environmental factors, including birth mode, feeding type, antibiotic treatment, and crowding conditions.[15,16] However, the "optimal" timing and mechanisms regulating the gain of new microbial species and functions in these two important niches are poorly understood.

There is robust evidence associating early-life gut and respiratory tract microbiomes with respiratory illness.[17–27] For example, gut microbiome maturation in the first year of life contributes to the protective "farm effect" against childhood asthma,[27] and early-life gut microbial dysbiosis has been consistently identified as an asthma risk factor.[18] There is also evidence linking dysbiosis of microbiota residing in the mucosal surfaces, including the nasopharynx, with immune modulation and severe infections.[28] Moreover, nasal and airway microbiome composition and maturation patterns have been repeatedly linked with future respiratory health.[17–26,29,30] Nonetheless, it remains unclear if it is merely the presence or abundance of specific microbes that matters, or whether the *timing and order of their arrival* in different body sites may also play a role. Indeed, emerging data hint that when the order or timing of colonization is altered, a normally commensal microbe could become pathogenic.[30,31] We therefore hypothesize that microbial colonization patterns in early life are crucial to the establishment and function of microbial communities as well as healthy infant respiratory development (Figure 1A, axis 1).

Breastfeeding and weaning from human milk as primary drivers of microbiome development

It is well established that gut microbial composition in infancy is influenced by breastfeeding and human milk.[6,32–35] For instance, breastfeeding is associated with lower bacterial diversity, higher proportions of *Bifidobacterium*, and a gradual maturation of the gut microbiota.[6,32] Cross-sectional associations with the nasopharyngeal microbiota have also been identified,[36,37] including positive associations with commensals of the respiratory tract, such as *Dolosigranulum* and *Corynebacterium*, and negative associations with some potentially pathogenic microbiota, namely *Staphylococcus*, *Veillonella*, *Prevotella*, *Rothia*, and *Gemella*, at the first months of life.[36] Breastfeeding imparts a competitive advantage to strains capable of utilizing the complex sugars in human milk (known as human milk oligosaccharides [HMOs]), exemplified by *Bifidobacterium* and *Bacteroides*.[38] Additionally, weaning from breastfeeding and the introduction of solid food were associated with significant changes in the microbiota and immune system maturation.[32,35,39] Animal models further show that restricting microbiome maturation during weaning stunts immune system development and increases susceptibility to enteric infection.[40] However, the consequences of early weaning on microbiome maturation and respiratory health are not thoroughly understood. In general, we propose that a more comprehensive understanding of microbial colonization as a mechanism underlying the health benefits of breastfeeding is needed (Figure 1A, axis 2). Elucidating how early weaning from human milk may disrupt this process is crucial and could prove essential in devising targeted early-life interventions to support optimal development.

Breastfeeding and respiratory health

It is also well established that breastfeeding is associated with lower rates of respiratory infections and immune-mediated diseases, both during and beyond infancy.[41–50] For example, in the CHILD Cohort Study, we have observed that breastfeeding is associated with lower rates of wheezing in the first year of life[48,49] and lower odds of asthma by 3 years of age,[49,50] consistent with other cohorts from around the world.[43–47] In addition, a recent study found that a longer period of exclusive breastfeeding was associated with decreased risk of childhood asthma.[51] However, the mechanisms (microbial or otherwise) underlying the association between breastfeeding and respiratory health remain poorly understood (Figure 1A, axis 3).

To examine whether infant microbiome colonization is one of the paths through which breastfeeding is associated with respiratory health (Figure 1A combining all 3 axes), we profiled nasal microbiome trajectories in the first year of life among 2,227 infants from the CHILD Cohort Study. For many of these infants, we also profiled their gut microbiome trajectories and analyzed the milk composition of their mothers. We developed designated computational approaches to analyze our extensive longitudinal, multi-omics dataset, allowing us to identify an accelerated gain of specific microbial species and functions in both niches among infants who experienced early cessation of breastfeeding and those later diagnosed with respiratory illness. We further used a rigorous machine learning framework to demonstrate that these patterns are clinically and biologically relevant, with the capacity to accurately predict preschool asthma years in advance. Finally, we devised a causal inference framework and demonstrated that the colonization patterns of both the nasal and gut microbiota mediate the protective effect of breastfeeding on asthma. Collectively, these results advance our understanding from mere associations to prediction and causality, providing new mechanistic links between breastfeeding, microbial colonization patterns, and infant health and development.

RESULTS

Profiling nasal and gut microbiota development in the first year of life across thousands of infants

We profiled the nasal microbiome composition of infants from the CHILD Cohort Study at age 3 months (median [interquartile range (IQR)] 3.4 [3.0–4.2]; n = 2,227) and 1 year (12.0 [11.5–12.8]; n = 1,868), using 16S rRNA gene sequencing (Figure 1A). We further conducted comprehensive profiling of the gut microbiome in a subset of 1,306 infants, integrating 16S rRNA and shotgun metagenomics sequencing data along with NMR and liquid chromatography-tandem mass spectrometry (LC-MS/MS) data for metabolomics profiling (STAR Methods; Figure S1). We coupled these data with the milk composition of their mothers (e.g., fatty acids, immunoglobulins, oligosaccharides; n = 803), as characterized in previous studies.[34,52,53]

Overall, nasal microbial communities were dominated by species from the genera *Moraxella*, *Streptococcus*, *Staphylococcus*,

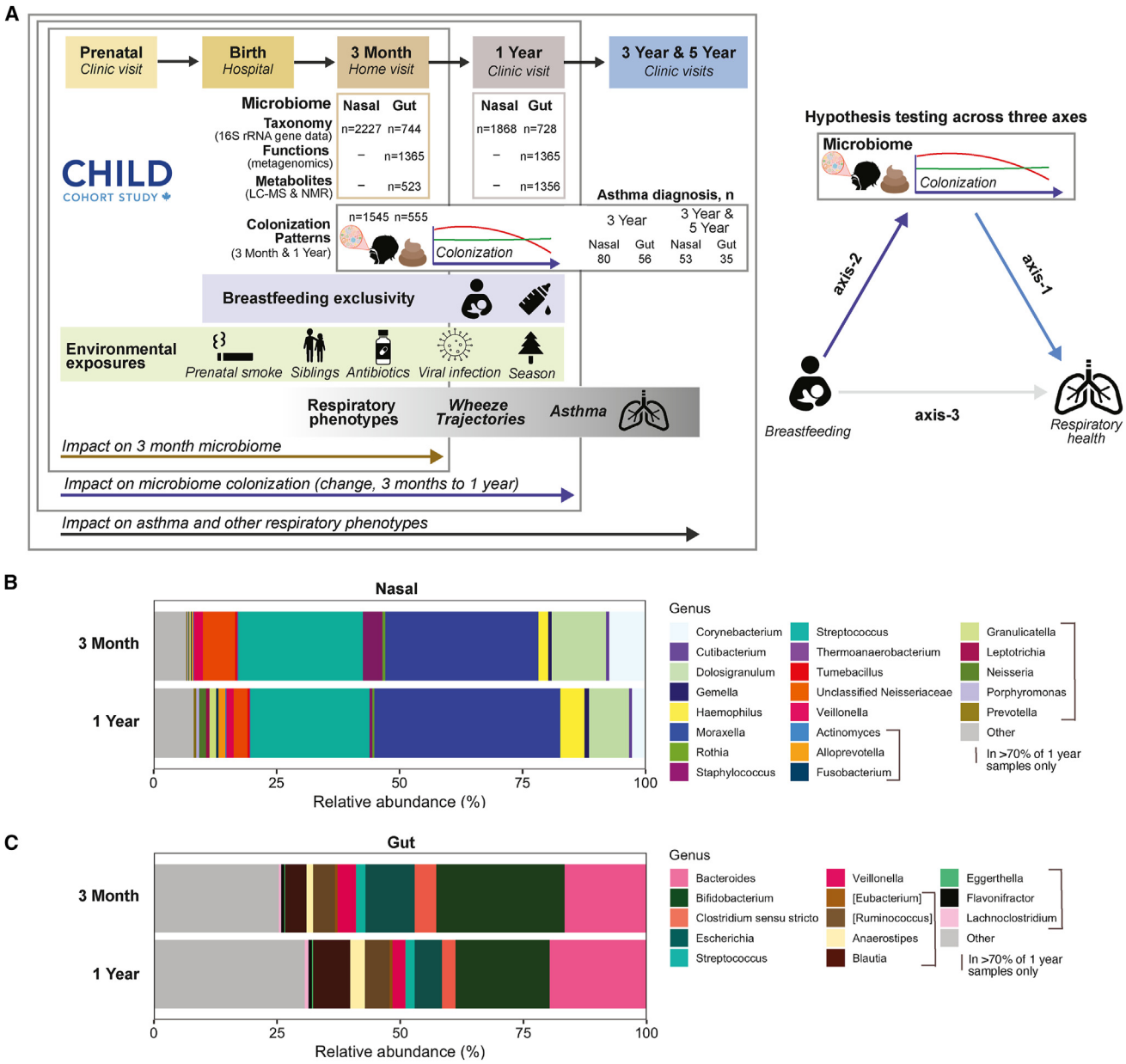


Figure 1. Study design and nasal and gut microbiota composition in early life in the CHILD Cohort Study

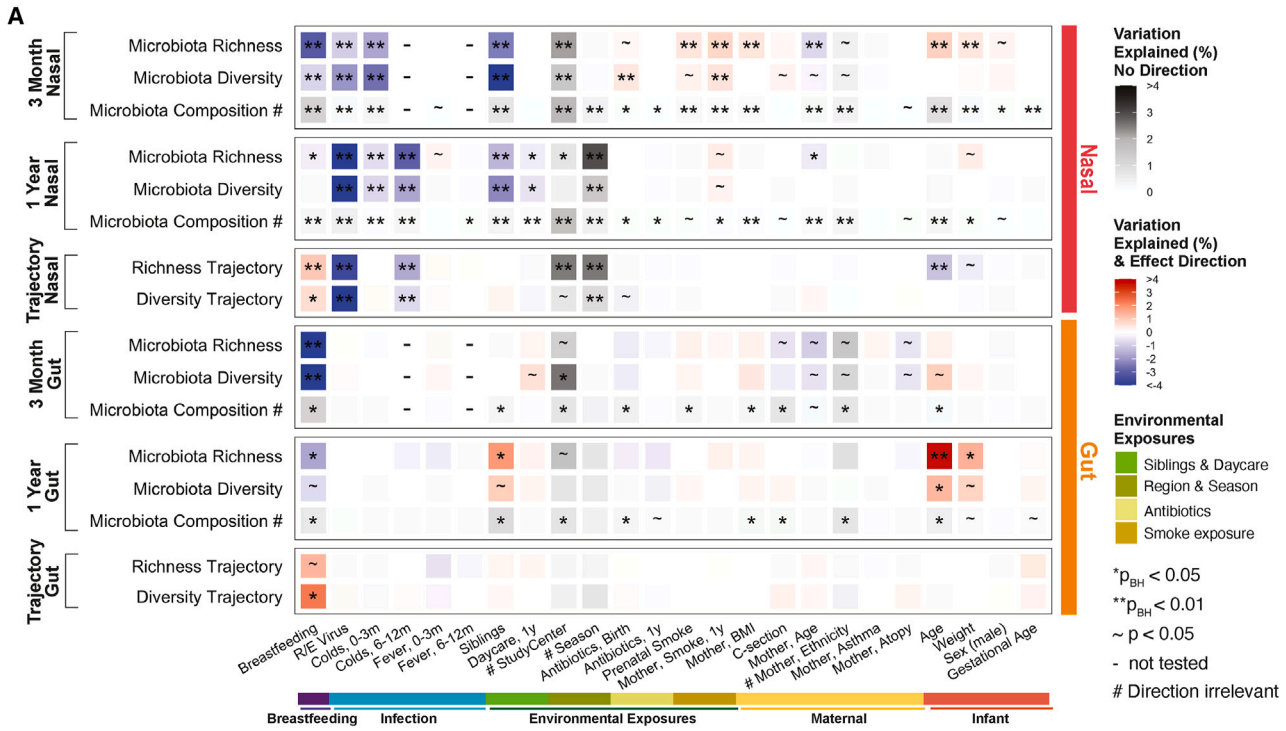
(A) Timeline of early-life exposures, infant nasal and gut microbiome data, and respiratory phenotypes in the CHILD Cohort Study, and hypothesis testing across 3 axes: (axis 1) microbiome development and respiratory health, (axis 2) breastfeeding and microbiome development, and (axis 3) breastfeeding and respiratory health. Sample sizes shown are after all preprocessing filters have been applied (see Figure S1).

(B and C) Average relative abundances (%) of genera (defined by 16S rRNA gene sequencing) present in at least 70% of samples at either 3 months and/or 1 year of age for (B) nasal microbiota (3 months [*n* = 2,227], 1 year [*n* = 1,869]) and (C) gut microbiota (3 months [*n* = 744], 1 year [*n* = 728]). Numbers in brackets represent the number of samples used to calculate the average relative abundances.

See also Figure S1 and Table S1.

Dolosigranulum, *Corynebacterium*, and *Haemophilus* (Figure 1B), with considerable inter-individual variability (Table S1). At 3 months, nasal communities primarily comprised *Streptococcus* (median 17.7% [IQR 5.7–39.9]), *Corynebacterium* (2.3% [0.5–9.1]), *Moraxella* (1.8% [0.1–70.4]), *Dolosigranulum* (1.8% [0.04–13.4]), and *Staphylococcus* (0.4% [0.07–2.0]). As reported previously,[34,53] the gut microbial communities at 3 months and 1 year were dominated by species from the genus *Bifidobacterium*, *Bacteroides*, *Escherichia*, and *Blautia* (Figure 1C).

A defining strength of the CHILD cohort is the extensive metadata encompassing breastfeeding status (e.g., duration, exclusivity, feeding mode), environmental exposures (e.g., antibiotics,



(legend on next page)

pre- and post-natal smoke exposure, birth mode, daycare attendance), viral infection, and structured clinical symptoms and diagnoses (Figure 1A). Breastfeeding status was categorized into 3 groups: "exclusive breast milk (BM)," "partial BM," and "no BM" at 3 months of age (the time of stool, nasal, and milk sample collection). Infants in the exclusive BM group received only their mother's milk from birth until the time of sample collection (n = 794), those in the partial BM group were fed a mixture of mother's milk and formula (n = 517), and those in the no BM group were not receiving mother's milk at the time of sample collection scheduled for age 3 months (n = 234). Infants who were never breastfed at all (n = 55) were excluded from analysis, as we wished to estimate the effects of early breastfeeding cessation. Breastfeeding and other sociodemographic characteristics are shown in Table S2.

Breastfeeding is the most consistent early-life factor associated with nasal and gut microbiota profiles

Next, we explored how nasal and gut microbiome profiles at 3 months and 1 year are associated with early-life factors (i.e., environmental exposures, infant, and maternal factors as well as feeding characteristics; Figure 2A). We found that breastfeeding status at the time of sampling (i.e., exclusive vs. cessation of breastfeeding before 3 months; any vs. no breastfeeding at 1 year) was the most consistent association in both strength and direction for microbiome richness, diversity, and composition across both niches (e.g., $p < 0.001$ for richness and diversity and $p < 0.05$ for composition at 3 months in both niches, with persistence to 1 year for richness and composition at $p < 0.05$; Figure 2A). Of note, the nasal and gut microbiome profiles were also variably associated with other early-life factors including environmental exposures (e.g., presence of siblings, maternal intrapartum antibiotics, season), maternal and infant characteristics (e.g., maternal smoking, BMI, infant weight), and rhinovirus/enterovirus infections, with generally smaller effect sizes (Figure 2A). However, unlike breastfeeding, these factors were not consistently associated with microbiome profiles across time and niches.

To explore how microbiome dynamics in the first year of life are affected by early-life factors, we evaluated microbiome richness and diversity trajectories, defined as the change in richness (observed operational taxonomic units [OTUs]) or diversity (Shannon index) between 3 months and 1 year, respectively (STAR Methods). Overall, we found that microbiota trajectories were associated with fewer environmental, maternal, and infant factors, as compared with the cross-sectional snapshots of these measures at 3 months or 1 year. However, breastfeeding remained significantly associated with most trajectory measures across both niches (Figure 2A). Further, nasal and gut microbiome richness and diversity, both cross-sectionally at 3 months and as trajectories, remained associated with breastfeeding after adjusting for covariates using multivariate regression (richness trajectory, nasal: $p < 0.001$ and gut: $p = 0.021$; diversity trajectory, nasal: $p = 0.004$ and gut: $p < 0.001$; Figure 2B). Specifically, we found a greater increase in richness and diversity over time for infants exclusively breastfed, compared with those who experienced early cessation of breastfeeding (richness trajectory [change in number of species], nasal: median [IQR] 24.0 [−3.0–55.0] for exclusively breastfed vs. 10.0 [−17.0–41.0] for non-breastfed, $p < 0.001$; gut: 20.0 [5.0–34.0] vs. 12.0 [−10.5–25.5], $p = 0.004$; diversity trajectory [change in Shannon index], nasal: 0.17 [−0.57–0.99] vs. 0.14 [−0.90–0.79], $p = 0.028$; gut: 0.87 [0.16–1.36] vs. 0.34 [−0.41–0.81], $p < 0.001$, Mann-Whitney U test; Figures 2C and 2D).

Notably, we found similar associations between these temporal measures and preschool asthma, namely: a greater increase in richness (nasal and gut) and diversity (gut) over time in healthy infants, compared with those diagnosed with asthma at 3 years of age (richness trajectory, nasal: 23.0 [−5.0–54.0] for healthy infants vs. 12.5 [−8.3–35.3] for those who developed asthma, $p = 0.046$; gut: 20.0 [3.0–33.0] vs. 8.5 [−17.8–23.3], $p = 0.001$; diversity trajectory, nasal: 0.22 [−0.60–1.02] vs. 0.16 [−0.54–0.80], $p = 0.90$; gut: 0.79 [−0.01–1.31] vs. 0.45 [−0.64–0.96], $p = 0.004$; Figures 2E and 2F). Interestingly, while breastfeeding was associated with both cross-sectional and temporal measures of nasal richness, associations with asthma at 3 years only reached significance for temporal measures (Figure S2).

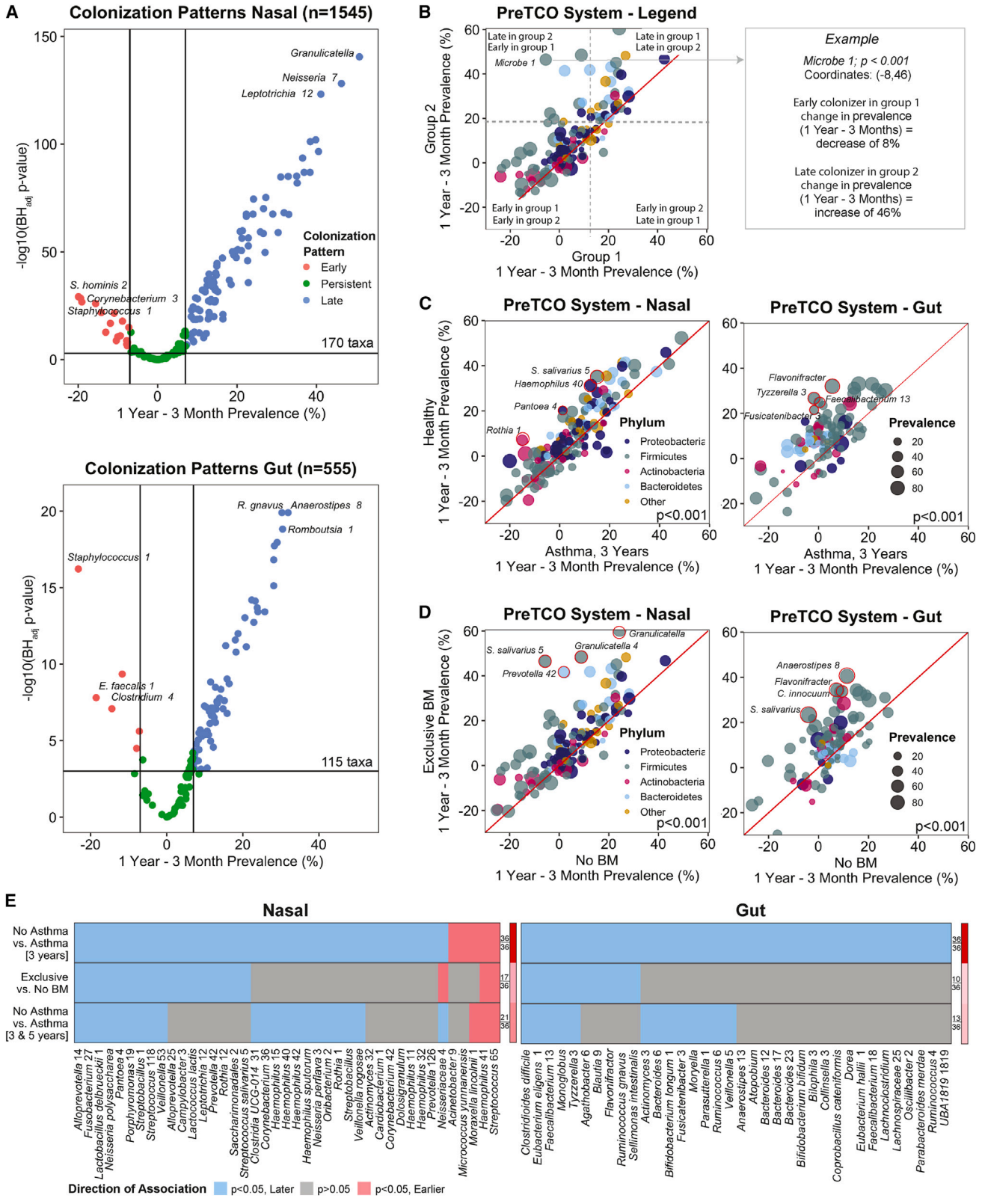
Based on these results, we hypothesize that temporal changes in diversity and richness from 3 months to 1 year are more biologically conserved, compared to the community state at a single time point. Thus, these changes likely provide a more robust representation of microbiome development that is less

Figure 2. Nasal and gut microbiome profiles and trajectories are associated with breastfeeding, preschool asthma, and other maternal, infant, and environmental factors

(A) Associations between early-life factors and infant nasal and gut microbiota analyzed by 16S rRNA gene sequencing at two time points (3 months [n = 2,227 for nasal and n = 744 for gut] and 1 year [n = 1,868 for nasal and n = 728 for gut]) and as a trajectory (change from 3 months to 1 year [n = 1,545 for nasal and n = 555 for gut]). Showing variation explained as the R^2 of the linear model (richness [observed OTUs], diversity [Shannon index] and trajectories [change in richness and diversity from 3 months to 1 year]), or R^2 of the redundancy analysis [microbiota composition]. p values were adjusted using Benjamini-Hochberg (BH) correction. Abbreviations: "Antibiotics, 1y," any antibiotics given within the first year of life; "Antibiotics, Birth," intrapartum antibiotics given to the mother; "BMI," maternal pre-pregnancy body mass index.

(B) Associations between breastfeeding at sample collection and measures of microbiota richness and diversity, tested using multivariate linear regression. Adjusted models include the following covariates: older siblings, antibiotics, R/E virus, colds, maternal asthma, pre-natal smoke exposure, Cesarean section, study center, and exact age at 3-month sample collection. Estimates refer to exclusively breastfed vs. no longer breastfed at 3 months for 3-month and trajectory models and breastfed vs. no longer breastfed for 1-year models. Partially breastfed infants were included to maximize sample sizes: nasal 3 months, nasal 1 year, and nasal trajectory (n = 1,545); gut 3 months, gut 1 year, and gut trajectory (n = 555). The same dataset was used for all models to ensure comparability.

(C–F) Within nasal and gut niches, microbiota richness and diversity compared between 3 months and 1 year of age (C and E) and microbiota richness and diversity trajectories (i.e., change between 3 months and 1 year) (D and F), for infants exclusively breastfed (exclusive, nasal n = 794, gut n = 271) and those no longer breastfed (no BM, nasal n = 234, gut n = 79) at the 3-month sample collection (C and D), and for infants who did (nasal n = 80, gut n = 56) and did not (nasal n = 1,236, gut n = 421) develop asthma by 3 years (E and F). Comparisons tested using Mann-Whitney U test. *$p < 0.05$; **$p < 0.001$. BM, breast milk.
See also Figure S2.



(legend on next page)

susceptible to "noise" and local environmental factors than a static community state. We further suggest that the process of acquiring new species and functions during early life is a more meaningful indicator of infant and child health, compared with the community state at any single point in time.

Microbiome colonization patterns and the "PreTCO system"

To better characterize nasal and gut microbiome changes in the first year of life, we defined three colonization patterns ("early," "persistent," and "late") and applied them to each taxon. To determine the most common colonization pattern for each taxon at the overall cohort level, we compared prevalence between 1 year and 3 months and assigned a *p* value using a McNemar test, allowing us to statistically identify colonization patterns as early (higher prevalence at 3 months), late (higher prevalence at 1 year), or persistent (similar prevalence at both time points) (STAR Methods; Figure 3A; Table S3). At the individual infant level, taxa were labeled based on their presence or absence at each time point (early, if present at 3 months only; late, if present at 1 year only; persistent, if present at both time points; or "absent," if absent at both time points).

Across the 170 nasal taxa detected, over half were late colonizers at the cohort level (53%, 91/170), for example, *Streptococcus salivarius* (*S. salivarius*) sp. 5, with a prevalence of 31% at 3 months and 64% at 1 year. Persistent colonizers accounted for 37% (63/170) of the total, including all *Moraxella* taxa. Only a few taxa were considered early colonizers (10%, 17/170), for example, five of the six *Staphylococcus* spp., including *Staphylococcus aureus* 10 (29% prevalence at 3 months, 15% at 1 year) (Figure 3A). Similarly, across the 115 gut taxa, the majority were late colonizers (57%, 65/115), for example, *Ruminococcus gnavus* (*R. gnavus*) (54% at 3 months, 84% at 1 year) and *Flavonifractor* sp. (46% at 3 months, 74% at 1 year). Persistent colonizers were 38% of the total, for example, *Escherichia coli* (80% at 3 months, 86% at 1 year) and *Bifidobacterium catenulatum* (30% at 3 months, 38% at 1 year). As in the nasal niche, only a few were early colonizers (5%, 6/115), for example *Staphylococcus* spp. 1 (33% at 3 months, 10% at 1 year) (Figure 3A). These results are in line with the concept of ecological succession, where an initial increase in complexity is followed by the colonization of additional microbes later in the ecosystem establishment process.

We further assessed gut colonization patterns using shotgun metagenomic data from the same infants (1,306 infants with both 3-month and 1-year samples). While we would not expect identical results from these different sequencing approaches, we found striking similarities between the microbial colonization patterns identified in each dataset, with analogous patterns of taxa categorized as late, persistent, and early colonizers (Figure S3). This congruence underscored a consistent phenomenon, emphasizing that the majority of microbiota are introduced later in the first year (87/115 or 76% late colonizers in the metagenomic data and 65/115 or 57% for 16S data). Further, the specific taxa identified as late colonizers showed consistency across different data types. For instance, *R. gnavus* was among the top three late colonizers in both 16S rRNA gene sequencing and shotgun metagenomic data, appearing in 36% of samples at 3 months and 95% at 1 year. This alignment between the two datasets reinforces the reliability of our microbial colonization insights.

We next sought to systematically compare the colonization patterns of infants with different health phenotypes and early-life exposures, both at the microbial community level and at the individual taxa level. To this end, we devised a computational approach we term "prevalence trajectories coordinate system" (PreTCO system; Figure 3B; STAR Methods). This approach allows us to compare the colonization patterns of different phenotypic groups visually and statistically (e.g., exposed vs. unexposed or cases vs. controls) both globally, across the entire microbial community, as well as locally, per taxa. We define the prevalence trajectories as a two-dimensional cartesian coordinate system with each point representing a taxon and its coordinates corresponding to the change in prevalence from 3 months to 1 year for two phenotypic groups (STAR Methods; Figure 3B).

A premature gain of microbial species is associated with early cessation of breastfeeding and asthma

Using the PreTCO system, we first compared the colonization patterns of healthy infants and those diagnosed with preschool asthma at 3 years of age. We found significant community-wide

Figure 3. Nasal and gut microbiota colonization patterns and their associations with breastfeeding and asthma

(A) Showing the difference in prevalence of microbiota between 3 months and 1 year in percentage, for infants with 16S rRNA gene sequence data for nasal samples (n = 170 taxa, n = 1,545 infants) or gut samples (n = 115 taxa, n = 555 infants; McNemar test). The horizontal line indicates the p value threshold ($p_{BH} < 0.001$), and vertical lines indicate the effect size thresholds (−7% and 7%) used to define early (more prevalent at 3 months), persistent (similar prevalence at both time points), and late (more prevalent at 1 year) colonizers. The three early and late colonizers with the highest effect sizes are annotated. BH_{adj}, Benjamini-Hochberg adjusted p values. Also see Table S3.

(B) Graphical legend for the prevalence trajectory coordinate system (PreTCO system).

(C and D) Prevalence trajectory coordinate analyses comparing the change in prevalence of microbiota (1-year to 3-month prevalence in percentage; STAR Methods) between healthy infants (nasal n = 1,236, gut n = 421), and those diagnosed with asthma at 3 years (nasal n = 80, gut n = 56) (C), and between exclusively breastfed infants (exclusive BM, nasal n = 794, gut n = 271) and those no longer breastfed at 3-month sampling (no BM, nasal n = 234, gut n = 79) (D). The four taxa with the largest effect size are labeled for each comparison. The overall prevalence across both time points is shown as point size. Also see Table S4.

(E) Nasal and gut microbiota that were significantly later (blue) or earlier (pink) colonizers in infants without vs. with asthma at 3 years, showing whether these taxa were also found later or earlier in infants (1) exclusively breastfed (nasal n = 794, gut n = 271) vs. no longer breastfed (nasal n = 234, gut n = 79) at 3-month sampling, (2) without asthma at 3 and 5 years (nasal n = 1,038, gut n = 347) vs. with asthma at 3 and 5 years (nasal n = 53, gut n = 35). The annotation bar to the right shows the fraction of the 36 nasal and 36 gut taxa found later in infants without asthma that were also found later in the other comparisons. Significance of later compared with earlier colonization was determined using a permutation test. Only microbiota with prevalence trajectories that differed between infants with and without asthma with $p < 0.05$ are shown. BM, breast milk.

See also Figures S3, S4, S5, and S6 and Tables S3 and S4.

differences in both the nasal and gut ecosystems between these two phenotypic groups ($p < 0.001$; Figure 3C). Specifically, at the community level (across all taxa), in both nasal and gut niches, there was a greater increase in prevalence from 3 months to 1 year in healthy infants (median [IQR] percent change in prevalence: 9.0% [1.0%–17.0%] for nasal and 8.8% [4.5%–14.5%] for gut), compared with infants later diagnosed with asthma (6.3% [−1.3%–13.1%] for nasal and 1.8% [−1.8%–7.1%] for gut). This result is further supported by a direct comparison of the colonization patterns, where healthy infants showed significantly higher proportions of late colonizers across both nasal and gut communities, as compared with infants diagnosed with preschool asthma (Mann-Whitney U test, nasal $p = 0.011$, gut $p = 0.006$; Figure S4A). Notably, we observed some consistent patterns for asthma at 5 years, particularly for children with preschool asthma diagnosed at 3 years that persisted at 5 years (Figure S5).

We found similar patterns when comparing infants with different breastfeeding patterns (Figure 3D). Specifically, at the community level for both the nasal and gut, there was a greater increase in prevalence from 3 months to 1 year in exclusively breastfed infants ($p < 0.001$, median [IQR] percent change in prevalence, 8.7% [0.9%–18.3%] for nasal and 10.0% [5.0%–15.7%] for gut), as compared with infants who were no longer breastfed (4.3% [−1.7%–12.8%] for nasal and 3.8% [1.3%–8.9%] for gut). This result is further supported by a direct comparison of the colonization patterns, where exclusively breastfed infants showed significantly higher proportions of late colonizers and lower proportions of early colonizers across both nasal and gut communities, as compared with those no longer breastfed at 3 months (Mann-Whitney U test, $p < 0.001$ for early, persistent, and late in nasal and early and persistent in gut, and $p = 0.003$ for late in gut; Figure S4B). We also analyzed infants with mixed feeding (receiving both human milk and formula), revealing monotonic trends in colonization patterns across the three feeding groups (exclusive, partial, and no BM), which indicate a dose-response relationship (Figure S4B): exclusively breastfed infants showed the highest percentage of late colonizers; followed by those partially breastfed; and finally, those who stopped breastfeeding before 3 months. The trend for early colonizers was the opposite. These results are consistent with our hypothesis that breastfeeding promotes a paced and gradual microbial colonization process, across multiple microbial niches (nasal and gut). In contrast, early breastfeeding cessation may lead to the premature acquisitionof microbial species.

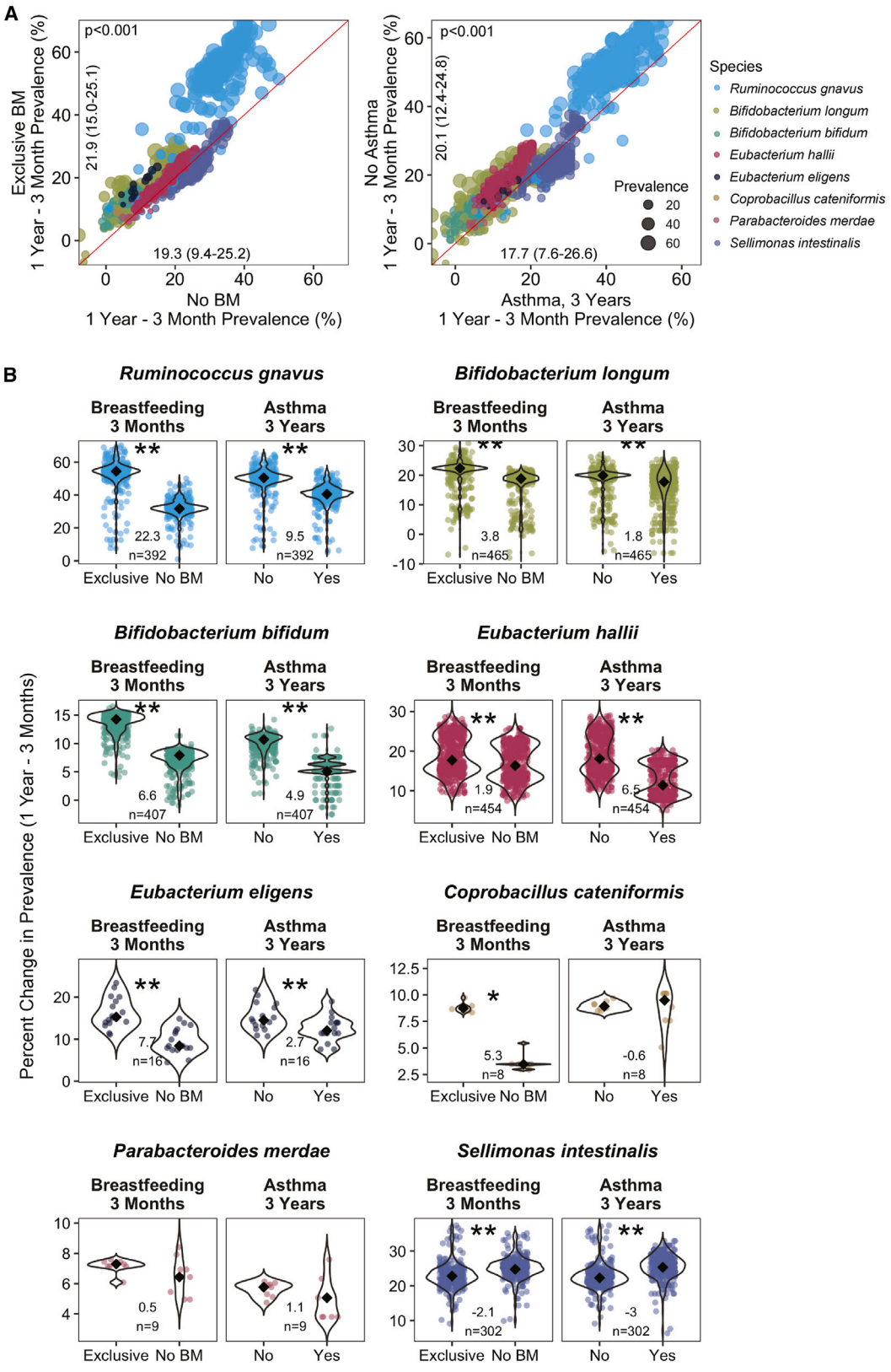
We next performed a permutation test using the PreTCO system, and assigned a p value to each taxon indicating whether its colonization pattern was significantly different between the two phenotypic groups (Table S4). Consistent with above results, more gut and nasal microbial taxa colonized significantly later in healthy infants (vs. infants diagnosed with asthma) and those exclusively breastfed (vs. no longer breastfed at 3 months), while only a few colonized earlier (i.e., 36 later and 5 earlier nasal taxa in healthy infants; 36 later and 0 earlier gut taxa with $p < 0.05$; Figures 3C and 3E; Table S4).

Further, we found a substantial overlap between specific taxa that exhibit later colonization patterns in exclusively breastfed infants (vs. no longer breastfed at 3 months) and healthy infants (vs. diagnosed with preschool asthma at 3 years). We focused on 36 nasal taxa and 36 gut taxa that were later colonizers in healthy infants and examined how consistently these colonization patterns were associated with breastfeeding. Overall, 47% (17/36) of the late nasal colonizers and 28% (10/36) of late gut colonizers were also associated with breastfeeding $p < 0.05$ (Figures S6B and S6C). For example, in the nasal microbial community, *S. salivarius*, a known commensal of the oral and nasal cavity, had a greater increase in prevalence from 3 months to 1 year (later colonization) in infants exclusively breastfed, compared with infants no longer breastfed (+47% vs. −6%, permutation test $p < 0.001$; Figure 3D), and in infants without asthma, compared with infants diagnosed with asthma at 3 years (+35% vs. +15%, $p = 0.004$; Figure 3C). Other nasal microbiota with late colonization patterns in both breastfeeding and no asthma groups included *Fusobacterium* sp., *Pantoea* sp., and *Leptotrichia* sp. (Figure 3E; Table S4). In the gut microbiome, we see similar patterns for taxa such as *R. gnavus* and *Flavonifractor* sp. (Figure 3E; Table S4).

Notably, we observed consistent colonization patterns for several taxa when we investigated persistent asthma (diagnosed at 3 years and still present at 5 years). Specifically, nine nasal taxa and five gut taxa, which represent 25% and 14% of late colonizers in healthy infants, respectively, consistently colonized later in infants who were exclusively breastfed and those who did not have asthma at 3 years, nor develop it by 5 years. These nasal taxa included *Alloprevotella* sp., *Fusobacterium* sp., *Pantoea* sp., *Porphyromonas* sp., and *Streptobacillus*; and gut taxa included *Faecalibacterium* sp., *Monoglobus* sp., and *Tyzzerella* sp. (Figures 3E and S6B). These results support our hypothesis that exclusive breastfeeding until at least 3 months selectively paces the colonization of specific microbiota in early life that are associated with infant health.[54]

Pangenome analysis reveals a premature gain of microbial functions associated with early cessation of breastfeeding and asthma

To unravel the underlying microbial mechanisms contributing to the protective effects of breastfeeding against asthma, we sought to move beyond microbiota composition and explore the functional potential of the implicated gut microbes. To achieve this, we analyzed 2,846 longitudinal shotgun metagenomic samples from the same infants (1,306 infants with both 3-month and 1-year samples) and a subset of corresponding metabolomics samples (529 infants with both 3-month and 1-year samples). We assessed the functional potential of the microbial communities under scrutiny using their pangenomes (i.e., the entire set of genes from all strains within a clade). Leveraging both the evolutionary genealogy of genes: non-supervised orthologous groups (EggNOG) database[55] and the Enzyme Commission (EC) classification scheme, we extracted 2,229 unique EggNOG functional annotations and 1,110 EC numbers associated with 73 species present in at least 10% of the total 2,846 samples (STAR Methods). We focused on eight species based on our 16S rRNA gene sequence results, specifically, those identified as later colonizers in infants who did not develop asthma ($p < 0.05$, Figure 3E), which had species-level identification and were found in the EC and EggNOG annotations:



(legend on next page)

R. gnavus (401 EC and 940 EggNOG), *Bifidobacterium bifidum* (419 and 797 functions), *Bifidobacterium longum* (*B. longum*; 481 and 713 functions), *Sellimonas intestinalis* (*S. intestinalis*; 305 and 128 functions), *Eubacteriu hallii* (487 and 574 functions), *Eubacterium eligens* (423 and 95 functions), *Coprobacillus cateniformis* (*C. cateniformis*; 288 and 21 functions), and *Parabacteroides merdae* (*P. merdae*; 360 and 15 functions). Of note, we hypothesize that the functions attributed to *B. longum* are not associated with *B. longum infantis*, which consumes HMOs, but rather with other *B. longum* subspecies that do not utilize HMOs.

We next calculated a complementary type of ''functional trajectories,'' describing the prevalence of *functions* assigned to these species. In this case, applying the PreTCO system analysis would reveal changes in functional prevalence, capturing the genetic diversity of the species, as represented by their pangenomes. This approach can potentially uncover functions of different subspecies or strains within the same clade (Figure S8). In this analysis we found that both exclusive breastfeeding at 3 months and the absence of asthma at 3 years exhibited a robust association with a delayed introduction of functions in these species ($p < 0.001$, Figures 4A and S7A). Examining each species separately, this pattern held true for all except one using EggNOG and five out of eight species using EC (EggNOG $p < 0.05$ for all species except *P. merdae*; EC $p < 0.05$ except *P. merdae*, *C. cateniformis*, and *S. intestinalis*; Figures 4B and S7B). *P. merdae* and *C. cateniformis* had less than 10 functional annotations in the EC data, limiting statistical power for our analysis.

Premature *R. gnavus* acquisition, along with tryptophan biosynthesis and its metabolites, links early breastfeeding cessation to asthma development

To identify and attribute microbial mechanisms underlying the protective effects of breastfeeding against asthma, we looked for species exhibiting overlapping late functions across these two phenotypes (i.e., exclusive breastfeeding and no asthma). Despite finding hundreds of functions temporally correlated with each phenotype across our eight species of interest (Figure 4), only a subset reached statistical significance (386 EC annotations for breastfeeding, 12 for asthma). Of these, only 8 EC annotations overlapped between both phenotypes (Figure 5A), with 22 overlapping late functions identified using EggNOG (Figure S7). Interestingly, all overlapping late functions originated from *R. gnavus*, indicating this species is a pivotal microbe linking breastfeeding with asthma. We therefore focused our attention on this microbe (Figure 5A).

First, we examined *R. gnavus* prevalence over time using the shotgun metagenomics data. We found that *R. gnavus* exhibited a delayed colonization pattern in exclusively breastfed infants (1-year to 3-month prevalence: 67.2% increase), compared with infants who experienced early cessation of breastfeeding (37.6%, $p < 0.001$) and in infants who did not develop asthma, compared with infants who did (59.8% vs. 49.4%, $p = 0.042$; Figure S8B). At the functional level, we similarly observed a larger temporal increase (from 3 months to 1 year) in the prevalence of total functions linked to *R. gnavus* in exclusively breastfed infants and those who did not develop asthma ($p < 0.001$, Figure 4B): across all EC functions linked to *R. gnavus*, the median increase in the presence of a function was 54.4% (IQR [51.5%–56.4%]) for exclusively breastfed infants, compared with 31.7% (IQR [30.2%–33.7%)] for infants who experienced early cessation of breastfeeding. Similarly, the median increase was 50.4% (IQR [48.3%–52.3%]) for infants who did not develop asthma, compared with 40.5% (IQR [39.2%–43.0%]) for infants who developed asthma.

Next, we examined the delayed functions in *R. gnavus*. Remarkably, all five enzymes required for tryptophan biosynthesis[56] (i.e., 4.1.3.27 anthranilate synthase, 2.4.2.18 anthranilate phosphoribosyltransferase, 5.3.1.24 phosphoribosylanthranilate isomerase, 4.1.1.48 indole-3-glycerol phosphate synthase, and 4.2.1.20 tryptophan synthase) were introduced significantly later in exclusively breastfed infants ($p < 0.001$ for all comparisons). Notably, these five enzymes were also delayed in infants without asthma, although only one (anthranilate phosphoribosyltransferase [AnPRT]; catalyzing the second step in tryptophan biosynthesis) reached significance and was among the eight overlapping functions between both phenotypes (Figure 5B). Other enzymes represented among these eight overlapping functions include 3.5.1.2 (glutaminase, $p = 0.008$) and 3.2.1.22 (alpha-galactosidase, $p = 0.004$) (Figures 5B and S8B). Notably, these eight functions attributed to *R. gnavus* exhibit a high correlation with its overall prevalence trajectory, achieving a median Spearman correlation of 0.72. In contrast, other functions of *R. gnavus*, which are not delayed in breastfed infants, show a median correlation of only 0.19 (Figure S8A; STAR Methods). These results indicate a functional diversity within *R. gnavus* members that may be influenced by early-life factors such as breastfeeding. To corroborate these results and move from suggestive (i.e., predicted functional potential) to conclusive evidence (i.e., measured functional outputs), we analyzed microbial metabolites in the same stool samples, focusing on tryptophan biosynthesis and metabolism (Figure 5C). As anticipated, in exclusively breastfed infants, both tryptophan and its metabolites, tryptamine and indole, exhibited

Figure 4. Timing of acquiring microbial functions in the first year of life is associated with breastfeeding and asthma

(A) PreTCO system analyses comparing the change in prevalence of enzyme-catalyzed reaction pathways, based on EC numbers from metagenomic data (1-year to 3-month prevalence in percentage; STAR Methods) between (1) infants exclusively breastfed ($n = 658$) and those no longer breastfed at 3 months ($n = 202$); and (2) healthy infants ($n = 1,075$), and those later diagnosed with asthma at 3 years ($n = 79$). Each point represents an EC annotation from the taxa listed in the key to the right, differentiated by color. The median (IQR) change in prevalence of each group is shown on the group's axis. Overall significance was tested using a Mann-Whitney U test. The overall prevalence across both time points is shown as point size. Replicated for EggNOG orthologs (see Figure S7).

(B) Percent change in prevalence of EC orthologs, stratified by species. Significance was tested using a Wilcoxon signed-rank test. Each point represents a single function from the focal species. As an effect size, the median of the difference in the trajectory measure (percentage change in prevalence) is shown. Sample sizes for this test are the number of functions (differs per species, as annotated in the plot). $^{*}p_{BH} < 0.05$; $^{**}p_{BH} < 0.001$. Replicated for EggNOG orthologs and sample size sensitivity analysis (see Figure S7).

See also Figure S7.

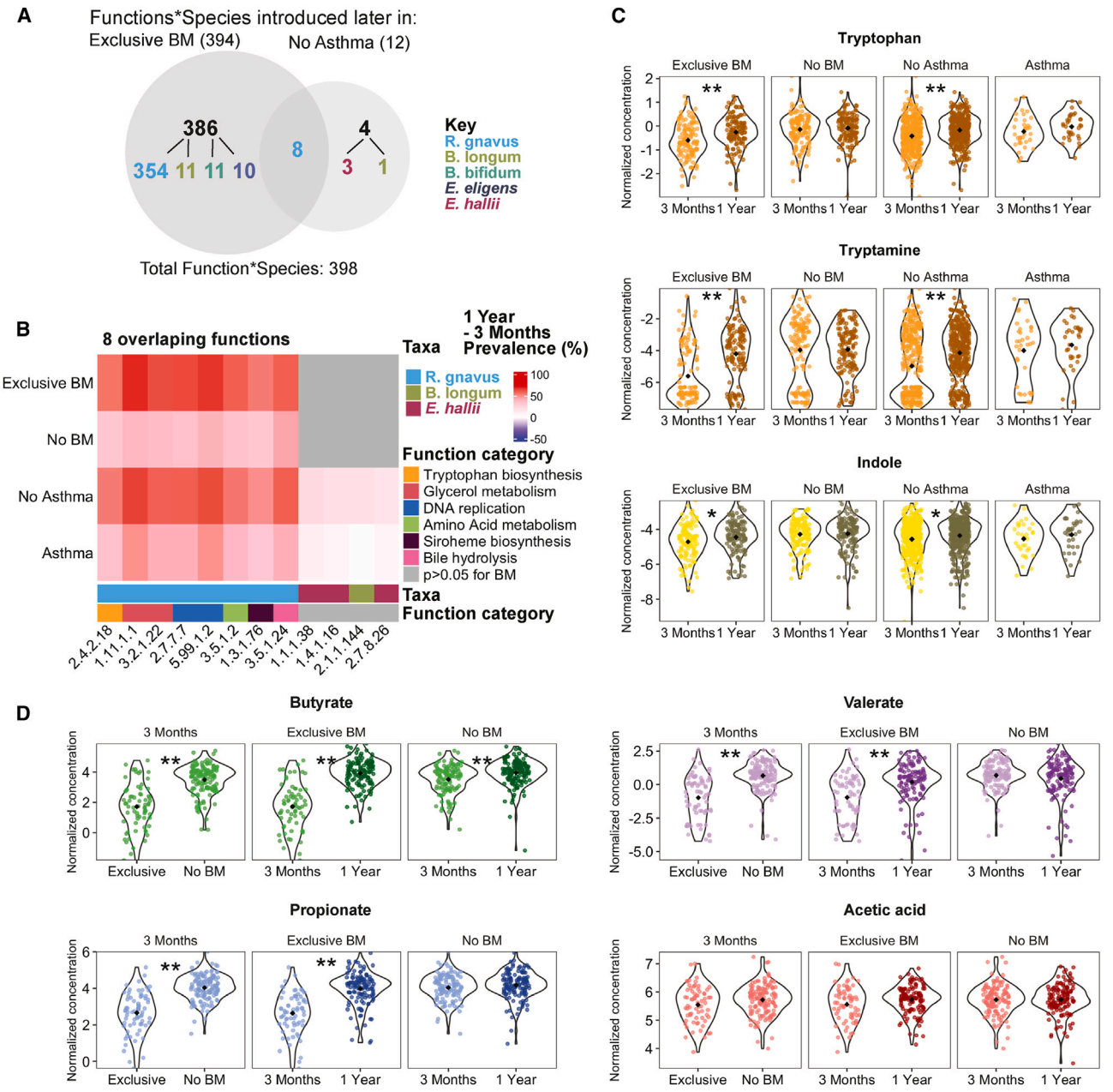


Figure 5. Premature acquisition of specific *R. gnavus* functions are linked to early breastfeeding cessation, increased asthma risk, and tryptophan metabolite variations

(A) Overlap in the EC annotations stratified by species introduced significantly later in exclusively breastfed infants (exclusive BM) compared with those no longer breastfed at 3 months sampling (no BM), and later in infants who did not develop asthma (no asthma) compared with those who did at 3 years ($p < 0.01$). Significance of individual pathway was tested using a permutation test.

(B) Percentage change in prevalence between 3 months and 1 year within each phenotype, for 12 EC that were significantly later in infants who did not develop asthma at 3 years ($p < 0.01$, permutation test). Only showing the percentage change in prevalence where $p < 0.01$ (gray indicates $p > 0.01$). EC names: 2.4.2.18, anthranilate phosphoribosyltransferase; 1.11.1.1, NADH peroxidase; 3.2.1.22, alpha-galactosidase; 2.7.7.7, DNA-directed DNA polymerase; 5.99.1.2, DNA topoisomerase; 3.5.1.2, glutaminase; 1.3.1.76, precorrin-2 dehydrogenase; 3.5.1.24, choloylglycine hydrolase; 1.1.1.38, malate dehydrogenase; 1.4.1.16, diaminopimelate dehydrogenase; 2.1.1.144, *trans*-aconitate 2-methyltransferase; 2.7.8.26, cobalamin synthase.

(C) Comparing normalized concentrations of tryptophan, tryptamine, and indole over time (3 months and 1 year) among exclusively breastfed infants ($n = 127$) and those no longer breastfed at 3-month sampling ($n = 140$) and healthy infants ($n = 393$) compared with those who develop asthma at 3 years ($n = 34$).

(legend continued on next page)

a significant increase at 1 year compared with 3 months, aligning with metagenomics results ($p < 0.001$ for tryptophan and tryptamine and 0.04 for indole). This pattern was similarly observed in healthy infants who did not develop asthma ($p < 0.001$ for tryptophan and tryptamine and 0.003 for indole). In contrast, no temporal changes in tryptophan, tryptamine, or indole levels were observed in infants who experienced early cessation of breastfeeding and those diagnosed with preschool asthma.

Early cessation of breastfeeding and the introduction of infant formula trigger a weaning response characterized by compositional and functional microbiome changes

Next, to better understand the impact of early weaning from human milk and the introduction of infant formula on the gut microbiome, we examined the functional aspects of these changes. In mice, the transition from nursing to solid foods initiates a robust immune response, marked by an increase in microbial colonization and the production of short-chain fatty acids (SCFAs), which are crucial for developing gut homeostasis and robust immune maturation.[57,58] Termed the "weaning reaction," this phenomenon has not yet been rigorously demonstrated in humans. Here, we wished to examine wheather a similar response is observed in response to weaning from human milk and the introduction of infant formula. We analyzed SCFAs in stool at 3 months ($N = 392$ infants) and observed significantly higher levels in infants no longer breastfed, compared with those exclusively breastfed (Figure 5D; $p < 0.001$ for butyrate, propionate, and valerate and 0.07 for acetate). Moreover, when examining temporal changes in these SCFAs between 3 months and 1 year, we detected significant temporal increases among exclusively breastfed infants (Figure 5D; $N = 340$ infants, $p < 0.001$ for butyrate, propionate, and valerate), whereas this distinctive pattern was absent in infants who discontinued breastfeeding before 3 months (Figure 5D). These results suggest a potential early "weaning response," reflected in an increase in microbial production of SCFAs, in infants who transition from human milk to formula before 3 months of age.

Microbiome colonization patterns and human milk composition in the first year of life accurately predict asthma at 3 years

To further establish the health implications of microbial colonization patterns and breastfeeding (including exposure to specific human milk components) in the first year of life, we next evaluated their ability to predict asthma at 3 years. We trained a machine learning model using gradient-boosted decision trees to differentiate between children diagnosed with asthma ($N = 80$ children for nasal and 56 for gut datasets) and healthy controls ($N = 1,327$ for nasal and 479 for gut). We evaluated the prediction accuracy using 10-fold cross-validation (i.e., held-out subjects unseen during training; STAR Methods). The composition of either the nasal or gut microbiome at a single time point (3 months or 1 year) had limited predictive power (average area under the receiver operating characteristic curve [auROC] values ranging from 0.56 to 0.65; Figure 6A, left). However, when we used the microbiota trajectories, the predictive accuracy was significantly increased (nasal: auROC = 0.79, 95% CI [0.70, 0.89], gut: auROC = 0.83, 95% CI [0.72, 0.93]; Figure 6A). When using the microbiome trajectories of the nasal and gut combined, the predictive accuracy was even higher (auROC = 0.90, 95% CI [0.80, 0.99]; Figure 6A). Human milk components (oligosaccharides, fatty acids, and immune components; $n = 632$) had some predictive power on their own (auROC = 0.71, 95% CI [0.55, 0.86]; Figure 6A), and they slightly improved the performance of microbiota trajectory models when these datasets were combined (Figure 6A, right). The highest model performance of all was obtained from using gut and nasal trajectories along with milk components (auROC = 0.93, 95% CI [0.86, 0.99]; Figure 6A). These accurate predictions of preschool asthma demonstrate the importance of microbial colonization in the first year of life, indicate a role for human milk components, and suggest utility for early risk stratification.

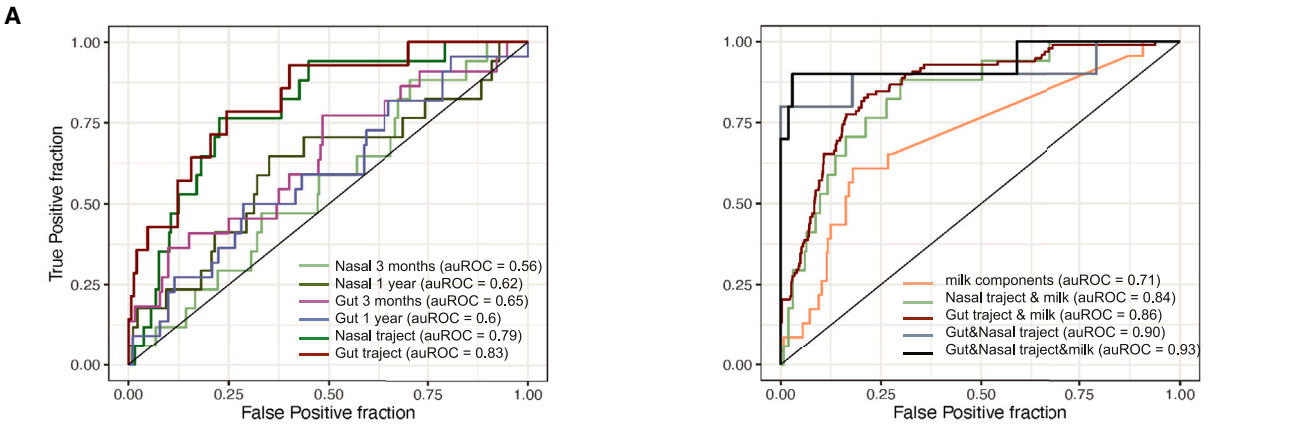
To elucidate which microbial taxa and human milk components drive the prediction accuracy of asthma at 3 years, we ranked the contribution of these features in our machine learning models. We found that microbes exhibiting later colonization in healthy infants, discovered in previous analyses, are the ones contributing the most to predicting asthma at 3 years (Figures 6B and 6C). Specifically, we found that 9 out of the top 10 contributing microbial features were also found in the PreTCO analysis (i.e., *S. salivarius*, *Dolosigranulum* sp., *Corynebacterium* sp., *Neisseriaceae* sp., *Prevotella* sp., *Fusobacterium* sp., *Moraxella* sp., *Leptotrichia* sp., and *Alloprevotella* sp. in nasal microbiota; Figure 6B). In terms of human milk components contributing to prediction accuracy, these included three omega 6 polyunsaturated fatty acids (C22:4n6, C20:3n6, and C18:4n6), saturated fatty acid C24:0, immunoglobulin A (IgA), and the HMOs lacto-N-neotetraose (LNnT) and lactose-N-hexaose (LNH) (Figure 6B).

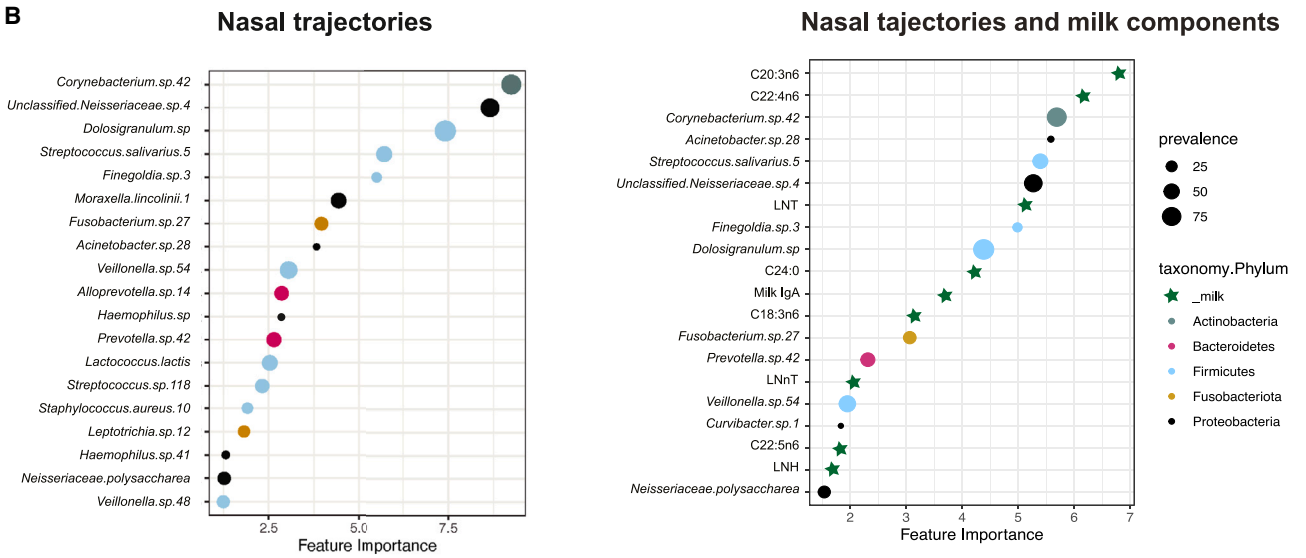
Microbial colonization is a key mechanism mediating the protective effect of breastfeeding against asthma

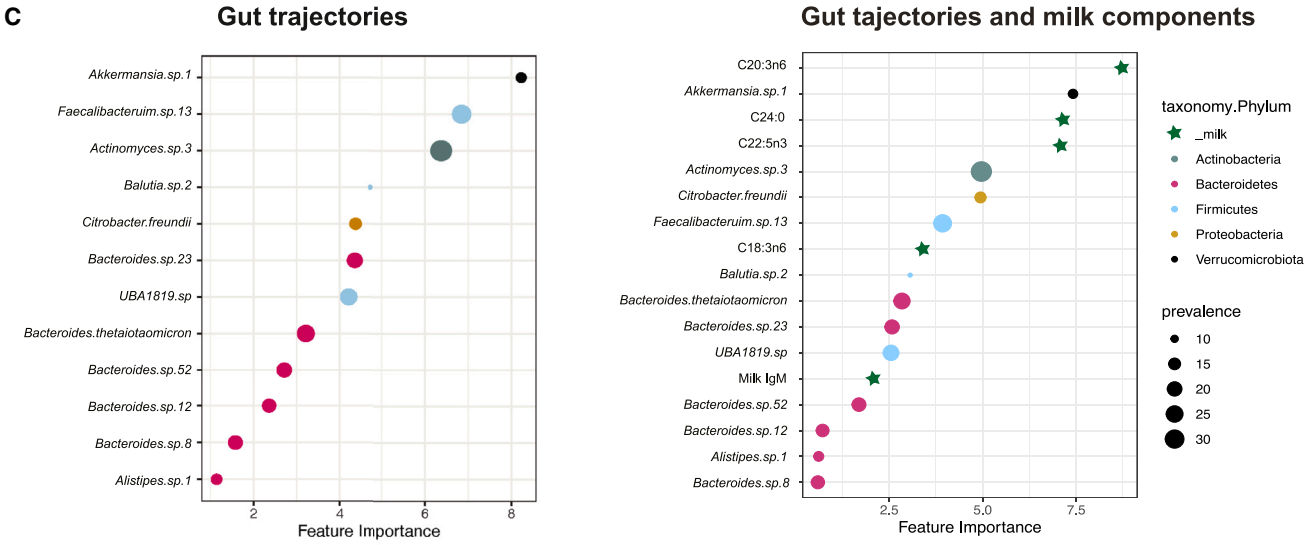
We have shown thus far that nasal and gut colonization patterns are affected by breastfeeding and are strong predictors of preschool asthma (Figures 2, 3, 4, 5, and 6). We further confirmed that these findings hold true after adjusting for potential confounders and used statistical methods to evaluate causality and mediation (Figure 7).

Next, we examined whether delayed microbial colonization patterns underlie the protective effect of breastfeeding against asthma later in life, using a multivariate mediation analysis and structural equation modeling (SEM) among infants who had both gut and nasal longitudinal data available ($n = 341$; 34

(D) Differences in SCFAs (butyrate, propionate, valerate, and acetic acid) between exclusively breastfed infants ($n = 69$) and those no longer breastfed at 3-month sampling ($n = 134$), suggesting an early weaning reaction in the latter group; tested using Mann-Whitney U test. Also showing change in SCFAs between 3 months and 1 year within exclusively breastfed infants ($n = 63$) and those no longer breastfed at 3 months ($n = 112$); tested using Wilcoxon signed-rank test for infants with data at both 3 months and 1 year. *$p_{BH} < 0.05$; **$p_{BH} < 0.001$.
See also Figure S8.







(legend on next page)

with asthma at 3 years, 307 healthy infants). For these analyses, indicators of the infants' nasal and gut microbiota colonization that were significantly associated with both breastfeeding and asthma in univariate analyses (PreTCO system for taxa and Mann-Whitney U test for summary metrics) were selected and modeled as latent variables (Figure S6A). In our final model, adjusted for potential confounders, we found that exclusive breastfeeding at 3 months was directly correlated with delayed (later) colonization of nasal (beta = 0.175, $p = 0.003$) and gut (beta = 0.224, $p < 0.001$) microbiota (Figure 7B). Specifically, exclusive breastfeeding was correlated with a greater increase in microbiota richness and diversity from 3 months to 1 year, a higher frequency of late colonizers, and delayed colonization of the 17 nasal taxa and 10 gut taxa included in the model. These specific nasal and gut microbiota trajectories were also correlated with a lower incidence of asthma at 3 years (nasal, beta = −0.284, $p < 0.001$; gut, beta = −0.274, $p < 0.001$).

Importantly, we found that the protective association of breastfeeding exclusivity with reduced asthma at 3 years was almost fully mediated by microbiota trajectories (82.8% mediation) with similar contributions from the nasal and gut microbiota (nasal: 37.3%, $p = 0.019$; gut: 45.5%, $p = 0.013$; Figure 7B; Table S6). Notably, there was no remaining direct effect of breastfeeding exclusivity on asthma in this mediation model ($p = 0.797$; Figure 7B; Table S6), indicating that regulation of microbiome development is a primary mechanism through which breastfeeding limits asthma development.

Finally, we performed a sensitivity analysis including infants with either nasal or gut samples (allowing for larger sample sizes) and found consistent results (Figure S9; Table S6). All models, including the combined nasal-gut model, had good model fit as indicated by a comparative fit index (CFI) > 0.9, root-mean-square error of approximation (RMSEA) < 0.05, and standardized root-mean residuals (SRMRs) < 0.08 (Table S7). Collectively, these analyses consistently show that, independent of the many covariates considered, nasal and gut microbiota trajectories, when considered alone or in combination, mediate the protective association of breastfeeding in early life on asthma at 3 years of age.

DISCUSSION

The associations of breastfeeding with microbiome composition and respiratory health have been extensively studied. However, they are rarely examined together, and research mostly focuses on gut microbiota and typically stops short of identifying mechanisms underlying these associations. In this work involving over 2,000 children, we provide evidence that the regulation of nasal and gut microbiome development serves as a key mechanism through which breastfeeding mitigates the development of asthma. We find that infants who undergo early cessation of breastfeeding (before 3 months), relying solely on formula feeding, experience an accelerated acquisition of microbial species, functions, and microbial-derived metabolites. These changes mirror the weaning response typically observed around 6 months of age during the natural shift from exclusive breastfeeding to solid food introduction but are happening prematurely at a much younger age. Remarkably, similar patterns of premature acquisition of these microbial species and functions are also observed in infants who later develop preschool asthma and appear to causally mediate the relationship between early weaning and respiratory health. Specifically in the gut microbiota, we observe an accelerated gain of microbial species, including *R. gnavus*, alongside functions and metabolites related to tryptophan biosynthesis, which have been previously linked to immune modulation and asthma.[59–61] Finally, we identified human milk components that enhanced our microbiota-based prediction models for asthma—including IgA, which is known to regulate microbial colonization and support microbial diversity.[62]

A critical time window for breastfeeding cessation

In mice, the transition from nursing to solid foods is characterized by a vigorous immune response involving a surge in microbial colonization and microbial production of SCFAs, which are essential for the development of gut homeostasis and robust immune maturation, affecting GTPase-activating protein (GAP) opening and immune programming.[57,58] This phenomenon, termed the weaning reaction, has not yet been rigorously demonstrated in human infants and may occur differently, given the inter-species differences in microbial taxa and SCFA levels at various stages of early life. Here, we examined whether a similar microbial metabolic response occurs in human infants when breastfeeding is ceased early and replaced with infant formula before 3 months of age. Our findings suggest that the weaning response may not solely be induced by the introduction of solid foods but also by early discontinuation of breastfeeding and exclusive reliance on formula feeding. In support of this, our metagenomics and metabolomics analyses demonstrate that the changes in microbial colonization and functional capacities identified in infants with early breastfeeding cessation coincide with increased levels of SCFAs, similar to the programmed weaning reaction identified in mice.

Recent research has demonstrated that *postponing* the weaning reaction is linked to an elevated risk of allergic inflammation

Figure 6. High prediction accuracy for asthma at 3 years based on nasal and gut microbiome colonization patterns and human milk components

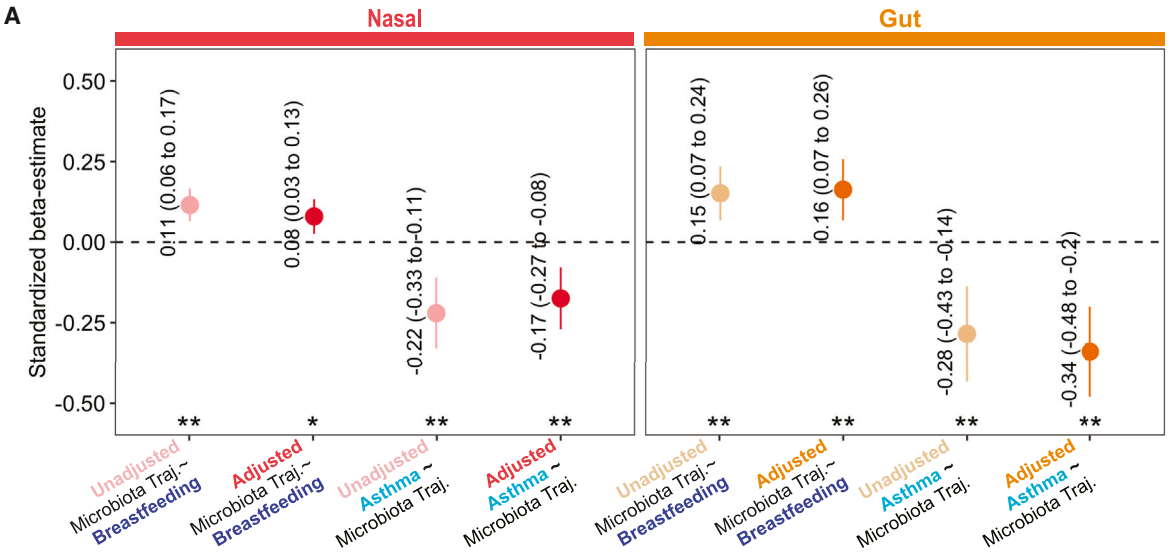
A machine learning model, gradient-boosted decision trees, trained to differentiate between children who were diagnosed with asthma at 3 years ($N = 80$ for nasal and 56 for gut) and healthy controls ($N = 1,327$ for nasal and 479 for gut). Prediction accuracy was evaluated by using held-out samples unused during training.

(A) Receiver operating characteristic curves using different combinations of predictor variables as shown (see Table S5 for true vs. predicted values).

(B) Feature importance plots for the "nasal microbiome trajectories" (left) and "nasal trajectories with milk components" (right) models of asthma prediction (green curves in A), colored by phylum; circle size corresponds to the prevalence of this feature in the data.

(C) Feature importance plots for the "gut microbiome trajectories" (left) and "gut trajectories with milk components" (right) models of asthma prediction (red curves in A), colored by phylum; circle size corresponds to the prevalence of this feature in the data.

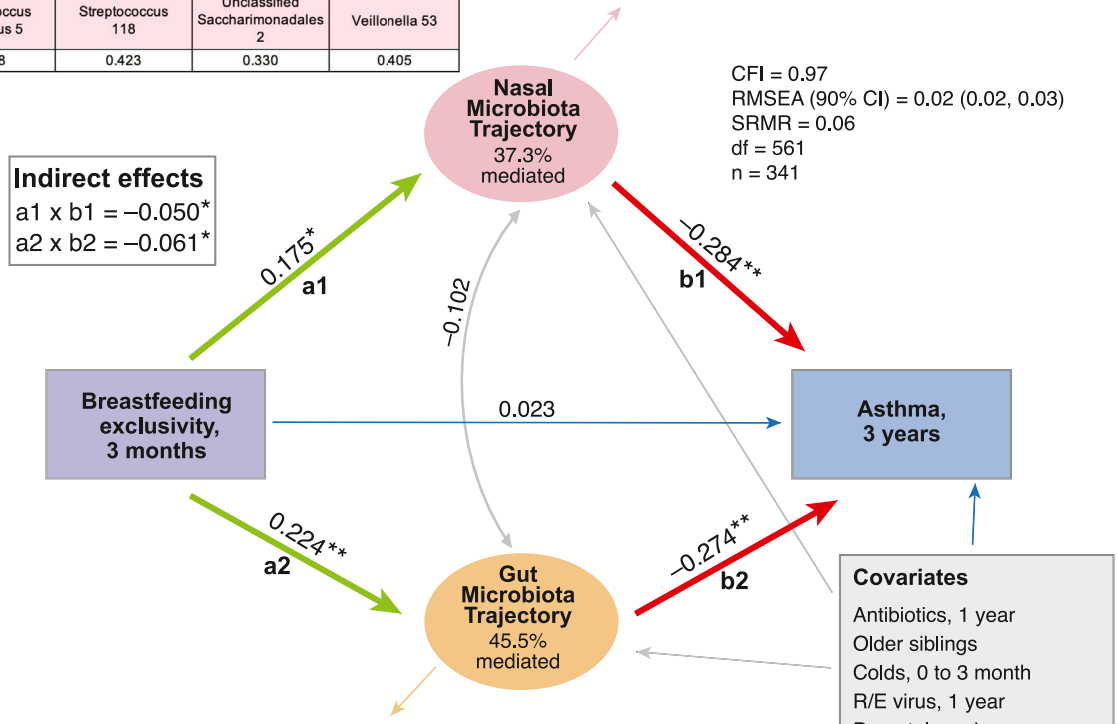
See also Table S5.



B

Nasal Microbiota Trajectory variable indicators - Standardized estimates

Richness Trajectory	Diversity Trajectory	Late Colonizers	Alloprevotella 14	Alloprevotella 25	Campylobacter 3	Fusobacterium 27	Lactobacillu delbrueckii 1
0.911	0.612	0.972	0.418	0.261	0.391	0.488	0.127
Lactococcus lactis	Leptotrichia 12	Neisseria polysaccharea	Pantoea 4	Porphyromonas 19	Prevotella 42	Rothia 12	Streptobacillus 1
0.242	0.328	0.485	0.327	0.497	0.486	0.326	0.349
Streptococcus salivarius 5	Streptococcus 118	Unclassified Saccharimonadales 2	Veillonella 53				
0.328	0.423	0.330	0.405				



Gut Microbiota Trajectory variable indicators - Standardized estimates

Richness Trajectory	Diversity Trajectory	Late Colonizers	Agathobacter 6	Blautia 9	Clostridioides difficile	Eubacterium eligens 1
0.972	0.894	0.910	0.571	0.336	0.003	0.413
Faecalibacterium 13	Flavonifractor	Monoglobus	Ruminococcus gnavus	Sellimonas intestinalis	Tyzzerella 3	
0.623	0.596	0.679	0.485	0.462	0.457	

(legend on next page)

and other immune pathologies.[40] Here, we reveal that an excessively *early* weaning reaction, triggered by the premature discontinuation of breastfeeding, may also heighten the risk of asthma. Thus, together with existing evidence, our new results suggest that weaning too early or too late can be equally detrimental to microbiome and immune development.[63] Further, our findings emphasize the importance of longitudinal analyses during dynamic periods of microbiome development, as our trajectory metrics (change over time and early/late classifications) were far more predictive of asthma than any cross-sectional measure. These results call for additional research with more frequent longitudinal measures of infant feeding and microbiome function throughout the weaning process to precisely define the optimal timing for this important process. Such research could inform potential updates to global recommendations for breastfeeding and weaning,[64] which are largely based on decades-old evidence that pre-date the modern era of microbiome research.

Premature *R. gnavus* acquisition links early breastfeeding cessation to asthma development

By examining the pangenome of the gut microbial communities in these infants, we identified members of *R. gnavus* as key time-dependent commensals—meaning that their beneficial role depends on the timing of their acquisition. Later colonization by members of *R. gnavus* (after 3 months) was observed in exclusively breastfed infants and those who did not develop asthma. This aligns with previous research showing that early *R. gnavus* colonization of the infant gut at 2 months of age is associated with allergic disease,[65] yet its presence at 6 months, alongside other mucin-degrading species, has been associated with reduced risk of allergic disease.[66] Evidence in mice suggests the mechanism for the protective effect of *R. gnavus* is linked to modulation of the T helper 2 (Th2) immune response via SCFA-stimulated signaling and Treg differentiation.[67]

Mucin-degrading species such as *R. gnavus* use alpha-galactosidase to break down non-digestible galacto-oligosaccharides, abundant in many non-human-milk dietary sources, and host mucin. The ability to ferment these oligosaccharides into SCFAs enables species like *R. gnavus* to establish themselves in the infant gut during the introduction of non-human-milk food sources.[68,69] Indeed, we found that the alpha-galactosidase enzyme in *R. gnavus* was introduced early (before 3 months) in infants who underwent early cessation of breastfeeding, as well as those who developed asthma, highlighting the potential importance of appropriately timed increased SCFA production during the critical weaning period.

Another enzyme of *R. gnavus* members that was introduced later with exclusive breastfeeding and non-asthma was AnPRT, which is involved in tryptophan biosynthesis.[70] We further found "prematurely elevated" levels of tryptophan and its metabolites in infants no longer breastfed at 3 months and those who develop asthma. Tryptophan metabolites have known associations with asthma,[71] and D-tryptophan produced by probiotic bacteria has been associated with gut microbial diversity and explored as a preventative pharmaceutical against allergic diseases including asthma.[72] Moreover, tryptophan metabolites including indole and tryptamine, produced by microbes residing in the gut,[61] are known to activate the aryl hydrocarbon receptor (AhR) that has emerged as a crucial player in asthma control, responding to environmental molecules and endogenous or dietary metabolites while regulating innate and adaptive immune responses.[73,74] Together, our findings suggest that exclusive breastfeeding limits colonization by members of *R. gnavus* until non-human-milk food sources are introduced, at which time these microbes can flourish and begin producing SCFAs and take part in tryptophan biosynthesis, which in turn stimulate Treg differentiation and activate the AhR. Disrupting the timing or sequence of these events leads to immune perturbation and predisposition to asthma.

Notably, the delayed acquisition of functions attributable to *R. gnavus* did not perfectly parallel the acquisition of *R. gnavus* species colonization, suggesting the importance of strain-specific functions. Indeed, previous research shows that the capacity of *R. gnavus* to metabolize mucin glycans and produce propionate is strain dependent.[68] In addition, reference database limitations may have led to over-attribution of certain functions to *R. gnavus* alone, where other microbes could also be contributing (e.g., infant gut microbiota such as *Escherichia coli*, *Clostridium sporogenes*, and *Lactobacillus* spp. may also contribute to tryptophan biosynthesis).[75]

Nasal (in addition to gut) microbiota is impacted by early breastfeeding cessation and linked to asthma

Previous research has focused on the weaning reaction phenomenon only in the context of the gut microbiota; however, we

Figure 7. Nasal and gut colonization patterns are associated with breastfeeding exclusivity and asthma in multivariate models and mediate the association between breastfeeding exclusivity and asthma

(A) Standardized beta-estimates with 95% confidence intervals for unadjusted and adjusted regression models for associations between a nasal or gut microbiota trajectory latent variable and (1) breastfeeding (nasal: exclusive $n = 794$, partial $n = 517$, no BM =234; gut: exclusive $n = 271$, partial = 205, no BM = 79), and (2) asthma (nasal: healthy $n = 1,236$, asthma $n = 80$; gut: healthy $n = 421$, asthma $n = 56$). The following potential confounders were included in adjusted models: antibiotics in the first year of life, older siblings, pre-natal smoke, birth mode, maternal asthma, study center, R/E virus at 1 year, colds in the first 3 months, and infant age at 3-month sampling. Variable selection for latent constructs was informed based on univariate associations with both asthma and breastfeeding (STAR Methods). $^{*}p < 0.05$, $^{**}p < 0.001$.

(B) A structural equation model showing the mediating effects of nasal and gut microbiota trajectories on the association between breastfeeding exclusivity at 3 months and asthma at 3 years. Standardized beta-coefficients are reported, adjusted for all covariates listed, with p values denoted by $^{*}p < 0.05$ and $^{**}p < 0.001$. Positive associations with microbiota trajectories indicate delayed microbiota colonization (later), whereas negative associations indicate earlier colonization. Variable selection for latent constructs is done similarly to (A) (STAR Methods). We included infants with data on both nasal and gut microbiota (34 infants with asthma and 307 infants without asthma diagnosed at 3 years), and breastfeeding exclusivity was an ordered variable (exclusive, partial, or no human milk at 3-month sampling). See Table S6 and Figure S9 for details, including both standardized and unstandardized estimates, and separate nasal and gut microbiota models.

See also Figure S9 and Tables S6 and S7.

hypothesize that a similar reaction to early breastfeeding cessation occurs in other body sites, such as the respiratory tract microbiome. In support of this, we identified similar patterns of an accelerated acquisition of nasal microbiota in infants who underwent early cessation of breastfeeding and those who later developed asthma. In the nasal microbiome, specific taxa that followed this pattern included *S. salivarius*, *Lactobacillus delbrueckii*, and *Fusobacterium*. All were found to be associated with respiratory health and disease in previous studies; however, their order of arrival to the nasal microbial niche was not rigorously examined. Both *Fusobacterium* and *Lactobacillus* have been associated with healthy respiratory phenotypes in previous research.[76,77] *S. salivarius* is particularly intriguing as it is being considered as a bacteriotherapy in the form of a nasal spray for treating recurrent respiratory infections.[78,79] In terms of early colonization, we found that in both the nasal and gut niches, *Staphylococcus* spp. were able to colonize the earlier (3 months) but generally not the more mature community. This corroborates previous research showing that *Staphylococcus* spp. are generally early colonizers in both nasal and gut niches.[30,80] We further found evidence that the early colonization of the nasal niche by specific taxa, such as *Acinetobacter* sp., may protect against asthma. Interestingly, previous research suggests that *Acinetobacter* is an abundant genus in human milk,[81,82] and its abundance in the nasal niche at 6 months of age may be beneficial and linked to the childhood allergy gap between distinct populations.[83]

Overall, our findings indicate that breastfeeding functions as a pacemaker, orchestrating the colonization process of microbial functions and species. Early cessation of breastfeeding seems to accelerate the premature acquisition of microbial species and functions, potentially increasing the risk of respiratory illness. Our robust findings across two body sites suggest a conserved mechanism and highlight the importance of a gradual succession toward a richer microbial community facilitated by breastfeeding. This gradual colonization, eventually leading to an appropriately timed weaning reaction, may be important to prevent an inappropriate response of the untrained immune system to microbiota or microbial compounds (e.g., immunogenic SCFAs), as opposed to the more fine-tuned response of a properly trained immune system later in life. The gradual co-adaptation of the immune system and microbial community is likely critical for the eventual acceptance of a larger diversity of microbial taxa as commensals later in life.

Strengths and limitations

Strengths of this study include our large sample size, use of human milk composition, shotgun metagenomics and metabolomics data, and access to rich longitudinal metadata and clinical asthma diagnoses from the general population of the CHILD cohort. We also had microbiome data from multiple niches at multiple time points, which allowed us to identify temporal patterns that were conserved across niches. Further, a defining strength of this study was our robust analytical approach that included stepwise computational analyses (from associations, to predictions, to informed causal inference), and the PreTCO system to assess colonization patterns at global and local scales. This computational approach utilized prevalence data, which allowed us to overcome challenges of compositionality and sparsity in our microbiome data. A main limitation of this study is its observational design, although we have identified specific microbiota that could be targeted for future mechanistic studies and biotherapeutic applications. Further, we acknowledge that *R. gnavus* may not be the only microbe capable of SCFAs production and tryptophan biosynthesis, and our findings are limited by the existing functional databases. Additionally, our nasal samples were not analyzed using shotgun metagenomic sequencing (owing to the very low biomass of these samples), therefore our machine learning and causal inference results focus on 16S rRNA sequence data partially for comparability between niches. Our human milk analyses were limited to a subset of components (long-chain fatty acids, HMOs, and a few bioactive proteins), and we acknowledge that other components could also be important (e.g., nutrients, metabolites, other proteins, or cells). Finally, we acknowledge that asthma is a complex and heterogeneous disease that is particularly challenging to diagnose in early childhood and that further research is warranted to understand the nature and persistence of our findings beyond 5 years of age.

Conclusions and future directions

We found evidence that microbial colonization patterns, across the nasal and gut niches, mediate the protective effect of breastfeeding on respiratory health. Specifically, we found that a paced colonization of specific microbiota mediated the effect of breastfeeding on preventing asthma later in childhood and implicated the production of specific microbial metabolites. Conversely, infants who undergo early cessation of breastfeeding (before or at 3 months) experience an accelerated acquisition of microbial species, functions, and metabolites. Thus, our results suggest that microbiome-targeted disease prevention strategies should not simply focus on supplementing with specific bacteria but rather aim to support appropriately timed beneficial microbiota community development in early life. Further research is needed to precisely define this critical time window and to develop human-milk-inspired and microbiome-targeted interventions for asthma prevention.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- Microbial colonization patterns
- Prevalence trajectories coordinate system (PreTCO System)
- Prevalence trajectories of microbes and their pangenomes
- Prediction model
- Multivariate analysis
- Structural Equation Models

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.07.022.

ACKNOWLEDGMENTS

We thank the CHILD Cohort Study (CHILD) participant families for their dedication and commitment to advancing health research. CHILD was initially funded by the Canadian Institutes of Health Research (CIHR) and AllerGen NCE. Visit CHILD at childcohort.ca. M.B.A. holds a Tier 2 Canada Research Chair in Early Nutrition and the Developmental Origins of Health and Disease and is a fellow of the CIFAR Humans and the Microbiome program. S.E.T. holds a Tier 1 Canada Research Chair in Pediatric Precision Health and the Aubrey J. Tingle Professor of Pediatric Immunology. L.B. is the UC San Diego Chair of Collaborative Human Milk Research endowed by the Family Larsson-Rosenquist Foundation in Switzerland. L.S. is funded by NIAID of the National Institute of Health under award number 1DP2AI185753. C.P. is funded by the University of British Columbia John Richard Turner Fellowship in Microbiology, the President's Academic Excellence Initiative PhD Award, and the University of British Columbia Four Year Doctoral Fellowship (4YF). D.L.Y.D. is funded by the CIHR Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award (CIHR CGS-D) and the University of British Columbia 4YF. Funding for the shotgun metagenomic and metabolomic gut microbiome profiles was provided by Genome British Columbia (274CHI) to S.E.T. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

AUTHOR CONTRIBUTIONS

L.S., K.F., and M.B.A. conceived the project, with input from all authors. L.S. and K.F. designed and implemented the computational methodology. L.S. and M.B.A. supervised the project. M.E.R., R.D., V.B., and E.S. enabled data access and analysis of the CHILD Cohort data. P.J.M. supervised clinical findings, clinical phenotypes, and CHILD Cohort data access. T.J.M. guided clinical findings. C.P., D.L.Y.D., and S.E.T. generated the gut 16S rRNA gene sequencing, the shotgun metagenomics, and metabolomics data. L.R., M.S., and M.G.S. generated the nasal gut 16S rRNA gene sequencing. L.B. generated HMO data. C.J.F. generated fatty acids data. J.S.M. generated human milk bioactive data. M.A.S. and M.L. generated immunological hypotheses and guided immunological findings. P.J.M. helped generate CHILD Cohort data. L.S., K.F., and M.B.A. drafted the manuscript with input from all authors.

DECLARATION OF INTERESTS

M.B.A. receives research funding from the Canadian and US governments, the Bill and Melinda Gates Foundation, and the Garfield Weston Foundation. She holds a Canada Research Chair in Early Nutrition and the Developmental Origins of Health and Disease and is a fellow of the CIFAR Humans and the Microbiome program. She has consulted for DSM Nutritional Products, serves on the scientific advisory board for TinyHealth, and has received speaking honoraria from Prolacta Biosciences. She has contributed without remuneration to online courses on breast milk and the infant microbiome produced by Microbiome Courses. L.B. receives research funding from the United States National Institutes of Health and the Bill and Melinda Gates Foundation. L.B. is the UC San Diego Chair of Collaborative Human Milk Research endowed by the Family Larsson-Rosenquist Foundation in Switzerland. V.B. is currently an employee of F. Hoffman-La Roche Ltd.; however, the published work was done prior to her employment and does not involve/promote any of Roche's materials or point of view.

Received: March 17, 2023
Revised: March 2, 2024
Accepted: July 12, 2024
Published: September 19, 2024

REFERENCES

1. Abrams, G.D., Bauer, H., and Sprinz, H. (1963). Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. Lab. Invest. *12*, 355–364.
2. Bouskra, D., Brézillon, C., Bérard, M., Werts, C., Varona, R., Boneca, I.G., and Eberl, G. (2008). Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. Nature *456*, 507–510. https://doi.org/10.1038/nature07450.
3. Charbonneau, M.R., Blanton, L.V., DiGiulio, D.B., Relman, D.A., Lebrilla, C.B., Mills, D.A., and Gordon, J.I. (2016). A microbial perspective of human developmental biology. Nature *535*, 48–55. https://doi.org/10.1038/nature18845.
4. Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., and Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature *591*, 633–638. https://doi.org/10.1038/s41586-021-03241-8.
5. Derrien, M., Alvarez, A.-S., and de Vos, W.M. (2019). The gut microbiota in the first decade of life. Trends Microbiol. *27*, 997–1010. https://doi.org/10.1016/j.tim.2019.08.001.
6. Stewart, C.J., Ajami, N.J., O'Brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross, M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., et al. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature *562*, 583–588. https://doi.org/10.1038/s41586-018-0617-x.
7. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227. https://doi.org/10.1038/nature11053.
8. Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe *17*, 690–703. https://doi.org/10.1016/j.chom.2015.04.004.
9. Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. *21*, 1228–1234. https://doi.org/10.1038/nm.3950.
10. Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. PLoS Biol. *5*, e177. https://doi.org/10.1371/journal.pbio.0050177.
11. Vatanen, T., Plichta, D.R., Somani, J., Münch, P.C., Arthur, T.D., Hall, A.B., Rudolf, S., Oakeley, E.J., Ke, X., Young, R.A., et al. (2019). Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. Nat. Microbiol. *4*, 470–479. https://doi.org/10.1038/s41564-018-0321-5.
12. Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D.T., Manara, S., Zolfo, M., et al. (2018). Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. Cell Host Microbe *24*, 133–145.e5. https://doi.org/10.1016/j.chom.2018.06.005.
13. Yassour, M., Jason, E., Hogstrom, L.J., Arthur, T.D., Tripathi, S., Siljander, H., Selvenius, J., Oikarinen, S., Hyöty, H., Virtanen, S.M., et al. (2018). Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. Cell Host Microbe *24*, 146–154.e4. https://doi.org/10.1016/j.chom.2018.06.007.

14. Moore, R.E., and Townsend, S.D. (2019). Temporal development of the infant gut microbiome. Open Biol. *9*, 190128. https://doi.org/10.1098/rsob.190128.

15. Liu, C.M., Price, L.B., Hungate, B.A., Abraham, A.G., Larsen, L.A., Christensen, K., Stegger, M., Skov, R., and Andersen, P.S. (2015). Staphylococcus aureus and the ecology of the nasal microbiome. Sci. Adv. *1*, e1400216. https://doi.org/10.1126/sciadv.1400216.

16. Man, W.H., de Steenhuijsen Piters, W.A.A., and Bogaert, D. (2017). The microbiota of the respiratory tract: gatekeeper to respiratory health. Nat. Rev. Microbiol. *15*, 259–270. https://doi.org/10.1038/nrmicro.2017.14.

17. van Nimwegen, F.A., Penders, J., Stobberingh, E.E., Postma, D.S., Koppelman, G.H., Kerkhof, M., Reijmerink, N.E., Dompeling, E., van den Brandt, P.A., Ferreira, I., et al. (2011). Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. J. Allergy Clin. Immunol. *128*, 948–955.e1. https://doi.org/10.1016/j.jaci.2011.07.027.

18. Arrieta, M.-C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., et al. (2015). Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. *7*, 307ra152. https://doi.org/10.1126/scitranslmed.aab2271.

19. Thorsen, J., Rasmussen, M.A., Waage, J., Mortensen, M., Brejnrod, A., Bønnelykke, K., Chawes, B.L., Brix, S., Sørensen, S.J., Stokholm, J., et al. (2019). Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. Nat. Commun. *10*, 5001. https://doi.org/10.1038/s41467-019-12989-7.

20. Johnson, C.C., and Ownby, D.R. (2017). The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. Transl. Res. *179*, 60–70. https://doi.org/10.1016/j.trsl.2016.06.010.

21. Stokholm, J., Blaser, M.J., Thorsen, J., Rasmussen, M.A., Waage, J., Vinding, R.K., Schoos, A.-M.M., Kunøe, A., Fink, N.R., Chawes, B.L., et al. (2018). Publisher Correction: maturation of the gut microbiome and risk of asthma in childhood. Nat. Commun. *9*, 704. https://doi.org/10.1038/s41467-018-03150-x.

22. Azad, M.B., and Kozyrskyj, A.L. (2012). Perinatal programming of asthma: the role of gut microbiota. Clin. Dev. Immunol. *2012*, 932072. https://doi.org/10.1155/2012/932072.

23. Sordillo, J.E., Zhou, Y., McGeachie, M.J., Ziniti, J., Lange, N., Laranjo, N., Savage, J.R., Carey, V., O'Connor, G., Sandel, M., et al. (2017). Factors influencing the infant gut microbiome at age 3–6 months: findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). J. Allergy Clin. Immunol. *139*, 482–491.e14. https://doi.org/10.1016/j.jaci.2016.08.045.

24. Gray, L.E.K., O'Hely, M., Ranganathan, S., Sly, P.D., and Vuillermin, P. (2017). The maternal diet, gut bacteria, and bacterial metabolites during pregnancy influence offspring asthma. Front. Immunol. *8*, 365. https://doi.org/10.3389/fimmu.2017.00365.

25. Koleva, P.T., Tun, H.M., Konya, T., Guttman, D.S., Becker, A.B., Mandhane, P.J., Turvey, S.E., Subbarao, P., Sears, M.R., Scott, J.A., et al. (2017). Sex-specific impact of asthma during pregnancy on infant gut microbiota. Eur. Respir. J. *50*, 1700280. https://doi.org/10.1183/13993003.00280-2017.

26. Douros, K., Moustaki, M., Tsabouri, S., Papadopoulou, A., Papadopoulos, M., and Priftis, K.N. (2017). Prenatal maternal stress and the risk of asthma in children. Front. Pediatr. *5*, 202. https://doi.org/10.3389/fped.2017.00202.

27. Depner, M., Taft, D.H., Kirjavainen, P.V., Kalanetra, K.M., Karvonen, A.M., Peschel, S., Schmausser-Hechfellner, E., Roduit, C., Frei, R., Lauener, R., et al. (2020). Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. Nat. Med. *26*, 1766–1775. https://doi.org/10.1038/s41591-020-1095-x.

28. Zeng, Y., and Liang, J.Q. (2022). Nasal microbiome and its interaction with the host in childhood asthma. Cells *11*, 3155. https://doi.org/10.3390/cells11193155.

29. Biesbroek, G., Tsivtsivadze, E., Sanders, E.A.M., Montijn, R., Veenhoven, R.H., Keijser, B.J.F., and Bogaert, D. (2014). Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. Am. J. Respir. Crit. Care Med. *190*, 1283–1292. https://doi.org/10.1164/rccm.201407-1240OC.

30. de Steenhuijsen Piters, W.A.A., Binkowska, J., and Bogaert, D. (2020). Early life microbiota and respiratory tract infections. Cell Host Microbe *28*, 223–232. https://doi.org/10.1016/j.chom.2020.07.004.

31. Tang, H.H.F., Lang, A., Teo, S.M., Judd, L.M., Gangnon, R., Evans, M.D., Lee, K.E., Vrtis, R., Holt, P.G., Lemanske, R.F., Jr., et al. (2021). Developmental patterns in the nasopharyngeal microbiome during infancy are associated with asthma risk. J. Allergy Clin. Immunol. *147*, 1683–1691. https://doi.org/10.1016/j.jaci.2020.10.009.

32. Pannaraj, P.S., Li, F., Cerini, C., Bender, J.M., Yang, S., Rollie, A., Adisetiyo, H., Zabih, S., Lincez, P.J., Bittinger, K., et al. (2017). Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. JAMA Pediatr. *171*, 647–654. https://doi.org/10.1001/jamapediatrics.2017.0378.

33. Forbes, J.D., Azad, M.B., Vehling, L., Tun, H.M., Konya, T.B., Guttman, D.S., Field, C.J., Lefebvre, D., Sears, M.R., Becker, A.B., et al. (2018). Association of exposure to formula in the hospital and subsequent infant feeding practices with gut microbiota and risk of overweight in the first year of life. JAMA Pediatr. *172*, e181161. https://doi.org/10.1001/jamapediatrics.2018.1161.

34. Fehr, K., Moossavi, S., Sbihi, H., Boutin, R.C.T., Bode, L., Robertson, B., Yonemitsu, C., Field, C.J., Becker, A.B., Mandhane, P.J., et al. (2020). Breastmilk feeding practices are associated with the co-occurrence of bacteria in mothers' milk and the infant gut: the CHILD cohort study. Cell Host Microbe *28*, 285–297.e4. https://doi.org/10.1016/j.chom.2020.06.009.

35. Ho, N.T., Li, F., Lee-Sarwar, K.A., Tun, H.M., Brown, B.P., Pannaraj, P.S., Bender, J.M., Azad, M.B., Thompson, A.L., Weiss, S.T., et al. (2018). Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat. Commun. *9*, 4169. https://doi.org/10.1038/s41467-018-06473-x.

36. Biesbroek, G., Bosch, A.A.T.M., Wang, X., Keijser, B.J.F., Veenhoven, R.H., Sanders, E.A.M., and Bogaert, D. (2014). The impact of breastfeeding on nasopharyngeal microbial communities in infants. Am. J. Respir. Crit. Care Med. *190*, 298–308. https://doi.org/10.1164/rccm.201401-0073OC.

37. Mika, M., Mack, I., Korten, I., Qi, W., Aebi, S., Frey, U., Latzin, P., and Hilty, M. (2015). Dynamics of the nasal microbiota in infancy: a prospective cohort study. J. Allergy Clin. Immunol. *135*, 905–912.e11. https://doi.org/10.1016/j.jaci.2014.12.1909.

38. Lawson, M.A.E., O'Neill, I.J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z., Chalklen, L., and Hall, L.J. (2020). Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. ISME J. *14*, 635–648. https://doi.org/10.1038/s41396-019-0553-2.

39. Bridgman, S.L., Konya, T., Azad, M.B., Sears, M.R., Becker, A.B., Turvey, S.E., Mandhane, P.J., Subbarao, P., CHILD Study Investigators, and Scott, J.A., et al. (2016). Infant gut immunity: a preliminary study of IgA associations with breastfeeding. J. Dev. Orig. Health Dis. *7*, 68–72. https://doi.org/10.1017/S2040174415007862.

40. Lubin, J.-B., Green, J., Maddux, S., Denu, L., Duranova, T., Lanza, M., Wynosky-Dolfi, M., Flores, J.N., Grimes, L.P., Brodsky, I.E., et al. (2023). Arresting microbiome development limits immune system maturation and resistance to infection in mice. Cell Host Microbe *31*, 554–570.e7. https://doi.org/10.1016/j.chom.2023.03.006.

41. OECD; The World Bank (2020). Infant and young child feeding. Preprint at Organisation for Economic Co-Operation and Development (OECD). https://doi.org/10.1787/67fe62a3-en.

42. Victora, C.G., Bahl, R., Barros, A.J.D., França, G.V.A., Horton, S., Krasevec, J., Murch, S., Sankar, M.J., Walker, N., Rollins, N.C., et al. (2016). Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet *387*, 475–490. https://doi.org/10.1016/S0140-6736(15)01024-7.

43. Miliku, K., and Azad, M.B. (2018). Breastfeeding and the developmental origins of asthma: current evidence, possible mechanisms, and future research priorities. Nutrients *10*, 995. https://doi.org/10.3390/nu10080995.

44. Xue, M., Dehaas, E., Chaudhary, N., O'Byrne, P., Satia, I., and Kurmi, O.P. (2021). Breastfeeding and risk of childhood asthma: a systematic review and meta-analysis. ERJ Open Res. *7*, 00504-2021. https://doi.org/10.1183/23120541.00504-2021.

45. Dell, S., and To, T. (2001). Breastfeeding and asthma in young children: findings from a population-based study. Arch. Pediatr. Adolesc. Med. *155*, 1261–1265. https://doi.org/10.1001/archpedi.155.11.1261.

46. Kull, I., Almqvist, C., Lilja, G., Pershagen, G., and Wickman, M. (2004). Breast-feeding reduces the risk of asthma during the first 4 years of life. J. Allergy Clin. Immunol. *114*, 755–760. https://doi.org/10.1016/j.jaci.2004.07.036.

47. Oddy, W.H., Holt, P.G., Sly, P.D., Read, A.W., Landau, L.I., Stanley, F.J., Kendall, G.E., and Burton, P.R. (1999). Association between breast feeding and asthma in 6 year old children: findings of a prospective birth cohort study. BMJ *319*, 815–819. https://doi.org/10.1136/bmj.319.7213.815.

48. Azad, M.B., Vehling, L., Lu, Z., Dai, D., Subbarao, P., Becker, A.B., Mandhane, P.J., Turvey, S.E., Lefebvre, D.L., Sears, M.R., et al. (2017). Breastfeeding, maternal asthma and wheezing in the first year of life: a longitudinal birth cohort study. Eur. Respir. J. *49*, 1602019. https://doi.org/10.1183/13993003.02019-2016.

49. Dai, D.L.Y., Petersen, C., Hoskinson, C., Del Bel, K.L., Becker, A.B., Moraes, T.J., Mandhane, P.J., Finlay, B.B., Simons, E., Kozyrskyj, A.L., et al. (2023). Breastfeeding enrichment of B. longum subsp. infantis mitigates the effect of antibiotics on the microbiota and childhood asthma risk. Med. *4*, 92–112.e5. https://doi.org/10.1016/j.medj.2022.12.002.

50. Klopp, A., Vehling, L., Becker, A.B., Subbarao, P., Mandhane, P.J., Turvey, S.E., Lefebvre, D.L., Sears, M.R., and CHILD Study Investigators, and Azad, M.B. (2017). Infant gut immunity: a preliminary study of IgA associations with breastfeeding. J. Dev. Orig. Health Dis. *190*, 192–199.e2.

51. Wilson, K., Gebretsadik, T., Adgent, M.A., Loftus, C., Karr, C., Moore, P.E., Sathyanarayana, S., Byington, N., Barrett, E., Bush, N., et al. (2022). The association between duration of breastfeeding and childhood asthma outcomes. Ann. Allergy Asthma Immunol. *129*, 205–211. https://doi.org/10.1016/j.anai.2022.04.034.

52. Moossavi, S., Sepehri, S., Robertson, B., Bode, L., Goruk, S., Field, C.J., Lix, L.M., de Souza, R.J., Becker, A.B., Mandhane, P.J., et al. (2019). Composition and variation of the human milk microbiota are influenced by maternal and early-life factors. Cell Host Microbe *25*, 324–335.e4. https://doi.org/10.1016/j.chom.2019.01.011.

53. Patrick, D.M., Sbihi, H., Dai, D.L.Y., Al Mamun, A., Rasali, D., Rose, C., Marra, F., Boutin, R.C.T., Petersen, C., Stiemsma, L.T., et al. (2020). Decreasing antibiotic use, the gut microbiota, and asthma incidence in children: evidence from population-based and prospective cohort studies. Lancet Respir. Med. *8*, 1094–1105. https://doi.org/10.1016/S2213-2600(20)30052-7.

54. Houghteling, P.D., and Walker, W.A. (2015). Why is initial bacterial colonization of the intestine important to infants' and children's health? J. Pediatr. Gastroenterol. Nutr. *60*, 294–307. https://doi.org/10.1097/MPG.0000000000000597.

55. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

56. Radwanski, E.R., and Last, R.L. (1995). Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. Plant Cell *7*, 921–934. https://doi.org/10.1105/tpc.7.7.921.

57. Al Nabhani, Z., Dulauroy, S., Marques, R., Cousu, C., Al Bounny, S., Déjardin, F., Sparwasser, T., Bérard, M., Cerf-Bensussan, N., and Eberl, G. (2019). A weaning reaction to microbiota is required for resistance to immunopathologies in the adult. Immunity *50*, 1276–1288.e5. https://doi.org/10.1016/j.immuni.2019.02.014.

58. Mann, E.R., Lam, Y.K., and Uhlig, H.H. (2024). Short-chain fatty acids: linking diet, the microbiome and immunity. Nat. Rev. Immunol. https://doi.org/10.1038/s41577-024-01014-8.

59. Bessede, A., Gargaro, M., Pallotta, M.T., Matino, D., Servillo, G., Brunacci, C., Bicciato, S., Mazza, E.M.C., Macchiarulo, A., Vacca, C., et al. (2014). Aryl hydrocarbon receptor control of a disease tolerance defence pathway. Nature *511*, 184–190. https://doi.org/10.1038/nature13323.

60. Zelante, T., Iannitti, R.G., Cunha, C., De Luca, A., Giovannini, G., Pieraccini, G., Zecchi, R., D'Angelo, C., Massi-Benedetti, C., Fallarino, F., et al. (2013). Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity *39*, 372–385. https://doi.org/10.1016/j.immuni.2013.08.003.

61. Gao, J., Xu, K., Liu, H., Liu, G., Bai, M., Peng, C., Li, T., and Yin, Y. (2018). Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism. Front. Cell. Infect. Microbiol. *8*, 13. https://doi.org/10.3389/fcimb.2018.00013.

62. Huus, K.E., Petersen, C., and Finlay, B.B. (2021). Diversity and dynamism of IgA-microbiota interactions. Nat. Rev. Immunol. *21*, 514–525. https://doi.org/10.1038/s41577-021-00506-1.

63. Chen, Y.-C., Chen, Y., Lasky-Su, J., Kelly, R.S., Stokholm, J., Bisgaard, H., Bønnelykke, K., Pedersen, C.-E.T., Chawes, B., Laranjo, N., et al. (2023). Environmental and genetic associations with aberrant early-life gut microbial maturation in childhood asthma. J. Allergy Clin. Immunol. *151*, 1494–1502.e14.

64. Meek, J.Y., and Noble, L.; Section on Breastfeeding (2022). Policy Statement: Breastfeeding and the Use of Human Milk. Pediatrics *150*, e2022057988. https://doi.org/10.1542/peds.2022-057988.

65. Chua, H.-H., Chou, H.-C., Tung, Y.-L., Chiang, B.-L., Liao, C.-C., Liu, H.-H., and Ni, Y.-H. (2018). Intestinal dysbiosis featuring abundance of Ruminococcus gnavus associates with allergic diseases in infants. Gastroenterology *154*, 154–167. https://doi.org/10.1053/j.gastro.2017.09.006.

66. Lee, M.-J., Kang, M.-J., Lee, S.-Y., Lee, E., Kim, K., Won, S., Suh, D.I., Kim, K.W., Sheen, Y.H., Ahn, K., et al. (2018). Perturbations of gut microbiome genes in infants with atopic dermatitis according to feeding type. J. Allergy Clin. Immunol. *141*, 1310–1319. https://doi.org/10.1016/j.jaci.2017.11.045.

67. Ahn, J.-R., Lee, S.-H., Kim, B., Nam, M.H., Ahn, Y.K., Park, Y.M., Jeong, S.-M., Park, M.J., Song, K.B., Lee, S.-Y., et al. (2022). Ruminococcus gnavus ameliorates atopic dermatitis by enhancing Treg cell and metabolites in BALB/c mice. Pediatr. Allergy Immunol. *33*, e13678. https://doi.org/10.1111/pai.13678.

68. Crost, E.H., Coletto, E., Bell, A., and Juge, N. (2023). Ruminococcus gnavus: friend or foe for human health. FEMS Microbiol. Rev. *47*, fuad014. https://doi.org/10.1093/femsre/fuad014.

69. Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., Gil, A., Vieites, J.M., Norin, E., Young, D., et al. (2011). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. Microbiology *157*, 1385–1392. https://doi.org/10.1099/mic.0.042143-0.

70. Scully, T.W., Jiao, W., Mittelstädt, G., and Parker, E.J. (2023). Structure, mechanism and inhibition of anthranilate phosphoribosyltransferase. Philos. Trans. R. Soc. Lond. B Biol. Sci. *378*, 20220039. https://doi.org/10.1098/rstb.2022.0039.

71. Lee-Sarwar, K.A., Lasky-Su, J., Kelly, R.S., Litonjua, A.A., and Weiss, S.T. (2020). Gut microbial-derived metabolomics of asthma. Metabolites *10*, 97. https://doi.org/10.3390/metabo10030097.

72. Kepert, I., Fonseca, J., Müller, C., Milger, K., Hochwind, K., Kostric, M., Fedoseeva, M., Ohnmacht, C., Dehmel, S., Nathan, P., et al. (2017). D-tryptophan from probiotic bacteria influences the gut microbiome and allergic airway disease. J. Allergy Clin. Immunol. *139*, 1525–1535. https://doi.org/10.1016/j.jaci.2016.09.003.

73. Cheng, Y., Jin, U.-H., Allred, C.D., Jayaraman, A., Chapkin, R.S., and Safe, S. (2015). Aryl hydrocarbon receptor activity of tryptophan metabolites in young adult mouse colonocytes. Drug Metab. Dispos. *43*, 1536–1543. https://doi.org/10.1124/dmd.115.063677.

74. Poulain-Godefroy, O., Bouté, M., Carrard, J., Alvarez-Simon, D., Tsicopoulos, A., and de Nadai, P. (2020). The aryl hydrocarbon receptor in asthma: friend or foe? Int. J. Mol. Sci. *21*, 8797. https://doi.org/10.3390/ijms21228797.

75. Agus, A., Planchais, J., and Sokol, H. (2018). Gut microbiota regulation of tryptophan metabolism in health and disease. Cell Host Microbe *23*, 716–724. https://doi.org/10.1016/j.chom.2018.05.003.

76. Rosas-Salazar, C., Shilts, M.H., Tovchigrechko, A., Schobel, S., Chappell, J.D., Larkin, E.K., Gebretsadik, T., Halpin, R.A., Nelson, K.E., Moore, M.L., et al. (2018). Nasopharyngeal Lactobacillus is associated with a reduced risk of childhood wheezing illnesses following acute respiratory syncytial virus infection in infancy. J. Allergy Clin. Immunol. *142*, 1447–1456.e9. https://doi.org/10.1016/j.jaci.2017.10.049.

77. Zhang, Q., Cox, M., Liang, Z., Brinkmann, F., Cardenas, P.A., Duff, R., Bhavsar, P., Cookson, W., Moffatt, M., and Chung, K.F. (2016). Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. PLoS One *11*, e0152724. https://doi.org/10.1371/journal.pone.0152724.

78. Passali, D., Passali, G.C., Vesperini, E., Cocca, S., Visconti, I.C., Ralli, M., and Bellussi, L.M. (2019). The efficacy and tolerability of Streptococcus salivarius 24SMB and Streptococcus oralis 89a administered as nasal spray in the treatment of recurrent upper respiratory tract infections in children. Eur. Rev. Med. Pharmacol. Sci. *23*, 67–72. https://doi.org/10.26355/eurrev_201903_17352.

79. Manti, S., Parisi, G.F., Papale, M., Licari, A., Salpietro, C., Miraglia Del Giudice, M., Marseglia, G.L., and Leonardi, S. (2020). Bacteriotherapy with Streptococcus salivarius 24SMB and Streptococcus oralis 89a nasal spray for treatment of upper respiratory tract infections in children: a pilot study on short-term efficacy. Ital. J. Pediatr. *46*, 42. https://doi.org/10.1186/s13052-020-0798-4.

80. Salminen, S., Endo, A., Isolauri, E., and Scalabrin, D. (2016). Early gut colonization with lactobacilli and Staphylococcus in infants: the hygiene hypothesis extended. J. Pediatr. Gastroenterol. Nutr. *62*, 80–86. https://doi.org/10.1097/MPG.0000000000000925.

81. Ruiz, L., García-Carral, C., and Rodriguez, J.M. (2019). Unfolding the human milk microbiome landscape in the omics era. Front. Microbiol. *10*, 1378. https://doi.org/10.3389/fmicb.2019.01378.

82. Biagi, E., Aceti, A., Quercia, S., Beghetti, I., Rampelli, S., Turroni, S., Soverini, M., Zambrini, A.V., Faldella, G., Candela, M., et al. (2018). Microbial community dynamics in mother's milk and infant's mouth and gut in moderately preterm infants. Front. Microbiol. *9*, 2512. https://doi.org/10.3389/fmicb.2018.02512.

83. Ruokolainen, L., Parkkola, A., Karkman, A., Sinkko, H., Peet, A., Hämäläinen, A.-M., von Hertzen, L., Tillmann, V., Koski, K., Virtanen, S.M., et al. (2020). Contrasting microbiotas between Finnish and Estonian infants: exposure to Acinetobacter may contribute to the allergy gap. Allergy *75*, 2342–2351. https://doi.org/10.1111/all.14250.

84. Wang, M., Huang, J., Liu, Y., Ma, L., Potash, J.B., and Han, S. (2017). COMBAT: a combined association test for genes using summary statistics. Genetics *207*, 883–891. https://doi.org/10.1534/genetics.117.300257.

85. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

86. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226. https://doi.org/10.1186/s40168-018-0605-2.

87. Gloor, G.B., and Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can. J. Microbiol. *62*, 692–703. https://doi.org/10.1139/cjm-2015-0821.

88. Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf.

89. Rosseel, Y. (2012). lavaan: an RPackage for Structural Equation Modeling. J. Stat. Software *48*, 1–36. https://doi.org/10.18637/jss.v048.i02.

90. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis (Springer International Publishing).

91. Palarea-Albaladejo, J., and Martín-Fernández, J.A. (2015). zCompositions-R package for multivariate imputation of left-censored data under a compositional approach. Chemom. Intell. Lab. Syst. *143*, 85–96. https://doi.org/10.1016/j.chemolab.2015.02.019.

92. Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

93. Wheeler, R.E., and Torchiano, M. (2016). Permutation tests for linear models in R. R package version 2. https://cran.r-project.org/web/packages/lmPerm/vignettes/lmPerm.pdf.

94. Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A.M., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife *10*, e65088. https://doi.org/10.7554/eLife.65088.

95. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.

96. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

97. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

98. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584. https://doi.org/10.7717/peerj.2584.

99. Subbarao, P., Anand, S.S., Becker, A.B., Befus, A.D., Brauer, M., Brook, J.R., Denburg, J.A., HayGlass, K.T., Kobor, M.S., Kollmann, T.R., et al. (2015). The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma. Thorax *70*, 998–1000. https://doi.org/10.1136/thoraxjnl-2015-207246.

100. Moraes, T.J., Lefebvre, D.L., Chooniedass, R., Becker, A.B., Brook, J.R., Denburg, J., HayGlass, K.T., Hegele, R.G., Kollmann, T.R., Macri, J., et al. (2015). The Canadian healthy infant longitudinal development birth cohort study: biological samples and biobanking. Paediatr. Perinat. Epidemiol. *29*, 84–92.

101. Azad, M.B., Robertson, B., Atakora, F., Becker, A.B., Subbarao, P., Moraes, T.J., Mandhane, P.J., Turvey, S.E., Lefebvre, D.L., Sears, M.R., et al. (2018). Human milk oligosaccharide concentrations are associated with multiple fixed and modifiable maternal characteristics, environmental factors, and feeding practices. J. Nutr. *148*, 1733–1742. https://doi.org/10.1093/jn/nxy175.

102. Cruz-Hernandez, C., Goeuriot, S., Giuffrida, F., Thakkar, S.K., and Destaillats, F. (2013). Direct quantification of fatty acids in human milk by gas chromatography. J. Chromatogr. A *1284*, 174–179. https://doi.org/10.1016/j.chroma.2013.01.094.

103. Miliku, K., Duan, Q.L., Moraes, T.J., Becker, A.B., Mandhane, P.J., Turvey, S.E., Lefebvre, D.L., Sears, M.R., Subbarao, P., Field, C.J., et al. (2019). Human milk fatty acid composition is associated with dietary, genetic, sociodemographic, and environmental factors in the CHILD Cohort Study. Am. J. Clin. Nutr. *110*, 1370–1383. https://doi.org/10.1093/ajcn/nqz229.

104. Dawod, B., Marshall, J.S., and Azad, M.B. (2021). Breastfeeding and the developmental origins of mucosal immunity: how human milk shapes the innate and adaptive mucosal immune systems. Curr. Opin. Gastroenterol. *37*, 547–556. https://doi.org/10.1097/MOG.0000000000000778.

105. Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G., and Neufeld, J.D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl. Environ. Microbiol. *77*, 3846–3852. https://doi.org/10.1128/AEM.02772-10.

106. Hoskinson, C., Dai, D.L.Y., Del Bel, K.L., Becker, A.B., Moraes, T.J., Mandhane, P.J., Finlay, B.B., Simons, E., Kozyrskyj, A.L., Azad, M.B., et al. (2023). Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease. Nat. Commun. *14*, 4785. https://doi.org/10.1038/s41467-023-40336-4.

107. Drall, K.M., Tun, H.M., Morales-Lizcano, N.P., Konya, T.B., Guttman, D.S., Field, C.J., Mandal, R., Wishart, D.S., Becker, A.B., Azad, M.B., et al. (2019). Clostridioides difficile Colonization is differentially associated with gut microbiome profiles by infant feeding modality at 3–4 months of age. Front. Immunol. *10*, 2866. https://doi.org/10.3389/fimmu.2019.02866.

108. Bridgman, S.L., Malmuthuge, N., Mandal, R., Field, C.J., Haqq, A.M., Mandhane, P.J., Moraes, T.J., Turvey, S.E., Simons, E., Subbarao, P., et al. (2022). Childhood body mass index and associations with infant gut metabolites and secretory IgA: findings from a prospective cohort study. Int. J. Obes. (Lond) *46*, 1712–1719. https://doi.org/10.1038/s41366-022-01183-3.

109. Zheng, J., Zhang, L., Johnson, M., Mandal, R., and Wishart, D.S. (2020). Comprehensive targeted metabolomic assay for urine analysis. Anal. Chem. *92*, 10627–10634. https://doi.org/10.1021/acs.analchem.0c01682.

110. Vergara, A., Wang, K., Colombo, D., Gheblawi, M., Rasmuson, J., Mandal, R., Del Nonno, F., Chiu, B., Scholey, J.W., Soler, M.J., et al. (2023). Urinary angiotensin-converting enzyme 2 and metabolomics in COVID-19-mediated kidney injury. Clin. Kidney J. *16*, 272–284. https://doi.org/10.1093/ckj/sfac215.

111. Edgar, R.C. (2013). Uparse: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods *10*, 996–998. https://doi.org/10.1038/nmeth.2604.

112. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596. https://doi.org/10.1093/nar/gks1219.

113. Kline, R.B. (2015). Principles and Practice of Structural Equation Modeling (Guilford Publications).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw nasal microbiome data (16S rRNA gene sequencing)	NCBI BioProject	PRJNA1127065
Raw gut microbiome data (16S rRNA gene sequencing)	NCBI BioProject	PRJNA657821
Raw gut microbiome data (shothun metagenomics)	NCBI BioProject	PRJNA838575
Gut metabolic profile data	MetaboLights	MTBLS7919
Human milk composition data	CHILD Cohort Study	https://childstudy.ca/childdb/
Software and algorithms		
Prevalence trajectories coordinate system (PreTCO System)	This paper	https://github.com/Shenhav-Lab/Microbial-colonization-programs
R	R Core Team	https://www.r-project.org/
ComBat	Wang et al.[84]	https://CRAN.R-project.org/package=COMBAT
Phyloseq package	McMurdie and Holmes[85]	https://doi.org/10.18129/B9.bioc.phyloseq
Decontam v. 1.10	Davis et al.[86]	https://doi.org/10.18129/B9.bioc.decontam
CoDaSeq package	Gloor and Reid[87]	https://github.com/ggloor/CoDaSeq
gbm package	Ridgeway[88]	https://CRAN.R-project.org/package=gbm
lavaan package	Rosseel[89]	https://CRAN.R-project.org/package=lavaan
ggplot2 package v3.5.1	Wickham[90]	https://CRAN.R-project.org/package=ggplot2
zCompositions package	Palarea-Albaladejo and Martín-Fernández[91]	https://CRAN.R-project.org/package=zCompositions
Vegan package	Dixon[92]	https://CRAN.R-project.org/package=vegan
lmPerm package	Wheeler and Torchiano[93]	https://CRAN.R-project.org/package=lmPerm
MetaPhlAn 3	Beghini et al.[94]	https://github.com/biobakery/MetaPhlAn
HUMAnN 3	Beghini et al.[94]	https://github.com/biobakery/humann
DADA2 v1.10.0	Callahan et al.[95]	https://github.com/benjjneb/dada2
QIIME 2 v2019.10	Bolyen et al.[96]	https://qiime2.org
cutadapt v2.6	Martin[97]	https://github.com/marcelm/cutadapt
VSEARCH v2.7.0	Rognes et al.[98]	https://github.com/torognes/vsearch
Adobe Illustrator	Adobe	https://www.adobe.com/products/illustrator.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Liat Shenhav (Liat.Shenhav@nyulangone.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Nasal and gut microbiome data have been deposited at NCBI BioProjects: PRJNA657821 (infant stool 16S data), PRJNA1127065 (nasal swab 16S data), and PRJNA838575 (shotgun metagenomic data). The metabolic profile data used in this study are available in MetaboLights: MTBLS7919. Datasets are publicly available as of the date of publication. Accession numbers are also listed in the key resources table. Other CHILD Study data, including human milk composition and participant metadata, are available by registration to the CHILD database (https://childstudy.ca/childdb/) and the submission of a formal request. The analysis pipeline is deposited in the GitHub repository (https://github.com/Shenhav-Lab/Microbial-colonization-programs). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study population

Women with singleton pregnancies were enrolled in the CHILD cohort study between 2008 and 2012 (n=3621) and remained eligible if they delivered a healthy infant >35 weeks gestation (n=3455).[99] Nasal swabs and infant stool were collected at a home visit scheduled for age 3 months and additional samples were collected at a clinic visit scheduled for age 1 year. Mothers gave written informed consent in accordance with the Declaration of Helsinki. The protocols were approved by the Human Research Ethics Boards at McMaster University, the Hospital for Sick Children, and the Universities of Manitoba, Alberta, and British Columbia. A total of 3326 participants contributed nasal swab samples, and from these, a representative set of 2725 3-month and 2336 1-year samples were selected for microbiome analysis (Figure S1). Gut microbiota 16S rRNA gene data were available for an asthma-enriched subset of 857 and 879 infants at 3-months and 1-year, respectively.[53] Gender was not captured among mothers or infants in the early years of the CHILD cohort, from which our data originate. Ongoing follow-up of this longitudinal cohort has been designed to capture this information in the future.

Early-life exposures

Infant feeding was reported by standardized questionnaires at 3, 6, and 12 months. At the time of milk sample collection, breastmilk (BM) feeding status was divided into 3 groups: 'Exclusive BM', 'Partial BM' and 'No BM' at 3 months of age (time of sample collection). Infants in the Exclusive BM group received only their mother's BM from birth until at least the time of sample collection (N = 794). Infants in the 'Partial BM' group were fed a mixture of BM and formula at the time of sample collection (N = 517); the infants in the 'No BM' group were breastfed to some extent in their first months of life, prior to the sample collection, but were no longer receiving BM at the time of sample collection (3 months; N = 234). Infants who were never breastfed were excluded due to small sample size (N=55).

Unless specified otherwise, analyses of breastfeeding at 3 months compared infants exclusively breastfed to those no longer breastfed at sampling. Human milk feeding at 1 year sample collection was classified as ''Yes'' (continuation of any human milk feeding) or ''No'' (ceased breastfeeding prior to 1 year sampling). Maternal age, infant sex, birth weight, gestational age, birth mode (Cesarean section or vaginal delivery), parity (any older siblings, yes vs. no), and intrapartum antibiotic use were documented from hospital records. Any colds and any fever from 0 to 3 months and from 6 to 12 months, any antibiotics in the first year of life, daycare (inferred based on the child regularly going to a location away from home), prenatal smoke exposure (second hand and maternal smoking), maternal ethnicity (4 category variable: Asian, Caucasian, First Nations, Other), and maternal asthma (Yes or No asthma ever) were reported by standardized questionnaire. Maternal atopy was diagnosed by skin prick test at 1 year postpartum. If the mother reported recalling her weight at the time, they became pregnant, maternal BMI was calculated using recalled weight, otherwise the mothers weight measured in the clinic at 1 year postpartum was used, or if available, prenatal records were used. Maternal height was measured in the clinic at 1 year postpartum.

Respiratory phenotypes

At ages 1, 3 and 5 years, the child was examined for evidence of asthma by clinical assessment as described previously.[99] Caregivers also completed questions from the International Study of Asthma and Allergies in Childhood (ISAAC) at these assessments. At each assessment, the child is examined by an expert physician for evidence of atopic dermatitis, allergic rhinitis, and asthma. Allergy skin prick tests and general anthropometrics are performed at all assessments. Blood pressure, waist circumference and skinfold thickness are measured at 3 and 5 years. Spirometry is performed in all children at age 5 years.

Sample collection

Collection of infant nasal swab, stool and milk samples has been described previously.[52,100] Briefly, during a home visit scheduled for 3-months of age, a sterile swab was inserted 3 mm into the infant's nostril and rotated. This was done in both nares using one swab. The swab was then placed in a sterile vial containing 3 mL universal transport medium. A second nasal swab sample was collected in a similar aseptic manner at the clinical assessment at 1 year of age. On the same days as the nasal swab collection, a soiled diaper was provided, which was refrigerated at home for up to 24 hours before being collected and processed by study staff.[100] Each

mother also provided one milk sample at this 3-month home visit, which was a mix of foremilk and hindmilk from multiple feeding during a 24-hour period collected as described previously.[52]

METHOD DETAILS

Milk composition analyses

Human milk oligosaccharides were measured using high-performance liquid chromatography with fluorescence detection at the University of California, San Diego, as previously described.[101] Fatty acids were measured by gas chromatography at the University of Alberta, as previously described.[102,103] Milk bioactive proteins were analyzed as follows: samples were thawed on ice and lipids removed by centrifugation at 10,000 g for 30 minutes at 4°C. The remaining whey was then centrifuged at 1,000 g for 10 minutes at 4°C. Milk bioactive proteins were selected based on the detectable immune factors present in human milk.[104] For analysis of milk IgA (E-Bioscience 88-50600-88), soluble TLR2 (sTLR2), TGF-β1, and TGF-β2, Sandwich ELISA was used(R and D systems). Remaining proteins were analyzed on a Luminex 200 system using Luminex assay multianalyte kits (R&D Systems LXSSAHM; ThermoFisher PPX-09 and MX2979v). Milk samples were randomized over the analysis plates, and the bioactive proteins below the lower limit of detection (LLOD) were excluded. Interplate variation was adjusted using median normalization of the log-transformed concentrations.

Microbiome analysis

DNA isolation

Nasal swab samples were thawed and vortexed well, and 250 uL of the suspension was then used for total nucleic extraction, which was performed using the NucliSens® easyMAG™method (BioMérieux, Quebec). Nucleic acid free UTM samples were included as extraction negative controls. DNA isolation from infant stool samples has been previously described.[53]

Microbiome 16S rRNA sequencing

The nucleic acid extracted from nasal swabs was analyzed at McMaster University by amplifying the V3 hypervariable region of the 16S rRNA gene using 341F (CCTACGGGAGGCAGCAG) and 518R (ATTACCGCGGCTGCTGG) Illumina adapted primers as described previously[105] with the addition of 0.05mg/ml RNAase to each reaction. Amplicons were then visualized on 1.5% agarose gel and positive samples were normalized using the SequalPrep normalization kit (ThermoFisher Scientific) and sequenced on an Illumina MiSeq platform. Infant stool was analyzed at the University of British Columbia by 16S rRNA gene sequencing of the V4 hypervariable region with F515/R806 primers on a MiSeq platform as previously described.[53]

Shotgun metagenomic sequencing

Shotgun metagenomic sequencing data were generated and quality filtered by Diversigen (Minneapolis, MN, USA) from fecal samples (average depth of 5 million reads per sample) as previously described.[106] Briefly, DNA extraction was performed using the MO BIO PowerSoil Pro kit, with DNA quality verified using Quant-iT PicoGreen. Libraries were then prepared, and sequencing data generated on an Illumina NextSeq using single-end 1 x 150 reads. Quality filtering included the removal of reads with a quality score <30, and length <50. Adapter sequences were also trimmed, and host reads were removed. Lastly, samples with fewer than 1 million reads were removed prior to downstream preprocessing.

Shotgun metagenomic data preprocessing

Using MetaPhlAn 3 and HUMAnN 3 from the bioBakery 3 pipeline, sequences were mapped and classified into taxonomic (species and strain level) and functional features within each sample, as previously described.[94,106] Using HUMAnN 3, sequences were mapped to Enzyme commission (EC) and EggNOG functional annotations[55] gene families were further stratified based on contribution from known and unknown species using MetPhlAn 3 and the ChocoPhlAn pangenome database (v3.0.13). For functional abundances stratified by species, additional filters were applied such that stratifications were retained only for species and overall functional annotations present in at least 10% of total samples and in at least 0.01% of total abundance. After additional sample filters (Figure S1), including the removal of samples from participants with missing breastfeeding data, infants never breastfed, and samples not collected within the expected age range, 1,365 3-month and 1365 1-year metagenomic samples were retained. For temporal analysis, 1306 infants with both a 3-month and 1-year sample were retained. In this final temporal dataset, species were only retained if they were present in over 10% of samples at either 3 months or 1 year.

NMR and LC-MS/MS metabolite quantification

Metabolic profiles were created from the infant stool samples at The Metabolomics Innovation Center (TMIC) in Edmonton, Alberta using nuclear magnetic resonance (NMR) analysis[107,108]liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis.[106,109,110] Briefly, analyte concentrations were determined using calibration curves of known concentrations in standard mixtures, and isotope-labeled internal standards were used to correct for technical variation. For all metabolites, precision was confirmed to be <5% and <10% coefficients of variability (CV), respectively. All 31 metabolites from the NMR analysis were kept for downstream analysis, and of the 590 metabolites from the LC-MS/MS analysis, 244 metabolites that were detected in fewer than 20% of our samples were excluded. Next, remaining values below the limit of detection (LOD) were imputed with a value of one-half the minimum concentration for each metabolite and all metabolite concentrations were log-transformed. Additionally, 132 low-variance metabolites (log(SD) less than −5) were excluded.

Also as previously described, PCA analysis and quantification of local outlier factor (lof) was used to identify technical sample outliers, and the "ComBat" R package[84] was used for batch normalization. The sample filtering process described above for metagenomic data was also applied to the final metabolomic dataset used for downstream analysis (Figure S1).

Microbiome 16S rRNA gene data preprocessing

Bioinformatics

Overlapping paired-end reads were processed with the *DADA2* v1.10.0 pipeline using the open-source software QIIME 2 v2019.10 (https://qiime2.org)[95,96] after primer sequences were removed from the nasal sequencing data using *cutadapt* v2.6.[97] Unique amplicon sequence variants were then clustered into operational taxonomic units (OTUs) using the *VSEARCH* v2.7.0 algorithm for closed-reference OTU picking by aligning sequences to the SILVA v138 rDNA reference database clustered at 99% sequence similarity.[98,111,112] A modified version of the SILVA v138 rDNA database taxonomy was used for taxonomic labeling (e.g., numbers were appended to differentiate distinct OTUs with the same lowest level of taxonomy identified). Downstream preprocessing was conducted in R and using the *Phyloseq* package.[85]

Decontamination

OTUs assigned to Eukaryota at the kingdom level, belonging to the family of mitochondria or order of chloroplast were removed (806 OTUs in nasal and 23 OTUs in gut samples). This left a remaining 17951 OTUs in nasal and 2505 OTUs in gut samples. Potential reagent contaminants were identified in nasal sequencing data using the prevalence method of *Decontam* v. 1.10.[86] Using all 70 PCR negative controls and all 3786 samples from the 34 sequencing runs containing PCR negative controls, 245 PCR associated contaminants were identified using the *Decontam* algorithm. Using all 51 extraction negative controls and all 4388 samples from the 35 sequencing runs containing extraction negative controls, 257 extraction associated contaminants were identified. All 478 OTUs identified as PCR and/or extraction contaminants were removed. Thirteen OTUs belonging to the genus *Halomonas* had a high read count in negative controls of the gut sequencing data and thus were removed from gut samples [median read-count (IQR), 213 (122-270) in negative controls and 0 (0-0) in samples].

Data filtering

Controls and samples with fewer than 8000 reads per sample were removed, leaving 5265 nasal and 1549 gut samples with 17167 and 2296 OTUs, respectively. We have rarefied to a minimum number of 8,000 reads per sample in order to retain a larger number of samples compared to using 10,000 reads (98 more nasal samples, 5265 vs. 5167 of 5681 total, and 66 more gut samples, 1549 vs. 1483 of 1725 total). Further, we found that 8000 reads were a sufficient depth to maintain accurate richness estimates based on the slopes of rarefaction curves, which were near zero (Figure S10 median slope 0.0008 (0.0003-0.0015) & 0.00008 (0.00002-0.0002) for nasal and gut, respectively). OTUs with less than 0.0001% mean relative abundance in samples or biological controls were removed, resulting in 4042 remaining OTUs in nasal and 1754 OTUs in gut samples. Using the threshold of 0.0001%, the majority of reads per sample were retained. A median (IQR) of just 0.10% (0.01-0.35%) and 0% (0-0.02%) of total sequences were removed per nasal and gut sample, respectively. Alpha-diversity metrics were calculated per dataset (i.e., gut and nasal). Prevalence-based analyses, rarefaction was additionally performed to the minimum depth (8000 reads/sample).

Sample filters

Figure S1 illustrates the sample filtering process and sample sizes for downstream statistical analyses of all datasets. For 16S rRNA gene data for instance, in addition to above mentioned filters, 560 nasal samples collected from various extra time-points from Toronto participants and 8 biological controls were removed, leaving 3-month (n=2510) and 1-year (n=2187) nasal samples from 2953 participants (Figure S1). Further, participants without information on breastfeeding status at the time of sample collection were removed (leaving 4212 nasal and 1525 gut samples). Additionally, samples collected for the 3-month time-point that were collected earlier than 1 month or later than or equal to 8 months of age, and samples collected for the 1-year time-point that were collected earlier than 8 months or later than or equal to 21 months of age were excluded (leaving 4181 nasal and 1518 gut samples). Lastly, infants that never received any breast milk were removed, leaving 4095 nasal (2227 3-month and 1868 1-year) and 1472 gut (744 3-month and 728 1-year) samples. These infants were removed because we would not have had sufficient power to examine them as a separate group and it was not biologically appropriate to group them with infants who were breastfed. This left 1545 infants with nasal microbiota data at both 3-months and 1-year, and 555 infants with gut microbiota data at both 3-months and 1-year to use for analyses of microbiota trajectories (e.g., paired analyses).

For our primary analyses of respiratory health, we included all infants who were subsequently diagnosed with asthma at 3 years (n = 80 (6.1%) for nasal and 56 (11.7%) for gut trajectories), as well as internal controls (n = 1236 for nasal and 421 for gut trajectories, Table S1), and excluded infants with lower certainty in their asthma diagnosis (i.e. possible asthma, n = 91 for nasal and 58 for gut) and those with missing asthma diagnosis data (n = 138 for nasal and 20 for gut).

QUANTIFICATION AND STATISTICAL ANALYSIS

Univariate analysis

We first assessed univariate associations between early-life factors thought to impact infant microbiota and/or infant health based on previous literature. To investigate the early-life factors associated with nasal and gut microbiota composition we used redundancy analysis.[92] OTUs with a prevalence over 5% were retained for this analysis, and abundances were centered log-ratio (CLR)

transformed using the *CoDaSeq* package[87] after zeros were imputed using a Bayesian-multiplicative replacement method.[91] The univariate models for 3-month and 1-year nasal microbiota included 210 OTUs and 306 OTUs with prevalence over 5% in the respective timepoints, and the models for 3 month and 1 year gut microbiota included 124 and 181 OTUs with prevalence over 5% in the respective timepoints.

Linear models[93] were used to assess univariate associations between early-life factors and microbiota richness (observed OTUs), diversity (Shannon index), richness trajectories (1 year minus 3-month richness), diversity trajectories (1 year minus 3-month diversity). The Benjamini-Hochberg procedure was used to adjust all p-values.

In the univariate results presented in Figure 2, we show variation explained as the R^2 of the linear model [richness (observed OTUs), diversity (Shannon index) and trajectories (change in richness and diversity from 3 months to 1 year)], or R^2 of the redundancy analysis [microbiota composition]. For linear models, the effect direction was extracted using beta-estimates. Color differentiation by direction is not shown for categorical variables (Study Center, Season and Ethnicity) or microbiota composition. Time-varying factors such as breastfeeding (exclusive, partial, and no [reference group]) were measured at each sample collection point. For trajectories, we show results from the time-point yielding the largest effect sizes; these are 3-month measurements for weight, age, and breastfeeding, and 1-year measurements for R/E virus (rhinovirus/enterovirus) and season.

Univariate associations of breastfeeding exclusivity at the 3-month sample collection and asthma at 3 and 5 years with nasal and gut microbiota richness, diversity and richness and diversity trajectories (1 year minus 3-month richness), early colonizers (percent of all OTUs that were only observed in an infant's 3 month sample), persistent colonizers (percent of all OTUs that were observed in both an infant's 3 month and 1 year sample), and late colonizers (percent of all OTUs that were only observed in an infant's 1 year sample), were also investigated using Mann-Whitney U tests. Using the Wilcoxon signed-rank test (for paired data), nasal and gut microbiota richness and diversity were also compared between 3 months and 1 year of age within infants exclusively breastfed and those no longer breastfed (No BM) at the 3-month sample collection, and within infants that did and did not develop asthma at 3 years. These same univariate tests were later applied to the static (3-month) and paired (3-month and 1-year) infant gut metabolomic datasets (Figure S1) to assess consistency in findings for microbial related metabolites (namely SCFAs, Tryptophan, Tryptamine and Indole).

Microbial colonization patterns

Prevalence was used rather than abundance to determine colonization patterns in order to 1) circumvent the constraints of compositionality and sparsity in microbiome abundance data 2) use a measure of the microbiome that is more indicative of whether or not a microbe has colonized the community (presence/absence) and 3) simplify the interpretations of microbiome trajectories in terms of colonization. For colonization pattern and prevalence trajectory tests (based on presence/absence data), OTUs that were present in over 10% of samples at either or both timepoints were retained for analysis, and one nasal OTU (*Streptococcus sp. 40*) present in over 95% of samples was removed, leaving 170 nasal and 115 gut OTUs. These filters were applied because prevalence tests require variability in prevalence among groups, which is impossible when overall prevalence is too high or too low. Notably, the sparsity of the nasal and gut microbiota datasets even after the abovementioned filters were applied would pose a challenge for tests of relative abundance; however, it became advantageous for our tests of prevalence trajectories (median, IQR prevalence of OTUs; Nasal, 14.7% 10.5-24.9%; Gut, 16.8%, 11.2-27.5%).

The overall difference in prevalence of microbiota between 3 months and 1 year (i.e., 1 year – 3 month) for infants with nasal samples (n=1545) or gut samples (n=555) at both timepoints was tested using the McNemar test. The Benjamini-Hochberg procedure was used to adjust p-values. Taxa were differentiated as late colonizers (more prevalent at 1 year) when the change in prevalence was greater than 7% and $p_{BH} < 0.001$, as early colonizers (more prevalent at 3 months) when change in prevalence was less than –7% and $p_{BH} < 0.001$, and otherwise were considered typically persistent colonizers. These p-value and effect-size thresholds were selected based on the appearance of the nasal and gut volcano plots (Figure 3A), and because even colonizers that we would consider "persistent" change slightly in prevalence between time-points, therefore a *p*-value threshold alone would not be sufficient. This provided general classifications of colonization patterns of microbiota over all infants. However, given that a specific microbe can be early (3 month only), persistent (both timepoints), late (1 year only) or absent at the individual level, we additionally compared these patterns between different early-life exposures and respiratory phenotypes.

Prevalence trajectories coordinate system (PreTCO System)

In Order to systematically compare between the colonization patterns of different health phenotypes, both at the microbial community level and at the individual taxa level, we devised a novel computational approach we term 'prevalence trajectories coordinate system' (PreTCO system, Figure 3B). We define the prevalence trajectories as a cartesian coordinate system in two dimensions, in which each point is defined by a pair of numerical coordinates. Each point represents a taxon and its coordinates correspond to the change in prevalence from 3-months to 1-year for two phenotypic groups. For example, when comparing nasal colonization patterns in infants diagnosed with asthma at 3 years and healthy infants (Figure 3C), the numerical coordinates of *Rothia 1* are (-15, 7). The first coordinate (i.e., -15) corresponds to the change (e.g., a 15% decrease) in prevalence between the 3-month and 1-year samples across all subjects from phenotype 1 (i.e., asthma). The second coordinate (i.e., 7) corresponds to the percent change (e.g., a 7% increase) in prevalence between the 3-month and 1-year samples across all subjects from phenotype 2 (i.e., healthy). The PreTCO system enables visualization of colonization patterns globally, with a perfect y = x linear relationship indicating no difference between groups and increasing deviations from the y = x linear relationship indicating a greater difference in prevalence trajectories between

groups. The overall difference in community colonization patterns, considering all taxa collectively, is reflected by evaluating the pairwise distance between the x and y coordinates using the Wilcoxon signed-rank test that generates a single p-value (i.e., the global difference in colonization patterns between the two phenotypes). This coordinate system also enables local evaluations of specific microbial taxa that differ in colonization patterns between the groups. To this end, we generate a null distribution using permutations and test the difference in prevalence trajectories of individual taxa between groups.

Of note, prior studies have relied on the concept of estimated 'microbiome age' to assess microbiome maturation patterns. While insightful, the microbiome age approach has its drawbacks. These include a lack of precision in identifying specific microbes simultaneously associated with multiple phenotypes, such as asthma and breastfeeding in this case. Moreover, it may not fully utilize longitudinal data when comparing repeated measures sampled from the same participants over time.

In contrast, the PreTCO system presents a novel perspective on microbial maturation. Rather than predicting a child's age based on the microbiome, PreTCO characterizes microbial colonization patterns over time, identifying commonalities across diverse phenotypes. This approach pinpoints microbial features with varying colonization patterns, offering a deeper understanding of maturation between two (or more) timepoints; in this case, from 3 months to 1 year. Unlike microbiome age, PreTCO allows exploration of dynamic changes within the microbiome community over time, enabling analysis of repeated measures from the same individual. Moreover, PreTCO accommodates various facets of microbiome data, including taxa abundance, presence/absence, and functional data such as EC numbers (see Figure 5). It empowers the comparison of temporal changes within phenotypic groups (e.g., breastfed vs. no longer breastfed, healthy vs. children diagnosed with asthma) and across groups (e.g., asthmatic and no longer breastfed vs. healthy and breastfed), providing valuable insights into microbiome development. Importantly, PreTCO addresses critical issues like sparsity and compositionality in trajectory analysis, enhancing the robustness of our findings.

The main exposures compared using the PreTCO were exclusive breast milk (nasal n=794, gut n=271) and no breast milk (nasal n=234, gut n=79) at 3-month sample collection, and the main respiratory phenotypes compared were no asthma (nasal n=1236, gut n=421), and asthma (nasal n=80, gut n=56) at 3 years. Additional comparisons tested were asthma at 3 and 5 years (nasal n=53, gut n=35, defined as probable asthma diagnosis at 3 years, and possible or probable asthma diagnosis at 5 years) compared to no asthma at 3 or 5 years (nasal n=1038, gut n=347).

For the main exposure (breastfeeding exclusivity at 3 months) and respiratory phenotype (asthma at 3 years), the PreTCO method was later applied to infant gut metagenomic data, to assess whether similar patterns of delayed functions could be observed for species identified as late colonizers based on 16S data. For this analysis, we selected 8 species identified as later colonizers in infants who did not develop asthma compared to those that developed asthma at 3 years and that had species-level identification and were found in the EC and EggNOG annotations: *Ruminococcus gnavus* (401 EC & 940 EggNOG), *Bifidobacterium bifidum* (419 & 797 functions), *Bifidobacterium longum* (481 & 713 functions), *Sellimonas intestinalis* (305 & 128 functions), *Eubacteriu hallii* (487 & 574 functions), *Eubacterium eligens* (423 & 95 functions), *Coprobacillus cateniformis* (288 & 21 functions), and *Parabacteroides merdae* (360 & 15 functions). In addition to using the same methods as described above for 16S data to assess the global and individual changes in prevalence of functional annotations, the change in prevalence of functional annotations was stratified by species and for each species, the Wilcoxon signed-rank test was used to generate p-value for the percent change in prevalence of functions between groups. The Benjamini-Hochberg procedure was used to adjust p-values.

Prevalence trajectories of microbes and their pangenomes

Throughout the manuscript we calculated two complementary types of trajectories: (1) trajectories describing changes in microbial prevalence, with OTUs for 16S data and species for shotgun metagenomics; and (2) trajectories describing the prevalence of functions assigned to these species, using shotgun metagenomics. Applying the PreTCO analysis to the first type of trajectories provides insights into the changes in microbial prevalence over time. For the second type of trajectories, applying the PreTCO reveals changes in functional prevalence, capturing the genetic diversity of the species' and representing their pangenomes (i.e., the entire set of genes from all strains within a clade). This approach potentially uncovers functions of different sub-species or strains within the same clade.

To discern the additional insights provided by assessing both types of trajectories, we examined their relationships, as depicted in Figure S8. While there is a moderate to strong correlation between the prevalence trajectories of a species and some of its functional trajectories, the correlation is not perfect. The significant variability among different functions, as illustrated in Figure 5A, highlights this point. This observed variability in functions within a species indicates that functional data not only mirror the prevalence of each species but also expose distinct genetic variations in subspecies that are not captured by species abundance data alone.

To exemplify this point, we focused on *R. gnavus*, and analyzed how its overall prevalence trajectory (derived from shotgun metagenomics and 16S) correlates with the trajectories of its functions (derived from shotgun metagenomics, n = 392 functions). Next, we compared the trajectories of these eight *R. gnavus* functions with the overall prevalence trajectories of the top ten "late" taxa in the gut from 16S data using a correlation analysis. Finally, we randomly selected a set of 6 functions that are not significantly delayed with breastfeeding and calculated the correlation between the overall *R. gnavus* prevalence trajectory and the trajectory of this random set of *R. gnavus* functions (Figure S8).

Prediction model

We trained a machine learning model, gradient-boosted decision trees, to differentiate between children that were diagnosed with asthma at 3 years (n = 80 for nasal and 56 for gut) and healthy controls (n = 1236 for nasal and 421 for gut). We trained and tested the

model using a 10- repeated cross validation. In each fold, the subjects were completely held out to avoid data leakage from train to test (each subject is represented by their trajectories/colonization patterns). We used the CLR transformed abundance to represent microbial composition of either the nasal or gut microbiome at a single time point (3 months; 170 and 115 features per time point for nasal and gut respectively) as well as prevalence for the microbial trajectories and colonization patterns (340 and 230 features for nasal and gut respectively). We evaluated and compared the prediction accuracy using the area under the receiver operating characteristic curve (auROC). Finally, we constructed a model combining the microbial trajectories of either the nasal or the gut with human milk components (HMOs, fatty acids, cytokines, antibodies). To elucidate which microbial taxa and human milk components drive the prediction accuracy of asthma at 3 years, we evaluated the contribution of these features to our machine learning model by ranking the features (i.e., microbial taxa and human milk components) by their contribution to the model performance. Sample sizes per model - nasal trajectories: n = 1407, nasal trajectories and milk n = 625, gut trajectories: n = 535, gut trajectories and milk: n = 283, nasal and gut trajectories combined: n = 378, nasal trajectories, gut trajectories and milk: n = 283. Models were generated using the *gbm* package.[88]

Multivariate analysis

We have shown that microbial colonization patterns are strongly associated with breastfeeding and asthma. We next confirmed that these findings hold true after adjusting for potential confounders. We specifically adjusted for likely confounders based on previous univariate analyses showing associations with both the primary outcome (asthma) and exposure (nasal and/or gut microbiota trajectories): antibiotics in the first year of life, any older siblings, prenatal smoke exposure, birth mode, maternal asthma, study center, R/E virus at 1 year, colds in the first 3 months and infant age at 3-month sample collection. Further, given that microbial communities work as a system, we constricted microbiome trajectories as latent constructs. Indicators of the infants' nasal and gut microbiota colonization that were significantly associated with both breastfeeding and asthma in univariate analyses (PreTCO system for taxa and Mann-Whitney U test for summary metrics) were selected and modeled as latent variables (Figure S6A). After adjusting for potential confounders using multivariate regression, nasal and gut microbiota trajectories remained associated with breastfeeding status at 3 months (standardized $beta_{adjusted}$ = 0.08, p = 0.004 for nasal and 0.16, p<0.001 for gut microbiota trajectories) and with asthma at 3 years (standardized-$beta_{adjusted}$ = –0.17, p < 0.001 for nasal and –0.34, p < 0.001 for gut microbiota trajectories) (Figure 7A).

Structural Equation Models

Structural equation modeling (SEM) was used to perform multivariate mediation in order to determine the indirect effects of breastfeeding exclusivity on asthma at 3 years through nasal and gut microbiota trajectories. SEM was conducted using the *lavaan* package.[89]

We created a latent variable representing the microbiome that includes the following information: (1) overall colonization pattern summary metrics for all the microbes (proportion of late colonizers, microbiota richness and diversity trajectories), (2) the prevalence of specific taxa found in our previous analyses. This way the latent variable is representing both the temporal trajectories across the entire community as well as in specific taxa found to have differential dynamics. Specific taxa were selected in the following way – in the analyses prior to the SEM we found 36 nasal taxa and 36 gut taxa that their colonization patterns significantly associated with asthma. We further found a substantial overlap between these specific taxa that exhibit later colonization patterns in exclusively breastfeed infants (vs. early breastfeeding cessation) and healthy infants (vs. infants diagnosed with asthma at 3 years). Overall, 47% (17/36) of the late nasal colonizers and 28% (10/36) of late gut colonizers were also associated with breastfeeding (p <0.05, permutation test). The prevalence trajectories of these specific taxa were selected. Covariate selection was also informed by statistical tests described above, and based on their potential importance to the outcome, asthma. The latent variables were scaled to have a variance of one. Given the inclusion of ordinal and dichotomous variables, diagonally weighted least squares model parameter estimates, robust standard errors and a mean and variance adjusted test statistic were used. A comparative fit index (CFI) > 0.9, root-mean-square error of approximation (RMSEA) <0.05, and standardized root-mean residuals (SRMR) <0.08 were considered indicators of good model fit[113] (Table S7). In addition to the removal of infants with an uncertain (''possible'') asthma diagnosis, only infants with data on both nasal and gut microbiota trajectories and with no missing data were included in the final model, limiting the sample size to 341 infants. Therefore, we also made separate models for nasal and gut microbiota trajectories to determine whether these findings held true for the larger subsets of infants with nasal microbiota (n=1316) or gut microbiota (n=477) at both timepoints (Figure S9). In this case, nasal and gut microbiota trajectories were modeled separately, and models were constructed in the same manner as described above, using the same covariates and gut and nasal microbiota trajectory latent variables, and also using the *lavaan* package.[89] Further, as a preliminary step prior to generating these final mediation models, multivariate regression was used to assess breastfeeding status at 3 months as a predictor of microbiota trajectories separately from the assessment of microbiota trajectories as a predictor of asthma at 3 years. Figures were generated using the R package ggplot2.[90]

Supplemental figures

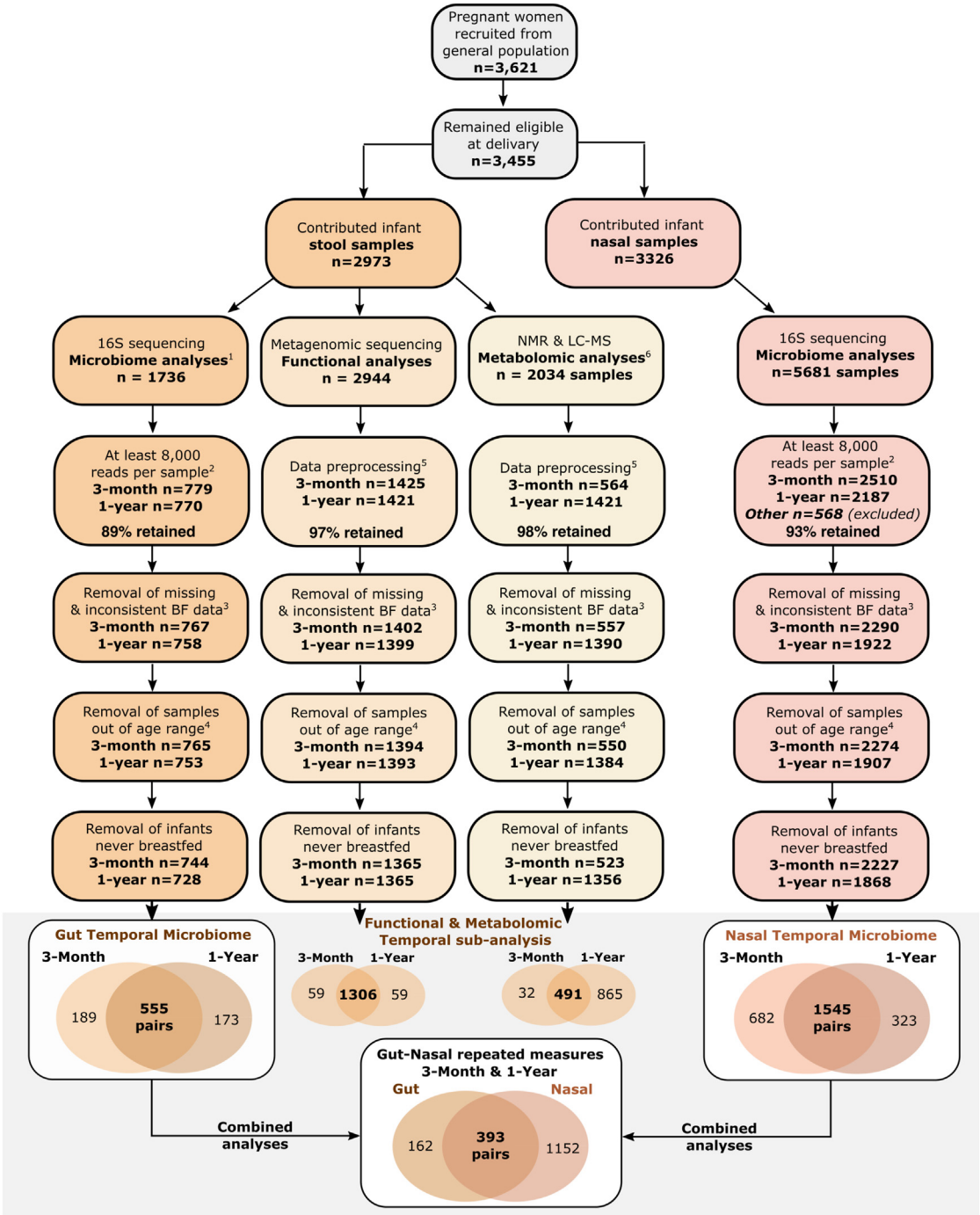


Figure S1. Study design and sample sizes from recruitment to analysis, related to Figure 1

[1]Used a subset of stool samples selected based on availability of infant blood samples and enrichment for infant atopy and/or wheeze at 1 year of age. [2]After removal of reads that did not pass DADA2 filters[95] and contaminant reads, samples with fewer than 8,000 reads were removed. [3]Participants were removed if they were missing values for implied breastfeeding duration or breastfeeding status at 3 months or were no longer breastfeeding when the breast milk sample was obtained (i.e., milk collected and stored at home). [4]Age range for 3-month visit defined as over 1 and less than 8 months and for 1 year visit defined as over 8 and less than 21 months. [5]As described in Hoskinson et al.[106] [6]Sample sizes indicated are for samples with either NMR and/or LC-MS data.

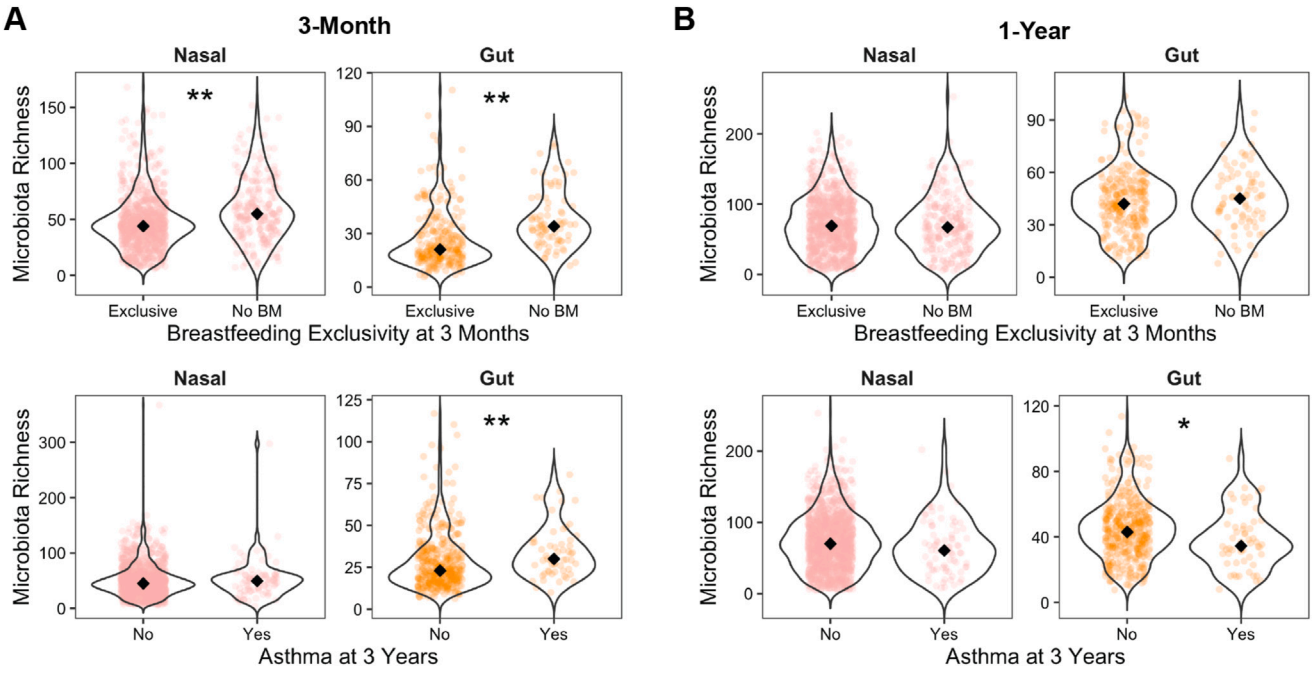


Figure S2. Microbiota richness at individual time points across nasal and gut microbial niches, related to Figure 2

(A and B) Microbiota richness at individual time points for the same subset of infants with both 3-month and 1-year data used to assess microbiota trajectory variables (see Figures 2B–2E).

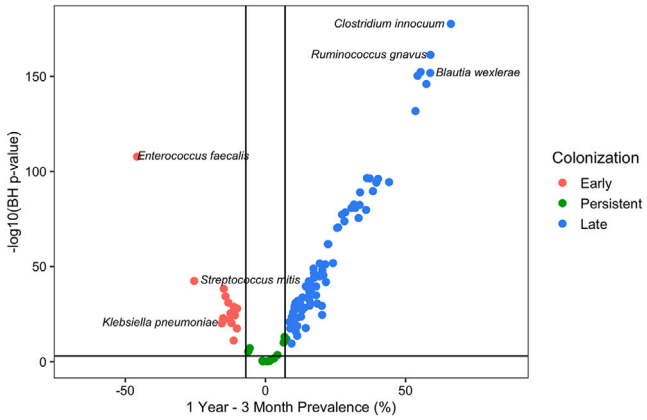


Figure S3. Replication of Figure 3A, using species-level shotgun metagenomic data

The difference in prevalence of microbiota between 3 months and 1 year (1-year to 3-month prevalence in percentage) for infants with gut samples at both time points (n = 1,306 infants) was assessed. The difference in prevalence between time points was tested using the McNemar test. The horizontal line indicates the p value threshold ($p_{(BH)} < 0.001$), and vertical lines indicate the effect size thresholds (−7% and 7%) used to define early (more prevalent at 3 months), persistent (similar prevalence at both time points), and late (more prevalent at 1 year) colonizers. The three early and late colonizers with the highest effect sizes are annotated. BH, Benjamini-Hochberg adjusted p values.

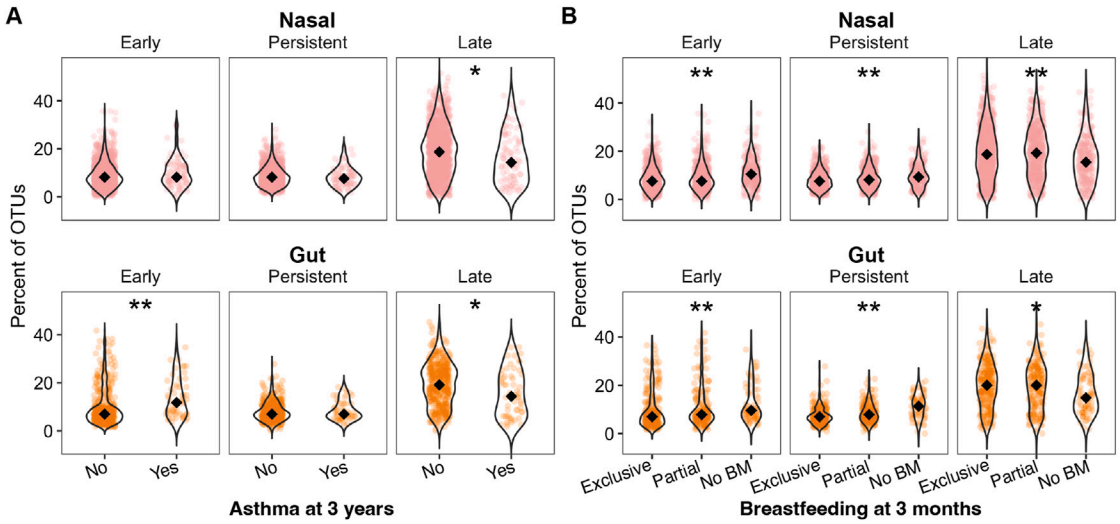


Figure S4. Distribution of early, persistent, and late colonization patterns in the nasal and gut niches across infant feeding groups and respiratory phenotypes, related to Figure 3

(A) Comparison of the proportion of late colonizers between healthy infants and those with asthma at 3 years.

(B) Comparison of the proportion of late colonizers among infant feeding groups. All infant feeding groups are shown, i.e., exclusive breastfeeding (exclusive), partial/mixed feeding (partial), and no breastfeeding (no BM) at 3-month sampling. *p* values correspond to the comparison between exclusive to no BM. The y axis shows the number of OTUs identified as early, persistent, or late colonizers per infant as a percentage of total OTUs. Comparisons are tested using the Mann-Whitney U test, $*p < 0.05$, $**p < 0.001$.

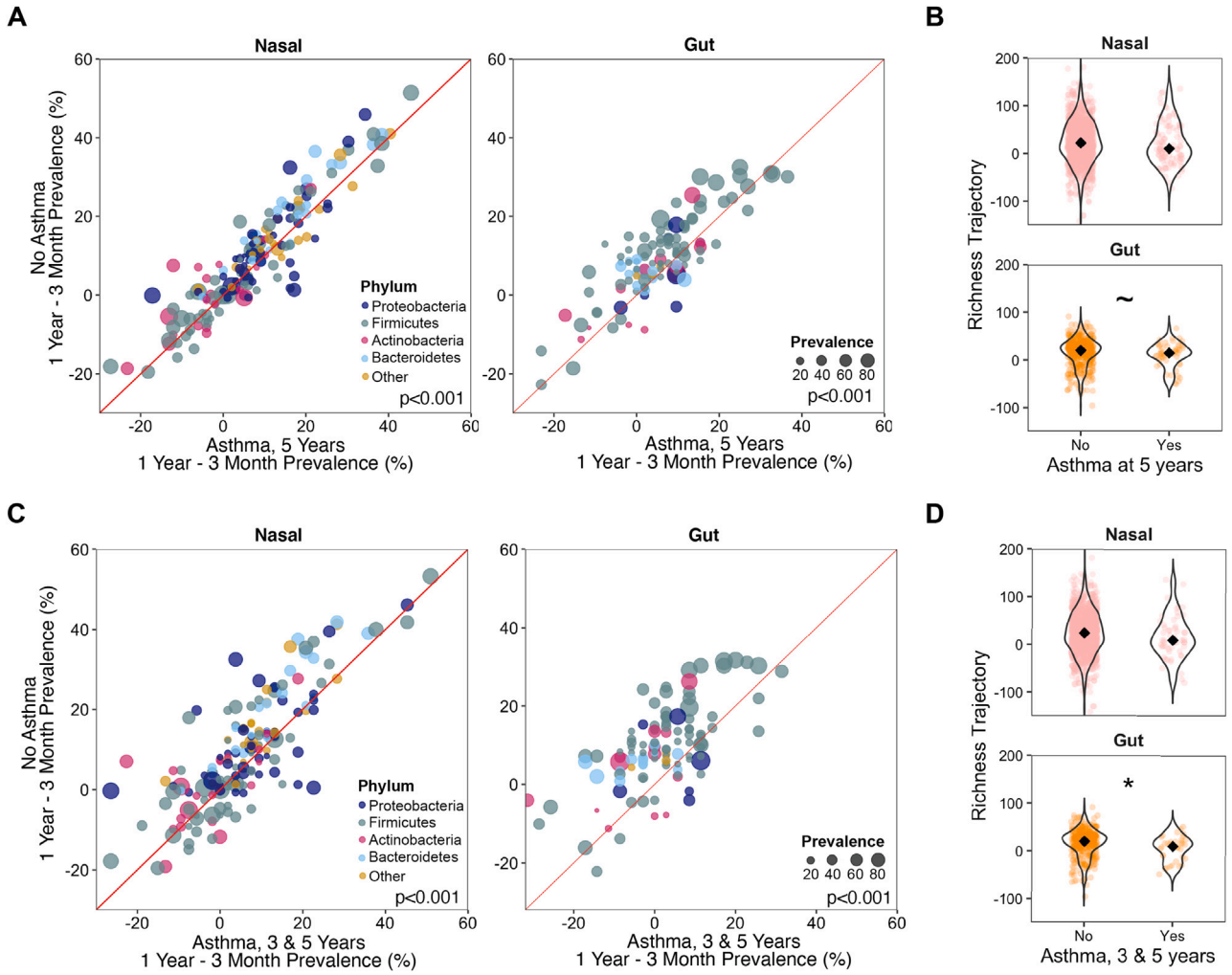


Figure S5. Microbiota richness and colonization patterns between infants with and without asthma at 5 years, and with and without asthma at both 3 and 5 years, related to Figure 3

Main analyses and figures refer to asthma at 3 years only (Figures 2F and 3C).

(A and B) Comparisons of the change in prevalence of microbiota (1-year to3-month prevalence in percentage) (A) and microbiota richness trajectories (B) between infants with (n = 99 for nasal, n = 52 for gut) and without (n = 1,142 for nasal, n = 409 for gut) asthma diagnosed at 5 years.

(C and D) Comparisons of the change in prevalence of microbiota (C) and microbiota richness trajectories (D) between infants with and without asthma at both 3 and 5 years. Infants with no asthma at both 3 and 5 years (n = 1,038 for nasal and n = 347 for gut) were compared to infants who had asthma at 3 years (probable asthma diagnosis), and possible or probable asthma at 5 years (n = 53 for nasal and n = 35 for gut). We found a strong association between the change in prevalence of nasal and gut microbial taxa and asthma status at 3 and 5 years (9.1% [0.8%–17.3%] for no asthma vs. 5.7% [0.0%–13.2%] for asthma in nasal samples; 8.4% [5.0%–14.3%] for no asthma vs. 2.9% [−2.9%–8.6%] for asthma in gut samples, $p < 0.001$). The trajectory of gut microbiome richness was associated with asthma at both 3 and 5 years of age ($p = 0.008$), whereas the nasal microbiome richness trajectory showed no significant association ($p = 0.11$). Comparisons were tested using the Mann-Whitney U test.

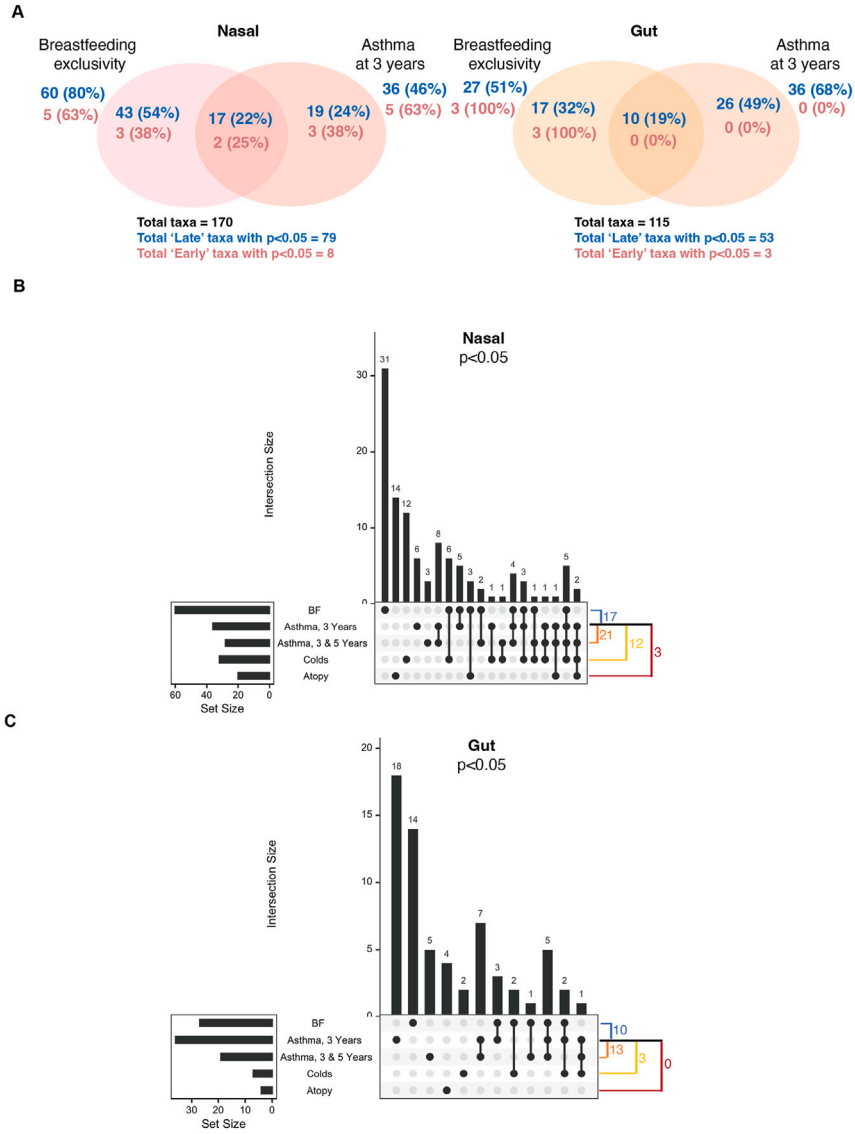
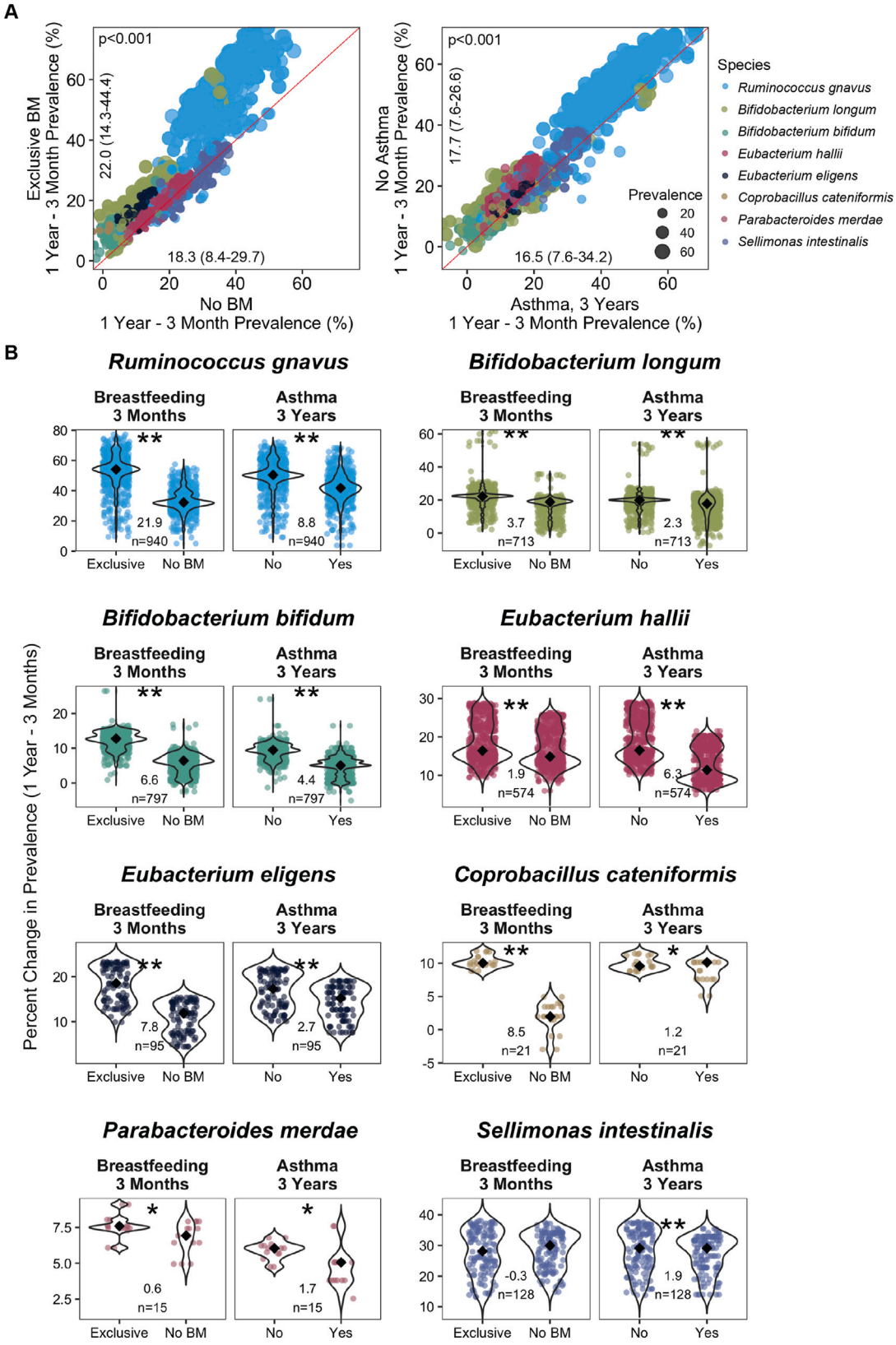


Figure S6. Overlap in associations between microbiota prevalence trajectories and breastfeeding and respiratory phenotypes, related to Figure 3

(A) Overlap in taxa with prevalence trajectories indicating significantly earlier (blue) or later (red) colonization in relation to breastfeeding status at 3 months (exclusive vs. no breastfeeding) and asthma status at 3 years (no asthma vs. asthma). Percentages are out of the total taxa associated with either breastfeeding exclusivity or asthma.

(B and C) Overlap among taxa that colonized later with exclusive breastfeeding (vs. no BM) and with healthy respiratory/allergy phenotypes (i.e., no asthma at 3 years vs. asthma; no asthma at 3 and 5 years vs. asthma at both 3 and 5 years; no colds at 0–3 months vs. colds; no atopy vs. atopy [no wheeze]) in nasal (B) and gut (C) niches. Using a *p* value threshold of 0.05. The numbers above the bars indicate the number of OTUs in the intersection, and the "set size" indicates the total number of associated OTUs ("later" colonizers) for each factor in the prevalence trajectory analysis (see also Table S4). The dendrograms show the total overlap with asthma for each variable.



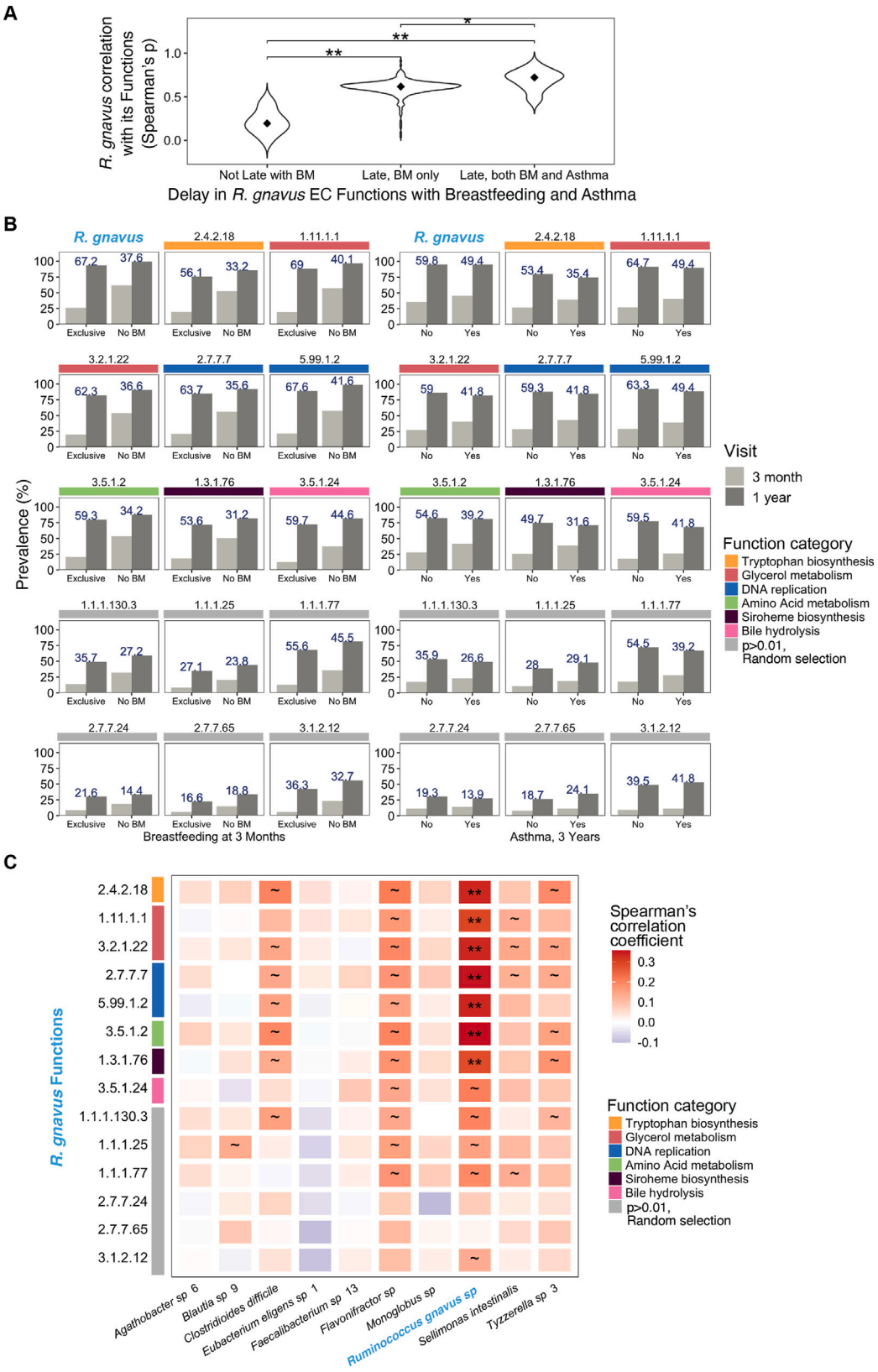
(legend on next page)

Figure S7. Timing of acquiring microbial functions in the first year of life associated with breastfeeding and asthma (EggNOG orthologs), related to Figure 4

Main analyses use EC functions (Figure 4).

(A) Prevalence trajectory coordinate system analyses comparing the change in prevalence of EggNOG orthologs (1-year to 3-month prevalence in percentage; STAR Methods) between (1) infants exclusively breastfed (n = 658) and those no longer breastfed (n = 202) at 3 months; and (2) healthy infants (n = 1,075) and those later diagnosed with asthma at 3 years (n = 79). The median (IQR) change in prevalence of each group is shown on the group's axis. Overall significance was tested using a Mann-Whitney U test. The overall prevalence across both time points is shown as point size.

(B) Percentage change in prevalence of EggNOG orthologs, stratified by species. Significance was tested using a Wilcoxon signed-rank test. Each point represents a single function from the focal species. As an effect size, the median of the difference in the trajectory measure (percentage change in prevalence) is shown. Sample sizes for this test are the number of functions (differs per species, as annotated in the plot). $^{*}p_{(BH)} < 0.05$; $^{**}p_{(BH)} < 0.001$. In both (A) and (B), we selected functions that were linked to species identified as late colonizers in infants who did not develop asthma (vs. asthma at 3 years) based on our 16S rRNA gene results ($p < 0.05$, Figure 3D).



(legend on next page)

Figure S8. The correlation between the species-based trajectories/colonization patterns and functional-based ones in *R. gnavus*, related to Figure 5

(A) Correlation between *R. gnavus* overall species trajectory and functional trajectories depends on associations with breastfeeding and asthma. Each point represents the trajectory of a different EC annotation assigned to *R. gnavus* and the strength of its correlation with the overall species trajectory of *R. gnavus* based on metagenomic data. Functions that are not significantly late in breastfed infants had the lowest Spearman correlation to the overall trajectory (30 functions; median of 0.19). Functions that are significantly late in breastfed infants have a higher correlation, with a median of 0.62 (max–min, 0.05–0.91). Finally, the eight functions we identified as late in both breastfed infants and those not diagnosed with preschool asthma exhibit comparable and even higher correlation, with a median of 0.72 (0.47–0.85). Functions were categorized as "not late with exclusive BM" ($n = 30$), "late with BM only not asthma" ($n = 392$), and "late with both BM and no asthma at 3 years" ($n = 8$) (based on a previous PreTCO analysis; see Figure 4). Mann-Whitney U test, $^{**}p < 0.001$, $^{*}p < 0.05$.

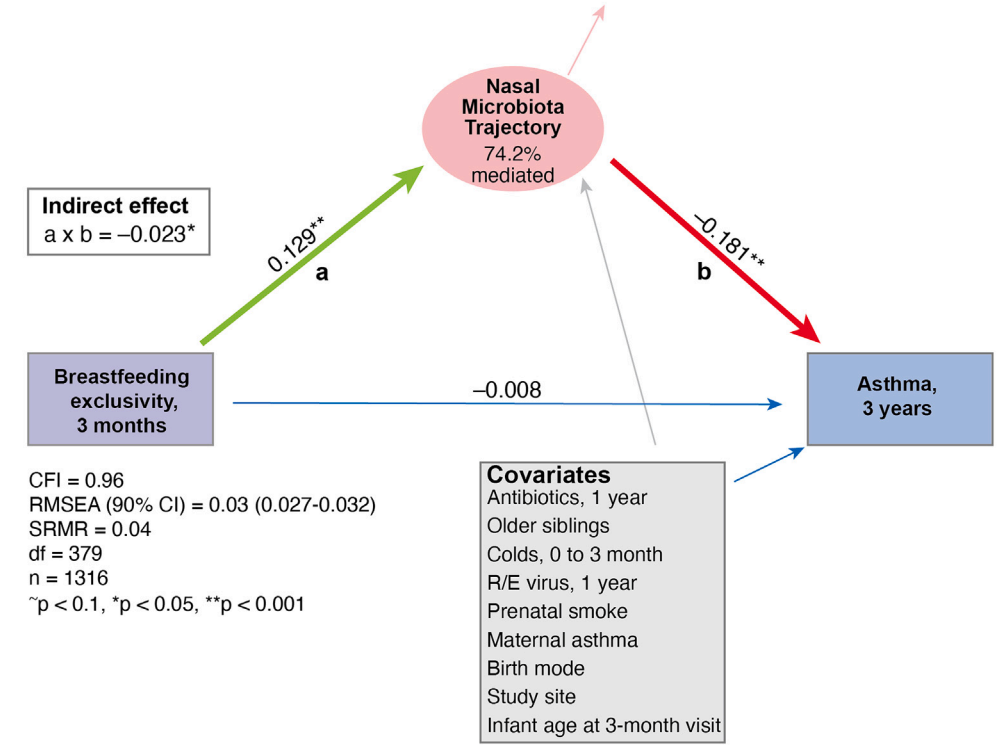
(B) Prevalence of *R. gnavus* species and selected *R. gnavus* functions across time for infants exclusively breastfed ($n = 658$), and not breastfed ($n = 202$) at 3 months (right), and healthy infants ($n = 1,075$) and those with asthma ($n = 79$) at 3 years. Prevalence was assessed from metagenomic data, and shown for eight functions (color annotations) with significant delay in their introduction in breastfed infants and healthy ones based on the PreTCO analysis ($p < 0.01$, see Figure 4A). An additional 6 functional trajectories (gray annotation) were randomly selected from the 30 not identified as delayed in breastfed infants ($p > 0.01$). Numbers shown in blue indicate the percentage change in prevalence (1 year–3 months).

(C) Comparing the trajectories of the identified eight *R. gnavus* functions (y axis) with the overall prevalence trajectories of the top 10 "late" taxa in the gut identified using 16S rRNA gene data. *R. gnavus* showed the strongest correlation with these functional trajectories (median [IQR] rho = 0.33 [0.29–0.34], pBH < 0.05), followed by *Flavonifractor* (rho = 0.17 [0.16–0.19], pBH < 0.05, Figure S8C). This correlation was anticipated as these functions are all mapped to *R. gnavus*. However, the moderate strength of these correlations suggests genetic variations within members of *R. gnavus* that are not captured by species abundance data alone. The functional trajectories (y axis) were selected based on a significant delay in their introduction in breastfed infants and healthy infants ($p < 0.01$) in the PreTCO analysis on functional data obtained from metagenomic sequencing (see Figure 4A). $^{**}p_{(BH)} < 0.001$, $^{*}p_{(BH)} < 0.05$, $\sim p < 0.05$. EC names: 2.4.2.18, anthranilate phosphoribosyltransferase; 1.11.1.1, NADH peroxidase; 3.2.1.22, alpha-galactosidase; 2.7.7.7, DNA-directed DNA polymerase; 5.99.1.2, DNA topoisomerase; 3.5.1.2, glutaminase; 1.3.1.76, precorrin-2 dehydrogenase; 3.5.1.24, choloylglycine hydrolase; 1.1.1.38, malate dehydrogenase; 1.4.1.16, diaminopimelate dehydrogenase; 2.1.1.144, *trans*-aconitate 2-methyltransferase; 2.7.8.26, cobalamin synthase. Overall, these results demonstrate that not all functional trajectories mirror the overall species trajectory. Additionally, these findings suggest the potential for functional diversity among *R. gnavus* members, with functions possibly changing over the course of early life. Whether these changes benefit respiratory development may depend on early-life factors such as breastfeeding. This hypothesis is further supported by extensive literature indicating strain-dependent functional adaptations within species. For example, the ability of *R. gnavus* to symbiotically utilize host mucin glycans is believed to be strain-dependent.[68]

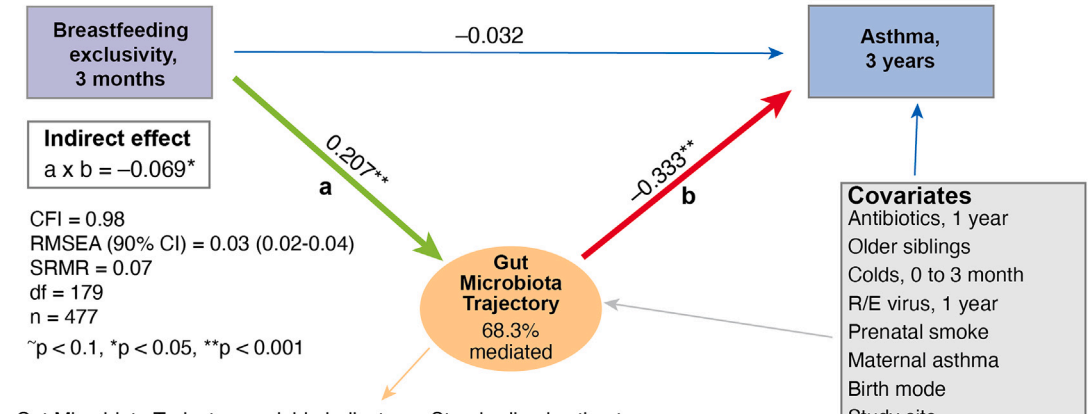
A

Nasal Microbiota Trajectory variable indicators - Standardized estimates

Richness Trajectory	Diversity Trajectory	Late Colonizers	Alloprevotella 14	Alloprevotella 25	Campylobacter 3	Fusobacterium 27
0.869	0.623	0.960	0.434	0.273	0.374	0.518
Lactobacillu delbrueckii 1	Lactococcus lactis	Leptotrichia 12	Neisseria polysaccharea	Pantoea 4	Porphyromonas 19	Prevotella 42
0.181	0.266	0.335	0.492	0.288	0.531	0.523
Rothia 12	Streptobacillus 1	Streptococcus salivarius 5	Streptococcus 118	Unclassified Saccharimonadales 2	Veillonella 53	
0.370	0.349	0.300	0.420	0.307	0.446	



B



Gut Microbiota Trajectory variable indicators - Standardized estimates

Richness Trajectory	Diversity Trajectory	Late Colonizers	Agathobacter 6	Blautia 9	Clostridioides difficile	
0.961	0.874	0.904	0.548	0.356	-0.006	
Eubacterium eligens 1	Faecalibacterium 13	Flavonifractor	Monoglobus	Ruminococcus gnavus	Sellimonas intestinalis	Tyzzerella 3
0.407	0.569	0.585	0.688	0.477	0.464	0.507

(legend on next page)

Figure S9. The trajectories of nasal and gut microbiota mediate the association between breastfeeding exclusivity and asthma in separate models, related to Figure 7

Separate structural equation models showing the mediating effects of nasal (A) and gut (B) microbiota trajectories implicated in asthma on the association between breastfeeding exclusivity at 3 months and asthma at 3 years. Standardized beta-coefficients are reported, with *p* values in parentheses. Positive associations with microbiota trajectories indicate delayed microbiota colonization (later), whereas negative associations indicate earlier colonization. Variable selection for latent constructs was informed based on univariate associations with both asthma and breastfeeding. Late colonizers (the proportion of microbiota classified as late colonizers), richness trajectories (change in observed OTUs), and diversity trajectories (change in Shannon index) were summary variables selected based on significant associations with asthma at 3 years and breastfeeding exclusivity (no vs. exclusive breastfeeding at 3 months) for nasal or gut microbiota in Mann-Whitney U tests ($p < 0.05$). Individual microbiota prevalence trajectories that were associated with asthma at 3 years and breastfeeding exclusivity in permutation tests ($p < 0.05$, see Table S4) were also included. This analysis was performed for infants who had data on nasal microbiota trajectories (A, 80 infants with asthma and 1,236 infants without asthma diagnosed at 3 years), and a second analysis was performed for infants who had gut microbiota trajectory data (B, 56 infants with asthma and 421 infants without asthma diagnosed at 3 years) breastfeeding exclusivity was an ordered variable (exclusive, partial, or no human milk at 3-month sampling). See Table S7 for details.

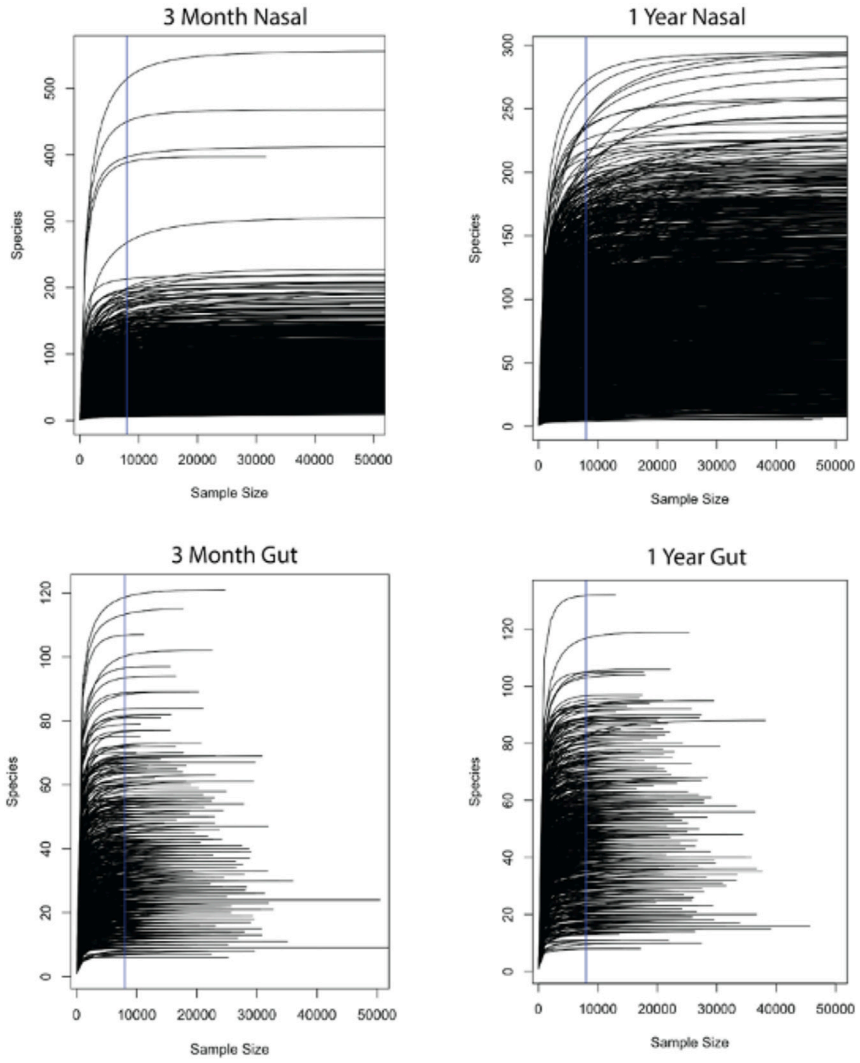


Figure S10. Rarefaction curves for all samples that were used to set a threshold for minimum number of reads, related to STAR Methods
The blue line is the rarefaction threshold of 8,000 reads per sample.