Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in the infant gut

Laursen, Martin Frederik; Sakanaka, Mikiyasu; von Burg, Nicole; Mörbe, Urs; Andersen, Daniel; Moll, Janne Marie; Pekmez, Ceyda Tugba; Rivollier, Aymeric; Michaelsen, Kim F.; Mølgaard, Christian; Lind, Mads Vendelbo; Dragsted, Lars Ove; Katayama, Takane; Frandsen, Henrik Lauritz; Vinggaard, Anne Marie; Bahl, Martin Iain; Brix, Susanne; Agace, William; Licht, Tine R; Roager, Henrik Munch

Published in:
Nature Microbiology

DOI:
10.1038/s41564-021-00970-4

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Laursen, M. F., Sakanaka, M., von Burg, N., Mörbe, U., Andersen, D., Moll, J. M., Pekmez, C. T., Rivollier, A., Michaelsen, K. F., Mølgaard, C., Lind, M. V., Dragsted, L. O., Katayama, T., Frandsen, H. L., Vinggaard, A. M., Bahl, M. I., Brix, S., Agace, W., Licht, T. R., & Roager, H. M. (2021). Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in the infant gut. Nature Microbiology, 6(11), 1367-1382. https://doi.org/10.1038/s41564-021-00970-4
Breastfeeding profoundly shapes the infant gut microbiota, which is critical for early life immune development, and the gut microbiota can impact host physiology in various ways, such as through the production of metabolites. However, few breastmilk-dependent microbial metabolites mediating host–microbiota interactions are currently known. Here, we demonstrate that breastmilk-promoted Bifidobacterium species convert aromatic amino acids (tryptophan, phenylalanine and tyrosine) into their respective aromatic lactic acids (indolelactic acid, phenyllactic acid and 4-hydroxyphenyllactic acid) via a previously unrecognized aromatic lactate dehydrogenase (ALDH). The ability of Bifidobacterium species to convert aromatic amino acids to their lactic acid derivatives was confirmed using monocolonized mice. Longitudinal profiling of the faecal microbiota composition and metabolome of Danish infants (n = 25), from birth until 6 months of age, showed that faecal concentrations of aromatic lactic acids are correlated positively with the abundance of human milk oligosaccharide-degrading Bifidobacterium species containing the ALDH, including Bifidobacterium longum, B. breve and B. bifidum. We further demonstrate that faecal concentrations of Bifidobacterium-derived indolelactic acid are associated with the capacity of these samples to activate in vitro the aryl hydrocarbon receptor (AhR), a receptor important for controlling intestinal homeostasis and immune responses. Finally, we show that indolelactic acid modulates ex vivo immune responses of human CD4+ T cells and monocytes in a dose-dependent manner by acting as an agonist of both the AhR and hydroxycarboxylic acid receptor 3 (HCA3). Our findings reveal that breastmilk-promoted Bifidobacterium species produce aromatic lactic acids in the gut of infants and suggest that these microbial metabolites may impact immune function in early life.

Human breastmilk is a well-adapted nutritional supply for the infant. Breastfeeding provides the infant with important short-term protection against infections and may also provide long-term metabolic and immunological benefits. These benefits may partly be mediated through the gut microbiota, since breastfeeding is the strongest determinant of gut microbiota composition and function during infancy. Human breastmilk contains human milk oligosaccharides (HMOs), which are complex, highly abundant sugars serving as substrates for specific microbes including certain species of Bifidobacterium. This co-evolution between bifidobacteria and the host, mediated by HMOs, to a large extent directs the colonization of the gut in early life, which has critical impact on the immune system. Depletion of specific microbes, including Bifidobacterium, in early life has been associated with increased risk of allergy and asthma development in childhood and has been suggested to compromise immune function and lead to increased susceptibility to infectious disease. Despite Bifidobacterium dominating the gut of breastfed infants and being widely acknowledged as beneficial, mechanistic insights into the contribution of these bacteria and their metabolites to immune function and development are limited and have mainly focused on short-chain fatty acids. Recent studies show that microbial aromatic amino acid metabolites including tryptophan-derived indoles via activation of the aryl hydrocarbon receptor (AhR) can fortify the intestinal barrier and protect against pathogenic infections and influence host metabolism, making this group of microbial metabolites of particular interest in the context of early life.

Here, we show that breastmilk-promoted Bifidobacterium species, via a previously unrecognized aromatic lactate dehydrogenase (ALDH), produce aromatic lactic acids including indolelactic acid (ILA) in substantial amounts in the infant gut and that faecal concentrations of this metabolite are correlated with the capacity of infant faeces to activate AhR. We furthermore demonstrate that ILA via AhR and hydroxycarboxylic acid receptor 3 (HCA3)-dependent pathways impact immune functions ex vivo, suggesting that...
breastmilk-promoted *Bifidobacterium* via production of aromatic lactic acids impact the immune system in early life.

**Results**

*Bifidobacterium* associate with aromatic lactic acids in the infant gut. To explore interactions between breastfeeding status, gut microbial composition and metabolism of aromatic amino acids in early life, we used 16S ribosomal RNA amplicon sequencing to infer gut microbiota composition and a targeted ultraperformance liquid chromatography–mass spectrometry (UPLC–MS) metabolomics approach to quantify 19 aromatic amino acids and derivatives thereof (Supplementary Tables 1 and 2) in faecal samples from 59 healthy Danish infants from the SKOTI cohort22. The SKOTI infants included were born full term, 9.1 ± 0.3 (mean ± s.d.) months of age at sampling and 40.7% were still partially breastfed (Supplementary Data 1a,b). After stratification of the 9-month-old infants on the basis of breastfeeding status (partially breastfed versus weaned), principal coordinates analysis (PCoA) of weighted UniFrac distances showed a significant separation across the first PC-axis ($r^2 = 0.093$, $P < 0.001$, Adonis test; Fig. 1a), which mirrored an increasing gradient in relative abundance of *Bifidobacterium* ($r^2 = 0.397$, $P < 0.001$, Adonis test; Fig. 1b). Other metadata (age, gender, mode of delivery, current formula intake and age of introduction to solid foods) did not explain gut microbiota variation to the same degree as breastfeeding status ($r^2 < 0.05$, $P > 0.03$, Adonis tests; Supplementary Data 1c) and no bacterial genera differed significantly according to these parameters (FDR-corrected $P > 0.1$, Mann–Whitney U-tests; Supplementary Data 1d).

Principal component analysis (PCA) of faecal amino acid metabolite concentrations (Supplementary Data 1e) also suggested a minor separation by breastfeeding status, which was largely driven by three aromatic lactic acids, 4-hydroxyphenyllactic acid (4-OH-PLA), phenyllactic acid (PLA) and indolelactic acid (ILA) (Fig. 1c). Correlation analysis revealed that *Bifidobacterium*, but no other bacterial genera, were significantly associated with faecal concentrations of all three aromatic lactic acids (4-OH-PLA, PLA and ILA), in addition to indolealdehyde (IAld) (Fig. 1d and Supplementary Data 1f). The *Bifidobacterium* species (Extended Data Fig. 1a and Supplementary Data 1g) enriched in the breastfed infants, *B. longum*, *B. bifidum* and *B. breve*, were positively associated with the faecal concentrations of aromatic lactic acids (4-OH-PLA, PLA and ILA) and IAld (cluster 1 in Fig. 1e) but negatively associated with the faecal concentrations of aromatic propionic acids, aromatic amino acids and, to a lesser degree, with aromatic acetic acids (cluster 2 in Fig. 1e). In contrast, postweaning *Bifidobacterium* species, including *B. adolescentis*, *B. animalis*/*pseudo*longum and *B. catenulatum* group3,23, were not significantly associated with aromatic lactic acids nor breastfeeding status (Fig. 1e). These associations were in agreement with the observation that the concentrations of the three aromatic lactic acids were higher in the faeces of breastfed than weaned infants (Extended Data Fig. 1b). Furthermore, the abundances of the three aromatic lactic acids in infant urine (Supplementary Figs. 1–3) showed similar positive associations with relative abundances of breastmilk-promoted *Bifidobacterium* species (Extended Data Fig. 1c). In addition, faecal and urinary levels of ILA were positively correlated ($r = 0.68$, $P < 0.0001$), showing that faecal levels of this metabolite are reflected systemically. Consistently, urine abundance of ILA, but not of PLA and 4-OH-PLA, were significantly higher in breastfed compared to weaned infants (Extended Data Fig. 1b). Together, this suggests that specific *Bifidobacterium* species produce aromatic lactic acids in the infant gut (Fig. 1f).

*Bifidobacterium* species produce aromatic lactic acids in vitro. To confirm the ability of *Bifidobacterium* species detected in infants to produce aromatic lactic acids, *Bifidobacterium* type strains were grown anaerobically in a medium containing all three aromatic amino acids with either glucose or HMOs as sole carbohydrate sources. Analyses of culture supernatants revealed that ILA, PLA and 4-OH-PLA were produced mainly by *B. bifidum*, *B. breve*, *B. longum* ssp. *longum*, *B. longum* ssp. *infantis* and *B. scardovii* (Fig. 2a), in accordance with the associations observed in the 9-month-old infants (Fig. 1e). Other *Bifidobacterium* species, namely *B. adolescentis*, *B. animalis* ssp. *lactis*, *B. animalis* ssp. *animalis*, *B. dentium*, *B. catenulatum*, *B. pseudocatenulatum* and *B. pseudolongum* ssp. *pseudolongum* produced only low amounts of these metabolites (Fig. 2a). The ability of *Bifidobacterium* species to produce high levels of the aromatic lactic acids was generally congeneral with the ability to use HMOs as a carbohydrate source (Fig. 2a), suggesting a link between breastmilk-promoted bifidobacteria and production of aromatic lactic acids. None of the downstream products of the aromatic lactic acids (Fig. 1f) was detected in any of the culture supernatants.

Identification of a responsible ALDH. Since it has been reported that a lactate dehydrogenase (LDH) in *Lactobacillus* species can convert phenylpyruvic acid to PLA24, we hypothesized that a corresponding enzyme would be present in *Bifidobacterium* species. Alignment and phylogenetic analysis of all genes annotated as *ldh* in the *Bifidobacterium* type strains included in this study, revealed four clusters (Fig. 2b). Whereas all *Bifidobacterium* genomes contain the *ldh* (here designated as type 1 *ldh*) responsible for conversion of pyruvic acid to lactic acid in the bifidobacterial fructose 6-phosphate shunt25,26, some species have an extra *ldh*, here designated as type 2, type 3 and type 4, respectively. In agreement with the in vitro fermentations (Fig. 2a), all prominent aromatic lactic acid-producing *Bifidobacterium* species contain the type 4 *ldh*, suggesting that this could encode a previously unrecognized ALDH. A further analysis of all available whole-genome sequenced *Bifidobacterium* strains showed that the type 4 *ldh* is universally present in *B. longum*, *B. bifidum*, *B. breve* and *B. scardovii* strains (Supplementary Table 3). Interestingly, genomic analysis of the *Bifidobacterium* type strains revealed that the type 4 *ldh* gene is part of a genetic element containing an amino acid transaminase gene (suspected to be responsible for converting the aromatic amino acids into aromatic pyruvic acids) and a haloacid dehalogenase gene (of unknown importance) (Supplementary Fig. 4), which has been indicated to constitute an operon in *B. breve*27. Cloning of the type 4 *ldh* gene from the type strain of *B. longum* ssp. *infantis* (DSM200888) into a vector transformed into *Escherichia coli* revealed that the expression of the type 4 *ldh* gene indeed resulted in the appearance of PLA, 4-OH-PLA and ILA in the culture supernatant (Fig. 2c). To verify the type 4 *ldh*-dependent production of aromatic lactic acids in *Bifidobacterium* species, we generated a type 4 *ldh* insertion mutant strain by homologous recombination in *B. longum* ssp. *longum* 105-A (Supplementary Fig. 5), a genetically tractable strain28,29 containing the type 4 *ldh* (Supplementary Fig. 6a). The type 4 LDH amino acid sequence of the 105-A strain had >98% identity to the homologues in type strains of *B. bifidum* and *B. breve* and *B. scardovii* (Supplementary Fig. 6b) but no non-bifidobacterial homologues were found by BLAST analysis (amino acid sequence identity cutoff 60%). Cultivation of the wild type (WT), the type 4 *ldh* mutant strain and a complemented type 4 *ldh* mutant strain in a medium containing the three aromatic amino acids confirmed that type 4 *ldh* disruption did not impair growth in a rich medium (Fig. 2d). ILA, PLA and 4-OH-PLA accumulated in the supernatant of the WT and of the complemented type 4 *ldh* mutant strains but not in the type 4 *ldh* mutant (Fig. 2e). Importantly, the type 4 *ldh* mutant was not significantly compromised in its ability to convert pyruvic acid to lactic acid (Fig. 2e), supporting the distinct role of type 4 *ldh* in converting aromatic pyruvic acids. Further, to demonstrate in vivo production of the indicated aromatic lactic acids, we monoclonized germ-free mice with either the WT or the type 4 *ldh* mutant.
Breastfeeding associates with faecal microbiota composition and aromatic amino acid metabolites in 9-month-old infants. a, b, PCoA plots of weighted UniFrac distances based on OTUs from faecal samples of 9-month-old infants participating in the SKOT cohort (n = 59). a, Samples are coloured according to breastfeeding status, with ellipses indicating 80% confidence interval (CI) of datapoints for partially breastfed (red, n = 24) and weaned (blue, n = 35) infants. b, Samples are coloured according to relative abundance of the genus Bifidobacterium. c, PCA plot of concentrations (nmol g⁻¹ faeces) of aromatic amino acids and their derivatives in SKOT faecal samples, coloured according to breastfeeding status, with ellipses indicating 80% CI of datapoints for breastfed (red, n = 24) and weaned (blue, n = 35) infants. Loadings (correlations between variables and the principal components) are shown with arrows, with annotations of the aromatic amino acids ILA, 4-OH-PLA and PLA shown. d, Heatmap illustrating Spearman’s rank correlation coefficients (two-sided tests) between the relative abundance of Bifidobacterium and concentrations of aromatic amino acids and their derivatives in SKOT faecal samples (n = 59). e, Heatmap illustrating hierarchical clustering (dendrogram on the right side) of Spearman’s rank correlation coefficients (two-sided tests) between the relative abundance of the different Bifidobacterium species and selected microbial-derived aromatic amino acid catabolites in SKOT faecal samples (n = 59). Box and whiskers plots are showing relative abundance (line, median; box, interquartile range (IQR); whiskers, minimum to maximum) of the Bifidobacterium species, stratified according to breastfeeding status, with statistical significance evaluated by two-sided Mann-Whitney U-test. f, The pathway of aromatic amino acid catabolism by gut microbes (modified from Smith and Macfarlane[110], Smith and Macfarlane[111] and Zelante et al.[113]). For all panels, asterisks indicate statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001.

4 ldh mutant strain and found a 20–60-fold increase in their concentrations in WT versus type 4 ldh mutant monoclonized mice (Extended Data Fig. 2). Purification and characterization of the recombinant type 4 LDH enzyme revealed that it had a mass of 33.9 kDa (Supplementary Fig. 7a), while the native molecular mass was estimated to be 71.9 kDa by size exclusion chromatography, indicating dimer formation in solution (Supplementary Fig. 7b). Lack of added metal ions or addition of ethylenediaminetetraacetic acid (EDTA) did not reduce enzymatic activity, the optimal pH was 8.0–8.5 and the enzyme was most stable at 37 °C (Supplementary Fig. 7c–e). Heterotrophic effects were neither observed for fructose 1,6-bisphosphate (an allosteric effector for type 1 LDH) nor for several intermediates for aromatic amino acid synthesis[26,27] (Supplementary Fig. 8). However, we found that phosphate served as a positive effector, suggesting that type 4 LDH is an intracellular enzyme (Supplementary Fig. 9a,b). Assay performed at the different
Fig. 2 | *Bifidobacterium* species produce aromatic lactic acids via an ALDH. **a.** In vitro production of ILA, PLA and 4-OH-PLA by *Bifidobacterium* species type strains in modified MRS medium (MRSc) with 2% (w/v) glucose or a mix of HMOs as sole carbohydrate source. For the type strains of *B. adolescentis*, *B. animalis* ssp. *lactis*, *B. dentium* and *B. catenulatum*, no or very poor growth (OD$_{600}$nm < 0.1) was observed with HMOs as the carbohydrate source. Means of three biological replicates and doubling times reported as mean ± s.e., Production of ILA, PLA and 4-OH-PLA by wild type, Type4$_{-ldh}$ mutant (Type4$_{-ldh}$ mutant) and the type 4 _ldh_ mutant (Type4$_{-ldh}$ mutant) in modified MRS medium (MRSc) with 2% (w/v) glucose or a mix of HMOs as sole carbohydrate source. Means of three biological replicates are shown. **b.** In vitro production of ILA, PLA and 4-OH-PLA by species type strains annotated as LDH-encoding genes in the *Bifidobacterium* species type strains annotated as LDH-encoding genes (Supplementary Fig. 9b). The catalytic rate (k$_{cat}$/K$_{m}$) was highest for indolepyruvic acid (194 s$^{-1}$·mM$^{-1}$), followed by 4-hydroxyphenylpyruvic acid (16 s$^{-1}$·mM$^{-1}$) and phenylpyruvic acid (11 s$^{-1}$·mM$^{-1}$), suggesting preference for indolepyruvic acid. The observed Hill coefficient (n$_{H}$ = 1.0–1.4) for all substrates indicate weak positive cooperativity under the conditions tested. Collectively, these results show that the type 4 _ldh_ gene (from now on denoted _aldh_) encodes an ALDH responsible for the production of ILA, PLA and 4-OH-PLA in *Bifidobacterium* species associated with breastfeeding. **Bifidobacterium** species govern aromatic lactic acid profiles during early infancy. To study the dynamics of *Bifidobacterium* species establishment and aromatic lactic acids in early infancy, we
established the Copenhagen Infant Gut (CIG) cohort including 25 healthy breastfed or mixed-fed infants, sampled every 2–4 weeks from birth until the age of 6 months (Supplementary Data 2a,b) for microbiome profiling (Supplementary Data 2c and Extended Data Fig. 3) and targeted metabolite quantification including aromatic lactic acids (Supplementary Table 4 and Supplementary Data 2e). PCoA using Bray–Curtis dissimilarities showed very similar results (Extended Data Fig. 8a). Finally, qPCR for microbiome profiling (Supplementary Data 2c and Extended Data Fig. 3) and targeted metabolite quantification including aromatic lactic acids (Fig. 4b). On the basis of quantitations of the aromatic lactic acids, Bifidobacterium was highly dominated by B. longum (38.5%), B. breve (9.1%), B. bifidum (7.9%), B. catenulatum group (6.4%) and B. dentium (1.7%) were found (Extended Data Fig. 3a,b), with the remaining Bifidobacterium species being assigned to B. scardovii (0.24%), B. adolescentis (0.15%) and B. animalis/pseudolongum (0.10%) (Supplementary Data 2e). PCoA using Bray–Curtis dissimilarities, revealed a separation of the communities across samples based on relative abundance of the five dominating Bifidobacterium species, B. longum, B. bifidum, B. breve, B. catenulatum group and B. dentium (Fig. 4a and Extended Data Fig. 4b–f). Community abundance of B. longum, B. bifidum and B. breve but not B. catenulatum group and B. dentium (Fig. 4a) matched the measured faecal concentrations of the aromatic lactic acids (Fig. 4b). On the basis of quantitative PCR (qPCR) estimated total bacterial load of all samples, we calculated absolute abundances of each bacterial taxon in the 16S rRNA amplicon dataset and defined infant-type Bifidobacterium species as the summarized abundance of B. longum, B. bifidum, B. breve and B. scardovii. We observed a significant increase in the absolute abundance of infant-type Bifidobacterium species from birth to around 6 months of age and this occurred concurrently with a progressive increase in the faecal concentrations of ILA, PLA and 4-OH-PLA and a progressive decrease in faecal abundances of HMO residuals (Extended Data Fig. 5). We confirmed by linear mixed models adjusting for subject and age that the absolute abundances of the infant-type Bifidobacterium species were positively associated with faecal levels of ILA, PLA and 4-OH-PLA and additionally negatively associated with abundances of HMOs residuals in faeces (Extended Data Fig. 5). Among all bacterial taxa detected, B. longum, B. bifidum and B. breve were most strongly associated with faecal levels of ILA, PLA and 4-OH-PLA (Supplementary Data 2g). These associations were also evident within individuals when using repeated measures correlations (Extended Data Fig. 6) and across individuals at each sampling point using Spearman’s rank correlations (Extended Data Fig. 7). Furthermore, re-analysing the microbiome data at the amplicon sequence variant (ASV) level further corroborate our findings regarding the relevance in early life and impact of breastfeeding, we mined a published metagenomic dataset from faecal samples from a cohort of 98 Swedish mother–infants pairs’ for bifidobacterial metagenome-assembled genomes (MAGs) containing the aldhl gene. This analysis revealed a

Fig. 3 | Kinetic characterization of the bifidobacterial ALDH (type 4 LDH). a–d. Substrate saturation curves of the type 4 LDH obtained for indolepyruvic acid (a), phenylpyruvic acid (b), 4-hydroxyphenylpyruvic acid (c) and pyruvic acid (d). Kinetic parameters, which were calculated by curve-fitting two independent experimental data to the Hill equation, are shown in the insets. The reaction was carried out in the presence of 100 mM phosphate.
Infant-type *Bifidobacterium* species determine faecal aromatic lactic acid concentrations during early infancy. a, b, PCA plots of Bray-Curtis dissimilarities (n = 234 (i)), coloured according to relative abundance of *Bifidobacterium* (a) or log$_{10}$-transformed concentration (nmol per g of faeces) of aromatic lactic acids (sum of ILA, PLA and 4-OH-PLA) (b) in faeces of infants participating in the CIG cohort. Dashed lined circles include communities dominated (relative abundance >50%) by *B. longum*, *B. bifidum*, *B. breve*, *B. catenulatum* group or *B. dentium* (*B. adolescentis*, *B. scardovii* and *B. animalis*/pseudolongum never dominated any of the communities; Extended Data Figs. 3 and 4). (i) Six samples were omitted from the analyses due to low read counts (<8,000), and one sample was omitted due to no available metabolomics data. c, Heatmap illustrating linear mixed-model coefficients (two-sided test, adjusted for subject and age) between the absolute abundance of *Bifidobacterium* species (cells per g of faeces) and faecal concentrations (nmol per g of faeces) of aromatic lactic acids (ILA, PLA and 4-OH-PLA, n = 240 (ii)) or faecal relative abundances of HMOs (2′F, 3′FL, 2′/3′-O-fucosyllactose; 3′SL/6′SL, 3′/6′-O-sialyllactose; LNT/LNNT, lacto-N-tetraose/lacto-N-neotetraose; n = 228 (iii)) in the CIG cohort. Infant-type *Bifidobacterium* species is the sum of absolute abundances of *B. longum*, *B. breve*, *B. bifidum* and *B. scardovii*. Statistical significance was evaluated by FDR-corrected P values indicated by asterisks with *P < 0.05*, **P < 0.01**, ***P < 0.001*** and ****P < 0.0001. (ii) One sample was omitted due to no available metabolomics data. (iii) Twelve samples were omitted due to the infants no longer being breastfed and one due to no available metabolomics data. d–f, Absolute abundance of infant-type *Bifidobacterium* species is indicated with grey background shading. d, Breastfed infants colonized early with infant-type *Bifidobacterium* species and with concurrent high concentrations of ILA, PLA and 4-OH-PLA through the first 6 months of life. e, Infants colonized late with infant-type *Bifidobacterium* species and with concurrent low concentrations of ILA, PLA and 4-OH-PLA. f, Infants with recorded oral antibiotics intake during the first 6 months of life. Similar dynamics of the remaining infants can be seen in Extended Data Fig. 10.
The tryptophan-derived metabolite ILA was consistently measured among the latter, infants CIG08 and CIG09 were twins, despite breastfeeding (Fig. 4e and Extended Data Fig. 10b). We noticed lower concentrations of the aromatic lactic acids, in particular of Bifidobacterium species, had relatively low faecal concentrations of aromatic lactic acids until age 172 days, when B. breve replaced B. dentium (Fig. 4c), consistent with the fact that B. dentium lacks the aldhl gene while B. breve contains it (Fig. 2b and Supplementary Table 3). Finally, in the three infants treated with antibiotics during our study, Bifidobacterium species abundances were temporarily decreased simultaneously with reduced concentrations of the aromatic lactic acids (Fig. 4f). Together, these results demonstrate that HMO-using infant-type Bifidobacterium species determine the abundance of aromatic lactic acids in the infant gut. Yet, the impact of early/late Bifidobacterium colonization, preterm delivery, exposure to antibiotics and formula supplementation with respect to bifidobacterial aromatic lactic acid production warrants further investigation.

Indolelactate modulates immune responses via AhR and HCA3. The tryptophan-derived metabolite ILA was consistently measured in the faeces of breastfed infants at 0–6 months (Supplementary Table 4) and 9 months of age (Supplementary Table 2). Microbial tryptophan catabolites have been found to contribute to intestinal and systemic homoeostasis, in particular by their ability to bind the AhR. Furthermore, aromatic lactic acids have been found to activate HCA3, which is involved in the regulation of immune function and energy homoeostasis. In accordance with previous reports, we observed modest but significant dose-dependent increases in agonistic activity of ILA in both rat and human AhR reporter gene cell lines (Supplementary Fig. 11). Furthermore, all three aromatic amino lactic acids, and especially ILA, showed very potent and dose-dependent agonistic activity towards the HCA3.

Subject- and age-adjusted β coefficients (versus FDR-corrected P values) measured in the same samples. Coloured circles/triangles mark taxa/metabolite measures that are significantly positively (red) or negatively (blue) associated with AhR activity within an FDR-corrected P of 0.1 (dashed line). ILA, indolelactic acid; Tyr, tyrosine.
inhibited IL-22 production, further corroborating that ILA acts through AhR to induce IL-22 production (Fig. 6b). We also isolated monocytes from human blood, where both AhR (ref. 45) and HCA3 (ref. 31) are expressed, stimulated the cells with *E. coli* lipopolysaccharide (LPS) and interferon-gamma (IFN-γ) to induce pro-inflammatory conditions, and assessed IL-12p70 production on ILA exposure. ILA reduced pro-inflammatory IL-12p70 production in a dose-dependent manner (Fig. 6c). Addition of CH-223191 blocked the ILA-induced inhibition of IL-12p70 production, confirming that ILA also acts through AhR in human monocytes (Fig. 6d). Furthermore, ILA-induced inhibition of IL-12p70 was prevented, when using knockdown of HCA3, by small interfering RNA (siRNA), supporting that ILA also acts as an anti-inflammatory agent via HCA3, in human monocytes (Fig. 6c). Thus, ILA affects human immune responses via AhR and HCA3-dependent pathways, suggesting that *Bifidobacterium*-derived ILA is a highly relevant AhR and HCA3 agonist that may impact immune responses in early life.

**Discussion**

The importance of intestinal commensal bacteria in regulation of the intestinal barrier function and immune development during infancy is well established.\(^{47,48}\) Yet, specifically the symbiotic role of the breastmilk-promoted *Bifidobacterium* species, which are highly abundant in many breastfed infants, remains largely unknown. Here, we identified an ALDH, which catalyses the last step of the conversion of aromatic amino acids into their respective aromatic propionic acids as end products.\(^{18}\) We show that only the infant-type *Bifidobacterium* species, adapted to breastfeeding by their HMO-transport and degradation genes providing them with a colonization advantage in infant gut,\(^{34–42}\), contain the ALDH gene. This fits the observation that *Bifidobacterium* species commonly isolated from the infant gut in vitro produce relative higher levels of ILA compared with adult- or animal-associated *Bifidobacterium* species.\(^{31}\) Our enzymatic assays showed strong adaptation of ALDH towards indolepyruvic acid, resulting preferentially in the formation of ILA. Importantly, our data suggest that the production of the AhR agonist ILA by breastmilk-promoted *Bifidobacterium* is a key determinant of AhR-dependent signalling in the gut during infancy. Previous studies have found that ILA decreases inflammation in intestinal cells through activation of AhR.\(^{35}\) Here, we show ex vivo that ILA impacts human primary immune cells via AhR- and HCA3-dependent pathways. The observed dose- and AhR-dependent stimulation of IL-22 secretion by ILA may reflect a mechanism by which infant-type *Bifidobacterium* species impact intestinal homoeostasis in early life, as IL-22 for example provides protection against gastrointestinal pathogens\(^{26–28}\) and promotes mucus production\(^{29,30}\) and epithelial function.\(^{31}\) Further, the AhR- and HCA3-dependent inhibitory effect of ILA on IL-12p70 secretion by monocytes may constitute a means by which infant-type *Bifidobacterium* species contribute to the regulation of the pro-inflammatory responses to LPS derived from Enterobacteriaceae species that also often colonize the neonatal/infant gut. While the functional implications of aromatic lactic acids remain to be fully characterized, the phenomenon observed here is probably of fundamental importance, since HCA3, which is only expressed in humans and other hominids, is involved in the regulation of immune functions and energy homeostasis.\(^{38,40}\) Furthermore, AhR signalling is involved in protection against gastrointestinal pathogens\(^{15}\) and plays a key role in enhancement of intestinal barrier function\(^{64,65}\), immune development\(^{66–68}\), attenuation of induced colitis\(^{69}\), autoimmunity\(^{70,71}\) and metabolic syndrome.\(^{72}\) In addition, ILA and PLA have been shown in vitro to have direct antibacterial\(^{73,74}\) and antifungal properties.\(^{75,76}\) Therefore, our findings provide a rationale for further investigation of the implications of aromatic lactic acids in infant health and immune development.
Methods

Human study populations and metadata. SKOT cohort. The discovery cohort consisted of a random subset of 59 healthy infants (30 male, 29 female) of the observational SKOT I cohort\(^1\). No statistical method was used to predetermine sample size. The infants were originally recruited from Copenhagen Frederiksberg regions by random selection from the National Danish Civil Registry\(^2\). Inclusion criteria were single birth and full-term delivery, absence of chronic illness and age of 9 months ± 2 weeks at inclusion. Mode of delivery, gender, age at sampling, use of medication, breast- and formula-feeding prevalence, as well as exclusive and total breastfeeding duration and age of introduction to solid foods were recorded by parental questionnaires (Supplementary Data 1a,b). Anthropometrics, full dietary assessment and other relevant metadata have been published previously\(^3\). Faecal samples were obtained at 9 months ± 2 weeks of age and were stored at −80 °C until DNA extraction, as described previously\(^4\). Urine samples were collected by the use of cotton balls placed in the infants’ disposable nappies from which the urine was squeezed into a sterile tube and stored at −80 °C. In cases of faeces in the nappy, the urine sample was discarded. The study protocol was approved by the Committees on Biomedical Research Ethics for the Capital Region of Denmark (H-KF-2007–0003) and The Data Protection Agency (2002-54-0938, 2007-54-026) approved the study. Informed consent was obtained from all parents of infants participating in the SKOT I study. The parents did not receive any compensation.

CIG cohort. The validation cohort, CIG, consisted of 25 healthy infants (12 male, 13 female), mostly vaginally born (23/25) and full-term delivered. No statistical method was used to predetermine sample size. However, on the basis of the original observations using the SKOT cohort, we estimated that 25 infants with multiple time points would be sufficient to demonstrate the dynamics between Bifidobacterium species and aromatic lactics acids. Infants in CIG were recruited through social media and limited to the Copenhagen region. Parents collected faecal samples approximately every second week, starting from the first week of life until age 9 months (between weeks 0, 2, 4, 6, 8, 10, 12, 16, 20 and 24), ending with a total of 269 samples. Parents were instructed to collect faecal samples from nappies into sterile faeces collection tubes (Sarstedt) and immediately store them at −18 °C in a home freezer until transportation to the Technical University of Denmark where the samples were stored at −80 °C until sample preparation. Gender, preterm versus full-term birth, mode of delivery, infant/maternal antibiotics, feeding patterns (breastmilk versus formula) and introduction to solid foods were recorded (Supplementary Data 2a,b). The Data Protection Agency (18/02459) approved the study. The Committees on Biomedical Research Ethics for the Capital Region of Denmark confirmed that the CIG study was not notifiable according to the Act on Research Ethics Review of Health Research Projects (paragraph 1, subsection 4), as the study only concerned the faecal microbial composition and activity and not the health of the children. Informed consent was obtained from all parents of infants participating in the CIG study. In addition, parents of twins gave informed consent to publish data from the twins although the parents themselves would be able to identify their children using indirect identifiers. The parents did not receive any compensation.

Gut microbiota analysis. 16S rRNA gene amplicon sequencing. Sample preparation and sequencing were performed as previously described\(^5\), using a subset of 59 faecal samples originating from infants participating in the SKOT I cohort and 241 faecal samples originating from infants participating in the CIG cohort (Supplementary Table 4). A total of 28 samples were missing due to insufficient sample material (n = 1), insufficient DNA extraction/loss of PCR product (n = 20), very low number of sequencing reads (n = 6) or resemblance of community to sequenced blank buffer DNA extraction negative controls (n = 1). Briefly, DNA was extracted from 250 mg of faeces or blank buffer negative controls (PowerLyzer PowerSoul DNA isolation kit, MoBio, Carlsbad, CA, USA) with a total of 16S RNA gene was amplified (30s at 98 °C, 24–30 cycles of 15 s at 98 °C and 30 s at 72 °C, followed by 5 min at 72 °C) using non-degenerate universal barcoded primers including sequencing adaptors (Supplementary Table 5)\(^6\) and then sequenced with the Ion OneTouch and Ion PGM platform with a 318-chip v.2. Sequences from SKOT and CIG were processed separately. After demultiplexing and trimming to the GreenGenes standard length, the taxonomy of these resulting OTUs (using the RDP database v.18) above the 97.5% of total community (Supplementary Data 2f). As for OTUs, the taxonomy for individual ASV analysis on the CIG cohort samples using the DADA2 pipeline v.1.14 (ref. 83) was estimated by total sum scaling. Taxonomy was assigned to the OTUs using the GreenGenes database v.13.8 (ref. 81). Estimating species composition in the CIG cohort, the OTUs were detected with identical taxonomy were collapsed and using a cutoff of average abundance cutoff of 0.1% of the total Bifidobacterium population (Supplementary Data 2f). For OTUs, the taxonomy for individual Bifidobacterium ASVs was confirmed by BLAST search against the 16S rRNA gene database at the National Center for Biotechnology Information (NCBI). The top BLAST hit indicated species annotation (Supplementary Data 1a and 2e). OTUs were collapsed into Bifidobacterium species (B. longum, B. bifidum, B. breve, B. infantis, B. adolescentis group, B. pseudolongum, B. pseudoscordovi) on the basis of the top BLAST hit (Supplementary Data 1a and 2e). In addition, to validate the findings from the OTU analysis, we performed ASV analysis on the CIG cohort samples using the DADA2 pipeline v.1.14 (ref. 83) with the demultiplexed and trimmed reads and the same cutoffs as for the OTU analysis (MAX_EE = 1, ASVs filtered to include only those with abundance across all samples above 0.005% of the total ASV counts), resulting in a total of 211 ASVs and 13 ASVs assigned to Bifidobacterium (using the RDP database v.18) above the abundance cutoff of 0.1% of the total Bifidobacterium population (Supplementary Data 2f). As for OTUs, the taxonomy for individual Bifidobacterium ASVs was confirmed by BLAST search against the 16S rRNA gene database at the National Center for Biotechnology Information (NCBI). Infant-type Bifidobacterium species were defined as the summarized abundance of B. longum, B. bifidum, B. breve and B. scordovi. CIG individuals were grouped on the basis of colonization with infant-type Bifidobacterium species, into those with early colonization (colonized within first month reaching average relative abundance >40% during first 6 months, n = 17) and late colonization (not达到 average relative abundance <0% during first 6 months, n = 5), as well as those associated with antibiotics (at least one episode of recorded oral antibiotics during the first 6 months of life, n = 3).

Quantitative PCR. Total bacterial load (universal primers) and absolute abundances of B. longum subspecies, B. longum sp. infantis (subspecies-specific primers), B. bifidum and B. breve (species-specific primers) were estimated by qPCR, using the primers listed in Supplementary Table 5. Each reaction was performed (in triplicates) with 5 μl of PCR-grade water, 1.5 μl of forward and reverse primer, 10 μl of SYBR Green Master Mix 2X (LightCycler 480 SYBR Green Master, Roche) and 2 μl of template DNA, in a total volume of 20 μl. Standard curves were generated from tenfold serial dilutions of linearized plasmid (containing 10^-6-10^-1 gene copies per μl), constructed by cloning a PCR-amplified 199-base pair (bp) fragment of the 16S rRNA gene (V3 region) of E. coli (ATCC 25922) or a 307-bp fragment of the Blon915 gene of B. longum sp. infantis (DSM 20088) or a 301-bp fragment of the BL0274 gene of B. longum sp. longum (DSM 20219) into a pCR4-Blunt-TOPO (Invitrogen) or pCR-Blunt-TOPO (Invitrogen) vector (Invitrogen). For B. bifidum (16S rRNA gene) and B. breve (geoEL gene), tenfold serial dilutions of DNA (containing 10^-6-10^-1 gene copies per μl) extracted from pure cultures of the type strains (DSM 20456 and DSM 20213) were used for standard curves. Plates were run on the LightCycler 480 Instrument II (Roche) with the programme including 5 min at 95 °C, 50–72 °C and 15 s at 72 °C and subsequent melting curve analysis including 5 min at 95 °C, 1 min at 65 °C and continuous temperature increase (ramp rate 0.1 °C s^-1) until 98 °C. Data were analysed with the LightCycler 480 Software (v.1.5) (Roche). Bacterial load data (using the universal primers) were used to estimate absolute abundances of each microbial taxon by multiplying with relative abundances derived from 16S rRNA gene amplicon sequencing.

Bifidobacterium strains and growth experiments. Aromatic lactic acid production by Bifidobacterium type strains. Bifidobacterium type strains (Supplementary Table 6) were cultivated on MRs (MRs containing 2% (w/v) glucose and supplemented with 0.05% (w/v) L-cysteine) agar plates for 48 h at 37 °C. Culturically, Single colonies were dissolved in 5.0 ml of preincubated Bifidobacterium broth and incubated for 24 h at 37 °C anaerobically with shake. The overnight (ON) cultures were washed (10,000g, room temperature, 5 min) and resuspended in sterile 0.9% NaCl water, diluted 1:20 in (triplicates) in preincubated MRs or MRs + HMOs (MRs broth without glucose but supplemented with 2.0% (w/v) HMO mixture and 0.05% (w/v) L-cysteine) and re-incubated at 37 °C anaerobically for 24 h, after which optical density OD_{620nm} was measured and the culture supernatants (16,000g, 5 min, 4 °C) were analysed by UPLC–MS. The individual HMOs were kindly donated by GlycoA/S, 2′-O-fucosylactose (2′FL), 3′-O-fucosylactose (3′FL), lacto-N-tetraose (LNT), lacto-N-neotetraose (LNNT), O-sialylactose (6 SL), O-sialylactose (3 SL), together representing the three types of structures found in human breastmilk (fucosylated, sialylated and neutral core). On the basis of the HMO composition in breastmilk\(^8\), these were mixed in a ratio of 53% 2′FL, 18% 3′FL, 13% LNT, 5% LNNT, 7% 6 SL and 4% 3 SL in sterile water to obtain a representative HMO mix used in the in vitro experiments at 2% (w/v).
Identification of aldhl gene/operon in Bifidobacterium strains, metagenomic data and homology searches. From the full genome sequences (available on NCBI Genome, https://www.ncbi.nlm.nih.gov/genome/) of Bifidobacterium type strains included in this study (Supplementary Table 6) all genes annotated as LDHs were identified using NCBI iBLASTn with default settings and a cutoff of 70% identity and 70% query coverage. Aligned genomic nucleotide sequences were translated and verified to match LDHs using reciprocal BLASTx against NCBI’s non-redundant database. In addition, the ALDH amino acid sequence (translated from the aldhl nucleotide sequence) of B. longum sp. longum 105-A was aligned (gap cost 10, gap extension cost 1) and subsequently a phylogenetic tree (Algorithm = Neighbor-Joining, Distance measure = Jukes-Cantor, 100 bootstrap replications) was constructed using ClC Main Workbench (Qiagen, Düsseldorf, Germany). The tree was visualized by use of the Figtree software v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). For identification of aldhl (type 4 aldhl) in Bifidobacterium strains, all complete human gut-associated Bifidobacterium genomes (n = 127) including plasmids were retrieved from NCBI Genome and aldhl genes were identified using NCBI iBLASTn with default settings and a cutoff of 70% identity and 70% query coverage. Aligned genomic nucleotide sequences were translated and verified to match LDHs using reciprocal BLASTx against NCBI’s non-redundant database. In addition, the ALDH amino acid sequence (translated from the aldhl nucleotide sequence) of B. longum sp. longum 105-A was aligned (gap cost 10, gap extension cost 1) and with the ALDH amino acid sequences of the B. longum sp. longum, B. longum sp. infantis, B. bifidum, B. breve and B. scardovii type strains and pairwise amino acid identity percentages were calculated in CLC Main Workbench. Potential non-bifidobacterial ALDH homologues were searched for by BLASTp analysis of the 105-A amino acid sequence against the non-redundant protein sequence database and the Swiss-Prot database using a cutoff of 60% amino acid sequence identity. Further, comparison of aldhl gene cluster/orientation in 12 Bifidobacterium type strains (Supplementary Fig. 4) was conducted by pairwise alignments in MBGD (Microbial Genome Database for Comparative Analysis; http://mbgd.genome.ad.jp/). The amino acid sequences of the gene cluster from B pseudolongum sp. pseudolongum type strain was collected from NCBI database (https://www.ncbi.nlm.nih.gov/genome/) and was used for comparison with that from B. animalis sp. animalis type strain. Using 193 infant samples collected at early (13 h) and subsequently 5 μl of plasmid DNA (80–100 ng μl–1) was mixed with 5 μl (5 pmol μl–1) pBAD forward (5′-ATGCCATAGGCTTTTAC-3′) or reverse (5′-GATTAACTGCTATCAGG-3′) sequencing primers (5 pmol μl–1) and shipped for sequencing at GATC (GATC-biotech). To remove the leader peptide in the recombinant protein the type 4 aldhl gene, 2 μl of LB-AMP broth was inoculated with a single recombinant colony or the non-transformed E. coli LMG194 (negative control) and grown at 37 °C ON at 250 r.p.m. In 3× triplicates, 100 μl of the ON cultures (2×3× 100 μl of transformant culture + 1×3× 100 μl of non-transformed E. coli LMG194 culture) were diluted 100-fold into 9.9 ml of prewarmed LB-AMP/LB broth and grown at 37 °C ON at 250 r.p.m. until OD600nm = 5,000 μl of culture was inoculated on ice for 3 min and incubation on ice for 2 min. A total of 900 μl of LB medium was added and cells were incubated at 37 °C for 1 h at 250 r.p.m., plating on LB-AMP agar plates and incubation at 37 °C ON. Transformants were picked, and plasmid DNA isolated and sequenced as described above. A transformation with correctly recombined plasmid was confirmed by colony PCR using the primers pairs Pr-598/Pr-599 and Pr-600/Pr-601 (Supplementary Fig. 5) amplified by PCR using a primer pair Pr-580/581 (Supplementary Table 7) and ligated with the BamHI-digested pBS423 fragment carrying PUC ori and a spectinomycin-resistance gene20. The resulting plasmid pMSK127 was introduced into B. longum sp. longum 105-A by electroporation to be integrated into type 4 aldhl locus by single crossover recombination (type 4 aldhl::pMSK127) and subsequently sequenced by genomic PCR with a primer pair (Pr-543/546) designed to anneal outside of the gene (Supplementary Fig. 5 and Supplementary Table 7). The amplified fragment was also sequenced to ensure the correct recombination event. Complementation plasmid pMSK128 was constructed by ligating PCR-amplified xfp (sulfolobus 5-phosphate/fructose-6-phosphate phosphoketolase) promoter region (Pxfp) and the type 4 aldhl coding region with PstI- and SalI-digested pBS838 (ref. 20). Using the In-Fusion cloning kit, by which type 4 aldhl was placed under the control of Pxfp. Primer pairs of Pr-598/Pr-599 and Pr-600/Pr-601 were used for amplifying Pxfp from pBS848 (ref. 20) and the type 4 aldhl gene from the B. longum sp. longum 105-A genome, respectively (Supplementary Table 7). The resulting plasmid was electroporated into type 4 aldhl::pMSK127 to give type 4 aldhl::pMSK127/pMSK128 (Pxfp::type 4 aldhl) (Supplementary Fig. 5).

Biochemical characterization of ALDH (type 4 LDH). Recombinant expression and purification. Type 4 LDH (BL105A_0985) was recombinantly expressed as a non-tagged form. The gene was amplified by PCR using the genomic DNA.
of *B. longum* sp. *longum* 105-A as a template and a primer pair of Pr-617 (5′-GGTGGGTGTGCTGAGTACACGACGCCCCCCGAG-3′) and Pr-635 (5′-AAGCTATGCTGACATGCTGATACACTACC-3′). Underlined bases indicate positions for In-Fusion cloning (Clontech). The amplified DNA fragment was inserted into the Ndel and XhoI site of pET23b(+) (Novagen) using an In-Fusion HD cloning kit (Clontech). The resulting plasmid was introduced into *E. coli* BL21 (DE3) ΔlonC carrying pRARE2 (ref. 17) and the transformant was cultured in LB medium supplemented with ampicillin (100 μg ml⁻¹) and chloramphenicol (7.5 μg ml⁻¹). When *ΔlonC*, reached 0.5, 5 μg ml⁻¹ β-lactamase was added at a final concentration of 0.02 mM to induce the protein expression.

The culture was incubated for 4 days at 18 °C, harvested by centrifugation and resuspended in 50 mM potassium phosphate buffer (KPB; pH 7.0) supplemented with 1 mM 2-mercaptoethanol (2-ME) and 200 μM phenylmethyl sullfonyl fluoride. Following cell disruption by sonication, the cleared lysate was saturated with ammonium sulfate (80%). The resulting precipitate was dissolved against 20 mM KPB (pH 7.0) containing 1 mM 2-ME and concentrated by Amicon Ultra 10 K centrifugal device (Merck Millipore). The sample was then loaded onto an Affigel blue column (Bio-Rad) preequilibrated with 20 mM KPB (pH 7.0) containing 1 mM 2-ME and eluted by the same buffer containing 1 M NaCl. The protein was concentrated by Amicon Ultra 100 K centrifugal device (Merck Millipore).

**Enzyme assay.** The standard reaction mixture contained 100 mM KPB (pH 8.0), 1 mM 2-ME, 0.1 mM β-NAAD and the substrate. The reaction was initiated by adding the enzyme and the mixture was incubated at 37 °C for an appropriate time, in which the linearity of the reaction rate was observed. The substrate concentration was hold between 0.1 and 5 mM. The reaction was stopped by adding an equal volume of 0.1 M NaCl in 20 mM Tris-Cl (pH 8.0) containing 1 mM 2-ME and Superdex 200 Increase 10/300 GL column (GE Healthcare; 10 mM KPB (pH 7.0) containing 50 mM NaCl and 1 mM 2-ME). Protein concentration was determined by measuring the absorbance at 280 nm based on a theoretical extinction coefficient of 26,470 M⁻¹ cm⁻¹.

**Metabolomics. Chemicals.** Authentic standards of the aromatic amino acids and derivatives (Supplementary Table 1) were obtained from Sigma-Aldrich, whereas isotope-labelled aromatic amino acids used as internal standards (1-phenylalanine (ring-d5, 98%), t-tyrosine (ring-d4, 98%), t-tryptophan (indole-d5, 98%) and indoleacetic acid (2,2-d2, 96%)) of the highest purity grade were obtained from Cambridge Isotope Laboratories.

**Extraction of metabolites from faecal samples.** Faecal samples (100–500 mg) from the SKOT (n = 59) and the CIG cohort (n = 267, data from two samples missing due to insufficient sample material (n = 1) and problems detecting the internal standards (n = 1)) were diluted 1:2 with sterile MQ water, vortexed for 10 s and centrifuged at 16,000g for 4 °C for 5 min. Subsequently, the supernatant liquor was transferred to a new tube and again diluted with MQ water to be 1:5 dilution of the faecal sample. The tubes were vortexed for 10 s and left at −20 °C for 10 min to precipitate the proteins. The supernatants were then centrifuged at 16,000g, 4 °C for 10 min and each supernatant (320 μl) was transferred to a new tube, which was dried with nitrogen gas.

**Extraction of metabolites from urine samples.** Urine samples (n = 49) from the SKOT cohort were thawed in a refrigerator and all procedures during the sample preparation were carried out at 0–4 °C using an ice bath. The subjects were randomized between analytical batches by placing all the samples from the each subject in the same 96-well plate. The run order of the samples was randomized within the analytical batch. Urine samples were centrifuged at 3,000g for 2 min at 4 °C. A total of 150 μl of each urine sample were added to separate wells and diluted with 150 μl of diluted (MQ water: formic acid (99.9:0.1, v/v)) internal standard mixture (100 μg ml⁻¹) and 240 μl of acetonitrile were added. The tubes were vortexed for 10 s and left at −20 °C for 10 min to precipitate the proteins. The supernatants were then centrifuged at 16,000g, 4 °C for 5 min and transferred to an LC vial, which was stored at −20 °C until analysis.

**Extraction of metabolites from in vitro fermentation samples.** Supernatants from in vitro fermentations were thawed at 4 °C, centrifuged at 16,000g, 4 °C for 10 min, before 80 μl was transferred to a new tube. To each sample, 20 μl of internal standard (40 μg ml⁻¹) and 300 μl of acetonitrile were added. The tubes were vortexed for 10 s and left at −20 °C for 15 min to precipitate the proteins. Following the centrifugation, the tubes were centrifuged at 16,000g, 4 °C for 10 min, 50 μl of each sample was diluted with 50 μl of sterile water and transferred to an LC vial (equalling a 1:10 dilution of the sample with internal standards having a concentration of 1 μg ml⁻¹), vortexed for 10 s, centrifuged at 16,000g, 4 °C for 5 min and transferred to an LC vial, which was stored at −20 °C until analysis.

**Metabolic profiling of faecal, caecal and in vitro samples using UPLC–MS.** Aromatic amino acids and derivatives (Supplementary Table 1) of faecal and in vitro samples were quantified by a semiquantitative UPLC–MS method94. In brief, samples were analysed in random order. For the analysis of the CIG faecal samples, a pooled quality control (QC) sample was injected for every ten samples. In all samples, 20 μl of sample with internal standards having a concentration of 1 μg ml⁻¹) were analysed with an electrospray interphase (Bruker Daltonics) operating in positive mode.
The analyses were separated on a Poroshell 120 SB-C18 column with a dimension of 2.1 x 100 mm and 2.7 μm particle size (Agilent Technologies) as previously published9. Aromatic amino acids and derivatives were detected by selected ion monitoring (SIM) with isotopic internal standards as listed in Supplementary Table 1. The recoveries of the internal standards varied but were, relative to each other, in general rather consistent (Supplementary Fig. 14) emphasizing that while the absolute concentrations may not be accurate due to lack of isotope-labelled internal standards for each single analyte, the relative metabolite concentrations were consistent across samples with the applied LC-MS method. Data were processed using QuantAnalysis v2.2 (Bruker Daltonics) and bracket calibration curves for every ten lumen samples were obtained for each metabolite. The calibration curves were established by plotting the peak area ratios of all of the analytes with respect to the internal standard against the concentrations of the calibration standards. The calibration curves were fitted to a quadratic function.

For untargeted metabolomics, the raw UPLC–MS data, obtained by analysis of the CIG faecal samples in positive ionization mode, were converted to mzXML files using Bruker Compass DataAnalysis 4.2 software (Bruker Daltonics) and preprocessed as previously reported9 using the R package XCMS (v.1.38.0; ref. 9). Noise filtering settings included that features should be detected in at least 50% of the samples. A data table was generated comprising mass-to-charge (m/z), retention time and intensity (peak area) for each feature in the every sample. The data were normalized to the total intensity and log-transformed. Subsequently, a retention time < 0.5 % in the QC samples and features with a retention time < 0.5 % in the QC samples were excluded from the data set. Parent ions with m/z HMO compounds of interest (2.1 FL/3FL, LNT/LN'T, 3.5 SL/6 SL) were searched in the cleaned dataset with 0.02 Da m/z and 0.02 min retention time tolerance. Subsequently, the identities of the features of interest were confirmed at level 1 (ref. 10) by tandem mass spectrometry and comparison to authentic standards (Supplementary Table 8). Of notice, HMO isomers could not be distinguished with the method applied due to identical retention times.

Metabolic profiling of urine samples using UPLC–MS. The samples were analysed by UPLC-QTOF-MS equipped with an electrospray ionization (ESI) (Waters Corporation). Reverse phase HSS T3 C18 column (2.1 x 100 mm, 1.8 μm) coupled with a pre-column (VanGuard HSS T3 C18 column (2.1 x 5 mm, 1.8 μm)) were used for chromatographic separation. A total of 5 μl of each well was injected into the mobile phase A (0.1% formic acid in MQ water), mobile phase B (10% 1 M ammonium acetate in methanol), mobile phase C (methanol) and mobile phase D (isopropanol). Mobile phase gradient during the run of 10 min was as follows: isopropanol (100% B), 0.75 min (100% A), 0.8 min (40% C, 60% B, 5% A, 5% D, 8 min (70% B, 30% D), 8.1 min (70% B, 30% D), 9 min (100% A) and 10 min (100% A). The flow rate gradient was as follows: start condition (0.4 ml min⁻¹), 0.75 min (0.4 ml min⁻¹), 6 min (0.5 ml min⁻¹), 6.5 min (0.5 ml min⁻¹), 8 min (0.6 ml min⁻¹), 8.10 min (0.4 ml min⁻¹), 9 min (0.4 ml min⁻¹) and 10 min (0.4 ml min⁻¹). ESI was operated in negative mode with 3.0 kV capillary probe voltage. The cone voltage and the collision energy were set at 30 kV and 5 eV, respectively. Ion spray and desolvation gas (nitrogen) temperature were 120 and 400 °C while sampling cone and desolvation gas flow rates were 50 and 1,000 l h⁻¹. Scan time set as 0.08 s with a 0.02 s interscan time for both modes. Data were acquired in centroid mode with mass range between 50 and 1,500 Da. Leucine-enkephalin (500 ng ml⁻¹) was infused as the lock-spray agent to calibrate the mass accuracy every 5 s with a 1 s scan time. Quality control samples were used to evaluate possible contamination, monitoring the changes in mass accuracy, retention time and instrumental sensitivity drifts99,100.

The raw data were converted to netCDF format using DataBridge Software v.3.5 (Waters) and imported into MZmine v.2.28 (ref. 5). A subset of samples was used to optimize the preprocessing parameters for the positive and negative data separately. Optimized preprocessing parameters are listed in Supplementary Table 9. Data preprocessing was used with the following steps: mass detection, chromatogram builder, chromatogram deconvolution, deisotoping, peak alignment and gap filling. After the preprocessing, each detected peak was represented by a feature, defined with width retention time, m/z and peak area. Parent ions m/z of the aromatic lactic acids (ILA, PLA and 4-OH-PLA) were searched in the cleaned dataset with 0.02 Da m/z and 0.02 s retention time tolerance. A linear regression model was used feature-wise to correct for batch differences and instrumental sensitivity drifts9. The aromatic lactic acids were confirmed at level 1 (ref. 10) by comparison to a variety of standards and by acquisition of an isotope cluster ion of the same experimental conditions (Supplementary Figs. 1–3).

Ex vivo stimulation of human immune cells. Humanuffy coats were acquired from the Copenhagen University Blood Center and used as anonymous donors. Use of theuffy coat material from healthy anonymous donors was approved by the Blood bank at Rigshospitalet, Copenhagen, under the jurisdiction of Region H. Prior written informed consent was obtained according to the Declaration of Helsinki. Blood samples were handled in accordance with guidelines put forward in the “Transfusion Medicine Standards” by the Danish Society for Clinical Immunology (www.dski.dk).

Isolation, cell culture and stimulation of T cells. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by density centrifugation on Lymphoprep and cryopreserved at −150 °C in FBS with 10% DMSO until the day of stimulation. For cultivation, PBMCs were thawed and cultured in CD4⁺ T cells isolated using EasySep Human CD4⁺ T Cell Isolation Kit (Stemcell, 17791) following the manufacturer’s protocol. In short, 2.5 x 10⁶ PBMCs were incubated for 5 min at room temperature in 500 μl of IMDM-medium containing 50 μl of CD4⁺ T cell isolation cocktail, followed by the addition of 50 μl of RapidSpheres. Subsequently, the volume was topped up to 2.5 ml with IMDM-medium, the cells placed in an EasySep magnet (Stemcell) and incubated at room temperature for 3–5 min. The pure CD4⁺ T cell fraction was obtained by pouring the enriched non-bound cell fraction into a new tube. Enriched CD4⁺ T cells were cultured in T⁺17-polarizing culture medium (IMDM supplemented with 10% FCS, 20 μM HEPES (pH 7.4), 50 μM 2-mercaptoethanol, 2 mM L-glutamine and penicillin–streptomycin (10,000 U ml⁻¹), 30 ng ml⁻¹ IL-1, 10 ng ml⁻¹ IL-1β, 0.5 ng ml⁻¹ TGFB-1, 10 ng ml⁻¹ IL-23, 25 μl ml⁻¹ ImmunoCult Human CD3/CD28 T cell activator for 3 d at 37 °C and 5% CO₂ in Falcon polystyrene 48-well plates (Thermo Fisher, 10059110). Each culture condition contained 0.2% DMSO with or without the indicated amounts of ILA and/or the AhR-inhibitor CH-223191. After 3 d of culture, supernatants were collected for ELISA and frozen down until further use. The ELISA to detect IL-22

1378 NATURE MICROBIOLOGY | VOL 6 | NOVEMBER 2021 | 1367-1382 | www.nature.com/naturemicrobiology
was performed in technical duplicates using the ELISA MAXI Deluxe Set Human IL-22 kit (Biolegend, 434504) following the supplied manufacturer’s protocol. In short, a Nunc MaxiSorb flat-bottom 96-well plate (Thermo Fischer, 44-2402-21) was coated for 12 h at 4 °C with IL-22 coating antibody followed by four rounds of washing with PBS + 0.05% Tween-20. The washed plate was blocked with assay diluent A buffer for 1 h at room temperature and 400 r.p.m., washed four more times with PBS + 0.05% Tween-20 and incubated with cell culture supernatants for 2 h at room temperature and 400 r.p.m. Serially diluted standard controls and blank samples were included on the plate. After 2 h, the plate was washed four more times with PBS + 0.05% Tween-20 and incubated with IL-22 detection antibody for 1 h at room temperature and 400 r.p.m. After four further washing steps with PBS + 0.05% Tween-20, Avidin-HRP was added for 30 min at room temperature and 400 r.p.m. To detect HRP activity, the plate was washed five times with PBS + 0.05% Tween-20 followed by an incubation with solution B for 10 min in the dark at room temperature. HRP activity was stopped after 20 min using 1 M H2SO4 and the optical density recorded (absorption at 450 nm) using a PowerWave HT Microplate Spectrophotometer (BioTek Instruments). Values below limit of detection (16 pg ml−1) of the kit were set to LOD/2. Sources and identifiers of all reagents used are given in Supplementary Table 10.

Isolation, cell culture and stimulation of monocytes. PBMCs were isolated by Ficoll-Paque (GE Healthcare) density centrifugation. Monocytes were isolated to >92% purity using the Pan Monocyte Isolation kit (Miltenyi Biotec). Cells were stained with diluted (1:25) CD14-PE-Cy7 (eBioscience) and diluted (1:25) CD16-FITC (Biolegend) to determine the purity of monocytes (Supplementary Fig. 15). Flow cytometry was performed 3 days after stimulation in a culture medium (RPMI 1640 (Lonza) containing 2 mm l-glutamine (Lonza), 10% heat-inactivated fetal bovine serum (Lonza), 1% penicillin/streptomycin (Lonza), 50 μM 2-DE (Sigma-Aldrich)) in a humidified 37 °C, 5% CO2 incubator. IFA (Sigma I5508), dissolved in a maximum of 0.1% DMSO, was added to the cells, with the final concentrations of ILA 5, 50 or 200 μM, respectively (0.1% DMSO) was added as a control. LPS (TLR4 ligand, e.coli O26:B6, Sigma L2654) at 100 ng ml−1 (final concentration) and IFN-γ (RD285-IF-100) at 10 ng ml−1 (final concentration) were then added to the cells and they were stimulated for 18 h. To determine the addition of AhR, monocytes were pretreated for 1 h before addition of above compounds with the AhR antagonist CH-223191 (Sigma C8124) at 10 μM. For HCA 3 silencing, where no specific antagonist is available, we performed reverse transfection using LipoFectamine RNAiMAX in Opti-MEM (Life Technologies) added 10 nM of scrambled siRNA (ScrRNA) (Thermo Fisher 4390846) or HCA-specific siRNA (Thermo Fisher 4427037), reaching a maximum knockdown of 79% on the basis of qPCR validation (HCA3 primer (Thermo Fisher 4448892) versus GAPDH) as compared to ScrRNA targeted cells. Monocytes were treated with the siRNA constructs for 24 h before stimulation as above. Viability of cells was >97% as analysed by flow cytometry (BD FACS Canto II) using SYTOX AAD. Supernatants of stimulated cells were harvested after 18 h and kept at −80 °C until analysis. IL-12p70 was quantified using Meso Scale Discovery kits as previously detailed34.

Statistics and reproducibility. All experiments were performed with full factorial (biological and technical) replication. Data collection and analysis were not performed blind to the conditions of the experiments. No data were excluded from the analyses, except in the CIG cohort, six samples were omitted from the PCoA performed blind to the conditions of the experiments. No data were excluded. One donor was excluded from the 200 μM indolelactate (ILA) stimulation of monocytes, since something went wrong during the stimulation. Statistical analyses were performed using QIIME v.1.9 (ref. 35), R v.3.1 (ref. 36) and GraphPad Prism v.8.1 (GraphPad Software). If data were normally distributed (evaluated by visual inspection and D’Agostino–Pearson test), parametric statistical analyses were performed when comparing two groups. For comparison of more than two groups, statistical significance was evaluated by one-way analysis of variance (ANOVA) or the non-parametric Kruskal–Wallis test. P<0.05 was considered statistically significant. When applicable, P values were corrected for multiple testing by the Benjamini–Hochberg false discovery rate (FDR) using a cutoff of 0.1.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All 16S rRNA gene amplicon sequencing data were deposited in the Sequence Read Archive (SRA) under the BioProjects PRJNA275694 (SKOT) and PRJNA554596 (CIG). The following databases were used: GreenGenes 16S rRNA database (https://greengenes.secondgenome.com/), RDP database (https://rdp.cme.msu.edu/), 16S rDNA gene sequence database at NCBI (https://ncbiinsights.ncbi.nlm.nih.gov/2020/02/21/rRNA-databases/), non-redundant protein sequence database at NCBI (https://www.ncbi.nlm.nih.gov/protein/), Swiss-Prot database (https://www.uniprot.org/), MBGD (http://mbgd.genome.ad.jp) and the genome database at NCBI (https://www.ncbi.nlm.nih.gov/genome/). Metabolomics data (concentrations of aromatic amino acid metabolites) from SKOT and CIG cohorts are available in Supplementary Data 1e and 2d. Source data are provided with this paper.

Code availability

No custom code was used in the analyses. R scripts are available on request.

Received: 3 February 2021; Accepted: 26 August 2021; Published online: 21 October 2021

References

1. Victora, C. G. et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 387, 475–490 (2016).
2. Horta, B. L., Loret de Mola, C. & Victora, C. G. Long-term consequences of breastfeeding on cholesterol, obesity, systolic blood pressure and type 2 diabetes: a systematic review and meta-analysis. Acta Paediatr. 104, 30–37 (2015).
3. Oddy, W. H. Breastfeeding, childhood asthma, and allergic disease. Ann. Nutr. Metab. 70, 26–36 (2017).
4. Laursen, M. E. et al. Infant gut microbiota development is driven by transition to family foods independent of maternal obesity. Mibphere https://doi.org/10.1128/mSphere.00691-15 (2016).
5. Bäckhed, F. et al. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe 17, 690–703 (2015).
6. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature 562, 583–588 (2018).
7. Bode, L. The functional biology of human milk oligosaccharides. Early Hum. Dev. 91, 619–622 (2015).
8. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. Science 352, 539–544 (2016).
9. Stokholm, J. et al. Maturation of the gut microbiome and risk of asthma in childhood. Nat. Commun. 9, 141 (2018).
10. Fujimura, K. E. et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat. Med. 22, 1187–1191 (2016).
11. Khoravi, A. & Mazmanian, S. K. Disruption of the gut microbiome as a risk factor for microbial infections. Curr. Opin. Microbiol. 16, 221–227 (2013).
12. Fukushima, S. et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature 469, 543–547 (2011).
13. Vatanen, T. et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. Nature 562, 589–594 (2018).
14. Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. Nat. Med. 22, 141 (2016).
15. Celziant-Barragan, L. et al. Lactobacillus reuteri induces gut intraepithelial CD4+CD8+ T cells. Science https://doi.org/10.1126/science.aah5825 (2017).
ARTICLES

17. Natividad, J. M. et al. Impaired aryl hydrocarbon receptor ligand production by the gut microbiota is a key factor in metabolic syndrome. *Cell Metab.* https://doi.org/10.1016/j.cmet.2018.07.001 (2018).

18. Dodd, D. et al. A bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* 551, 648–652 (2017).

19. Guo, X. et al. Innate lymphoid cells control early colonization resistance against intestinal pathogens through ID2-dependent regulation of the microbiota. *Immunity* 42, 731–743 (2015).

20. Quintana, F. J. et al. Control of Treg and TH17 cell differentiation by the aryl hydrocarbon receptor. *Nature Microbiology* 4, 589–598 (2019).

21. Irukayama-Tomobe, Y. et al. Aromatic d-amino acids act as natural agonists for aryl hydrocarbon receptor in culture medium are highly potent agonists of human hydroxycarboxylic acid receptor, GPR109B. *Proc. Natl Acad. Sci. USA* 116, 106–109 (2019).

22. Karni, J. et al. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat. Med.* 24, 1070–1080 (2018).

23. Krishnan, S. et al. Gut microbiota-derived tryptophan metabolites modulate inflammatory response in hepatocytes and macrophages. *Cell Rep.* 23, 1099–1111 (2018).

24. Madsen, A. L., Schack-Nielsen, L., Larnkjaer, A., Mølgaard, C. & Michaelsen, K. F. Determinants of blood glucose and insulin in healthy 9-month-old term Danish infants; the SKOT cohort. *Diabet. Med.* 27, 1350–1357 (2010).

25. Katoh, K. et al. Age-related changes in the composition of gut bifidobacterium species. *Curr. Microbiol.* 74, 987–995 (2017).

26. Koga, Y. et al. Age-associated effect of kestoce on Foscoecibacter prasorunzsis and symptoms in the atopic dermatitis infants. *Pediatr. Res.* 80, 844–851 (2016).

27. Li, X., Jiang, B., Pan, B., Mu, W. & Zhang, T. Purification and partial characterization of Lactobacillus species SK007 lactate dehydrogenase (LDH), catalyzing pyruvate into lactic acid. *PLoS ONE* 6, e23992 (2011).

28. Koide, S., Iwata, S., Matsuzawa, H. & Ohta, T. Crystallization of allosteric l-tartaric dehydrogenase from *Thermus caldophilus* and preliminary crystallographic data. *Biochem. J.* 109, 6–7 (1991).

29. Takashi, M., So, I., Hiroshi, S., Haruhiko, M. & Takahisa, O. Sequence and characteristics of the bifidobacterium longum gene encoding l-tartaric dehydrogenase and the primary structure of the enzyme: a new feature of the allosteric site. *Gene* 85, 161–168 (1989).

30. Bottacini, F. et al. Global transcriptional landscape and promoter mapping of the gut commensal *Bifidobacterium breve* UC2003. *BM C Genomics* 18, 219 (2017).

31. Hiyarama, Y. et al. Development of a double-crossover markerless gene deletion system in *Bifidobacterium longum*: functional analysis of the α-galactosidase gene for raffinose assimilation. *Appl. Environ. Microbiol.* 78, 4984–4992 (2012).

32. Matsumura, H., Takeuchi, A. & Kano, Y. Construction of *Escherichia coli–Bifidobacterium longum* shuttle vector transforming *B. longum* 105-A and 108-A. *Biosci. Biotechnol. Biochem.* 61, 1211–1212 (1997).

33. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies 3 Preprint at bioRxiv https://doi.org/10.1101/2021.01.20.427420 (2021).

34. Bokdash, J. Z. & Marusich, L. R. Repeated measures correlation. *Front. Psychol.* 8, 456 (2017).

35. Alfa, M. J. et al. An outbreak of necrotizing enterocolitis associated with a novel *Clostridium* species in a neonatal intensive care unit. *Clin. Infect. Dis.* 35, S101–S105 (2002).

36. Bae, M. J. et al. Conditions of bifidobacterial colonization in preterm infants: a prospective analysis. *Pediatr. Gastroenterol. Nutr.* 44, 577–582 (2007).

37. Moles, I. et al. Bacterial diversity in meconium of preterm neonates and evolution of their fecal microbiota during the first month of life. *PLoS ONE* 8, e66986 (2013).

38. Arboleya, S. et al. Establishment and development of intestinal microbiota in preterm neonates. *EMBS Microbiol.* 79, 763–772 (2012).

39. Korpela, K. et al. Intestinal microbiota development and gestational age in preterm neonates. *Sci. Rep.* 8, 24538 (2018).

40. Peters, A. et al. Metabolites of lactic acid bacteria present in fermented foods are highly potent agonists of human hydroxyacylsarcosine acid receptor 1. *PLoS Genet.* 15, e1008145 (2019).

41. Ahmed, K. et al. Deorphanization of GPR109B as a receptor for the β-oxidation intermediate 3-OH-octanoic acid and its role in the regulation of lipolysis. *J. Biol. Chem.* 284, 21928–21933 (2009).

42. Irukayama-Tomobe, Y. et al. Molecular agonists for aryl hydrocarbon receptor in culture medium are highly potent agonists of human hydroxycarboxylic acid receptor, GPR109B. *Proc. Natl Acad. Sci. USA* 116, 107–117 (2019).

43. Quintana, F. J. et al. Control of Treg and TH17 cell differentiation by the aryl hydrocarbon receptor. *Nature* 453, 63–71 (2008).

44. Veldhoen, M., Hirota, K., Christensen, J., O’Garra, A. & Stockinger, B. Natural agonists for aryl hydrocarbon receptor in culture medium are essential for optimal differentiation of Th17 T cells. *J. Exp. Med.* 206, 43–49 (2009).

45. Veldhoen, M. et al. The aryl hydrocarbon receptor links T,17-cell-mediated autoimmune to environmental toxins. *Nature* 453, 106–109 (2008).
Nature Microbiology

101. Smart, K. F., Aggio, R. B. M., Van Houtte, J. R. & Villas-Bôas, S. G. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11, 395 (2010).

van der Kloot, F. M., Bobeldijk, L., Verheij, E. R. & Jelluma, H. R. Analytical error reduction from single point calibration for accurate and precise metabolomic phenotyping. J. Proteome Res. 8, 5132–5141 (2009).

Smart, K. F., Aggio, R. B. M., Van Houtte, J. R. & Villas-Bôas, S. G. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography–mass spectrometry. Nat. Protoc. 5, 1709–1720 (2010).
Extended Data Fig. 1 | *Bifidobacterium* species composition and aromatic lactic acids in partially breastfed and weaned infants aged 9 months from the SKOT cohort. a, Pie chart of the average *Bifidobacterium* species composition given as percent of total *Bifidobacterium* abundance (n = 59 infants). Legend includes the average percent of each species compared to the total faecal microbiota community (See also Supplementary Data 1g). b, Left panel: Box and whiskers plot (line: median, box: IQR, whiskers: min-max) of faecal abundance of indolelactic acid (ILA), phenyllactic acid (PLA) and 4-hydroxyphenyllactic acid (4-OH-PLA) in partially breastfed (n = 24, red) and weaned (n = 35, blue) infants. Right panel: Box and whiskers plot (line: median, box: IQR, whiskers: min-max) of urine abundance of ILA, PLA and 4-OH-PLA in partially breastfed (n = 19, red) and weaned (n = 30, blue) infants. Statistical significance was evaluated by two-sided Mann–Whitney U-test. c, Heatmap of Spearman's Rank correlation coefficients (two-sided tests) between the relative abundances of ILA, PLA and 4-OH-PLA measured in urine and the faecal relative abundance of *Bifidobacterium* species or faecal concentrations of ILA, PLA and 4-OH-PLA of the same infants (n = 49 infants). For all panels asterisks indicate statistical significance: *p < 0.05, **p < 0.01, ***p < 0.001 and ****p < 0.0001.
Extended Data Fig. 2 | In vivo production of aromatic lactic acids in previously germ-free mice colonized with either B. longum 105-A WT or type 4 \textit{ldh} mutant. \textbf{a}, CFU counts of \textit{B. longum} 105-A WT or type 4 \textit{ldh} mutant from caecal content of mice monoclonized with either the WT (\(n=21\)) or type 4 \textit{ldh} mutant strain (\(n=29\)). Bars and error bars indicate median ± IQR. \textbf{b–d}, caecal concentrations of the aromatic lactic acids (indolelactic acid (ILA), phenyllactic acid (PLA) and 4-hydroxyphenyllactic acid (4-OH-PLA)) in mice monoclonized with either the WT (\(n=21\)) or type 4 \textit{ldh} mutant strain (\(n=29\)). Line and error bars indicate median ± IQR. For all panels statistical significance was evaluated by two-sided Mann–Whitney U-tests.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Faecal microbiota composition of the CIG infants. a, Average relative abundance of the dominant faecal microbial taxa (average relative abundance >0.1%) across all samples (comprising 97.5% of all microbial taxa detected). b, Temporal development of the average gut microbiota composition and Shannon diversity index (marked with circles) across all individuals. c, Intra-individual temporal development of gut microbiota composition and Shannon diversity index (marked with circles). Dietary patterns and consumption of antibiotics are indicated for each individual. If nothing else is indicated, infants were singletons, vaginally born at term.
Extended Data Fig. 4 | Beta diversity in CIG cohort. a–f, Principal coordinates analysis (PCoA) plots of Bray–Curtis dissimilarities, based on all OTUs detected in CIG faecal samples (n = 234), coloured according to a, subject and b–f, relative abundances of Bifidobacterium longum, Bifidobacterium bifidum, Bifidobacterium breve, Bifidobacterium catenulatum group or Bifidobacterium dentium, respectively.
Extended Data Fig. 5 | Temporal development in abundance of infant-type *Bifidobacterium* species, aromatic lactic acids, and human milk oligosaccharides in faeces from infants in the CIG cohort. Scatter plots of a, age against absolute abundance of infant-type *Bifidobacterium* species (defined as the sum of absolute abundances of *B. longum*, *B. breve*, *B. bifidum* and *B. scardovii*) in faeces or b–d, age against faecal concentrations of aromatic lactic acids (ILA, PLA and 4-OH-PLA, n = 240) or e–g, age against relative faecal abundance of human milk oligosaccharides (2′FL/3FL, LNT/ LNnT and 3′SL/6′SL, n = 228) during the first 6 months of life in the CIG cohort. A local polynomial regression (LOESS) fit is shown with coloured mean line and 95% CI shaded in grey. Statistical significance was evaluated by two-sided repeated measures correlations (r_{rm} is the repeated measures correlation coefficient).
Extended Data Fig. 6 | Associations between Infant-type \textit{Bifidobacterium} species and aromatic lactic acids and human milk oligosaccharides in faeces from infants in the CIg cohort. Scatter plots of the relationship between faecal absolute abundance of infant-type \textit{Bifidobacterium} species, and a–c, faecal concentrations of aromatic lactic acids (ILA, PLA and 4-OH-PLA, \(n=240\)) or d–f, relative faecal abundance of human milk oligosaccharides (2′FL/3FL, LNT/LNnT and 3′SL/6′SL, \(n=228\)) in the CIg cohort. Statistical significance was evaluated by two-sided repeated measures correlations (\(r_{rm}\) is the repeated measures correlation coefficient). Linear regression curve fits are shown with coloured mean line and 95% CI indicated in grey shading.
Extended Data Fig. 7 | Correlations between relative abundance of bacterial taxa and concentrations of the aromatic lactic acids at each sampling point in the CIg cohort. a, Spearman’s rank correlations between relative abundance of faecal bacterial genera (average relative abundance > 1%) and faecal concentrations of the aromatic lactic acids (ILA, PLA and 4-OH-PLA) at each sampling point. b, Spearman’s rank correlations between relative abundance of faecal *Bifidobacterium* species (average relative abundance > 0.1%) and faecal concentrations of the aromatic lactic acids at each sampling point. Infant-type *Bifidobacterium* species is the sum of the relative abundances of *B. longum*, *B. breve*, *B. bifidum* and *B. scardovi*. For both panels statistical significance was evaluated by uncorrected p-values (two-sided tests) indicated by asterisks with *p* < 0.05, **p** < 0.01, ***p** < 0.001 and ****p** < 0.0001.
Extended Data Fig. 8 | Validation of associations between absolute abundance of *Bifidobacterium* species and faecal abundances of the aromatic lactic acids and human milk oligosaccharides in the CIG cohort. Heatmaps illustrating linear mixed model coefficients (adjusted for subject and age) between the absolute abundance of *Bifidobacterium* species estimated by a, 16S rRNA amplicon sequence variant analysis or b, species/subspecies-specific qPCR and faecal concentrations of aromatic lactic acids (ILA, PLA and 4-OH-PLA, n = 240) or relative faecal abundances of human milk oligosaccharides (2′FL/3FL, 3′SL/6′SL and LNT/LNnT, n = 228) in the CIG cohort. For both panels statistical significance was evaluated by false discovery rate-corrected p-values (two-sided tests) indicated by asterisks with *p < 0.05, **p < 0.01, ***p < 0.001 and ****p < 0.0001.
Extended Data Fig. 9 | Relative abundance of high quality metagenome-assembled genomes (MAGs) with aldhl (type 4 ldh) in 4 months and 12 months old infants and mothers. Metagenome data were retrieved from Bäckhed et al [reference]. Line, boxes and whiskers indicate median, IQR and ±1.5*IQR and statistical significance was evaluated by two-sided Mann–Whitney U-tests.
Extended Data Fig. 10 | Absolute abundances of *Bifidobacterium* species and concentrations of aromatic lactic acids in faeces of infants in the CIG cohort. a,b. Absolute abundance of *Bifidobacterium* species (average relative abundance >1% of total community) and concentrations of indolelactic acid (ILA), phenyllactic acid (PLA) and 4-hydroxyphenyllactic acid (4-OH-PLA) in faeces of individuals from the Copenhagen Infant Gut (CIG) cohort. Values of bacterial counts below $10^6$ cells/g faeces and metabolite concentrations below 1 nmol/g faeces are not shown. Infant-type *Bifidobacterium* species is the sum of the absolute abundances of *B. longum*, *B. breve*, *B. bifidum* and *B. scardovii* and is indicated with grey background shading. a. Infants early colonized with infant-type *Bifidobacterium* species (colonized within first month reaching average relative abundance >40% during first 6 months). b. Infants with late colonization of infant-type *Bifidobacterium* species (not detectable or on average <0.5% of total community within the first 3 months of life).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed
☐ □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ □ A statement whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ □ The statistical test(s) used AND whether they are one- or two-sided
☐ □ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ □ A description of all covariates tested
☐ □ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ □ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ □ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever possible.
☐ □ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ □ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ □ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

No software was used

Data analysis

Quantification of aromatic amino acids in faeces and in vitro fermentation samples was performed using QuantAnalysis version 2.2 (Bruker Daltonics, Bremen, Germany). The raw urine metabolome data were converted to netCDF format using DataBridge Software version 3.5 (Waters, Manchester, UK) and imported into MZmine version 2.28. The raw untargeted faecal metabolome data were converted to mzXML files using Bruker Compass DataAnalysis 4.2 software (Bruker Daltonics). Gut microbiota data were processed by CLC Genomic Workbench v8.3. CLCbio, Qiagen, Aarhus, DK), QIIME (v1.967), and DADA2 pipeline (v1.14). qPCR data were processed by the LightCycler® 480 Software v1.5. Alignment of lactate dehydrogenase genes was constructed in CLC Main Workbench (v7.6.3, CLCbio, Qiagen, Aarhus, DK) and the phylogenetic tree was visualized by use of the FigTree software v1.4.3 [http://tree.bio.ed.ac.uk/software/figtree/]. The kinetic parameters (Kcat, KDS, and Hill coefficient nH) were calculated by curve-fitting the experimental data to the Hill equation, using GraphPad Prism v8.1. Bifidobacterium metagenome assembled genomes (MAGs) were identified by MGIs and IGGdb (v.1.0.087). Statistical analyses were performed using QIIME (v1.967), R (v3.189), GraphPad Prism (v8.1), and Matlab R2014b or R2015b (Mathworks, Inc.). The follow R packages were used: ggplot2 [v3.3.3], rcorr (v0.4.3), maaslin2 [v1.0.0], gplots (v3.1.1).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All 16S rRNA gene amplicon sequencing data were deposited in the Sequence Read Archive (SRA) under the BioProjects PRJNA273694 [SKOT] and PRJNA554596 [CIG]. The following databases were used: GreenGenes 16S rRNA database [https://greengenes.secondgenome.com/], RDP database [https://rdp.cme.msu.edu/], 16S rRNA gene sequence database at NCBI [https://ncbiinsights.ncbi.nlm.nih.gov/2020/02/21/rrna-databases/], non-redundant protein sequence database at NCBI [https://www.ncbi.nlm.nih.gov/protein/], Swiss-Prot database [https://www.uniprot.org/], MIGS (Microbial Genome Database for Comparative Analysis; http://migs.genome.ad.jp/), and the genome database at NCBI [https://www.ncbi.nlm.nih.gov/genome/]. Metabolomics data (concentrations of aromatic amino acid metabolites) from SKOT and CIG cohorts are available in Supplementary Data 1e and 2d. Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/ic-reports-ic-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
No sample size was estimated. Based on the original correlations observed between Bifidobacterium and aromatic lactic acids in 59 infants [SKOT], we estimated that 25 infants [CIG] with multiple time points would be sufficient to demonstrate the dynamics between Bifidobacterium species and aromatic lactic acids.

Data exclusions
In the CIG cohort, 6 samples were omitted from the Principal coordinate analyses due to low read counts (<8000). For correlation analyses between Bifidobacterium species and HMO residues in feces, 12 samples with no reported breastfeeding were excluded. One donor was excluded from the 200 μM indolelactate (ILA) stimulation of monocytes, since something went wrong during the stimulation.

Replication
All experiments were performed with full factorial (biological and technical) replication.

Randomization
No randomisation was performed as the study was observational.

Blinding
No blinding was performed as the study was observational.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | ChIP-seq |
| ☐ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

Antibodies

Antibodies used
CD14-PE-Cy7 [E Biosciences, catalog number: 25-0149, clone name: 61D3] and CD16-FITC [Biolegend, catalog number: 302006, clone name: 3G8].

Validation
The 61D3 monoclonal antibody reacts with human CD14, a 53-55 kDa GPI-linked glycoprotein. CD14 is expressed on monocytes, interfollicular macrophages and some dendritic cells (manufacturer’s website: https://www.thermofisher.com/antibody/product/CD14-Antibody-clone-61D3-Monoclonal/25-0149-42)
CD16 is known as low affinity IgG receptor III (FcγRIII). It is expressed as two distinct forms (CD16a and CD16b). CD16a (FcγRIIIA) is a 50-65 kD polypeptide-anchored transmembrane protein. It is expressed on the surface of NK cells, activated monocytes, macrophages, and placentral trophoblasts in humans (manufacturer's website: https://www.biologend.com/en-us/products/fc-anti-human-cd16-antibody-567)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Rat AHR Reporter Cells (H4IIIE) was a personal gift from Dr. Michael S. Denison, Department of Environmental Toxicology, Meyer Hall, University of California, Davis California 95616, United States

Human AHR Reporter Cells from Indigo Biosciences (PA, USA, catalogue number: IB0001)

oAMP Hunter™ eXpress GPR109B CHO-K1 GPCR Assay from DiscoverX (Fremont, CA, USA, catalogue number: 95-0141E2K2P2M)

Authentication

None of the cells were tested for authentication.

Mycoplasma contamination

H4IIIE cells were tested negative for mycoplasma contamination using the MycoAlert® Mycoplasma Detection Kit. The human Aryl Hydrocarbon Receptor (AhR) Reporter Assay System and the oAMP Hunter™ eXpress GPCR Assay were commercially available assays obtained from Indigo Biosciences and DiscoverX, respectively. We did not test these cells for mycoplasma contamination, but relied on the internal quality control of the manufacturers.

Commonly misidentified lines

(See iCLAC register)

No commonly misidentified cell lines were used in the study.

Animals and other organisms

Policy information about studies involving animals, ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Germ-free Swiss Webster mice (Tac:SW, originally obtained from Taconic Biosciences, NY, USA) were bred and housed at the National Food Institute, Technical University of Denmark. In two separate experiments, pregnant GF mice were randomly be colonised with either B. longum 125-A wild-type (n=4) or adh (type 4 idh) mutant (n=5) by a single gavage one week prior to giving birth. The mono-colonised offspring (wildtype=21, 12 males and 9 females, adh=29, 18 males and 11 females) were euthanised at 4 weeks of age.

Wild animals

No wild animals were used in the study.

Field-collected samples

No field-collected samples were used in the study.

Ethics oversight

All mouse experiments were approved by the Danish Animal Experiments Inspectorate (License number 2015-15-0201-00553) and carried out in accordance with existing Danish guidelines for experimental animal welfare.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics

The discovery cohort consisted of a random subset of 59 healthy infants (30 male, 29 female) participating in the SKOT I study. Inclusion criteria were single birth and full term delivery, absence of chronic illness and age of 19 months ± 2 weeks at inclusion. The validation cohort, CIG, consisted of 25 healthy infants (12 male, 13 female), vaginally born (23/25) and full-term (23/25) delivered.

Recruitment

Infants in SKOT I were originally recruited by postal invitations to randomly selected parents of infants on the basis of extractions from the Danish National Civil Registration System. We used a random subset of this cohort for this discovery. For validating the results, infants in CIG were recruited through social media and limited to the Copenhagen region. These recruitment might have biased the recruitment towards an over-representation of breastfed, healthy infants from families with a high socioeconomic status. According to the GDPR rules, no information was available regarding covariate relevant population characteristics of the healthy anonymous donor buffy coat material provided by the Blood bank at Rigshospitalet, Copenhagen. The only criteria was 'no intake of pain killers' on the day of blood draw.

Ethics oversight

The SKOT study protocol was approved by the Committees on Biomedical Research Ethics for the Capital Region of Denmark (H- KF-2007-0033) and The Data Protection Agency (2002-54-0938, 2002-54-026) approved the study. The Committees on Biomedical Research Ethics for the Capital Region of Denmark confirmed that the CIG study was not notifiable according to the Act on Research Ethics Review of Health Research Projects (§ 1, subsection 4), as the study only concerned the faecal microbial composition and activity and not the health of the children. Informed consent was obtained from all parents of infants participating in the study. In addition, parents of CIG twins gave informed consent to publish data from the twins although the parents themselves would be able to identify their children using indirect identifiers. The Data Protection Agency (18/02/436) approved the study. Use of the buffy coat material from healthy anonymous donors was approved by the Blood bank at Rigshospitalet, Copenhagen, under the jurisdiction of Region H.

Note that full information on the approval of the study protocol must also be provided in the manuscript.