

Title	Gut microbiome features associated with Bifidobacterium colonization predict personalized probiotic persistence patterns
Authors	Goswami, Sourav;Ansari, Alisha;Sharaf, Chetan;O'Toole, Paul W.;Shanahan, Fergus;Ahuja, Vineet;Ghosh, Tarini Shankar
Publication date	2026-04-23
Original Citation	Goswami, S, Ansari, A, Sharaf, C, O'Toole, P W, Shanahan, F, Ahuja, V & Ghosh, T S 2026, 'Gut microbiome features associated with Bifidobacterium colonization predict personalized probiotic persistence patterns', Nature Communications, vol. 17, no. 1, 5678, pp. 1-21. https://doi.org/10.1038/s41467-026-72289-9
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.scopus.com/pages/publications/105043514720 - 10.1038/s41467-026-72289-9
Rights	© 2026, the Author(s). This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/ . - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-10-27 05:01:57
Item downloaded from	https://hdl.handle.net/10468/19234



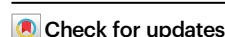
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Gut microbiome features associated with *Bifidobacterium* colonization predict personalized probiotic persistence patterns

Received: 11 September 2025

Accepted: 9 April 2026

Published online: 23 April 2026

Sourav Goswami^{1,4}, Alisha Ansari^{1,4}, Chetan Sharaf¹, Paul W. O'Toole², Fergus Shanahan², Vineet Ahuja³ & Tarini Shankar Ghosh¹✉

Bifidobacteria are key health-associated members of the human gut microbiome and are widely used as probiotics, but their colonization success varies substantially between individuals, partly due to baseline microbiome composition. We analyzed 51,244 gut microbiomes from 149 cohorts (45 countries) to identify non-*Bifidobacterium* taxa associated with the *Bifidobacterium* features. We observed several consistent and age-/life-style-specific association patterns of different non-*Bifidobacterium* taxa with the different *Bifidobacteria*. Multiple *Bifidobacteria* showed positive associations with butyrate-producing *Firmicutes* and *Collinsella*; negative associations involved pathobiont-*Firmicutes* and specific *Bacteroidota* taxa. *B. adolescentis* and *B. breve* showed the strongest positive and negative associations, respectively, with health-associated adult gut microbiome members. We quantified these relationships as Association-Scores, stratified by age/life-style/sequencing-strategy/disease, which were significantly reproducible after adjusting for multiple microbiome-linked, host life-style/clinical covariates, and predictable using species-specific genomic functions. We used these Association-Scores to derive microbiome-level Receptive-Scores that quantify how permissive a baseline microbiome is to increases or persistence of a given *Bifidobacterium*. In an external dataset of eight *Bifidobacterium* interventions ($n = 1633$ gut microbiomes), Receptive-Scores combined with baseline abundance of the administered *Bifidobacteria* significantly predicted post-treatment persistence/increase in 69.23% of trial-probiotic pairs. Together, this work identifies microbiome features governing *Bifidobacterium* colonization and provides tools to predict personalized probiotic responses.

Bifidobacteria are one of the earliest colonizers of the human gut, where they feed on human milk oligosaccharides and exhibit health-associated activities¹. Benefits include promotion of pathogen resistance, immune maturation, anti-inflammation, and microbiome maturation, including the promotion of butyrate producers². After

infancy, *bifidobacteria* have also been associated with beneficial influences including enhanced nutrient absorption; immune regulation; protection against infection, inflammation and cancer; bone-health; cardio-metabolic health and fat metabolism and psychological welfare^{3,4}. Consequently, various species belonging to the genus

¹Indraprastha Institute of Information Technology (IIIT-Delhi), New Delhi, India. ²Schools of Medicine and Microbiology, and APC Microbiome Ireland, University College Cork, Cork, Ireland. ³Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi, India. ⁴These authors contributed equally: Sourav Goswami, Alisha Ansari. ✉e-mail: tarini.ghosh@iiitd.ac.in

Bifidobacterium have been explored as potential probiotics (independently or as part of dietary supplements and functional foods; as single microbe or in consortia)^{3,5}. Probiotics are defined as live microorganisms, which when administered in adequate amounts confer a health benefit on the host⁶. *Bifidobacterium* species, such as *B. animalis*, *B. longum*, *B. breve*, *B. bifidum*, and *B. adolescentis*, have been assessed for probiotic efficacy in a diversity of disorders^{7–9}. However, a continuing challenge is the inconsistency and variability of *Bifidobacterium*-derived probiotic products in terms of efficacy and the extent of colonization of the administered *Bifidobacterium* taxa^{10,11}. To rationally select *Bifidobacterium* strains for future trials, systematic reviews and cross-cohort data analysis of endogenous *Bifidobacterium* species across different age groups and lifestyle patterns have been conducted^{12–14}. These reveal variability in *Bifidobacterium* taxa across different age-groups and lifestyle patterns (especially industrialization). For example, *B. breve/bifidum* are predominant in the infant microbiome; *B. adolescentis* is more prevalent in the adult gut and; *B. longum* is present in all age-groups but with specific strains associated with infants and adults¹³. In general, bifidobacterial species are enriched in the westernized populations when compared with non-westernized, although relative abundances of *Bifidobacterium* are similar in infancy¹². Overall, a lower *Bifidobacterium* abundance has been found in elderly subjects. Other than these general associations, research trials have shown that persistence of individual *Bifidobacterium* strains varies widely even within individuals of a similar age group and lifestyle patterns^{10,11,15}. This variability has been attributed to characteristics of the host's baseline gut microbiome.

A more comprehensive understanding of the degree to which the baseline gut microbiome (particularly non-bifidobacterial taxa) influences the representation of different *Bifidobacterium* members is needed. Large-scale data-driven studies are required to systematically investigate how individual non-bifidobacterial members of the baseline gut microbiome associate with different *Bifidobacterium* taxa in hosts of different age groups and lifestyles, and if genome-specific functions influence such association patterns. Moreover, leveraging such data to stratify individuals at baseline according to their likelihood of responding to different probiotic *Bifidobacterium* treatments is an intriguing possibility.

In this work, to address these issues, we systematically investigate 51,244 gut microbiomes across 45 countries to investigate the non-bifidobacterial taxa associated with the bifidobacterial features, across 149 cohorts of different age-groups, industrialization-patterns, disease-groups and microbiome-profiling strategies. The non-bifidobacterial community composition shows a strong influence on the bifidobacterial composition of the gut microbiome, with the association patterns of the individual non-bifidobacterial taxa with the different bifidobacterial taxa showing distinct patterns depending upon study-cohort characteristics (with certain association patterns consistently observed across cohorts). We quantify these association patterns as association-scores, further stratified by age group, lifestyle, disease and profiling strategy (amplicon/16S and whole genome sequencing/WGS) and investigate their reproducibility after adjusting for multiple host-/microbiome-associated covariates. We then utilize these association scores to formulate microbiome-level receptive scores for the bifidobacterial features. Receptive-scores quantify how accommodating a given microbiome is to post-administration increase or persistence of a specific *Bifidobacterium*, given its baseline microbiome composition and host context (age group, lifestyle, disease status, and 16S or WGS sequencing-type). Combining baseline abundance of the administered *Bifidobacterium* along with the corresponding baseline receptive score significantly predicts post-treatment persistence/increase across 69.23% trial-probiotic pairs from 8 additional intervention studies (encompassing 1633 gut microbiomes).

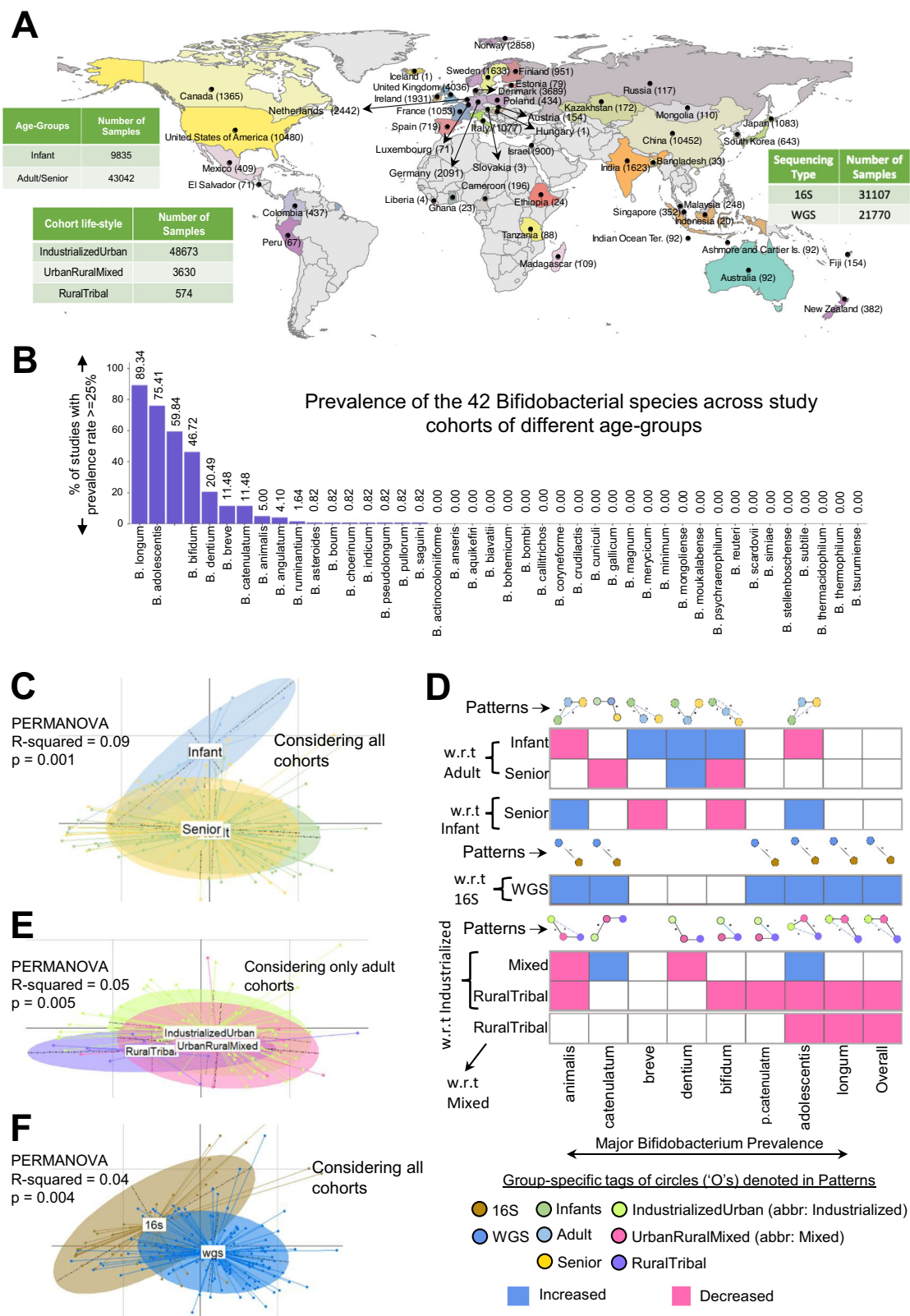
Results

Global bifidobacterial prevalence patterns are strongly associated with study-cohort characteristics and the composition of non-bifidobacterial taxa

We initiated the investigation creating a global collection of 46,218 gut microbiomes collected from 122 study-cohorts (subset of the of 51,244 gut microbiomes utilized overall), spread across 45 countries (spanning six continents) (“Methods”, Supplementary Data 1 and Fig. 1A; also described in ref. 16). We specifically focused on gut microbiomes from subjects in three distinct age-groups, namely infant/new-born (age < 2 y, $n = 9123$ gut microbiomes), adult (18 y < Age < 60 y; $n = 27,055$ gut microbiomes) and seniors (age ≥ 60 y, $n = 10,040$ gut microbiomes). For every cohort, we also collected the information pertaining to its cohort-type (IndustrializedUrban, UrbanRuralMixed, and RuralTribal), as described earlier^{16,17}. For each gut microbiome, we computed the relative abundances of the different species-level taxa along with three other microbiome-specific properties, namely Shannon index, Pielou's evenness index, and the estimated microbial load (see “Methods”). Forty-two *Bifidobacteria* were detected in at least one microbiome. Amongst these, eight, namely *B. longum*, *B. adolescentis*, *B. pseudocatenulatum*, *B. bifidum*, *B. dentium*, *B. breve*, *B. catenulatum*, and *B. animalis* demonstrated strong associations with the gut microbiomes with prevalence rates of > 25% in > 5% of studies (in at least one age-group), in line with previously reported observations^{12–14} (Fig. 1B). Both age-group and lifestyle (especially westernization/industrialization versus traditional lifestyle) have significant impact on the gut microbiome, which were also observed in our investigation^{18,19}. We replicated several previously observed trends pertaining to the prevalence of these taxa. For example, there was significant variation in the prevalence patterns of the different bifidobacterial taxa between the infant and adult/senior cohorts (PERMANOVA- R -squared = 0.09, $P = 0.001$) (Fig. 1C). *B. breve/bifidum* were especially prevalent in infant cohorts, and *B. adolescentis/animalis/catenulatum* were more prevalent in the adult/seniors (as reported earlier¹³) (Figs. 1B–D and 2). The prevalence patterns also showed significant variations across studies belonging to the cohort-types (PERMANOVA- R -squared = 0.05, $P = 0.005$, Fig. 1E). In general, there was an overall reduced prevalence of bifidobacterial taxa in the RuralTribal as compared to westernized (IndustrializedUrban) cohorts¹² (Fig. 1D, E).

Notably, we observed significant variations in the bifidobacteria prevalence patterns depending upon the profiling strategies (16S-v/s-WGS) of study-cohorts (PERMANOVA- R -squared = 0.04, $P = 0.004$, Fig. 1F), with multiple taxa like *B. longum/adolescentis/catenulatum/animalis* showing significantly lower detection in 16S cohorts as compared to WGS (Figs. 1D and 2). These observations could be a reflection of previously reported methodological limitations of the 16S approaches that impede efficient profiling of bifidobacterial amplicons²⁰.

Next, we asked whether variations in the resident gut microbiomes (specifically those of non-bifidobacterial taxa) had a significant association with the bifidobacterial features. For this purpose, in each cohort, we investigated whether the variations within the non-bifidobacterial taxa across the corresponding gut microbiomes (using two different measures, Bray–Curtis and Kendall) were significantly associated with the relative abundance of the bifidobacterial features (See “Methods”). Variations in the non-bifidobacterial gut microbiome were significantly associated with the overall *Bifidobacterium* detection rates in 74 and 58% of the study cohorts, when the community variations were profiled using the Kendall and Bray–Curtis measures, respectively (Fig. 2). Consistency of associations varied across species and age-groups, with the consistency of associations gradually decreasing from infant > adults > senior (Supplementary Fig. 1A, B). Five major bifidobacteria (*B. longum/breve/bifidum/dentium/pseudocatenulatum*) and the overall *Bifidobacterium* detection showed significant associations with non-bifidobacterial taxa in >50% of the



infant cohorts, and five of these bifidobacterial features (except *B. pseudocatenulatum* abundance) showing gradual decrease from infant to adult to senior cohorts. These patterns putatively indicate that the strength of the associations of the bifidobacteria with the non-bifidobacterial taxa decreases with age (despite the prevalence patterns of the different species being relatively similar between the adult and senior cohorts; Fig. 1C). Among the species, *B. longum* (showing

the highest prevalence across study-cohorts) also exhibited the most consistent associations with the non-bifidobacterial taxa across all age groups. This was followed by *B. bifidum*, *B. breve* and *B. pseudocatenulatum*, with *B. breve* and *B. bifidum* much stronger associations in the infant cohort.

These patterns suggest that certain non-bifidobacterial lineages may either influence or be influenced by the presence of specific

Fig. 1 | Bifidobacterium prevalence patterns are shaped by age, cohort lifestyle, and sequencing modality across global study populations. **A** Geographical distribution of the study cohorts used in this investigation is illustrated on a world map. Each country is highlighted in a distinct color, representing the different age groups for which gut microbiome data were available. The total number of microbiomes from each country is also indicated on the map. The total number of microbiomes from the different age groups are also indicated. Also shown are the number of 16S and WGS gut microbiomes and the number of microbiomes from different cohort lifestyle (IndustrializedUrban, UrbanRuralMixed, and RuralTribal). **B** Bar plots showing the percentage of studies in which each of the 42 *Bifidobacterium* taxa had a prevalence of $\geq 25\%$. Bars are ordered in decreasing order of the percentage of studies where a *Bifidobacterium* taxon shows prevalence $\geq 25\%$. Taxa detected in $\geq 25\%$ prevalence in at least 5% of cohorts were considered as consistently represented and selected for further analyses. **C** Principal coordinate analysis plot showing the variation in the overall prevalence pattern of the major bifidobacterial taxa between the infant, adult and senior study-cohorts (considering all cohorts). Each point indicates a study-cohort. The results of the PERMANOVA analysis (R -squared, p -value), highlighting the significance of these variations, are also indicated. **D** Heatmap summarizing the results of linear regression models examining variation in the prevalence of the nine bifidobacterial features (abundances of the eight major bifidobacterial taxa, and overall *Bifidobacterium* detection), across different pairwise comparisons. These include pairwise comparisons

between different age-groups, cohort lifestyle, and sequencing types, highlighting differences in species-specific and overall *Bifidobacterium* prevalence patterns. Please note comparisons between any two groups (derived from a specific study-cohort specific metadata, either age-group or sequencing-type or cohort-life-style) is performed after adjusting for the remaining two metadata. For each of the three metadata, the overall patterns for each individual species (or overall detection) are also indicated, where the statistical significance ($p < 0.05$) has been indicated as * notation. These are the two-sided p -values of the associations obtained using linear regression (two-sided), with adjustment for covariates as described. No multiple testing correction was applied, as analyses were performed independently across features and comparisons. **E** Principal coordinate analysis plot showing the variation in the overall prevalence pattern of the major bifidobacterial taxa between the adult cohorts belonging to the different cohort lifestyle (IndustrializedUrban, UrbanRuralMixed, and RuralTribal). Each point indicates a study-cohort. The results of the PERMANOVA analysis (R -squared, p -value), highlighting the significance of these variations, are also indicated. **F** Principal coordinate analysis plot showing the variation in the overall prevalence pattern of the major bifidobacterial taxa between the adult cohorts belonging to the different sequencing-types (16S or WGS). Each point indicates a study-cohort. The results of the PERMANOVA analysis (R -squared, p -value), highlighting the significance of these variations, are also indicated.

bifidobacteria. To investigate this further, we analyzed association patterns of the non-bifidobacterial taxa, and the three microbiome properties of Shannon index, Pielou's evenness index and estimated microbial load (only for WGS datasets), with the bifidobacterial features across diverse cohorts.

Non-bifidobacterial taxa exhibit distinct association patterns with bifidobacteria, varying by subject age, life-style (industrialization), and microbiome profiling method

Our objective here was to identify which non-bifidobacterial taxa showed consistent association with the individual with the bifidobacterial features across gut microbiomes and how did these association patterns vary in subjects from different age-groups (infant/adult/senior) and life-style patterns (IndustrializedUrban/UrbanRuralMixed/RuralTribal). We performed these in three steps.

First, we removed sparsely detected taxa across gut microbiomes (described in Supplementary Note 1 and Supplementary Fig. 2). Removal of such taxonomic noise is necessary as they can drastically decrease the sensitivity of the subsequent association analyses. Next, for each bifidobacterial feature, we identified the sets of non-bifidobacterial taxa consistently identified as the top-predictors across the study cohorts. For this, we built study-cohort-specific random forest (RF) models using subsets of consistent gut-associated taxa (identified in the first step) to predict the corresponding bifidobacterial features and selected consistent top-predictors across each model (see "Methods" and Supplementary Data 2–4).

Different non-bifidobacterial taxa showed varying and cohort-specific predictive power for different bifidobacterial features, as indicated by their feature importance scores (Supplementary Data 2–4 and Supplementary Figs. 3–11). These patterns varied by profiling method (16S/WGS), age-group (infant/adult/senior), and lifestyle (Industrialized urban/others). For example, for overall *Bifidobacterium* detection across age groups, while some non-bifidobacterial taxa were consistently identified as top-predictors across all cohorts of an age group, others were specific to either 16S or WGS datasets (Supplementary Fig. 3). Amongst the microbiome properties, while Shannon/Pielou's evenness indices showed high predictive power, the predictive power of estimated microbial load was relatively lower.

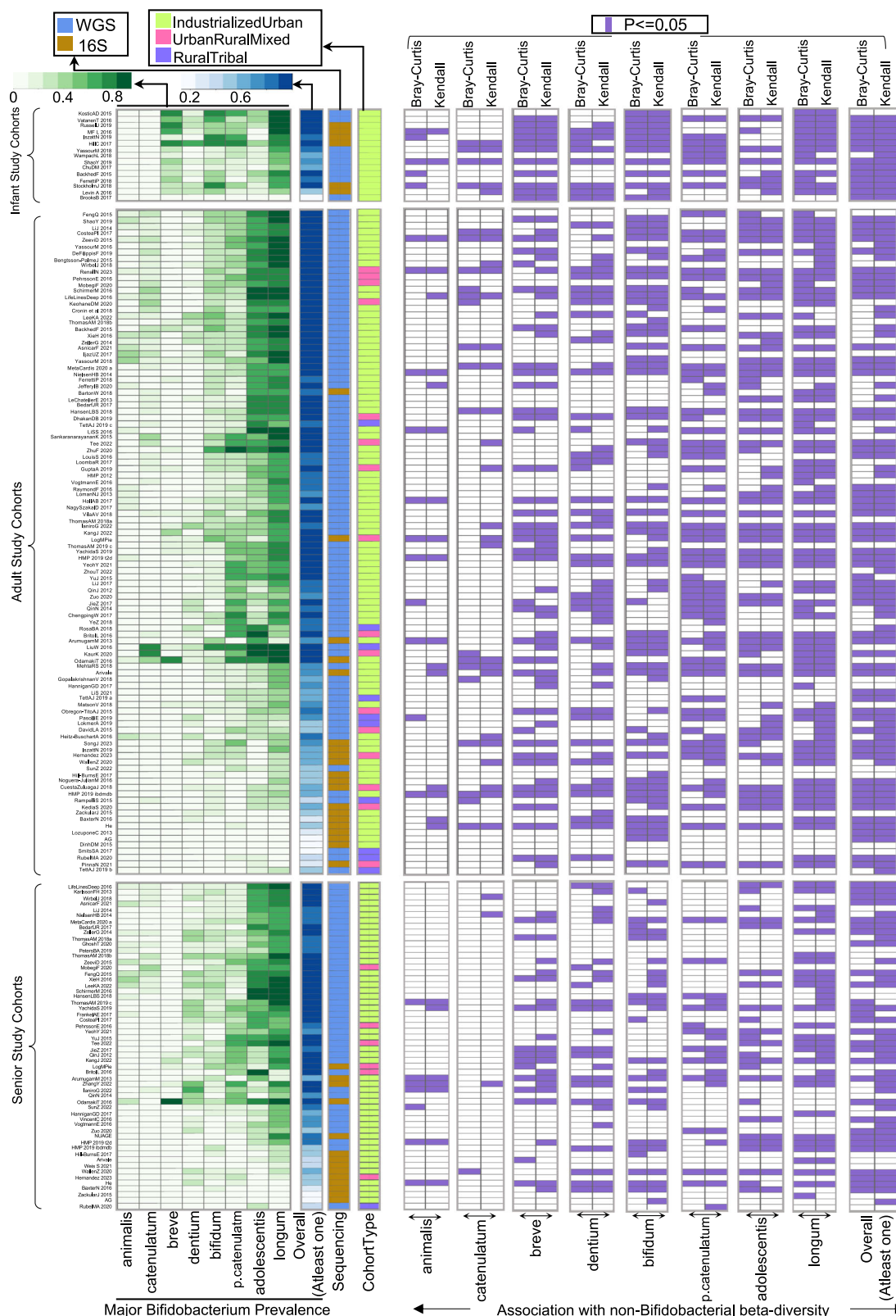
There were also clear differences in the top-predictor taxa lists and their feature importance scores across infant, adult and senior cohorts. RuralTribal cohorts displayed distinct feature importance patterns compared to IndustrializedUrban cohorts, with

UrbanRuralMixed cohorts showing intermediate profiles. Feature importance patterns of non-bifidobacterial taxa for predicting the abundance of the eight bifidobacteria also showed strong cohort-specific associations (Supplementary Figs. 4–11). These patterns were significantly linked to sequencing-type, age group, and life-style across all 8, 7, and 6 bifidobacteria taxa, respectively (Supplementary Figs. 12–14). Thus, the RF models revealed that the predictive power of the different non-bifidobacterial taxa for the various bifidobacterial features differs drastically with cohort characteristics, namely sequencing-type, age-group, and lifestyle, with cohorts sharing similar traits exhibiting more similar association patterns.

Then, we investigated the directionalities of these associations. For a given bifidobacterial feature and cohort with specific characteristics (sequencing-type: 16S/WGS; age-group: infant/adult/senior and cohort-life-style: IndustrializedUrban/UrbanRuralMixed/RuralTribal), we selected the corresponding set of top-informative (top-predictor) non-bifidobacterial taxa (listed in Supplementary Data 2–4 and Supplementary Figs. 4–11), and then computed the Spearman correlations between the abundances of these taxa (and the values of Shannon-index/Pielou's evenness/estimated microbial load) and the corresponding bifidobacterial features across all study-cohorts of that type (Supplementary Data 5–10). These Spearman correlation patterns were then utilized to compute association-scores of the different non-bifidobacterial taxa with different bifidobacterial features for the given cohort-type (Fig. 3A). The association-scores were quantitative measures of the extent of association of each non-bifidobacterial taxa with each bifidobacterial feature in microbiome cohorts with specific characteristics, stratified based on cohort characteristics (profiling strategy, age-group and cohort-life-style).

Analysis of these association scores revealed several notable patterns. Four *Actinobacteria* phyla (i.e., the same phylum as *Bifidobacterium*) members, namely *Collinsella aerofaciens*, *Collinsella intestinalis*, *Enorma massiliensis*, and *Olsenella scatoligenes* (all belonging to the order *Coriobacteriales*), showed the most consistent positive associations with multiple bifidobacterial features cohorts spanning diverse age-groups and lifestyle patterns. We also observed similar patterns for *Lactobacillus ruminis* (another health-associated *Lactobacilli*) and *Intestinibacter bartlettii*, along with Shannon/Pielou's evenness indices (Fig. 3B).

Association patterns for other non-bifidobacterial taxa differed by age and lifestyle and varied depending upon the specific bifidobacterial feature. In infant gut microbiomes, taxa, such as *Eggerthella*



lenta and *Gordonibacter pamelaee* (both *Actinobacteria* phylum, like *Bifidobacterium*), along with another *Bifidobacteriaceae* family *Aeriscardovia aeriphila* and a human milk-associated putative probiotic *Lactobacilli* (*L. rhamnosus*)²¹, showed consistent positive associations with multiple bifidobacterial features. Similar pattern was observed for multiple *Bacteroides* (*B. uniformis/vulgatus/fragilis/thetaiotamicon/stercosis*) and *Parabacteroides* (*P. distasonis/merdae*), which are

typically observed in normal gut microbiomes. However, the above associations weakened/disappeared/in certain cases even reversed in adult and senior microbiomes.

Several butyrate-producing *Firmicutes* taxa from genera, including *Coprococcus*, *Eubacterium*, *Ruminococcus*, *Roseburia*, *Gemmiger*, *Dorea*, and *Fusicatenibacter*, previously linked to gut health^{16,22,23}, showed consistent positive association scores with multiple

Fig. 2 | Prevalence patterns of bifidobacterial taxa across study-cohorts, and associations with the non-bifidobacterial gut microbiome. (Left panel) Heatmap showing the prevalence patterns of the eight major bifidobacterial taxa and the fraction of microbiomes with at least one bifidobacterial species detected (with relative abundance ≥ 0.0001) across the different study cohorts. The study cohorts have been divided based on the age groups (infant, adult, and senior). Also indicated are the sequencing types (WGS or 16S) and the cohort type (IndustrializedUrban, UrbanRuralMixed, and RuralTribal). Panels on the right indicate the associations of the non-bifidobacterial community compositions with the relative abundances of the eight different bifidobacterial taxa, as well as the overall *Bifidobacterium* detection rate. The associations were computed in each study-cohort using PERMANOVA with the non-bifidobacterial community variations using two different distance measures, namely Kendall and Bray–Curtis. The eight major bifidobacterial taxa displayed strong variations in their prevalence, both with

respect to each other, as well as with respect to the age-group and cohort-type of the study-cohorts. Many of these variations replicated previous observations. The pattern of association of the non-Bifidobacterial community with the representation of the different Bifidobacterial taxa strongly correlated with their prevalence rates across study cohorts of different age groups and cohort type. (Right panel) Heatmap showing, for each cohort, whether significant associations (PERMANOVA $p \leq 0.05$; purple) were detected between non-bifidobacterial community composition and nine bifidobacterial features using PERMANOVA utilizing either Bray–Curtis or Kendall distances as the non-bifidobacterial community beta-diversity measures. Statistical significance of associations was assessed using PERMANOVA, and $p \leq 0.05$ were considered significant. Multiple testing correction was not applied, as PERMANOVA was conducted separately within each study-cohort to assess independent associations, and the analyses were not intended for aggregated inference across a unified hypothesis space.

Bifidobacterium species across age groups and cohort types. Notable exceptions were *B. dentium* (a known pathobiont) and *B. breve* in adult and senior gut microbiomes, which exhibited consistent negative associations with these health-associated core gut lineages, suggesting unfavorable health links. In contrast, *B. adolescentis* and *B. catenulatum* displayed consistent positive associations with these putatively beneficial lineages. These two bifidobacteria members also showed consistent negative associations with multiple pathobiont *Firmicutes*, like *Ruminococcus gnavus*, *Flavonifractor plautii* and *Clostridium bolteae/clostridioforme/hathewayi/symbiosum/ramosum*, further highlighting their health-associated links.

Another group exhibiting strong age-dependent variations in association patterns comprised oral-associated taxa from the genera *Streptococcus*, *Rothia*, and *Haemophilus*. These taxa were generally negatively associated with multiple bifidobacterial features in infant gut microbiomes, whereas the pattern reversed in adults, where *B. dentium* showed consistent positive associations with multiple oral taxa.

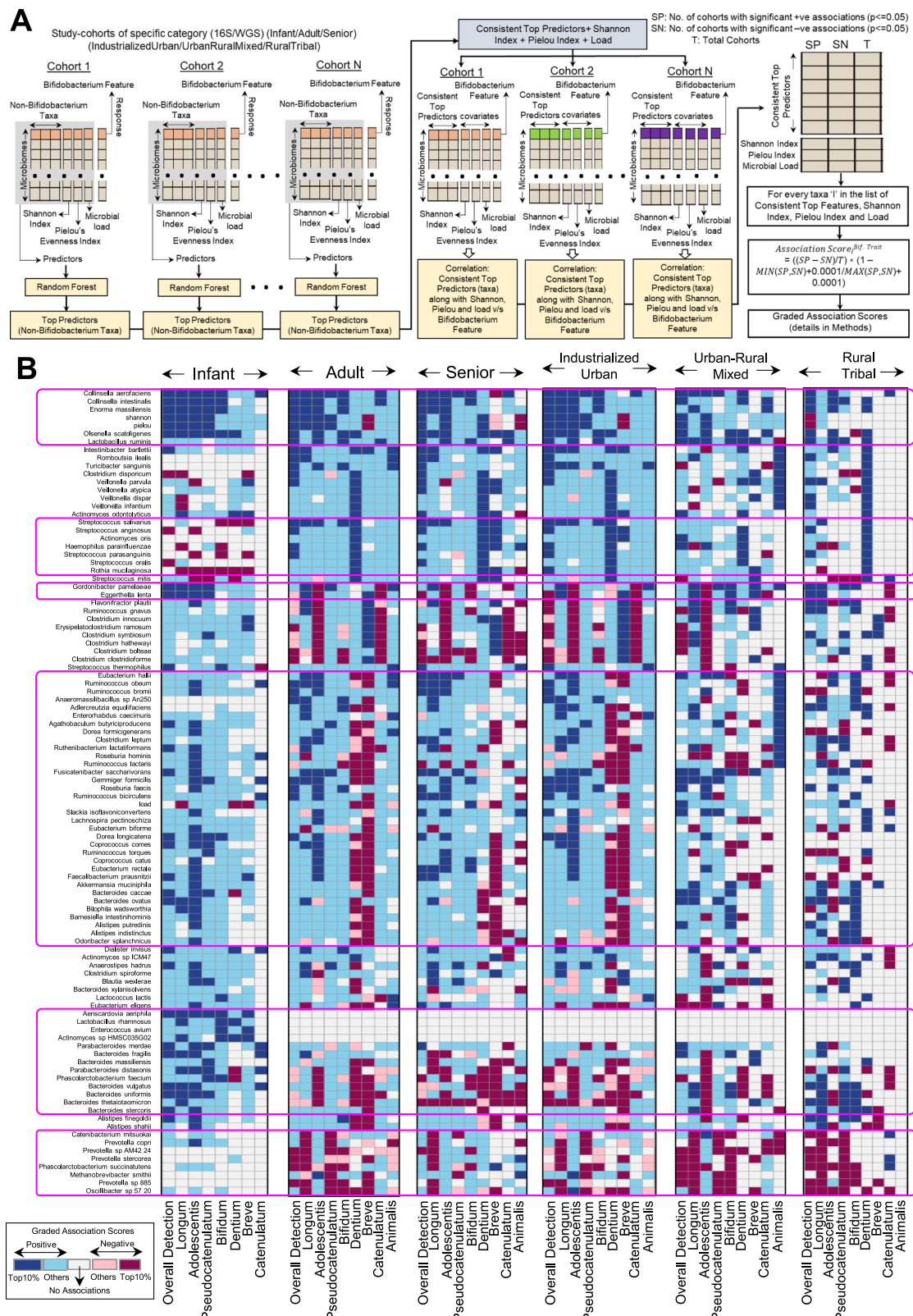
Association scores also varied along the lifestyle and industrialization gradient from IndustrializedUrban to RuralTribal gut microbiomes. In RuralTribal cohorts, associations with *Firmicutes* were generally attenuated, with reduced strengths observed for both positive and negative relationships (as compared to IndustrializedUrban cohorts). A consistent feature of RuralTribal cohorts was negative association scores for *Prevotella* species (*copri*, *stercorea*, *sp. AM.42_24*, *sp. 885*), *Catenibacterium mitsuokai*, *Phascolarctobacterium succinatutens*, *Oscillibacter sp. 57_20*, and *Eubacterium bifforme*. UrbanRuralMixed cohorts showed intermediate patterns, with consistent negative associations for *Prevotella*, *Catenibacterium*, *Oscillibacter*, and *Phascolarctobacterium*, resembling patterns observed in RuralTribal groups. The strengths of associations with *Firmicutes* were generally lower than IndustrializedUrban cohorts but higher than RuralTribal groups. The consistent negative associations of *B. adolescentis* with pathobiont *Firmicutes* were however, reproduced as in the IndustrializedUrban groups. We computed the Association-Scores separately for the 16S cohorts (Supplementary Data 11 and 12). Association-scores for commonly detected non-bifidobacterial taxa with individual bifidobacterial feature showed significant positive correlations between the two sequencing-types (16S/WGS) in 22 of the 34 combinations (65%) across the four cohort categories (Supplementary Fig. 15). Shannon/Pielou's-evenness indices showed consistent positive associations with four bifidobacterial features (overall *Bifidobacterium* detection and *B. longum/adolescentis/pseudocatenulatum* abundances) in both the Infant and adult gut microbiomes from IndustrializedUrban cohorts (irrespective of sequencing-type), but weaker associations in microbiomes from RuralTribal cohorts (Fig. 3 and Supplementary Data 5–12). We next investigated the effects of different host phenotypes and associated factors on these patterns, starting with the effect of disease groups.

Control and disease groups show similar association patterns between non-bifidobacterial taxa and bifidobacteria

Given the limited availability of patient gut microbiomes of different diseases from non-Industrialized populations, here, we specifically investigated the variation of non-bifidobacterial association patterns (with different bifidobacterial features) in the adult/senior gut microbiomes from IndustrializedUrban cohorts. To investigate this in a comprehensive manner, we further supplemented our initial global gut microbiome collection (of 46,218 gut microbiomes, 122 study-cohorts) with data from 27 study-cohorts (an additional 5026 gut microbiomes) containing matched diseased and non-diseased gut microbiomes (resulting in a total of 13,913 gut microbiomes from 69 study-cohorts with matched-diseased-control sub-groups) (see “Methods”, Supplementary Note 2 and Supplementary Data 1). For ensuring robustness of findings, we focused on six diseases, namely, colorectal cancer (CRC), type-II-diabetes (T2D), cardiovascular disease, inflammatory bowel disease, COVID and Parkinson's, for which ≥ 7 study-cohorts were available in our data collection. For each of these diseases, we computed the association scores of the non-bifidobacterial taxa with the different bifidobacterial features as described in the previous section (Fig. 3A and “Methods”), and compared the same with those obtained with respect to the non-diseased controls from the same study-cohorts (Fig. 4 and Supplementary Data 13).

For most bifidobacterial features (except *B. catenulatum/animalis* abundance) and most diseases, association scores of non-bifidobacterial taxa with bifidobacterial features in non-diseased controls showed significant positive correlations with those observed in disease cohorts, indicating largely conserved association patterns (Supplementary Fig. 16 and Supplementary Data 14). The strongest reproducibility was observed for the abundances of *B. adolescentis*, *B. longum*, *B. dentium*, and *B. breve*, as well as for overall *Bifidobacterium* detection. This was followed by *B. pseudocatenulatum* and *B. bifidum*, which showed significant positive correlations between control-derived association scores and those from four disease groups. This indicates that, for most of the diseases and the majority of bifidobacteria, non-bifidobacterial taxa showing higher association-scores with a given bifidobacterial feature in controls also showed higher association-scores with the same feature across the majority of disease-groups.

We observed several group-agnostic consistent patterns in this investigation that mimicked those we observed for the overall association scores (Figs. 3 and 4). *B. adolescentis*, followed by *B. longum* and the overall *Bifidobacterium* detection showed consistent positive associations with health-associated butyrate-producing *Firmicutes* along with Shannon/Pielou's evenness indices. Similar patterns were observed to a lesser extent for *B. pseudocatenulatum* and *B. bifidum*. *B. adolescentis* also showed multiple negative associations with the pathobiont *Firmicutes* (Fig. 4; also seen in Fig. 3), suggesting the strongest association of these taxa with health-associated microbiome members. The particularly strong associations seen in disease-specific



contexts further underscore their potential role beneficial probiotic organisms (especially *B. adolescentis/longum*) capable of contributing to the amelioration of diverse pathological conditions (Fig. 4).

At the other end of the spectrum were *B. breve* and the pathobiont *B. dentium*, that showed the opposite pattern with clear links to a putatively detrimental microbiome (irrespective of control/disease-groups): consistent negative associations with health-associated

Firmicutes and overall microbiome diversity and evenness and to a lesser extent with normal-gut-associated *Bacteroidetes* members; multiple positive associations with disease-associated taxa (*B. breve* with pathobiont Firmicutes; *B. dentium* with oral-taxa that have been linked to multiple diseases²⁴).

Another association pattern robustly observed irrespective of host-status (control-v/s-disease-groups) was consistent positive

Fig. 3 | Association patterns of the non-bifidobacterial taxa and three microbiome properties (Shannon index, Pielou's evenness index and estimated microbial load with different bifidobacterial features), quantitatively captured as association-scores, stratified by profiling-strategy or sequencing-type (16S/WGS), age-group (infant/adult/senior), and lifestyle category (IndustrializedUrban/UrbanRuralMixed/RuralTribal). **A** Schematic illustration of the methodology used to compute the association score for a given non-bifidobacterial taxon, along with Shannon index, Pielou's evenness index and estimated microbial load with a specific bifidobacterial feature within cohorts of a defined category. Finally, the overall association scores are graded based on the consistency of association patterns to compute graded association scores (following the approach as mentioned in Supplementary Note 2). **B** Association-score patterns for different

non-bifidobacterial taxa and three microbiome properties with various bifidobacterial features across cohorts of different categories. Only those non-bifidobacterial taxa with showing an association score with absolute value ≥ 0.25 with at least two *Bifidobacterium*-feature-cohort-type combinations are shown here. For ease of display for driving insights and identifying specific patterns, association-scores for each *Bifidobacterium*-feature-cohort-type combinations were further graded with the top 10% positive association-scores shown in dark blue and the top 10% negative association-scores shown in dark pink, as denoted in the legend (refer to Supplementary Note 2). The complete association-scores list and the association-scores for 16S cohorts are provided in Supplementary Data 11 and 12.

associations of especially *B. longum/adolescentis/bifidum* abundances with other *Coriobacteriaceae* members (Figs. 3 and 4). We next investigated the effects of the different microbiome and host-associated characteristics on the association patterns between the non-bifidobacterial taxa and the nine bifidobacterial features.

Pattern of associations with non-bifidobacterial taxa remains consistent for the dominant bifidobacteria after adjusting for microbiome properties, diet, medications, and clinical characteristics of subjects

We first investigated the effects of microbial load. Data pertaining to the quantitative microbiome profiling (QMP) was already available in the MetaCardis cohort (in the form of sample-specific microbial load)²⁵ (Supplementary Note 3). We next evaluated whether accounting for microbial load significantly alters the observed associations. For this, we first obtained estimated quantitative microbiome profiles (or absolute abundance profiles) of each taxon by multiplying the taxa abundances with the sample-specific microbial load (as suggested in the original study). We then recomputed associations between non-bifidobacterial taxa and bifidobacterial features using the absolute abundances ("Methods" and Supplementary Note 3).

The resulting absolute abundance-derived correlations were highly consistent with the corresponding association scores obtained using relative abundances. Specifically, for WGS-sequenced adult and senior industrialized cohorts, significant positive correlations were observed between absolute abundance-derived associations and relative-abundance-based association-scores for all nine bifidobacterial features in adults and for seven features in seniors (Supplementary Fig. 17). Moreover, for the five bifidobacterial features that showed the strongest and most consistent patterns across controls and disease-groups (Fig. 4, Supplementary Figs. 16 and 17), correlations were uniformly high (≥ 0.50), indicating that the observed associations were robust and largely independent of whether relative or absolute abundances are used.

We then expanded our investigations to all cohorts and evaluated whether the three microbiome properties, Shannon/Pielou's evenness indices and estimated microbial load for WGS datasets²⁵, confounded the observed associations between non-bifidobacteria taxa and bifidobacteria. For this, we recomputed microbiome-property-adjusted association scores, wherein study-specific correlations between non-bifidobacterial taxa and the nine bifidobacterial features used to derive the association scores were recalculated as partial correlations, adjusting for the three microbiome properties (see Supplementary Note 3, "Methods", Supplementary Figs. 18A, 19, and Supplementary Data 15). Partial correlations quantify the strength of association between two variables while controlling for one or more confounding factors. Correlating the microbiome-property-adjusted association scores with the overall association scores for all cohort types revealed significant positive correlations in 91 of the 95 tested scenarios, indicating that non-bifidobacterial taxa showing unadjusted higher association scores with a given bifidobacterial feature also showed higher microbiome-property-adjusted association scores with the same

property (Fig. 5A and Supplementary Fig. 18B). Thus, overall microbiome properties exert limited influence on these association patterns.

We next investigated the effects of diet, medications, and host clinical markers. Across the investigated cohorts, a subset (listed in Supplementary Data 16) contained metadata pertaining to these three metadata categories. For each cohort-category combination, both the metadata schemas and measured variables varied substantially between cohorts. To address this heterogeneity, we summarized variations for a given metadata category in a cohort using robust principal component analysis and principal coordinate analysis, retaining the top three components from each method (which together captured on average 73.47% of total variation across all cohort-category combinations) (Supplementary Fig. 20).

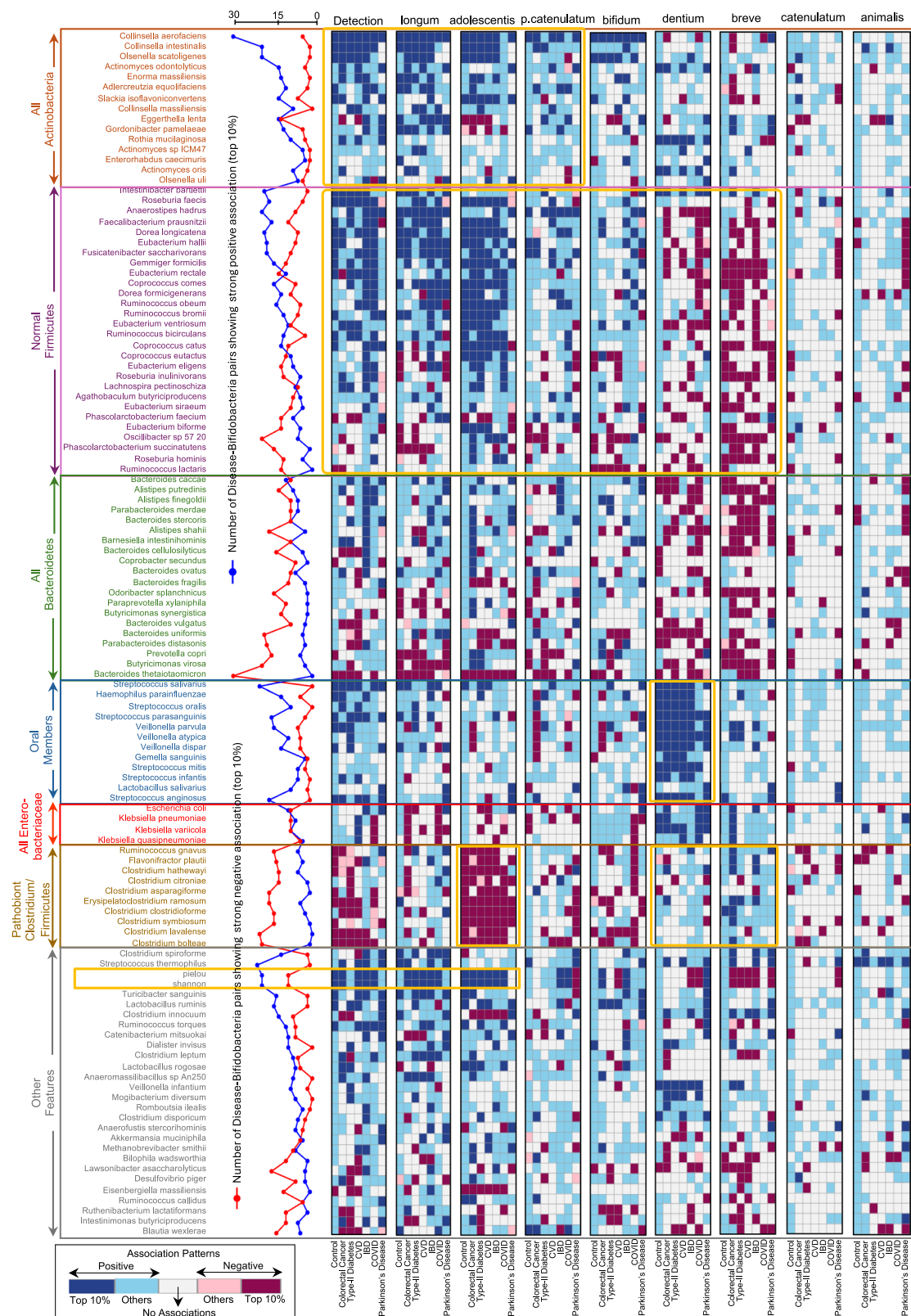
We then recalculated associations between non-bifidobacterial taxa and bifidobacteria features (in each cohort-category combination) as partial correlations, adjusting for these six metadata components. This yielded metadata-adjusted association-scores that account for major confounding variation (Supplementary Figs. 18C, 21, Supplementary Note 3, and Supplementary Data 17). Correlating these metadata-adjusted association scores for the different metadata with the overall association scores (for the specific cohort types) yielded significant positive correlations in 46 out of the 48 scenarios (Fig. 5B), indicating significant similarities in the adjusted and unadjusted association patterns.

The similarities were especially strong for five bifidobacterial features (*B. longum/adolescentis/breve/dentium* relative abundances and the bifidobacterial detection rates) (average correlation = 0.56). The other four bifidobacterial taxa showed significant correlations across a majority of scenarios, albeit with weaker association strengths. These observations indicate that these metadata have a limited influence on the majority of associations between non-bifidobacterial taxa and bifidobacteria.

Careful investigations of the association patterns across the four variants of association-scores (investigated herein; overall, across diseases, microbiome-property-adjusted; diet/medication/clinical-marker-adjusted) further highlighted the reproducibility of the associations of some of the bifidobacteria with some of the major non-bifidobacterial taxonomic groups (described earlier) (Fig. 5C). For *B. animalis* and *B. bifidum*, which did not show consistent pattern with any broader non-bifidobacterial groups, we observed consistent positive associations with specific taxa consortia (*B. animalis* with known probiotic *Streptococcus thermophilus*, *Gordonibacter pame-laeae*, and *Intestinibacter barlettii*; *B. bifidum* with *Collinsella aerofaciens/intestinalis*, *Olsenella scatoligenes* and health-associated *Lactobacillus ruminis*) (Fig. 5C).

Post-treatment increase of *Bifidobacterium* probiotics can be significantly predicted by microbiome-level receptive-scores, computed using the association-scores and baseline abundances of the administered *Bifidobacterium* probiotic

We next investigated the translational applicability of these association scores. We collated data from eight independent trials in this sub-



analysis, encompassing a total of 1633 gut microbiomes from 507 subjects (Supplementary Data 1 and 18; described in detail in Supplementary Note 4)^{11,15,26–31}. While seven of these trials directly administered different live *Bifidobacterium* probiotics (either alone or in combination), one of the trials had utilized *B. adolescentis*, *B. animalis* and galacto-oligosaccharide (GOS), a prebiotic aimed at increasing bifidobacterial species composition in general in infants.

Across the trials, the association patterns of the non-bifidobacterial taxa were reasonably reproduced (significant positive correlations between 0.25–0.69 for WGS cohorts (see “Methods”, Supplementary Note 5, Fig. 6A, and Supplementary Data 19A). We also observed strong reproducibility of the association patterns for the 16S adult/senior subgroup cohorts, one Infant-specific cohort¹⁵ (Supplementary Note 5, Supplementary Fig. 22A, B, Supplementary Data 19B and 20).

Fig. 4 | Associations of the non-bifidobacterial taxa groups with the different bifidobacterial features show consistent patterns between control and six different disease groups. Association scores computed for different non-bifidobacterial taxa with various bifidobacterial features across cohorts of six different diseases viz. colorectal cancer (CRC), type-II diabetes (T2D), cardiovascular diseases (CVD), inflammatory bowel disorder (IBD), COVID and Parkinson's disease) and apparently non-diseased individuals obtained by investigating the matched-control-sub-cohorts for each of the 69 studies corresponding to these six diseases. Only those diseases were considered where the data was available for at least seven study-cohorts, and only those non-bifidobacterial taxa showing consistent association patterns (absolute value of association-score ≥ 0.25 to capture both consistent positive and negative associations) with different bifidobacterial features in more than two *Bifidobacterium*-feature-disease combinations are shown here. In this figure, the heatmap shows the graded association-scores, which is computed

using the disease association-scores by grading the scores based on the strength of associations as elaborated in Supplementary Note 2. The deep blue and deep raspberry colors indicate the strong (top 10%) positive and negative associations, respectively, whereas sky blue and light pink cells indicate other (apart from top 10%) positive and negative associations, respectively. The complete graded association-scores list is provided in Supplementary Data 13. The line plot shows the number of bifidobacteria-category (diseases/controls) pairs showing strong association (top 10% association scores across a particular category), where the blue line indicates strong positive associations and the red line indicates strong negative associations. Non-bifidobacterial taxa are colored depending upon the groupings based on mostly their taxonomic lineage as well as their associations with human health. Interestingly, we observe certain non-bifidobacterial group-specific patterns with different bifidobacteria, which have been highlighted on the heatmap panel.

We next investigated the translational applicability of our findings. In other words, we explored whether the cohort-type-stratified association scores of the various non-bifidobacterial taxa for the different bifidobacteria (Fig. 3A) could be utilized to formulate microbiome-level quantitative measures for stratifying or ranking individuals based on their gut microbiome's receptivity to different *Bifidobacterium* probiotics. We referred to these measures as microbiome-level receptive scores. Given a target bifidobacterial feature and microbiomes of a specific category (based on host-age, host life-style and 16S/WGS profiling), the corresponding receptive scores were computed using a two-step approach (Fig. 6B). In the first step, given a set of microbiomes of a specific category, we considered the abundances of all non-bifidobacterial taxa and computed cumulated weighted sums, where in the abundances of the non-bifidobacterial taxa were weighted by their corresponding association scores computed previously for microbiome-types of that category (Fig. 3B) (the abundances of only those non-bifidobacterial taxa with matching association scores were considered). This approach ensured that microbiomes with a greater abundance of non-bifidobacterial taxa, which globally show consistent positive associations with a target *Bifidobacterium*, received higher scores (and those having greater abundance of microbes showing consistent negative associations received lower scores). Ranking these scores across a population-cohort facilitates effective ranking of individuals based on how well their current microbiome is expected to support the target *Bifidobacterium* species. We referred to this the expected-representation-score for that *Bifidobacterium*.

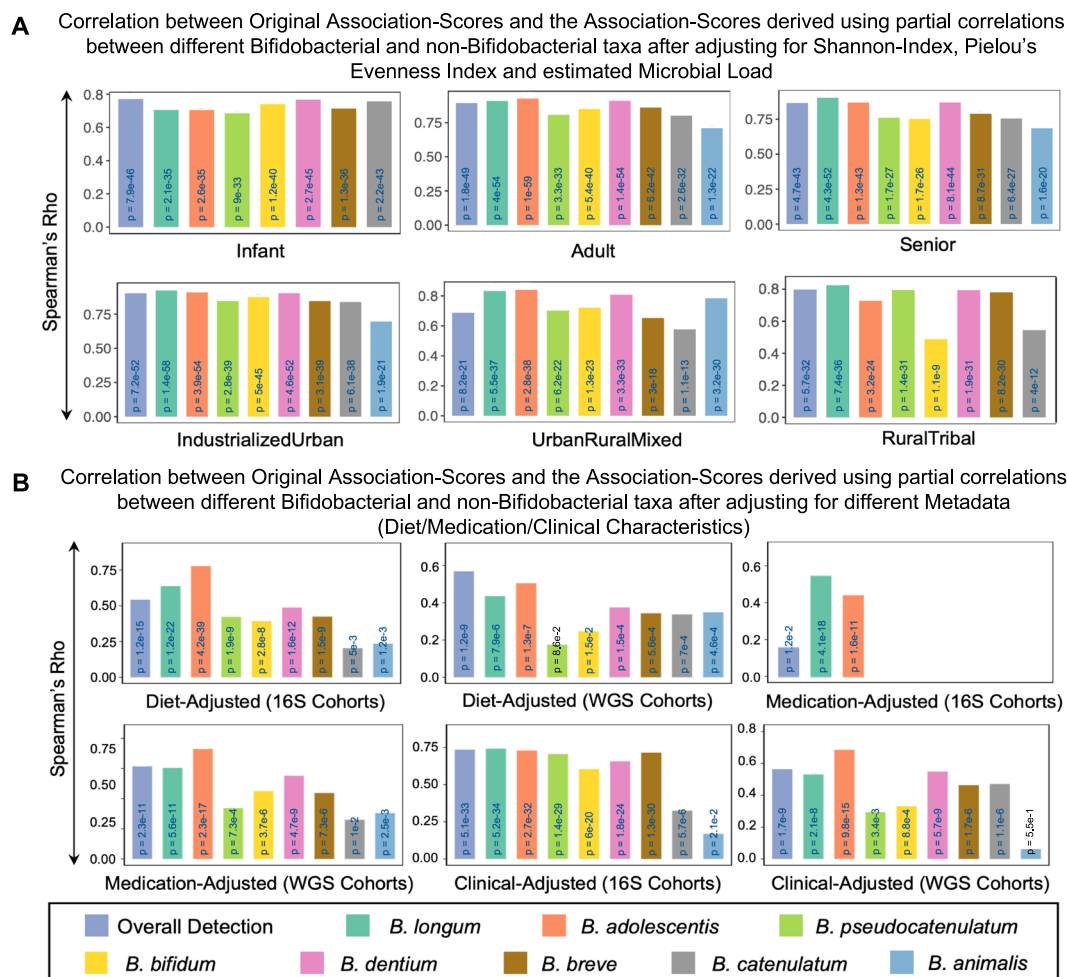
Expected-representation-scores quantify how supportive a community is for a given *Bifidobacterium*, based on the composition of other resident taxa. However, the receptivity of a community for a particular *Bifidobacterium* feature (in terms of the future success of its administration/supplementation in a community, measured by persistence or increase) also depends on the existing abundance of that *Bifidobacterium*. Communities already rich in a specific *Bifidobacterium* may have limited capacity to accommodate more members of the same lineage and might not show a noticeable increase/persistence after its administration^{32,33}. Therefore, it's equally important to rank individuals based on the current (or the baseline) abundance of the target *Bifidobacterium* in their gut. Thus, for every gut microbiome from a specific population group (based on cohort characteristics as described above) and a target *Bifidobacterium* feature, we can calculate a receptive score as the difference between the ranked expected-representation-score and the ranked baseline abundance of the target *Bifidobacterium* feature. A higher receptive-score suggests that the microbiome can accommodate a higher representation of the target *Bifidobacterium* probiotic, thus identifying individuals whose gut environment is conducive to increase of the target *Bifidobacterium*, considering both its current levels and the supportive microbial context.

Next, we validated the concept/applicability of receptive scores in the eight validation *Bifidobacterium* intervention trials (Supplementary Data 18). The first trial to be investigated was by Maldonado-Gomez et

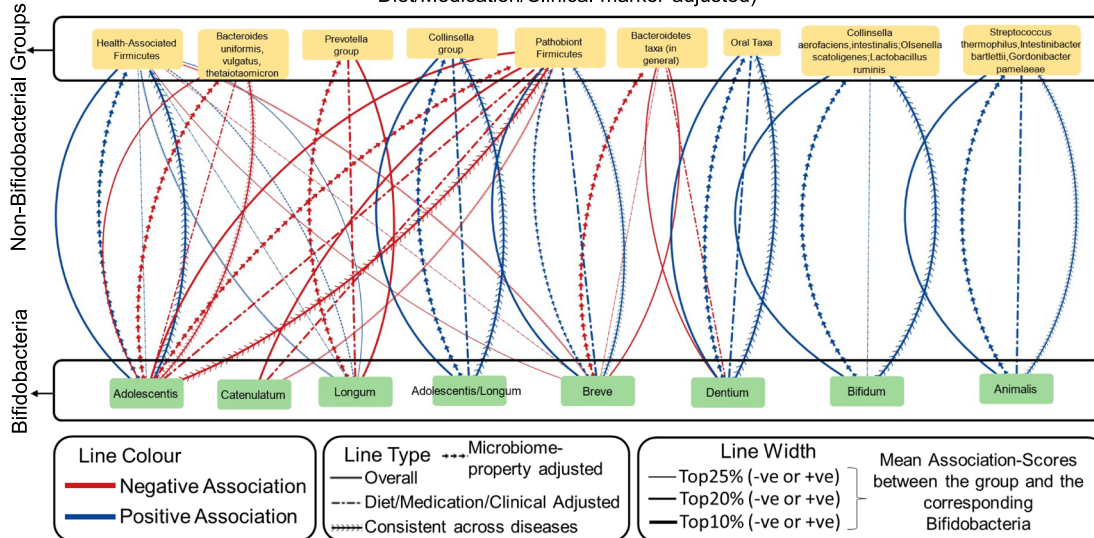
al., which had investigated the persistence of *Bifidobacterium longum* AH1206 in 23 individuals¹¹. In only 30% individuals (the persisters), the strain had shown long-term persistence, and this pattern was associated with variations in the baseline microbiome. We thus checked whether the receptive-scores for *B. longum* (using the corresponding *B. longum* association-scores in microbiome types profiled using WGS in adult individuals with IndustrializedUrban life-styles) in the baseline gut microbiomes of this specific trial dataset could distinguish future persisters and non-persisters. Notably, the computed baseline receptive scores were significantly higher for persisters compared to non-persisters (Mann-Whitney *U*-test p : 0.018; Fig. 6C). This provided the first validation, suggesting that baseline receptive scores can effectively identify individuals likely to exhibit future persistence of the target probiotic.

Encouraged by this initial validation, we investigated all eight *Bifidobacterium*-associated trials (See Supplementary Note 4 and Supplementary Data 18)^{15,26,28–31,34}. For each of the trials, we identified the target *Bifidobacterium*-related feature (or features) that was (were) expected to be increased. The feature investigated was either the ranked abundance of any of the eight *Bifidobacterium* taxa or the overall *Bifidobacterium* detection. To evaluate whether the baseline receptive score for a given target property was predictive of its future increase, we utilized two approaches. First, we employed robust linear regression (RLR) models across each trial (see "Methods" and Fig. 6D), where we tested whether the receptive-score was associated with the property's post-treatment increase, after adjusting for its cohort-ranked baseline abundance as a covariate to control for its potential confounding effects. This was because both the receptive-score and the ranked post-treatment increase are derived from the ranked baseline abundance (Fig. 6D). Secondly, for each trial-*Bifidobacterium* probiotic combination, we developed a linear regression-based validation approaches to predict post-treatment increases in the corresponding *Bifidobacterium* feature. Predictions were based on baseline abundance and the baseline receptive score for the relevant feature, using leave-one-observation-out cross-validation (LOOCV), in which the target subject was excluded from model training.

Across the 13 trial-*Bifidobacterium*-feature combinations (in eight trials), the RLR-models showed significant positive relationship between the baseline receptive-score and an increase of the target *Bifidobacterium*(s) (RLR coefficient > 0 ; $p \leq 0.05$) in 9 combinations (in 11 combinations, positive RLR coefficient with $p \leq 0.1$). Thus, overall, in 9 out of 13 trial-Bifidobacterium-feature combinations (69.23%), we observed that higher baseline receptive scores are significantly associated with increased post-treatment increase/persistence of the target *Bifidobacterium*. Similarly, in LOOCV, a linear regression model showed significant positive correlations between predicted and observed increases in *Bifidobacterium* features in 69.23% of combinations (9/13 combinations for LOOCV) (Supplementary Note 4, Fig. 6D, and Supplementary Data 21). To further assess the robustness of these associations, we recomputed both the RLR and



C Association patterns between Bifidobacteria with some of the major non-Bifidobacterial taxonomic groups across four variants of Association-Scores (Overall, Across Diseases, Microbiome-property-adjusted; Diet/Medication/Clinical-marker-adjusted)



LOOCV analyses for the same trial-*Bifidobacterium*-feature combinations by additionally including the association scores derived from microbiome properties, namely, Shannon/Pielou's evenness indices, as well as estimated microbial load (in WGS cohorts). The RLR models, incorporating these additional features, retained a significant positive association between the receptive score and post-treatment increases in the target *Bifidobacterium* features in 9 out of 13 combinations, while

the LOOCV-based linear regression models showed significant correlations between predicted and observed increases in 10 out of 13 combinations, including the Gomez trial, indicating similar performance (Supplementary Data 21). Collectively, these results demonstrated the ability of the receptive scores as an efficient predictor of *Bifidobacterium* increase or persistence across trials (Supplementary Data 21).

Fig. 5 | Association patterns between non-*Bifidobacterium* taxa and *Bifidobacterium* features are conserved across host clinical phenotypes and after adjustment for diet, medications, and overall microbiome properties. **A** Bar plots showing Spearman correlations between the original association-scores and the association-scores computed using study-specific partial correlations of non-bifidobacterial versus bifidobacterial features, after adjusting for microbiome-level properties, Shannon index, Pielou's evenness index and estimated microbial load. Results are shown across age groups (infant, adult, and senior) and cohort lifestyle (IndustrializedUrban, UrbanRuralMixed, and RuralTribal). Bars correspond to individual *Bifidobacterium* taxon and overall *Bifidobacterium* detection as denoted in the color legends. **B** Bar plots showing Spearman correlations between overall association-scores and association-scores adjusted for host metadata, including diet, medication, and clinical characteristics, stratified by sequencing-type (16S and WGS). Spearman's rank correlation (two-sided) was used to assess associations between scores. $p \leq 0.05$ were considered significant. Multiple testing correction was not applied, as each correlation was evaluated independently within

predefined strata, and the analyses were intended to assess consistency of associations rather than to identify novel discoveries across a common hypothesis set. All p -values are displayed within each bar; statistically significant associations ($p \leq 0.05$) are highlighted in blue, while non-significant ones are shown in black. **C** Network diagram summarizing the most consistent association patterns between bifidobacteria and non-bifidobacteria (both broad taxonomic groups as well as individual taxa (that are conserved across host clinical phenotypes and after adjusting for host diet, medications, and overall microbiome properties)). The network was constructed by integrating association patterns from four association-score variants: overall, stratified by controls and disease-groups, microbiome property adjusted, and diet, medication, and clinical adjusted. Distinct edge shapes indicate associations that were consistently observed across the individual variants. Edge widths represent the mean positive or negative association scores between bifidobacteria and non-bifidobacteria, stratified by levels of consistency (top 10%, top 20%, and top 25%). Edge colours indicate the direction of the association, with blue denoting positive and red denoting negative associations.

In addition, there were also groups of trials administering the same bifidobacteria (as probiotic) on cohorts of the similar types (three trials administering *B. animalis* and two trials administering *B. adolescentis* on adult/senior IndustrializedUrban cohorts)^{26,28,29}. Within cohort-pairs of the same category, we further assessed the robustness of the receptive score framework using cross-cohort validations of linear regression models (trained on one cohort and test on the other) to predict the post-treatment increase of the administered probiotic using the cohort-stratified baseline receptive-score of that bifidobacteria and its baseline abundance as predictors (Fig. 6E). Seven of the eight such pairwise cross-cohort validations yielded significant positive correlations between predicted and actual abundance of the administered bifidobacteria (Fig. 6F–I and Supplementary Fig. 23), indicating the translational applicability of the concept of receptive-scores to predict response in individuals across novel unseen cohorts. Together, these results highlight the translational applicability of the identified association patterns, where in the cohort-type-stratified baseline receptive scores, along with baseline abundance of the target bifidobacteria, can be used to predict its future persistence or increase in subjects.

Two of the trials, RamakrishnanM_2025 and BaZ_2021, also contained metadata pertaining to the change in different health-linked clinical parameters pertaining to a subset of subjects^{29,31}. For eight of the 10 subjects constituting the RamakrishnanM_2025 trials, there was a negative correlation between receptive scores (for *B. adolescentis*) and an increase in gastrointestinal symptoms (Diarrhea incidence: $R = -0.76$; $p = 0.02$ [significant]; abdominal pain: $R = -0.58$; $p = 0.13$; faecal urgency: $R = -0.41$; $p = 0.3$) (Supplementary Fig. 24A–E and Supplementary Note 6). While we could not attain statistical significance because of the limited cohort size, there was a significant positive correlation between receptive score and overall reduction in all gastrointestinal symptoms ($R = 0.73$; $p = 0.038$). Similarly, for the BaZ_2021 *B. animalis* trial, higher receptive scores at baseline were associated with reduction in multiple markers of inflammation, like CD3–CD8 ratio, interleukin-2, tumor-necrosis-factor, and interleukin-6. Thus, for the limited set of subjects where the clinical response data was available, higher baseline receptive scores were associated with better clinical responses (Supplementary Fig. 24F–I and Supplementary Note 6).

Specific genome-level functional features of the different non-bifidobacterial taxa link to their association scores with the different bifidobacterial lineages

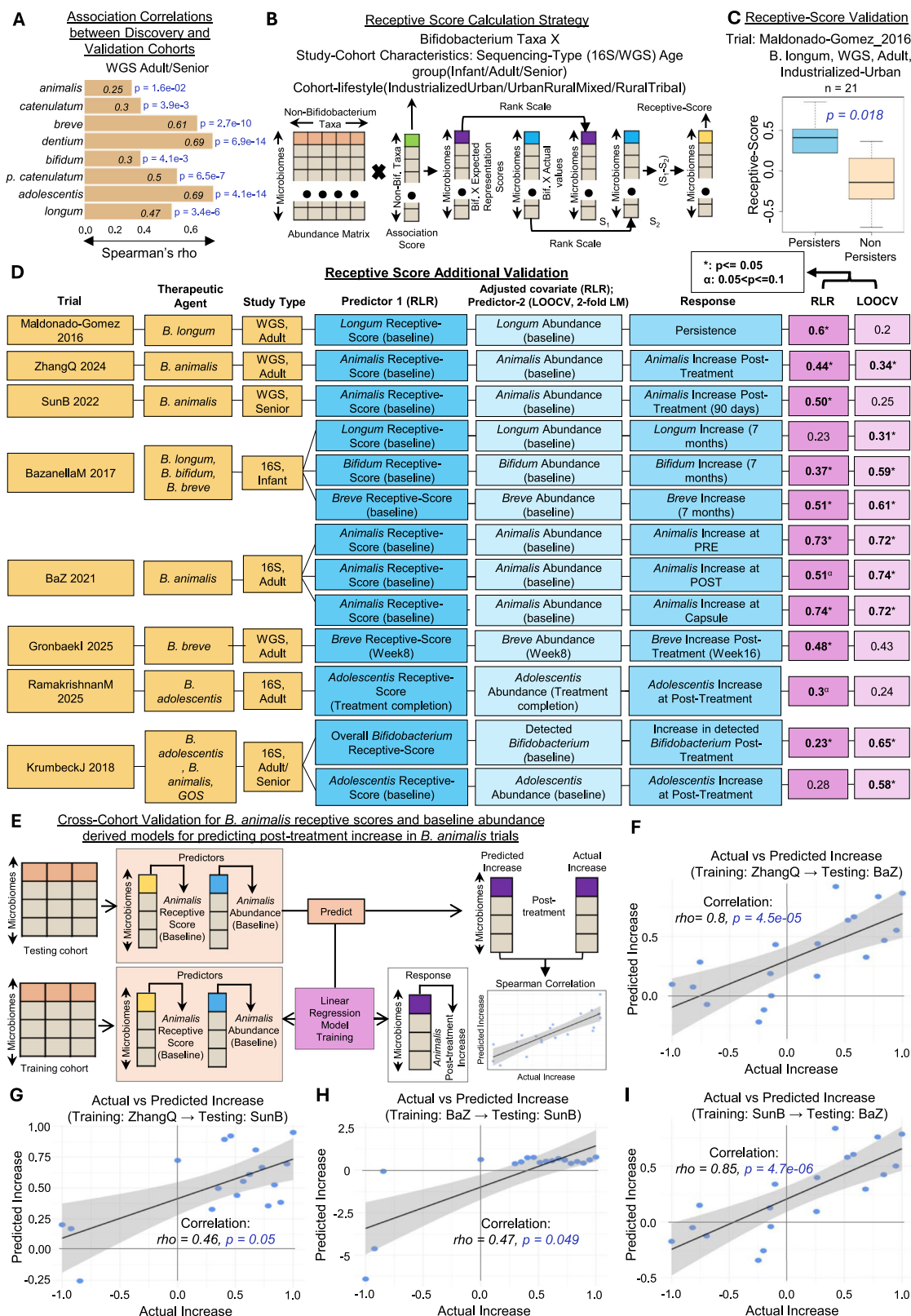
Associations between microbes are reflections of their functional interactions, which may promote better colonization of a *Bifidobacterium* in a community^{32,33}. Thus, next, we attempted to identify the functional features of the non-bifidobacterial taxa members that were associated with the relative abundance of the different bifidobacteria.

Using a systematic approach, we collected a set of 24,980 genomes belonging to 66 consistently occurring non-bifidobacterial gut microbial taxa for which association-scores were obtained against multiple bifidobacterial features and profiled their genome-level functional signatures, annotating them using the eggNOG mapper pipeline and finally obtained a list of 16,077 non-redundant functional-groups (see “Methods”, Supplementary Note 7, Figs. 3B and 7A, and Supplementary Fig. 25).

Subsequently, for each of the nine bifidobacterial features, we devised two-step RF models to predict the corresponding association scores of the non-bifidobacterial taxa based on these functional groups (see Fig. 7A, Supplementary Fig. 25, and “Methods”).

These two-step RF models, for each *Bifidobacterium* feature, resulted in distinct sets of informative functional-groups and demonstrated strong predictive power for the prediction of the corresponding association-scores with significantly positive out-of-box correlations between the actual and predicted association-scores (ranging 0.33–0.70) all 9 features (Fig. 7B). These results indicated that association strengths of the non-bifidobacterial taxa with respect to the different bifidobacterial features are strongly linked to their genome-level functions. Applying further filtration criteria, we finally identified a set of 162 functional groups that showed significant association with multiple *Bifidobacterium* features (and were contributed by multiple non-bifidobacterial taxa) (Supplementary Data 22). The biochemical and mechanistic bases of many identified functional groups across *Bifidobacterium* species were not immediately evident. We focused on a subset of 68 groups showing positive associations with either *B. adolescentis*/*B. longum* or *B. breve*/*B. dentium* showed the strongest positive and negative associations with health-associated non-bifidobacterial gut microbiome, respectively. Deeper investigations identified certain patterns providing into the cross-feeding interactions in the gut community (described in detail in Supplementary Note 8 and Supplementary Fig. 26).

Functional groups showing positive links with *B. adolescentis*/*longum* abundance were predominantly observed in the genomes of the health-promoting members (majority of them butyrate) of the gut microbiome (Supplementary Note 8 and Supplementary Fig. 26). These included functional groups like Shikimate dehydrogenase and pantothenate kinase, which are associated with the production of essential aromatic acids (e.g., tryptophan) and the vitamin pantothenate (and coenzyme-A, a precursor for short-chain fatty acids or SCFAs like butyrate). Notably, we identified many functional-groups which are associated with the production of SCFAs or their precursors. These included proteins like alpha-glucoside sugar transporters (involved in the internalization of simple sugars), GH32 CAZyme (breakdown of fructo-oligosaccharides into fermentable substrates for butyrate synthesis), threonine biosynthesis functional-groups (precursor for SCFA propionate), and aspartate aminotransferase (conversion of



aspartate to butyrate precursors oxaloacetate and glutamate). The identification of multiple functional groups associated with simple sugar import and SCFA production has important implications for understanding the mechanistic basis of these consistent associations between non-bifidobacterial taxa and bifidobacteria (discussed later).

In contrast, non-bifidobacterial functional groups positively associated with *B. breve*/*B. dentium*, were enriched for proteins

involved in the transport of metal ions (iron, zinc, and manganese) (predominantly observed in oral taxa) and glycine betaine, a trimethylamine (TMA) precursor (Supplementary Fig. 26). Metal ion uptake/sequestration are strategies employed by certain microbes, particularly pathogens and pathobionts, to outcompete commensals by limiting micronutrient availability, thereby disrupting gut homeostasis.

Fig. 6 | Reproducibility of association scores and utility of baseline *Bifidobacterium* receptive scores for personalized prediction of future increase/persistence. **A** Bar plots showing the Spearman correlations between the association scores of the different bifidobacterial taxa with eight major bifidobacteria computed for the WGS-profiled adult/senior cohorts with IndustrializedUrban population-type in the discovery dataset with the same association scores recomputed within the cohorts of the same microbiome-type in the validation dataset. Overall, these results indicate a strong and significant (exact *p* values have been displayed above each horizontal bar as well as provided in the corresponding source data) reproducibility of the association patterns of the different non-*Bifidobacterium* taxa with the different *Bifidobacterium* taxa in the discovery and validation cohorts. **B** Strategy for computing the microbiome-level receptive-score for a given *Bifidobacterium* feature, given the microbiome-type of the specific cohort (characterized by the specific age-group, cohort-life-style and microbiome-profiling strategy). **C** Boxplots showing the significantly higher baseline *B. longum* receptive-scores for the persisters compared to the Non-Persisters in the Maldonado-Gomez et al. cohort, along with the *p*-value. For the boxplot, *p* value of pairwise group comparisons using the Mann–Whitney *U* test (two-sided) is also indicated. In the boxplot, the center line is the median; box limits are the upper and lower quartiles; and whiskers are 1.5 times the interquartile range (outlier points are not included in the visualization). **D** Horizontal flow diagrams illustrate significant associations between baseline receptive scores and post-treatment increases in target *Bifidobacterium*-associated features. Post-treatment increase is defined as the rank-scaled difference between immediate post-treatment and the baseline values. Associations were assessed using RLRs (two-sided; “Methods”), with

receptive-score as the predictor, post-treatment increase as the outcome, and baseline feature levels as confounders (specified per model). No multiple testing correction was applied, as analyses were performed independently for each trial-feature combination. Each diagram presents results for seven *Bifidobacterium*-focused trials, covering thirteen trial-*Bifidobacterium* combinations. The RLR and LOOCV column highlights significant positive association (* denotes $p \leq 0.05$ and coefficient > 0 ; α denotes $0.05 < p \leq 0.1$ and coefficients > 0). The exact *p*-values for all comparisons are provided in Supplementary Data 21. **E** Approach adopted for the cross-cohort validation using the linear regression model designed to predict post-treatment increase of *B. animalis* based on receptive scores and baseline abundance of *B. animalis*. **F–I** Results of the cross-cohort validation of these linear regression models with SunB and BaZ as the testing cohorts. For each scatterplot, individual points represent microbiomes, where the number of microbiomes (*n*) are as follows: **F** (*n* = 19), **G** (*n* = 18), **H** (*n* = 18), and **I** (*n* = 19). The best-fit line indicates the trend of association between variables. Shaded regions around the (y v/s x) linear-regression-estimated best-fitted lines represent 95% confidence intervals of the estimated mean value of the variable on the y-axis, calculated as the fitted value $\pm 1.96 \times$ standard error (s.e.) of the linear regression-derived fit based on the variable on the x-axis. Correlation coefficients (Spearman’s rho) and corresponding two-sided *p* values are shown to quantify the strength and significance of the relationship. No multiple testing correction was applied, as each analysis was conducted independently for predefined hypothesis-driven trial-feature combinations, thereby limiting the risk of inflated false discovery rates and rendering additional correction unnecessary.

Despite these specific patterns that explained some of the association patterns, the mechanistic basis of the other functional groups needs experimental investigations. We have provided a list of the non-bifidobacterial protein sequences belonging to these functional groups in the GitHub repository containing the results of this study (https://github.com/Sourav-MiRe/Project_Bifidobacterium; <https://doi.org/10.5281/zenodo.19311870>). This list represents a set of non-bifidobacterial functional genes that can be evaluated further to probe the intestinal ecology of commensal microbes.

Discussion

Although variation of baseline gut microbiome has been implicated as a key factor in determining personalized colonization/persistence patterns of different probiotic species^{11,35}, systematic identification of the determinants (both taxonomic and functional) of colonization/persistence/high-representation patterns of health-associated taxa has been very limited.

Our study is an effort in this direction. By leveraging a large dataset of over 50,000 gut microbiomes from 149 cohorts, we identify how the different non-bifidobacterial microbes link with the abundances of different bifidobacteria and overall detection across individuals of different age-groups, industrialization status and six major diseases (quantified and reported as cohort-stratified association scores). To further measure the effect of various microbiome properties and host-associated factors on these association patterns, we also measure the extent of these associations after adjusting for these factors (for a subset of adult/senior Industrialized cohorts) for which the data was available.

In this sub-group of cohorts (adult/senior, IndustrializedUrban), association-scores are strongly correlated: non-bifidobacterial taxa that interact more strongly with a given *Bifidobacterium* also show stronger associations overall, largely independent of host status, diet, medication use, clinical markers, or microbiome composition. Importantly, many strong non-bifidobacterial-bifidobacterial associations are consistently reproduced across host statuses and dietary/medication contexts, suggesting genuine microbe–microbe interactions with limited host influence.

Some of these associations have mechanistic bases. The positive associations of *B. adolescentis/longum* with butyrate-producing *Firmicutes*, as well as other *Coriobacteriaceae* members like *Collinsella*

(also known producers of SCFAs)^{36,37}, could be due to specific cross-feeding interactions^{14,32}. Bifidobacteria produce acetate and degrade dietary and host-associated glycans into simple sugars (e.g., fucose, galactose), which can then be internalized by these saccharolytic gut fermenters for their growth and as precursors for SCFA production. SCFAs feed the human intestinal epithelial cells and stimulates the production of mucins, acting as sources of glycan fueling Bifidobacteria growth, thus resulting in symbiotic cross-feeding cycle³⁸. Notably, many of these associated functions, e.g., those involved the internalization/transport of sugars and involved in SCFAs and SCFA precursors synthesis, are also identified in our machine-learning-based identification of functional groups associated with increased *B. adolescentis/longum* relative abundances (Supplementary Fig. 26).

Interestingly, *Bacteroidota* taxa, like *B. thetaiotaomicron* and *P. copri*, also have similar metabolic activities in the gut as bifidobacteria (acetate production, dietary/host glycan digestion into simple sugars), and thus may compete with bifidobacteria for similar metabolic roles. This could partly explain the consistent negative associations observed between these taxa and certain bifidobacteria in the adult gut microbiomes (from IndustrializedUrban to RuralTribal cohorts).

Similarly, preclinical studies and human trials have also shown the administration of bifidobacterial probiotics to be associated with decrease in the host levels of TMA and its downstream metabolite trimethylamine-N-oxide (TMAO) and the abundances of TMA-producing pathobiont *Firmicutes*, like *Clostridium asparagiforme/hathewayi/citroniae*^{39,40}. Consistent with the negative associations of these groups with *B. adolescentis/longum*, we also identified glycine-betaine transporters that were negatively associated with *B. adolescentis* abundance. Glycine-betaine is a precursor of TMA/TMAO, metabolites that have been consistently linked to diseases, such as atherosclerosis, neuroinflammation, and colorectal cancer^{23,41,42}. However, the mechanistic bases of such links still need to be investigated.

Association patterns also exhibited pronounced cohort-type-specific trends. Positive associations between *Bifidobacterium* species and butyrate-producing *Firmicutes* were strongest in adult gut microbiomes from IndustrializedUrban cohorts, weakened in senior and UrbanRuralMixed cohorts, and were largely absent in RuralTribal cohorts. Similarly, the negative associations with *Bacteroidetes*

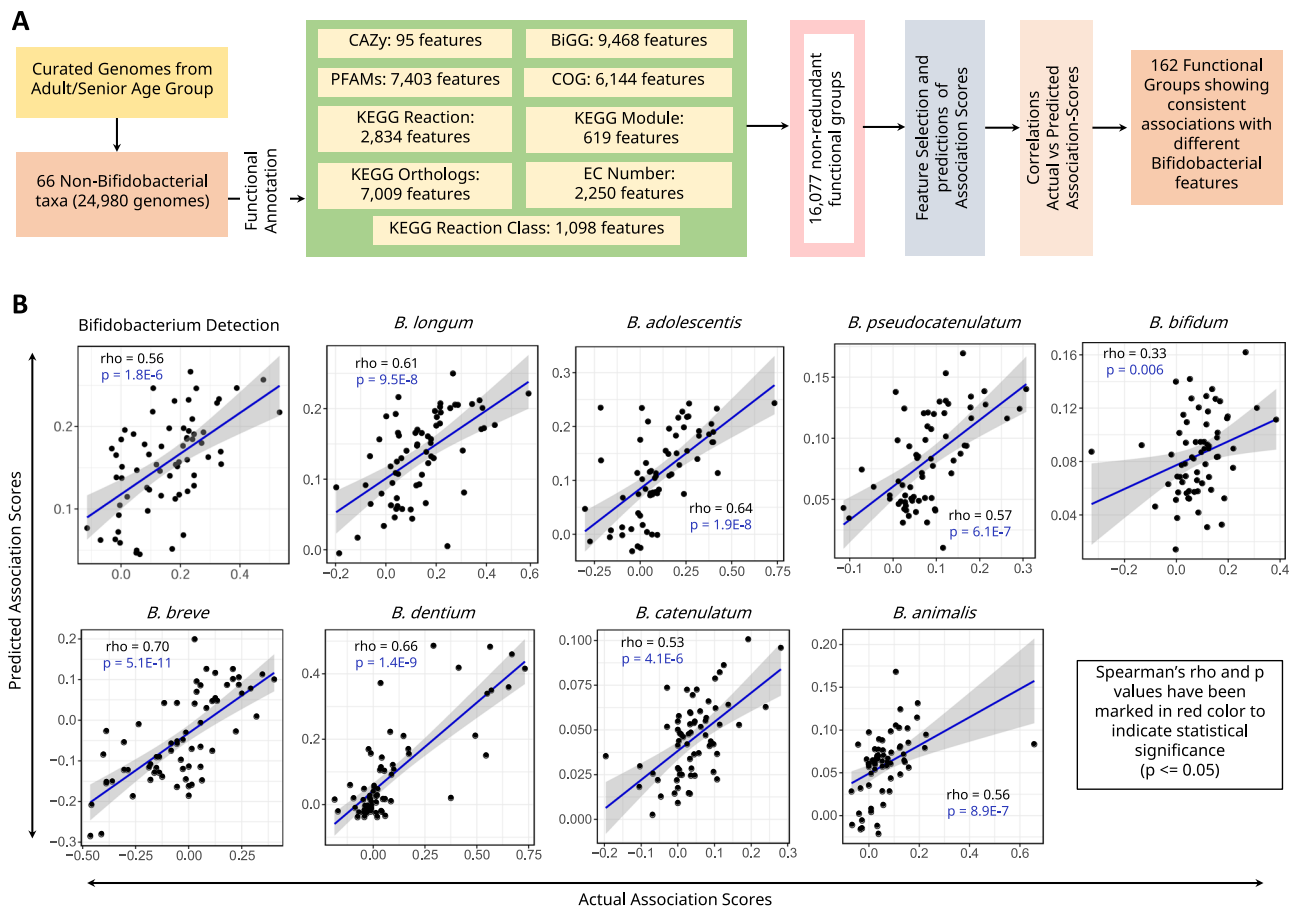


Fig. 7 | Non-bifidobacterial taxa-derived functional groups show predictive ability of different bifidobacterial association scores. **A** Schematic overview of generating non-bifidobacterial taxa derived 162 functional groups showing consistent associations with bifidobacterial taxa, starting from curation of high-quality genomes (including MAGs), followed by functional annotations, random-forest based feature selection, prediction of association-scores, and finally correlating original vs. predicted association-scores to identify the consistent set of functional groups **B** Here in this panel, the scatter plots show the actual vs predicted association-scores ($n = 66$) along with the best fit curve which shows the extent of positive correlation in our case. The Spearman's rho and p value have been shown in red color (where $p \leq 0.05$) on the individual plots, and interestingly, for all nine *Bifidobacterium* features, the actual and predicted association scores show

significant positive correlation. For each scatterplot, individual points represent non-bifidobacterial taxa. The best-fit line indicates the trend of association between variables. Shaded regions around the (y v/s x) linear-regression-estimated best-fitted lines represent 95% confidence intervals of the estimated mean value of the variable on the y-axis, calculated as the fitted value $\pm 1.96 \times$ standard error (s.e.) of the linear regression-derived fit based on the variable on the x-axis. Correlation coefficients (Spearman's rho) and corresponding two-sided p values are shown to quantify the strength and significance of the relationship. No multiple testing correction was applied, as the analyses were performed on a limited number of predefined, hypothesis-driven feature sets and evaluated independently for each bifidobacterial feature, minimizing the likelihood of spurious associations.

lineages observed in adult/senior gut microbiomes were not detected in infants, where the direction of association was reversed, indicating age-dependent restructuring of *Bifidobacterium*–microbiome interactions. By investigating a large collection of diverse gut microbiomes and computing cohort-type-stratified association scores, our analysis thus extends the understanding of baseline-gut-microbiome-to-bifidobacterial association patterns.

The key highlight of this study is the development of microbiome-level receptive scores, that highlight the translational relevance of our findings. These scores provide a quantitative measure of the microbiome's capacity to support colonization and persistence of the administered bifidobacteria. Using multiple complementary frameworks, we show the utility of receptive scores in predicting post-intervention increase/colonization/persistence of the administered bifidobacteria as well as for stratifying individuals in personalized probiotic interventions.

Microbiome-based therapies (including probiotics) have two different levels of responses, first, colonization/persistence/abundance-increase of the administered health-associated microbe(s) and second, the clinical efficacy. These may occur synergistically or independently.

While clinical outcome data were available for only a limited subset of the trials investigated here, validation in these few cohorts with accompanying clinical metadata suggests that receptive scores may also be predictive of clinical efficacy. However, substantially larger numbers of intervention trials with detailed clinical measurements will be required to rigorously evaluate this potential.

As a data-driven analysis of publicly available microbiome datasets, this study has certain limitations. While we report certain metadata-adjusted association scores for a minority of cohorts, broader stratification by these variables was not possible due to their unavailability, particularly in cohorts, especially from infants and non-Industrialized populations. In addition, despite incorporating cohorts spanning a wide lifestyle gradient, there remained a relative paucity of studies from non-Industrialized populations, especially within infant and senior age groups. Lastly, while we have investigated the association patterns at the species-taxonomic-level, multiple *Bifidobacteria* have shown consistent strain-wise variations in their colonization patterns and thus may show variabilities of association patterns with different non-bifidobacterial lineages⁴³. This forms the objective of one of our future investigations.

Despite these limitations, the identification of consistent cohort-type-stratified association patterns between non-bifidobacterial and bifidobacterial lineages, the utilization of these association patterns to build microbiome-level bifidobacterial receptive-scores and the robust validation of these scores to predict personalized patterns of persistence/increase of the administered *Bifidobacterium* taxa in diverse intervention trials, demonstrate the applicability of our global analysis in formulation of approaches to predict personalized responses to bifidobacterial-probiotic interventions. Beyond microbiome receptivity, our findings further provide insights into how the gut microbiome likely operates as a trophic network that is shaped by cross-feeding and metabolic dependencies, wherein cooperativity and competitive interactions shape structure community assembly and influence probiotic colonization outcomes. Our data identify taxa whose interactions can be further explored at functional resolution level in future studies that integrate multi-omics measurements, including metagenomics, metabolomics, flux-based modeling, and artificial in vitro consortia. Controlled perturbation and longitudinal intervention studies can also help establish causal links between ecological network structure and probiotic persistence dynamics.

Methods

Collation of discovery study cohorts

The collation of the datasets was performed integrating two sources (listed in Supplementary Data 1). First, was the dataset collation of 124 study cohorts from a recently published study by our group¹⁶. This group of studies majorly contained gut microbiomes from adult and senior populations. This was supplemented by 33 additional study-cohorts from infants/adult/senior available either as part of the curatedMetagenomicData repository (<https://github.com/waldronlab/curatedMetagenomicData>), or downloaded as part of extensive literature searches and processed for inclusion in the analysis^{44–49}.

The individual microbiomes were categorized into three age-based groups according to the age of the corresponding hosts: infants/newborns (≤ 2 years), adults (18–60 years), and seniors (> 60 years). In certain studies, while individual-level age data were unavailable, the study-cohorts were reported to fall within specific age ranges, as indicated in the respective publications^{50–56}. In such cases, all microbiomes from those cohorts were assigned to the appropriate age group (infant, adult, or senior) based on the available cohort-level information. Additionally, each microbiome was assigned to one of three lifestyle groups, namely IndustrializedUrban, UrbanRuralMixed, or RuralTribal, that are reflective of the general lifestyle of the study cohort. This cohort-life-style classification was performed using the nationality and geographic location of the participants using methodologies described in previous studies from our group^{16,17}. For each study-cohort, we also recorded the profiling strategy employed, namely 16S rRNA gene sequencing (16S) or whole genome shotgun (WGS) sequencing. Additional metadata, such as gender, disease or study condition, birth mode, and others, were available for some cohorts. However, these variables were not included in the current analysis due to their limited availability across cohorts or because they have already been investigated in previous studies^{12–14}.

Taxonomic profiling

For datasets obtained using the curatedMetagenomicData repository, the taxonomic profiles of all shotgun datasets (or cohorts) generated using metaphlan3 were already available⁵⁷. For this reason, to maintain uniformity of profiling, we profiled the taxonomic composition of all WGS microbiome datasets of cohorts of not obtained from curatedMetagenomicData using metaphlan3. Microbiome datasets using 16S were all profiled using SPINGO⁵⁸. The taxonomic profiled obtained for each study-cohort were subdivided into sub-parts, one containing the abundances of the non-bifidobacterial taxa (which was

renormalized after removing the abundances of the different bifidobacterial species), and the second containing the abundances of only the bifidobacterial species.

Filtering out sparsely detected species

We divided the study cohorts into two bins based on their sequencing type. For each bin, we selected a subset of taxa detected consistently across multiple cohorts using the strategy summarized below (and described in detail in Supplementary Note 1).

Profiling of different bifidobacterial across study cohorts

To first profile the prevalence pattern, the abundance matrix corresponding to the different bifidobacterial species were first converted to a detection matrix to investigate cohort-specific prevalence patterns for the different bifidobacterial species. In the detection matrix, a given *Bifidobacterium* species with a relative abundance ≥ 0.0001 was marked as 1 (or detected) and 0 (not-detected) otherwise. The threshold of 0.0001 was determined based on previous studies using simulated metagenomes, especially using the Metaphlan approach, which have observed that, for abundance values ≥ 0.0001 , there was a linear relationship between the actual abundance values, indicating reliable taxonomic assignments. While the detection profiles were utilized for investigating the association patterns of different *Bifidobacterium* species with different cohort-specific characteristics like sequencing-type, cohort lifestyle, and host age-group, the abundance matrix was utilized for investigating the association of the non-bifidobacterial community members with the overall *Bifidobacterium* detection as well as with the relative abundances of the eight major *Bifidobacterium* taxa.

Association of the detection pattern of different bifidobacteria with the cohort sequencing-type, cohort-life-style, and host age-group categories

To investigate the association patterns of the eight bifidobacterial taxa with different study-cohort-specific characteristics, we utilized linear regression models. We utilized three different regression models for this purpose, each with a specific objective and investigating the effect of a specific cohort characteristic after adjusting for the effects of all others. This is described in Supplementary Note 9.

Association of the overall non-bifidobacterial beta-diversity with *Bifidobacterium* relative abundances across study-cohorts

We utilized the PERMANOVA approach for this purpose. For a given *Bifidobacterium* taxon i in a study-cohort x , we first computed the non-bifidobacterial beta-diversity by computing the all-v/s-all sample distances using the renormalized relative abundance matrices of the non-bifidobacterial taxa across the samples. This was performed using the Bray–Curtis and Kendall distance matrices and provided a quantitative measure of how different the sample was respect to all the other samples in the cohort x when considering only the non-bifidobacterial taxa. Subsequently these variations were correlated with the abundance differences of bifidobacterial species i using PERMANOVA. This was performed using the `adonis2` function of the `vegan` R package (v2.6-4). Significant association obtained using PERMANOVA indicates that variation of the species i across the different samples is strongly correlated with community composition of the non-bifidobacterial taxa, indicating a relationship between the two.

Computation of three microbiome properties–Shannon index, Pielou's evenness index and estimated microbial load

Microbiome diversity and structural properties were quantified using three complementary measures: Shannon index, Pielou's evenness index, and estimated microbial load. Detail for the computation of these properties has been described in Supplementary Note 3.

Study-specific random forest models to identify the top predictive non-bifidobacterial taxa for the different bifidobacterial features across the cohorts

For each study-cohort, we built RF-based machine-learning models for the prediction of each of the bifidobacterial features (overall *Bifidobacterium* detection and the abundances of the eight bifidobacterial taxa), using the abundance profiles of the cohort-type-specific lists of non-Bifidobacterial taxa identified in Supplementary Note 1 as well as using three microbiome properties viz. Shannon Index, Pielou's evenness index and estimated microbial load (computed as described in Supplementary Note 3). For each model (formulated for a given bifidobacterial feature in a given cohort), we identified the top predictor non-bifidobacterial taxa in the following manner. Taxa were sorted in decreasing based on their feature importance scores. Based on this ordering, for any set of top I taxa, we computed the fractional cumulative importance (FCI) value as:

$$FCI_I = \frac{\text{Sum of feature importance scores of all predictors from rank 1 to } I}{\text{Sum of feature importance scores of all predictors}} \quad (1)$$

For each model, the optimal I was identified the minimum value for which FCI _{I} exceeded 0.90 (accounted for 90% of the model performance). The top predictor taxa were identified as the top I (I corresponding to this optimal value) taxa based on their feature importance scores.

Finally, for a given bifidobacterial feature and cohort-type (16S/WGS, infant/adult/senior, IndustrializedUrban/UrbanRuralMixed/RuralTribal), the consensus top informative taxa were identified as those which were identified as top predictors in at least 20% of the cohorts of that type.

Computation of association scores for the different bifidobacterial features

The consensus top predictors identified above, along with the three microbiome properties (Shannon index, Pielou's evenness index and estimated microbial load), were then tested for their association with the corresponding *Bifidobacterium* features. For a given cohort-type (categorized based on profiling strategy: 16S or WGS; age-group: infant or adult or senior and; lifestyle: IndustrializedUrban, or UrbanRuralMixed or RuralTribal) and *Bifidobacterium* feature, for each taxon in the list of the corresponding top predictors, we computed the Spearman correlation of the taxon with the *Bifidobacterium* features (either the total number of *Bifidobacterium* or the abundance of any of the eight *Bifidobacterium* species) in each of the study-cohorts belonging to the specific cohort-type. We counted the total number of study cohorts where the given taxa had a positive correlation and a negative correlation as SP and SN, respectively. Given the total number of study-cohorts for that specific cohort-type as T , the association-score for the given feature with the specific *Bifidobacterium* feature in a given cohort-type (categorized above) as:

$$\text{Association score} = \frac{(SP - SN)}{T} \times \left(1 - \frac{\text{MIN}(SP, SN) + 0.0001}{\text{MAX}(SP, SN) + 0.0001} \right) \quad (2)$$

where, MAX(SP, SN) refers to the maximum of SP and SN and MIN(SP, SN) refers to the minimum of the two values.

Computation of association scores for different disease groups

The detailed description regarding the computation of association-scores of the different non-bifidobacterial taxa with different bifidobacterial features, including the collation of additional study-cohorts with matched diseased-control samples, selection of specific diseases for this investigation and final computation of disease-specific association-scores is described in Supplementary Note 2.

Computation of quantitative microbiome profiles (QMP) using experimental microbial load data provided in the MetaCardis study-cohort

Microbiomes from the MetaCardis cohort (WGS) were selected by matching study identifiers between species abundance profiles, experimental microbial load measurements, and associated metadata, and the common microbiomes (1245 adult microbiomes and 503 senior microbiomes) were retained. Species-level relative abundance profiles were converted to quantitative microbiome profile by multiplying relative abundances with sample-specific microbial load measurements²⁵. Microbiomes were stratified into adult and senior groups based on metadata annotations, and analyses were restricted to those non-bifidobacterial taxa for which the association score was pre-computed to ensure comparability.

For adult and senior microbiomes separately, pairwise Spearman rank correlations were computed between quantitative metagenomic profiles of microbial species and *Bifidobacterium* features, and these association profiles were subsequently compared with previously derived species-specific association scores using Spearman correlations between QMP-derived association coefficients and the previously computed relative abundance-based corresponding association scores for each *Bifidobacterium* taxon. Correlation coefficients and exact p -values were obtained for adult and senior groups independently. All data processing and statistical analyses were performed in R.

Computation of partial correlation-based association scores after adjusting for different microbiome properties, as well as diet, medications, and clinical characteristics of subjects

Details about the computation of partial correlation after adjusting different covariates has been described in Supplementary Note 3.

Description of the eight *Bifidobacterium*-associated intervention trials cohorts for external validation and validation of receptive scores

Here, we collected data from eight *Bifidobacterium*-associated intervention trials (listed in Supplementary Data 18 and see Supplementary Note 4). While one trial Bazanella et al., was on infants/new-borns¹⁵, the rest seven were on adult/senior populations. Four of these cohorts were 16S sequenced^{15,29,31,34}, and the four were profiled using WGS^{11,26,28,30}. For each of the trials, besides investigating the reproducibility of the association patterns (identified in our analyses of the discovery cohorts described above), we also identified specific *Bifidobacterium*-associated property whose response was to be investigated in our validation of the concept of receptive scores (described below and in the subsequent sections).

Six of these trials used single *Bifidobacterium* probiotics: three using *B. animalis*^{26,28,29}; one using *B. breve*³⁰; one using *B. longum*¹¹, and one using *B. adolescentis*³¹. For each of these trials, the target *Bifidobacterium* feature (described later) to be investigated were the abundances of the administered *Bifidobacterium* taxa. One of the trials used a combination of three bifidobacteria (*B. longum*, *B. breve*, and *B. bifidum*)¹⁵. Thus, here, there were three targeted *Bifidobacterium* features, corresponding to the three administered bifidobacterial taxa. One of the trials utilized *B. adolescentis*, *B. animalis* and galactooligosaccharide (GOS) that is expected to enrich the already resident gut *Bifidobacterium*³⁴. Thus, the target feature to be investigated here was the overall *Bifidobacterium* detection and *B. adolescentis* in a gut microbiome.

Investigating the reproducibility of the association scores

The detailed methodology for investigating the reproducibility is described in Supplementary Note 5.

Formulation of receptive scores

For a given *Bifidobacterium* feature (the feature could be the relative abundances of any of the eight bifidobacteria or the overall detection of *Bifidobacterium* taxa) and taxonomic composition profile of a microbiome of a given category (based on age-group, cohort-life-style, and microbiome-profiling-strategy), the receptive-score was computed as follows. We first considered the relative abundances of all non-*Bifidobacterium* taxa whose association-scores for the specific *Bifidobacterium* feature were available (i.e., computed from our investigation of the discovery cohorts). Using these values, we first computed an expected-representation-score for that feature in the given microbiome as the weighted cumulated abundance of these non-bifidobacterial taxa, where the abundances of each non-bifidobacterial taxa was multiplied by a weight, which was simply the association-score obtained for those taxa with the specific *Bifidobacterium* feature in microbiomes of the same category (as described above). For a matrix with multiple microbiome profiles from a specific cohort, this was performed using an inner multiplication of this relative abundance profile matrix with the corresponding association-score vector (containing the association-scores for the various non-bifidobacterial taxa arranged in the same order as the columns of the relative abundance profile matrix).

This expected-representation-score for each microbiome was then rank-scaled to values between 0 and 1 by considering all the baseline microbiomes of a cohort as below. For a given array of values $\mathbf{x}$ of length k , denoted by $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$, the values are first ranked. Next, the corresponding rank-scaled values, denoted by $x_R = \{x_{1R}, x_{2R}, \dots, x_{kR}\}$, of this set were obtained as:

$$x_{iR} = \frac{\text{Rank}(x_i) - \min(\text{Rank}(\mathbf{x}))}{\max(\text{Rank}(\mathbf{x})) - \min(\text{Rank}(\mathbf{x}))} \quad (3)$$

The transformation was essentially an approach of ranking the values and then converting the ranks to values between 0 and 1. We also applied the same rank-scale functions on the baseline values of the target *Bifidobacterium* feature (i.e., actual-ranked-baseline-score). The receptive-score for the *Bifidobacterium* feature was then computed as:

$$\text{Expected Representation Score} - \text{Actual Ranked Baseline Score} \quad (4)$$

Correlating baseline receptive-score, ranked baseline abundance with post-treatment increase

For each of the eight trials, we identified the specific *Bifidobacterium*-associated feature to be investigated. For the three trials involving *B. animalis*, the target feature was the abundance of *B. animalis*. For the *B. longum* supplementation Maldonado-Gomez et al. trial and the *B. breve* supplementation trial of Grønbæk et al., the target features were *B. longum* and *B. breve* relative abundance, respectively. For the Krumbeck et al. trial, that involved the administration of *B. adolescentis*, *B. animalis* and GOS that promotes the growth of all *Bifidobacterium*, we observed no significant change in *B. animalis* abundances across time-points. Consequently, the target feature chosen here were the overall *Bifidobacterium* detection and one with *B. adolescentis*. For the Bazanella et al trial that involved supplementation of a consortia of three *Bifidobacterium* taxa (*longum/breve/bifidum*), there were three target features, namely the relative abundances of *B. longum*, *B. breve*, and *B. bifidum*. Another trial involving supplementation of *B. adolescentis* by Ramakrishnan et al was considered, where the target feature was the abundance of *B. adolescentis*.

The objective here was to investigate whether the receptive score was significantly associated with post-treatment increase of the corresponding target features. Since both the receptive score as well as the ranked post-treatment increase were calculated with respect to ranked baseline values of the target *Bifidobacterium* feature, it was important to

adjust for this confounding effect, as it may lead to spurious correlations. We utilized two approaches. First, we did that using RLR models as:

$$\text{rlm}(\text{ranked post treatment increase} \sim \text{baseline Receptive Score} + \text{Actual Ranked Baseline Score}) \quad (5)$$

Here, the predictors were the baseline-receptive-score and the actual-ranked-baseline-score. The ranked-post-treatment-increase was simply the differences between the rank-scaled relative abundances at the immediate post-treatment time-point and the baseline (computed using the rank-scale function described above). The RLR models were built using the `rlm` function of the MASS package (v 7.3.60) of R.

Second approach utilized the linear regression-based validation: LOOCV

In the LOOCV approach, each subject within a given trial was iteratively held out as a test observation. For a given iteration, a linear regression model was trained on all remaining subjects using the predictors described above. The trained model was then used to generate a prediction for the ranked post-treatment increase of the held-out subject. This procedure was repeated until each subject had served exactly once as the test observation, yielding a complete set of out-of-sample predictions for all subjects within the trial. The predicted ranked post-treatment increases obtained from LOOCV were then compared to the corresponding observed values across subjects. We also created similar variants of LOOCV and RLR variants, where the Association-Scores utilized to obtain the microbiome-level receptive-scores also included those pertaining to Shannon index, Pielou's evenness index and estimated microbial load (for WGS datasets).

Cross-cohort validation in *B. animalis*-associated trials

In the cross-cohort validation, for each of the *B. animalis*-associated trials, we trained linear regression models to predict ranked-post-treatment-increase using the baseline-receptive-score and the actual-ranked-baseline-score (i.e., the baseline *B. animalis* relative abundance rank-scaled) on that cohort and test the same on the remaining two cohorts.

Cross-cohort validation in *B. adolescentis*-associated trials

In the cross-cohort validation, for each of the *B. adolescentis*-associated trials, we trained linear regression models to predict ranked-post-treatment-increase using the baseline-receptive-score and the actual-ranked-baseline-score (i.e., the baseline *B. adolescentis* relative abundance rank-scaled) on that cohort and test on the other cohort.

Assessment of clinical effects in intervention trials

The detailed methodology for investigating the clinical effect is described in Supplementary Note 6.

Curation of reference genomes and metagenome-assembled genomes accompanied by a detailed methodological framework for functional investigation

We have curated 28,991 high-quality (completeness > 90% and contamination < 5%) genomes along with metagenome-assembled genomes encompassing 66 non-bifidobacterial taxa for which association scores were obtained against multiple bifidobacterial features (Fig. 3B), and the functional profiles for these 66 consistent taxa were used for the functional investigation. The detail for generating mean functional profiles and identification of taxa-specific functional groups has been described in Supplementary Note 7.

Statistics and reproducibility

For each statistical test in this manuscript, the type of test, the size (n) of the test, and statistical summaries or measures of dispersion are clearly defined in the figure legends or in the accompanying text in the "Methods" section or wherever it is required.

Reporting summary

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

Data availability

This paper analyzes existing, publicly available data, accessible at the European Nucleotide Archives (ENA) and curated Metagenomic Data (see “Methods”). Study details are listed in Supplementary Data 1. However, the metadata for two cohorts used in this study (Arivale⁵⁹ and LifeLinesDeep_2016⁶⁰) are subject to access restrictions and are not publicly available; these datasets can be accessed through the respective original publications (<https://doi.org/10.1038/s41598-019-43058-0> and <https://doi.org/10.1136/bmjopen-2014-006772>). All the data generated in the form of results in our study have been deposited at https://github.com/Sourav-MiRe/Project_Bifidobacterium; Code Ocean repository (<https://codeocean.com/capsule/7280337/tree>) and Zenodo <https://doi.org/10.5281/zenodo.19311870>, and it is publicly available⁶¹. The Source data corresponding to the figures has been provided. Source data are provided with this paper.

Code availability

All original code has been deposited at https://github.com/Sourav-MiRe/Project_Bifidobacterium; Code Ocean repository (<https://codeocean.com/capsule/7280337/tree>) and Zenodo <https://doi.org/10.5281/zenodo.19311870>, and it is publicly available⁶¹.

References

1. Stuivenberg, G. A., Burton, J. P., Bron, P. A. & Reid, G. Why are bifidobacteria important for infants? *Microorganisms* **10**, 278 (2022).
2. Zhao, S., Lau, R., Zhong, Y. & Chen, M. H. Lactate cross-feeding between *Bifidobacterium* species and *Megasphaera indica* contributes to butyrate formation in the human colonic environment. *Appl Environ. Microbiol.* **90**, e01019-23 (2023).
3. Chen, J., Chen, X. & Ho, C. L. Recent development of probiotic bifidobacteria for treating human diseases. *Front. Bioeng. Biotechnol.* **9**, <https://doi.org/10.3389/fbioe.2021.770248> (2021).
4. Hidalgo-Cantabrana, C. et al. Bifidobacteria and their health-promoting effects. *Microbiol. Spectr.* **5**, <https://doi.org/10.1128/microbiolspec.bad-0010-2016> (2017).
5. Poch, M. & Bezkorovainy, A. Growth-enhancing supplements for various species of the genus *Bifidobacterium*. *J. Dairy Sci.* **71**, 3214–3221 (1988).
6. Hill, C. et al. The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 506–514 (2014).
7. Sharma, M., Wasan, A. & Sharma, R. K. Recent developments in probiotics: an emphasis on *Bifidobacterium*. *Food Biosci.* **41**, 100993 (2021).
8. Quigley, E. M. M. & Shanahan, F. Probiotics in health care: a critical appraisal. *Annu. Rev. Med.* **76**, 129–141 (2025).
9. Sarkar, A. & Mandal, S. Bifidobacteria—insight into clinical outcomes and mechanisms of its probiotic action. *Microbiol. Res.* **192**, 159–171 (2016).
10. Ghosh, T. S. & Valdes, A. M. Evidence for clinical interventions targeting the gut microbiome in cardiometabolic disease. *BMJ* e075180, <https://doi.org/10.1136/bmj-2023-075180> (2023).
11. Maldonado-Gómez, M. X. et al. Stable engraftment of *Bifidobacterium longum* AH1206 in the human gut depends on individualized features of the resident microbiome. *Cell Host Microbe* **20**, 515–526 (2016).
12. Pasolli, E. et al. Bifidobacteriaceae diversity in the human microbiome from a large-scale genome-wide analysis. *Cell Rep.* **43**, 115027 (2024).
13. Derrien, M., Turroni, F., Ventura, M. & Van Sinderen, D. Insights into endogenous bifidobacterium species in the human gut microbiota during adulthood. *Trends Microbiol.* **30**, 940–947 (2022).
14. Ladeira, R., Tap, J. & Derrien, M. Exploring *Bifidobacterium* species community and functional variations with human gut microbiome structure and health beyond infancy. *Microbiome Res. Rep.* **2**:N/A-N/A, <https://doi.org/10.20517/mrr.2023.01> (2023).
15. Bazanella, M. et al. Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome. *Am. J. Clin. Nutr.* **106**, 1274–1286 (2017).
16. Goel, A. et al. Toward a health-associated core keystone index for the human gut microbiome. *Cell Rep.* **44**, <https://doi.org/10.1016/j.celrep.2025.115378> (2025).
17. Shete, O. & Ghosh, T. S. Normal gut microbiomes in diverse populations: clinical implications. *Annu. Rev. Med.* **76**, 95–114 (2025).
18. Shanahan, F., Ghosh, T. S. & O’Toole, P. W. The healthy microbiome—what is the definition of a healthy gut microbiome? *Gastroenterology* **160**, 483–494 (2021).
19. Yatsunenkov, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
20. Matsuo, Y. et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* **21**, 35 (2021).
21. Riaz Rajoka, M. S. et al. Identification, characterization, and probiotic potential of *Lactobacillus rhamnosus* isolated from human milk. *LWT* **84**, 271–280 (2017).
22. Ghosh, T. S., Shanahan, F. & O’Toole, P. W. Toward an improved definition of a healthy microbiome for healthy aging. *Nat. Aging* **2**, 1054–1069 (2022).
23. Ghosh, T. S., Shanahan, F. & O’Toole, P. W. The gut microbiome as a modulator of healthy ageing. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 565–584 (2022).
24. Manghi, P. et al. Meta-analysis of 22,710 human microbiome metagenomes defines an oral-to-gut microbial enrichment score and associations with host health and disease. *Nat. Commun.* **17**, 196 (2025).
25. Nishijima, S. et al. Fecal microbial load is a major determinant of gut microbiome variation and a confounder for disease associations. *Cell* **188**, 222–236 (2025).
26. Zhang, Q. et al. Efficacy of *Bifidobacterium animalis* subsp. *lactis* BL-99 in the treatment of functional dyspepsia: a randomized placebo-controlled clinical trial. *Nat. Commun.* **15**, 227 (2024).
27. Looijesteijn, E. et al. A double-blind intervention trial in healthy women demonstrates the beneficial impact on *Bifidobacterium* with low dosages of prebiotic galacto-oligosaccharides. *Front Nutr.* **11**, <https://doi.org/10.3389/fnut.2024.1440319> (2024).
28. Sun, B. et al. *Bifidobacterium lactis* probio-M8 adjuvant treatment confers added benefits to patients with coronary artery disease via target modulation of the gut-heart/-brain axes. *mSystems* **7**, e00100-22 (2022).
29. Ba, Z. et al. Matrix effects on the delivery efficacy of *Bifidobacterium animalis* subsp. *lactis* BB-12 on fecal microbiota, gut transit time, and short-chain fatty acids in healthy young adults. *mSphere* **6**, <https://doi.org/10.1128/msphere.00084-21> (2021).
30. Grønbaek, I. M. B. et al. The effects of probiotic treatment with *Bifidobacterium breve*, Bif195 for small intestinal Crohn’s disease and the gut microbiome: results from a randomised, double-blind, placebo-controlled trial. *Gut Pathog.* **17**, 19 (2025).
31. Ramakrishnan, M. et al. Two-week supplementation of *Bifidobacterium adolescentis* iVS-1 reduces symptoms associated with lactose intolerance in lactose maldigesters. *Gut Microbes Rep.* **2**, 2508199 (2025).

32. Turrone, F. et al. Glycan utilization and cross-feeding activities by bifidobacteria. *Trends Microbiol.* **26**, 339–350 (2018).
33. Xiao, M. et al. Cross-feeding of bifidobacteria promotes intestinal homeostasis: a lifelong perspective on the host health. *NPJ Biofilms Microbiomes* **10**, 47 (2024).
34. Krumbeck, J. A. et al. Probiotic *Bifidobacterium* strains and galactooligosaccharides improve intestinal barrier function in obese adults but show no synergism when used together as synbiotics. *Microbiome* **6**, 121 (2018).
35. Zmora, N. et al. Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell* **174**, 1388–1405.e21 (2018).
36. Qin, P. et al. Characterization a novel butyric acid-producing bacterium *Collinsella aerofaciens* subsp. *Shenzhenensis* subsp. Nov. *Microorganisms* **7**, 78 (2019).
37. Bag, S., Ghosh, T. S. & Das, B. Complete genome sequence of *Collinsella aerofaciens* isolated from the gut of a healthy indian subject. *Genome Announc.* **5**, <https://doi.org/10.1128/genomea.01361-17> (2017).
38. Willemsen, L. E. M., Koetsier, M. A., Deventer, S. J. H. & van Tol, E. A. F. Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E1 and E2 production by intestinal myofibroblasts, <https://doi.org/10.1136/gut.52.10.1442> (2003).
39. Zhou, Z. et al. Probiotic *Bifidobacterium* reduces serum TMAO in unstable angina patients via the gut to liver to heart axis. *Liver Res.* **9**, 57–65 (2025).
40. Matsumoto, M., Kitada, Y., Shimomura, Y. & Naito, Y. *Bifidobacterium animalis* subsp. *lactis* LKM512 reduces levels of intestinal trimethylamine produced by intestinal microbiota in healthy volunteers: a double-blind, placebo-controlled study. *J. Funct. Foods* **36**, 94–101 (2017).
41. Boysen, A. K. et al. Glycine betaine uptake and metabolism in marine microbial communities. *Environ. Microbiol.* **24**, 2380–2403 (2022).
42. Piccinno, G. et al. Pooled analysis of 3,741 stool metagenomes from 18 cohorts for cross-stage and strain-level reproducible microbial biomarkers of colorectal cancer. *Nat. Med.* **31**, 2416–2429 (2025).
43. Ma, B. et al. Strain-specific alterations in gut microbiome and host immune responses elicited by tolerogenic *Bifidobacterium pseudolongum*. *Sci. Rep.* **13**, 1023 (2023).
44. Iszatt, N. et al. Environmental toxicants in breast milk of Norwegian mothers and gut bacteria composition and metabolites in their infants at 1 month. *Microbiome* **7**, 34 (2019).
45. Hill, C. J. et al. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET cohort. *Microbiome* **5**, 4 (2017).
46. Levin, A. M. et al. Joint effects of pregnancy, sociocultural, and environmental factors on early life gut microbiome structure and diversity. *Sci. Rep.* **6**, 31775 (2016).
47. Laursen, M. F. et al. Infant gut microbiota development is driven by transition to family foods independent of maternal obesity. *mSphere* **1**, <https://doi.org/10.1128/msphere.00069-15> (2016).
48. Russell, J. T. et al. Genetic risk for autoimmunity is associated with distinct changes in the human gut microbiome. *Nat. Commun.* **10**, 3621 (2019).
49. Stokholm, J. et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun.* **9**, 141 (2018).
50. Song, J. S., Kim, J. O. R., Yoon, S. M., Kwon, M. J. & Ki, C. S. The association between gut microbiome and hypertension varies according to enterotypes: a Korean study. *Front. Microbiomes* **2**, <https://doi.org/10.3389/frmbi.2023.1072059> (2023).
51. Hill-Burns, E. M. et al. Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov. Disord.* **32**, 739–749 (2017).
52. Baxter, N. T., Ruffin, M. T., Rogers, M. A. M. & Schloss, P. D. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med.* **8**, 37 (2016).
53. Tee, M. Z. et al. Gut microbiome of helminth-infected indigenous Malaysians is context dependent. *Microbiome* **10**, 214 (2022).
54. Kang, J. T. L. et al. Long-term ecological and evolutionary dynamics in the gut microbiomes of carbapenemase-producing Enterobacteriaceae colonized subjects. *Nat. Microbiol.* **7**, 1516–1524 (2022).
55. Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
56. Zackular, J. P., Rogers, M. A. M., Ruffin, M. T. IV & Schloss, P. D. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev. Res.* **7**, 1112–1121 (2014).
57. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. (eds Turnbaugh, P., Franco, E., & Brown, C. T.). *eLife*, <https://doi.org/10.7554/eLife.65088> (2021).
58. Allard, G., Ryan, F. J., Jeffery, I. B. & Claesson, M. J. SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinform.* **16**, 324 (2015).
59. Zubair, N. et al. Genetic predisposition impacts clinical changes in a lifestyle coaching program. *Sci. Rep.* **9**, 6805 (2019).
60. Tigchelaar, E. F. et al. Cohort profile: lifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics, <https://doi.org/10.1136/bmjopen-2014-006772> (2015).
61. Goswami, S. & Ansari, A. Sourav-MiRe/Project *Bifidobacterium*: initial release. Published online April 1, 2026. <https://doi.org/10.5281/zenodo.19364458>

Acknowledgements

T.S.G. acknowledges the Department of Biotechnology, Ministry of Science and Technology, Government of India for the Ramalingaswami Re-entry Fellowship (BT/HRD/35/02/2006). T.S.G. acknowledges IIIT-Delhi for the Discovery Track Investigator grant. S.G. acknowledges IIIT-Delhi for the Institute-Fellowship. A.A. acknowledges the Department of Science and Technology (DST-INSPIRE) for the INSPIRE fellowship.

Author contributions

Conceptualization, T.S.G., methodology, formal analysis, validation, T.S.G., S.G., A.A., C.S., investigation, T.S.G., S.G., A.A., P.W.O.T., writing-original draft, T.S.G., S.G., A.A., writing-review and editing, T.S.G., P.W.O.T., F.S., V.A., supervision, T.S.G., funding acquisition, T.S.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains Supplementary material available at <https://doi.org/10.1038/s41467-026-72289-9>.

Correspondence and requests for materials should be addressed to Tarini Shankar Ghosh.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026