

Supplementary Information

Supplementary Notes 1-9, Supplementary Figures 1-26, and Supplementary References

(PDF 21.5 MB)

Supplementary Note 1

Determining subset of consistent gut-associated taxa for cohorts of different Sequencing-Types and Age-Groups

A key challenge of gut microbiomes is the presence of a large number of microbes that are atypical and allochthonous. These microbes are mostly derived from the diet and environment and not regular residents of the gut microbiome. This results in large number of sparsely taxa that tends to inflate the microbial diversity in the gut microbiome. Thus, for identifying consistent and reproducible microbial markers across study-cohorts, it is important to remove this noise. We previously implemented specific protocols to eliminate sparsely detected microbial noise and identify a minimal set of consistently detected microbes that maximally represent the gut microbial community across all investigated microbiomes¹. Here also, we adopted a similar strategy as described below to investigate the cohorts included in this investigation. The objective of this investigation was to select the smallest possible list of microbes that could cumulatively account for the maximum fraction of the studied gut microbiomes across all study-cohorts.

To identify these taxa subsets, we divided our study-cohorts into four groups, based on different combinations of age-groups (infant and adult/senior) and Sequencing-Types (16S and WGS). For each group, we adopted the following approach.

To get the list of taxa for a specific study-group (16S-Infant or WGS-Infant or 16S-Adult/Senior or WGS-Adult/Senior), we defined two terms or thresholds for cohorts of that group:

1. Threshold X: Percentage of microbiomes in a given study-cohort where a given taxa is detected
2. Threshold Y: Percentage of study-cohort where a given tax needs to be detected in at least X% of the microbiomes

Now the next goal was to find a list of N taxa that is in at least X% of the microbiomes in at least Y% of the study-cohorts, such that:

- a. N is reasonably small
- b. N taxa cumulatively represent as much of the gut microbial community composition as possible across all investigated cohorts

Subsequently, we used a grid search approach where the values of each of X and Y were varied from 0.05 to 0.95 at a difference of 0.05 (i.e. 0.05, 0.1, 0.15, etc till 0.95). Each combination of X and Y provided different sets of microbes and the sizes of these microbial sets are noted in **Supplementary Figure 2**, where the sets were determined separately for cohorts of different categories. These categories were Infant-16S, Infant-WGS, Adult/Senior-16S and Adult/Senior-WGS)

The next task was to investigate how good was to determine how maximally each microbial set represented the gut microbial community across all gut microbiomes from the corresponding category of study-cohorts. This was judged by computing two scores:

- Score 1: The proportion of microbiomes where the selected species accounted for more than 90 percent of the entire microbiome community
- Score 2: The proportion of microbiomes where the selected species accounted for less than 70 percent of the entire microbiome community

Our objective was to identify for each of the four cohort categories the smallest affordable set of microbes with the highest possible Score 1 and the lowest possible Score 2. Using this logic, we investigated each microbial set for Score 1 and Score 2 and selected the following microbial taxa sets as described below. For Infant 16S cohorts, we shortlisted 164 taxa ($x=0.35$, $y=0.15$). This set cumulatively contributed to greater than 90% relative abundance in 100% of microbiomes in Infant cohorts sequenced using 16S approach (Score 1) and less than 70% in none (Score 2). For Adult/Senior 16S cohorts, we shortlisted the microbial set with 194 taxa ($x=0.4$, $y=0.1$). This set cumulatively accounted for greater than 90% relative abundance in 90.9% of the microbiomes of cohorts belonging to this category and less than 70% in none. Similarly for the Infant cohorts sequenced using WGS, we shortlisted 226 taxa ($x=0.05$, $y=0.1$), which accounted for $\geq 90\%$ abundance in 93.3% microbiomes in cohorts of this category and less than 70% in none. Finally, in this manner, for the Adult/Senior WGS cohorts, we shortlisted 139 taxa ($x=0.35$, $y=0.1$), which accounted for which the Score 1 was 90.3% and Score 2 was 2.03%.

Supplementary Note 2

Investigation of the association patterns of the non-Bifidobacterial taxa with the Bifidobacteria in different Diseased-Groups

Collation of datasets from different diseases and disease-matched controls

Since a majority of the study-cohorts investigated here were from the IndustrializedUrban cohorts and from the adult/senior age-groups (with minimal datasets containing patient gut microbiomes from non-Industrialized populations), we specifically investigated the variation of non-Bifidobacterial association patterns (with different Bifidobacterial properties) in the adult/senior gut microbiomes from IndustrializedUrban cohorts. Further, to gain a comprehensive understanding of the disease-associated alterations in these association patterns, we further supplemented our original data collection of 46,218 gut microbiomes from 122 study-cohorts with an additional 27 study-cohorts containing a total of 5,026 gut microbiomes from different diseases with matched non-diseased controls (**Supplementary Data 1**)^{2–28}.

Addition of these cohorts resulted in a total of 69 study-cohorts containing gut microbiomes from six different diseases along with the microbiomes from matched study-specific control (non-diseased) individuals from Adult/Senior age-groups and IndustrializedUrban population. These six diseases were Colorectal Cancer (CRC), Type-II Diabetes (T2D), Inflammatory Bowel Disorder (IBD), Cardiovascular Diseases (CVD), Parkinson's Disease and COVID. We considered six diseases as they contained data from at least seven study-cohorts, thus increasing the reliability of the computed associations.

Out of these 69 study-cohorts thus selected, 34 and 35 datasets were from 16S and WGS profiling strategy respectively, forming a pooled set of 13,913 microbiomes which were finally utilized for disease-wise Association-Scores, thereby enabling the investigation of the replicability of the patterns across different profiling strategies (see **Supplementary Data 1**).

Computation of Association-Scores for the different Bifidobacterial features

The consensus top predictors identified while computing the overall Association-Scores were investigated for their association with the corresponding *Bifidobacterium* features. For a given disease category (it can be one of the six diseases or the pooled controls) 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 feature (either the total number of

Bifidobacterium or the abundance of any of the eight *Bifidobacterium* taxa) in each of the study-cohorts belonging to the specific disease category. 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 taxon with the specific *Bifidobacterium* feature in a given disease category (categorized above) as:

$$\text{Association-Score} = ((\text{SP} - \text{SN}) / T) * (1 - (\text{MIN}(\text{SP}, \text{SN}) + 0.0001) / (\text{MAX}(\text{SP}, \text{SN}) + 0.0001))$$

(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 graded association-scores based on cohort-stratified quantitative Association-Scores

Although, Association-Scores quantified the strength of association between each non-*Bifidobacterium* taxon and *Bifidobacterium* feature across cohorts stratified by profiling strategy, age group, cohort life-style and disease categories, grading of association-scores was performed to identify strongly consistent associations while filtering out weaker or less stable signals, with the top 10% of association-scores designated as robust associations. To derive graded Association-Scores, the Association-Scores, computed from the correlation-based quantification of consistencies, were first stratified by categories (whether it's age-category or life-style or disease categories or host-associated metadata categories). For each category, the empirical distribution of all corresponding association-scores was used to define data-driven thresholds, with the 10th and 90th percentiles serving as lower and upper cut-offs, respectively. Association values exceeding the upper threshold and with positive values (>0) were classified as strong positive associations (+2), whereas values below the lower threshold and with negative values (<0) were classified as strong negative associations (-2). Associations falling between these thresholds were assigned a directional weak score (+1 or -1) based solely on their sign. This grading procedure was applied separately within each category to account for category-specific variability in association strengths. The resulting graded Association-Scores, thus encodes both the direction and relative strength of associations in a discrete, comparable

framework, facilitating comparison across categories to identify the conserved association patterns.

Investigation of association patterns across different diseases as well as with non-diseased state

For each of the nine *Bifidobacterium* features, spearman correlation was calculated using the Association-Scores across all the six diseases as well as controls to check if there is pattern of association among the disease groups as compared to control. These inter-disease association patterns have been shown as heatmaps in **Supplementary Figure 16**.

Supplementary Note 3

Investigation of the association patterns of the non-Bifidobacterial groups with Bifidobacteria after adjusting for microbiome/host-associated covariates

Investigating the effect of different microbiome-properties on the different non-Bifidobacterial v/s Bifidobacterial association patterns: In terms of their association strengths, Shannon Index and Pielou's-Evenness Index 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 Industrialized Urban cohorts (irrespective of sequencing-type), but weaker associations in microbiomes from Rural Tribal cohorts (**Figure 3; Supplementary Data 5-12**).

We then investigated the effect of different microbiome properties on the different association patterns of the non-Bifidobacterial groups with Bifidobacteria. For this purpose, for each microbiome sample, we had computed Shannon Index (using the diversity function of the vegan package (v 2.7-2) and Pielou's-Evenness Index in R. Pielou's-Evenness Index was derived from the Shannon Index by normalizing entropy with respect to observed species richness, computed as the number of taxa with non-zero abundance per microbiome. Specifically, Pielou's-Evenness Index was calculated as the ratio of Shannon Index to the natural logarithm of species richness, with undefined values assigned to samples containing fewer than two observed taxa. The estimated Microbial Load was computed as described in ²⁹. Predicted Microbial Load and quantitative microbiome profiles were estimated using a machine-learning-based approach implemented in the MLP R package²⁹. Species-level abundance profiles generated using MetaPhlAn3³⁰ were used as input to a pretrained model which was trained on the MetaCardis dataset as mentioned by the authors. This trained model was applied to infer sample-specific Microbial Load as well as quantitative microbiome profiles (QMP), thereby transforming relative abundance estimates into absolute abundance-scaled representations. We specifically computed the estimated Microbial Load only for WGS datasets, given that the model primarily developed for such data.

The next step was to compute the Association-Scores adjusted for the different microbiome properties. The overall methodology for this approach is summarized in **Supplementary Figure 18A**. For this purpose, we utilized the concept of partial-correlations. Partial correlations are measurement of strength of associations between two properties after adjusting/controlling the effect of one (or more) confounding variables. We utilized this

concept. For each study-cohort, we first re-computed the associations between the non-Bifidobacterial taxa and Bifidobacterial feature as partial correlations after adjusting for the different microbiome-properties (Shannon Index and Pielou's-Evenness Index for 16S cohorts; Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for WGS cohorts). The significance of each correlation (using the partial correlation estimates and N of the number of samples) was computed as for the original Association-Scores (to recompute SP and SN values as described in **Figure 3A**; refer to **Supplementary Figure 18A**). These recomputed SP and SN values were then utilized to calculate the overall microbiome-property-adjusted Association-Scores (listed in **Supplementary Data 15**). The microbiome-property-adjusted Association-Scores were computed in cohort-stratified as in **Figure 3A** and **Supplementary Figure 18A**. We next correlated these microbiome-property-adjusted Association-Scores with the overall unadjusted Association-Scores. In 91 of the 95 tested scenarios, the correlations were significantly positive, 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, thus indicating that overall microbiome properties exert minimal influence on these association patterns (**Figure 5A**; **Supplementary Figure 18B**).

Investigating the effect of host diet, medication and clinical status on the different non-Bifidobacterial v/s Bifidobacterial association patterns

Across the investigated cohorts, a subset contained the relevant metadata (**Supplementary Data 16**), and both the schemas and the measured variables varied corresponding to each of the metadata substantially between cohorts. To address this heterogeneity, we first grouped metadata into three categories, diet, medication, and clinical markers. Given a metadata category, we performed dimensionality-reduction of the subject-specific metadata profiles within each cohort using robust principal component analysis (rPCA) and principal coordinate analysis (PCoA) for each cohort where the metadata was measured. The top three components of each approach (rPC1-3 for rPCA; PCo1-3 for PCoA) represented the components capturing the maximum variation of each metadata profile (**Supplementary Figure 18C**). Across each metadata-cohort combination, the percentage of the total variation captured by either rPC1-3 or PCo1-3 averaged 73.47% (**Supplementary Figure 20**). We computed metadata-adjusted Association-Scores by recalculating study-specific associations between the abundances of

different non-Bifidobacterial taxa and Bifidobacteria features as partial correlations, after adjusting for the six metadata-derived components (rPC1-3 and PCoA1-3) (that captured the maximum variation corresponding to that metadata) (**Supplementary Figure 18C**; **Supplementary Figure 21**; the different adjusted Association-Scores listed in **Supplementary Data 17**).

We subsequently investigated the effects of the different metadata categories on the association patterns by correlating these adjusted-Association-Scores for the different metadata with the overall Association-Scores (for the specific cohort-types) (**Supplementary Figure 18C**). In 46 out of the 48 scenarios, there was a significant positive correlation between the two (**Figure 5B**). Although the absolute values varied widely depending upon the specific Bifidobacterial feature and the metadata, the five Bifidobacterial features (*B. longum/adolescentis/breve/dentium* abundances and the Bifidobacterial detection rates) showed especially strong associations (average correlation = 0.56), further validating that the association patterns of the non-Bifidobacterial members with these five Bifidobacterial features remains stable even after adjusting for microbiome characteristics, diet, medications, host clinical characteristics and disease. The other four Bifidobacteria showed significant correlations across a majority of scenarios albeit with weaker association-strengths.

Supplementary Note 4

Descriptions of Unseen Validation Datasets utilized for investigating the reproducibility of the Association-Scores

We investigated eight datasets from different intervention trials that relate to previously published studies (detailed in **Supplementary Data 18**) to confirm the patterns of Association-Scores as a validation.

Below are the summary of the datasets and the different analyses conducted on the datasets. We also investigated the reproducibility of the Association-Scores in these validation datasets that is summarized in **Supplementary Note 5**.

1. **Maldonado-Gomez_2016**: This study was a double-blind, placebo-controlled crossover trial performed in the United States, including 23 healthy adult participants aged from 22 to 38 years³¹. Participants followed an industrialized-urban life-style and received a daily dose of *Bifidobacterium longum* subsp. *longum* AH1206 during two separate 7-week testing phases, each consisting of a baseline, a 14-day treatment period, and a 28-day follow-up. Whole-genome shotgun sequencing (WGS) of stool microbiomes was performed and the taxonomic profiling was done using MetaPhlAn3³⁰ to get the species-level taxa abundance profiles for the microbiomes and for our analysis, we utilized data for 21 subjects (baseline), as the metadata was available for 21 subjects only including the baseline. A notable aspect of the study was the stratification of subjects into “Persisters” and “Non-Persisters” based on whether the administered *B. longum* strain established long-term colonization. Prior analyses in the study revealed that baseline microbiome features particularly the abundance of carbohydrate metabolism genes, differentiated these two groups, suggesting that functional traits of the gut microbiome influence probiotic engraftment potential.

For the above-mentioned groups (Persisters and Non-Persisters), we applied our Receptive-Score (see **Methods**) methodology to the baseline WGS profiles of the 21 individuals. Using *B. longum*-specific Association-Scores trained in industrialized adult cohorts, we computed Receptive-Scores and tested their ability to distinguish Persisters from Non-Persisters. Indeed, baseline Receptive-Scores were significantly higher among Persisters (Mann-Whitney U-test $p = 0.018$), confirming that our framework has the potential to identify individuals with greater probiotic receptivity (**Figure 6C**). We also demonstrated that microbiome-level Receptive-Scores, computed using Association-

Scores and baseline abundances of non-*Bifidobacterium* taxa, significantly predicted the persistence of *B. longum* using robust linear regression model (**Coeff = 0.6, p = 0.036**), and Leave-One-Out-Cross-Validation (Coeff = 0.2, p = 0.36) (**Figure 6D; Supplementary Data 21**). Additionally, this dataset was used also included in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Figure 6A**).

2. **ZhangQ_2024**: This study was a randomized, parallel-group, placebo-controlled clinical trial carried out in China involving 200 adult individuals from IndustrializedUrban life-style, with gut microbiomes profiled using whole-genome shotgun (WGS) sequencing³. The taxonomic profiling was done using MetaPhlAn3 to get the species-level taxa abundance profiles for the microbiomes. The study investigated the impact of *Bifidobacterium animalis* subsp. *lactis* BL-99 supplementation. Participants first underwent a 2-week run-in phase with dietary restrictions on probiotic-containing foods, followed by 8 weeks of treatment (BL-99, PPI, or placebo) and an 8-week follow-up. Fecal samples were collected, of which we utilized 350 microbiomes from 175 subjects at baseline and post-treatment.

This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Figure 6A**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, significantly predicted the post-treatment increase of *B. animalis* using robust linear regression model (**Coeff = 0.44, p = 3.37E-06**), and Leave-One-Out-Cross-Validation (**Coeff = 0.34, p = 0.0016**) (**Figure 6D; Supplementary Data 21**). Third, the dataset was used to assess the robustness of the Receptive-Score framework through a cross-cohort validation approach: a cohort-specific linear regression model was trained to predict *B. animalis* increase based on baseline Receptive-Score and baseline abundance, and then tested on other intervention cohorts (**Figure 6E-I; Supplementary Figure 23**). This approach confirmed the generalizability and predictive power of the Receptive-Score methodology across independent datasets.

3. **SunB_2022**: This study was a randomized, double-blind, placebo-controlled clinical trial carried out in China to examine the effects of *Bifidobacterium animalis* subsp. *lactis* Probio-M8 on patients with coronary artery disease (CAD)³². A total

of 71 adult/senior participants were randomized to receive either the probiotic (37 subjects) or a placebo (34 subjects) over a six-month period, during which the probiotic group was administered daily sachets of Probio-M8. Fecal samples were collected from a subset of 41 individuals (24 from the probiotic group and 17 from the placebo group) at three time points: baseline (day 0), midpoint (day 90), and end of treatment (day 180). We focused on the 24 probiotic-group individuals, contributing 72 microbiomes profiled using whole-genome shotgun (WGS) sequencing and the taxonomic profiling was done using MetaPhlAn3 to get the species-level taxa abundance profiles for the microbiomes. The study aimed to assess microbiota and metabolic changes, especially in the context of adjunctive probiotic therapy for CAD.

This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Figure 6A**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, predicted the post-treatment increase of *B. animalis* at 90 days (RLR: **Coeff = 0.5, p = 0.04**; LOOCV: **Coeff = 0.25, p = 0.3**) (**Figure 6D; Supplementary Data 21**). Third, the dataset was used to assess the robustness of the Receptive-Score framework through a cross-cohort validation approach: a cohort-specific linear regression model was trained to predict *B. animalis* increase based on baseline Receptive-Score and baseline abundance, and then tested on other intervention cohorts (**Figure 6E-I; Supplementary Figure 23**). This approach confirmed the generalizability and predictive power of the Receptive-Score methodology across independent datasets.

4. ***BazanellaM_2017***: This study was a double-blind randomized, placebo-controlled trial involving 106 healthy infants in Germany (IndustrializedUrban cohort), where participants were allocated to receive bifidobacteria-supplemented formula, placebo formula, or exclusive breastfeeding during their first year of life³³. Microbiome profiling was conducted using 16S rRNA sequencing and the taxonomic profiling was done using SPINGO³⁴ to get the species-level taxa abundance profiles for the microbiomes. From this dataset, we selected subjects with samples available at 7, 12, and 24 months, in overall resulting in 106 subjects and 712 microbiome samples.

This dataset was utilized in several validation analyses. First, we utilized the dataset in computing the associations between the relative abundances of non-*Bifidobacterium* taxa and each of seven *Bifidobacterium* species (excluding *B. animalis*). These association

patterns were then compared to those derived from discovery cohorts within the same study category (Infant–IndustrializedUrban–16S). We observed significant positive correlations ($p \leq 0.05$), expect in *B. pseudocatenulatum* for these association patterns (spearman coefficient ranging from 0.1 to 0.73) (**Supplementary Figure 22A**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, predicted the increase of *B. longum*, *B. bifidum*, *B. breve* at 7 months. The significant increase, predicted at 7 months for *B. breve* was (RLR: **Coeff = 0.51, $p = 2.50E-03$** ; LOOCV: **Coeff = 0.61, $p = 1.07E-05$**), for *B. bifidum* it was (RLR: **Coeff = 0.37, $p = 0.019$** ; LOOCV: **Coeff = 0.59, $p = 2.21E-05$**). For *B. longum* it was (RLR: **Coeff = 0.23, $p = 0.199$** ; LOOCV: **Coeff = 0.31, $p = 0.04$**) (**Figure 6D; Supplementary Data 21**).

5. **BaZ_2021**: This study was a randomized, four-period, crossover model conducted in a partially blinded fashion with adult participants from USA (IndustrializedUrban life-style)³⁵. Participants received four treatments in randomized order: (A) yogurt smoothie without BB-12 (YS), (B) yogurt smoothie with BB-12 added post-fermentation (POST), (C) yogurt smoothie with BB-12 added pre-fermentation (PRE), and (D) BB-12 capsule (CAP), with each treatment lasting 4 weeks and a 2-week washout period in between. Measurements included anthropometrics, biochemical markers, physical activity, and immune endpoints. Microbiome profiling was conducted using 16S rRNA sequencing and the taxonomic profiling was done using SPINGO to get the species-level taxa abundance profiles for the microbiomes. For our analysis, we took 34 subjects for baseline, PRE, POST and CAP covering 111 microbiomes. This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Supplementary Figure 22B**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, significantly predicted the increase of *B. animalis* at PRE (RLR: **Coeff = 0.73, $p = 0.007$** ; LOOCV: **Coeff = 0.72, $p = 0.0004$**), at POST (RLR: **Coeff = 0.51, $p = 0.07$** ; LOOCV: **Coeff = 0.74, $p = 0.0007$**), and at CAP (RLR: **Coeff = 0.74, $p = 0.02$** ; LOOCV: **Coeff = 0.72, $p = 0.0005$**) (**Figure 6D; Supplementary Data 21**). Third, the dataset was used to assess the robustness of the Receptive-Score framework through a cross-cohort validation approach: a cohort-specific linear regression model was trained to predict *B. animalis* increase based on baseline Receptive-Score and baseline abundance,

and then tested on other intervention cohorts (**Figure 6E-I; Supplementary Figure 23**). This approach confirmed the generalizability and predictive power of the Receptive-Score methodology across independent datasets.

6. **GronbaekI_2025**: This study was a randomized, double-blind, placebo-controlled investigation³⁶. Participants were given either Bif195 or placebo capsules each day for an 8-week period, followed by another 8 weeks without treatment. Stool samples were gathered during every visit for microbiome evaluation, along with surveys measuring disease severity and quality of life. This dataset employed WGS sequencing data from adult/senior participants in Denmark (DNK), belonging to an IndustrializedUrban life-style cohort. The taxonomic profiling was done using MetaPhlAn3 to get the species-level taxa abundance profiles for the microbiomes. The therapeutic agent used was *Bifidobacterium breve*. We took a total of 31 probiotic subjects encompassing 124 microbiomes across four time-points: Baseline, Week 4, Week 8, and Week 16.

This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Figure 6A**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, significantly predicted the increase of *B. breve*. We used Week 8 as the baseline, since it marked the completion of the intervention and computed Receptive-Scores using Association-Scores and Week 8 abundances of non-*Bifidobacterium* taxa. This predicted the post-treatment increase of *B. breve* at Week 16 (RLR: **Coeff = 0.48, p = 0.002**; LOOCV: **Coeff = 0.43, p = 0.12**) (**Figure 6D; Supplementary Data 21**).

7. **RamakrishnanM_2025**: This study was a randomized, placebo-controlled trial examining the effectiveness of *B. adolescentis* VS-1 in alleviating symptoms of lactose intolerance in confirmed lactose maldigesters³⁷. Participants were screened using a hydrogen breath test and randomized to either probiotic (11 subjects) or placebo (10 subjects). Those in the treatment group received *B. adolescentis* iVS-1 supplementation over the intervention period, with stool samples collected pre- and post-treatment for microbiome profiling. This dataset includes 16S sequencing data from adult participants in the USA, categorized under the IndustrializedUrban life-style cohort. The taxonomic profiling was done using

SPINGO to get the species-level taxa abundance profiles for the microbiomes. We used data from 11 subjects who took probiotics in the treatment group, resulting a total of 33 microbiomes across three time-points. This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Supplementary Figure 22B**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using Association-Scores and pre-treatment abundances of non-*Bifidobacterium* taxa, predicted the post-treatment increase of *B. adolescentis* (RLR: **Coeff = 0.3, p = 0.06**; LOOCV: **Coeff = 0.24, p = 0.51**) (**Figure 6D**; **Supplementary Data 21**). Since the probiotic intervention was ongoing until the second time point, we treated the first (pre-treatment) time point as the effective baseline, as it reflects the point when the treatment was completed. Although the RLR, and LOOCV outcomes did not reach statistical significance, this was probably due to the limited statistical power associated with the small sample size ($n = 11$). Third, the dataset was used to assess the robustness of the Receptive-Score framework through a cross-cohort validation approach: a cohort-specific linear regression model was trained to predict *B. adolescentis* increase based on baseline Receptive-Score and baseline abundance, and then tested on other intervention cohort (**Supplementary Figure 23**). This approach confirmed the generalizability and predictive power of the Receptive-Score methodology across independent datasets.

8. **KrumbeckJ_2018**: This study was a double-blind, randomized, parallel-group clinical trial carried out in the Chicago, USA involving obese individuals aged between 18 to 65 years, from IndustrializedUrban life-style³⁸. The participants were randomized into 6 groups: (i) Placebo (lactose only), (ii) Prebiotic (galacto-oligosacharides only), (iii) Prebiotic (*B. adolescentis* IVS-1 only), (iv) Prebiotic (*B. animalis* subsp. *lactis* BB-12), (v) Synbiotic combination (IVS-1 + GOS) and (vi) Synbiotic combination (BB-12 + GOS). The treatments were administered daily for 3 weeks, with participants consuming sachets containing the assigned preparation mixed with water. The taxonomic profiling was done using SPINGO to get the species-level taxa abundance profiles for the microbiomes. We used data from 105 subjects, resulting a total of 210 microbiomes across two time-points (baseline and treatment). This dataset was utilized in several validation analyses. First, we utilized the dataset in evaluating whether the association patterns identified in the discovery cohorts could be reliably reproduced in the validation cohorts (**Supplementary Figure 22B**). Second, we demonstrated that microbiome-level Receptive-Scores, computed using

Association-Scores and baseline abundances of non-*Bifidobacterium* taxa, predicted the post-treatment increase of *B. adolescentis* (RLR: **Coeff** = 0.28, **p** = 0.27; LOOCV: **Coeff** = **0.58**, **p** = **0.00028**) and another case where we predicted the post-treatment increase of detected *Bifidobacterium* (RLR: **Coeff** = **0.23**, **p** = **0.04**; LOOCV: **Coeff** = **0.65**, **p** = **3.56E-14**) (**Figure 6D**; **Supplementary Data 21**). Third, the dataset was used to assess the robustness of the Receptive-Score framework through a cross-cohort validation approach: a cohort-specific linear regression model was trained to predict *B. adolescentis* increase based on baseline Receptive-Score and baseline abundance, and then tested on other intervention cohort (**Supplementary Figure 23**). This approach confirmed the generalizability and predictive power of the Receptive-Score methodology across independent datasets.

Supplementary Note 5

Detailed methodology for investigating the reproducibility of the Association-Scores in the Validation Intervention Cohorts

Four of the validation intervention trial cohorts contained microbiomes of a specific type, namely from adult/senior individuals from IndustrializedUrban populations and sampled using WGS approach^{31,32,36,39}. This microbiome group also dominated our discovery cohorts. Thus, first, we investigated here whether the Association-Scores computed from our discovery-cohorts showed similar patterns in this validation cohorts. Thus, we recomputed the Association-Scores of each of the non-Bifidobacterial taxa with respect to each of the eight *Bifidobacterium* features (as described above) in specifically these four validation cohorts and subsequently correlated these values with the overall Association-Scores obtained for each of the corresponding eight features in the discovery cohorts. Since the four validation cohorts corresponded to adult/senior individuals from IndustrializedUrban with WGS profiling approach, for the correlation analysis, we took the mean of the overall Association-Scores computed from adult and senior individuals, from IndustrializedUrban populations using WGS (**Supplementary Data 19A; Figure 6A**). We performed the same analysis for 16S using three validation cohorts of adult/senior individuals from IndustrializedUrban populations^{40–42}. Association-Scores were recomputed within these cohorts and correlated with discovery-cohort scores, using the mean Association-Scores derived from adult and senior IndustrializedUrban cohorts profiled with 16S sequencing (**Supplementary Data 19B; Supplementary Figure 22B**).

Only one of our validation cohorts was from Infants (IndustrializedUrban population using 16S profiling)⁴³. Thus, to investigate the reproducibility of the Infant-associated association patterns, we first computed the spearman correlations of each of the non-Bifidobacterial taxa with seven *Bifidobacterium*-associated features (with the exception of *B. animalis* as it did not yield any consistent pattern of associations in the discovery cohort). We then correlated this Spearman correlation patterns with the corresponding Association-Scores for each non-*Bifidobacterium* taxa (detected commonly across both datasets) with each of the seven *Bifidobacterium* features computed for the 16S-sequenced microbiomes from Infants from IndustrializedUrban populations (**Supplementary Data 20; Supplementary Figure 22A**).

Supplementary Note 6

Detailed methodology for the assessment of Clinical Effect in the Validation Intervention Cohorts

Given the established relevance of *Bifidobacterium* to host health, we further explored whether the microbiome-based Receptive-Score was associated with clinical outcomes in validation intervention cohorts where clinical metadata were available. Among the eight intervention cohorts considered, two studies contained sufficiently detailed clinical measurements with matching sample annotations in microbiome and clinical outcome data and were therefore included in this analysis.

RamakrishnanM_2025: Gastrointestinal symptom outcomes

The RamakrishnanM_2025 study was a randomized, placebo-controlled trial evaluating the effect of *Bifidobacterium adolescentis* iVS-1 supplementation on symptoms of lactose intolerance in confirmed lactose maldigesters⁴⁴. Adult participants from the USA (IndustrializedUrban life-style) were randomized to receive either iVS-1 or placebo, with stool samples collected before, during, and after the intervention. Microbiome profiling was performed using 16S rRNA sequencing, and species-level taxonomic profiles were generated using SPINGO.

For the present analysis, we focused on the probiotic-treated group and included subjects for whom both microbiome and clinical data were available across all three time points, resulting in ten individuals. Clinical metadata comprised gastrointestinal symptom scores, including abdominal pain, bloating, diarrhea, fecal urgency, flatulence, and a composite symptom score.

To assess clinical change, symptom differences were calculated as the change between post-treatment and pre-treatment time points ($T3 - T1$). These symptom changes were then correlated with the Receptive-Score computed at the during-treatment time point ($T2$). In addition, we derived a summary metric capturing overall symptom improvement by counting the number of symptoms that showed a reduction (negative change) per individual. This symptom reduction was also correlated with the Receptive-Score (**Supplementary Figure 24A-E**).

BaZ_2021: Immunological blood markers

The BaZ_2021 study was a randomized, four-period crossover intervention conducted in adult participants from the USA (IndustrializedUrban life-style)⁴⁰. Participants received multiple formulations of *Bifidobacterium animalis* subsp. *lactis* BB-12, including yogurt-based and capsule-based interventions, with washout periods between treatments. Microbiome profiling was performed using 16S rRNA sequencing, and species-level abundances were inferred using SPINGO.

For our analysis, we included 20 participants with available baseline and PRE intervention data having clinical metadata. Clinical metadata encompassed a broad range of immune-related measurements. To quantify clinical response, changes in each clinical parameter were computed as the difference between the PRE intervention phase and baseline. These clinical changes were then correlated with the baseline Receptive-Score, allowing us to assess whether baseline microbiome receptivity to *Bifidobacterium* was associated with subsequent clinical responses to BB-12 supplementation (**Supplementary Figure 24F–I**).

Supplementary Note 7

Database of reference genomes and metagenome-assembled genomes along with the detailed methodology for the functional investigation

Creation of the database

We collated a genome collection of 60,573 high quality genomes (with completeness > 90% and contamination < 5%), comprising of reference genomes and metagenome assembled genomes (MAGs) from adult/senior age-group as most of the MAGs were available for this age-group only, encompassing 1671 different human microbiome associated taxa. The genomes were collated from two major genome repositories: NCBI-RefSeq database containing 26,026 reference genomes and 34547 MAGs from two different repositories^{45,46}. We retained 24,980 genomes belonging to 66 consistently occurring non-Bifidobacterial taxa for which Association-Scores were obtained against multiple Bifidobacterial features (**Figure 3B**) and finally these 24,980 genomes were utilized for generating the functional profiles for 66 non-Bifidobacterial taxa.

Genome annotation

All selected high-quality genomes were functionally annotated using the eggNOG-mapper pipeline (v2), which utilizes the eggNOG database (version 5.0)^{47,48}. Protein-coding genes were predicted with Prodigal, integrated within the eggNOG-mapper workflow⁴⁷. The resulting annotation files were processed using an R script to construct functional profile matrices by quantifying the occurrence of each functional feature within a genome. These matrices quantitatively capture the presence of functional signatures across genomes. Functional profiles were generated for nine different functional schemas resulting in a total of 36,917 functional features (Features break-up: CAZy: 95, KEGG Module: 619, KEGG Ortholog: 7009, KEGG Reactions: 2834, KEGG RClass: 1098, BiGG: 9468, EC Number: 2250, PFAMs: 7403, COG: 6144)⁴⁹⁻⁵³.

Mean Functional Abundance Profile and Conservation Profile of unique non-Bifidobacterial species-level taxa and grouping based on functional redundancy

The genome-level functional profiles (24,980 genomes X 36917 functional features) were then aggregated to obtain species-taxa-level functional features (66 taxa X 36917 functional

features) using two approaches. We computed the Mean Functional Abundance profile by computing the mean abundance of each functional feature across all genomes of a given species to obtain a 66 species-level-taxa X 36917 functional feature profile matrix. Similarly, a Conservation profile (66 species-level-taxa X 36917 functional features) was computed by obtaining the detection percentage of each functional feature across all genomes of each of the 66 species-level-taxa.

Subsequently, we reduced the dimensions of the Mean Functional Abundance matrix by removing functionally redundant features originating from different schema in a step-wise manner. First, we removed functional features that did not show $\geq 90\%$ conservation across the genomes of at least one of the 66 taxa. Subsequently, the functional features showing zero Euclidean distance in terms of their mean abundance across all taxa were grouped using a Greedy approach, where each functional feature was sequentially compared with all other features preceding it in the profile. If any of the preceding functional-features showed exact similarity of the mean abundance profile with the given feature, there were grouped as one functional-group. Else, it was assigned its own functional-group. Finally, we obtained a set of 16,077 functional groups from 66 non-Bifidobacterial taxa.

Machine-learning based identification of genome-specific functional groups of different taxa that are associated with their Association-Scores with different features

The next goal was to investigate whether the functional-groups of non-Bifidobacterial taxa had the ability to predict the different Bifidobacteria-linked Association-Scores. For this purpose, we adopted a two-stage Random Forest (RF)-based approach. In the first stage, we trained nine Initial models to predict the Association-Scores for each of the eight Bifidobacterial taxa and the overall *Bifidobacterium* detection based on the mean abundance of the 16,077 functional-groups as predictors. The importance-scores for the 16,077 functional-groups were obtained for each of the nine models and the functional-groups sorted based on their scores. Based on these sorted lists of functional-groups, for each of the model, we next selected the most informative features as the top set of features that cumulatively accounted for 90% of the total importance scores for the corresponding model.

In the second stage, the nine models were re-trained using these selected set of most informative features corresponding to each model. Finally, the predictive performance of each of these nine refined models were judged based on the correlation between the actual and the predicted Association-Scores.

Identifying functional features associated to different Bifidobacterial Association-Scores

The union set of informative features for the nine models encompassed 5,780 unique functional groups. From this list, we further selected only those functional-groups that satisfied the following criteria: appearing amongst top 50 features in at least one model; present in the list of informative features of at least two of the nine models; present in >1 non-Bifidobacterial taxa; showing significant spearman correlation of its taxa-specific mean abundances with at least two Association-Scores. Finally, 162 functional-groups were obtained that satisfied all the above criteria (**Supplementary Data 22; Supplementary Figure 25**).

Supplementary Note 8

Investigation of non-Bifidobacterial functional features associated with the different Bifidobacterial features

As expected, functional groups showing positive links with *B. adolescentis/longum* relative abundance were predominantly observed in the genomes of the health-promoting members of the gut microbiome. Some of these functional groups were also showing negative associations with relative abundances of *B. breve/dentium* (Supplementary Figure 26).

One such functional group, showing positive associations with *B. longum/B. pseudocatenulatum* and a negative association with *B. breve*, was pantothenate kinase. This enzyme was predominantly detected in several beneficial taxa, including *Faecalibacterium prausnitzii*, *Gemmiger formicilis*, commensal *Ruminococcus* species (*R. bromii/lactaris/torques*), *Fusicatenibacter saccharivorans*, as well as (interestingly) the pathobiont *Flavonifractor plautii*. Pantothenate kinase is a key enzyme in pantothenate biosynthesis, an essential vitamin for the host and a precursor for coenzyme A, which in turn supports the production of beneficial gut microbiome-derived metabolites such as short-chain fatty acids (SCFAs)⁵⁴.

We identified multiple functional groups associated with *B. longum/adolescentis* that are linked to the production of short-chain fatty acids (SCFAs), including butyrate and propionate and other key metabolites like tryptophan. For example, the GH32 CAZyme, positively associated with *B. longum/pseudocatenulatum*, was predominantly found in gut-resident butyrate producers such as *Intestinibacter bartlettii*, *Ruminococcus obeum*, *Fusicatenibacter saccharivorans*, *Anaerostipes hadrus*, *Coproccoccus comes*, and *Dorea longicatena*. GH32 enzymes catalyze the breakdown of fructo-oligosaccharides into fermentable substrates that promote butyrate biosynthesis⁵⁵. Similarly, the aspartate aminotransferase functional group, predominant in gut-associated Firmicutes, namely *Eubacterium halii*, *Ruminococcus obeum*, *Fusicatenibacter saccharivorans*, *Anaerostipes hadrus*, and *Gemmiger formicilis*, catalyzes the reversible conversion of aspartate to glutamate and oxaloacetate, key precursors for butyrate biosynthesis^{56,57}. We also observed functional groups involved in threonine biosynthesis, a known precursor of propionate⁵⁶. Another functional group associated with increased relative abundance of *B. adolescentis/pseudocatenulatum* and predominantly found in normal-gut-associated

Firmicutes was the alpha-Glucoside sugar transporter, involved in the transport of sugar molecules⁵⁸. Similarly, shikimate dehydrogenase, participates in aromatic amino acid biosynthesis, and contributes to the production of tryptophan, an essential host amino acid and precursor of multiple health-associated metabolites. Together, these findings indicate that non-*Bifidobacterium* functional groups positively associated with health-linked *Bifidobacterium* taxa are mechanistically connected to the production of beneficial microbiome-derived metabolites.

In contrast, non-Bifidobacterial functional groups positively associated with *B. breve*/*B. dentium* and negatively associated with *B. adolescentis/longum*, were enriched for proteins involved in the transport of metal ions (iron, zinc, and manganese) (predominantly observed oral taxa) and glycine betaine, a trimethylamine (TMA) precursor (**Supplementary Figure 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 homeostasis⁵⁹.

Supplementary Note 9

Linear Regression Model based approach to compute the association of the detection pattern of different *Bifidobacteria* with the cohort SequencingType, Cohort-Type and host age-group categories

For a given *Bifidobacterium* species *i*, the first linear regression models were implemented as:

$$\text{Model 1: } \text{lm}(\text{Prevalence} \sim \text{SequencingType} + \text{Age-Group} + \text{CohortType}) \quad (6)$$

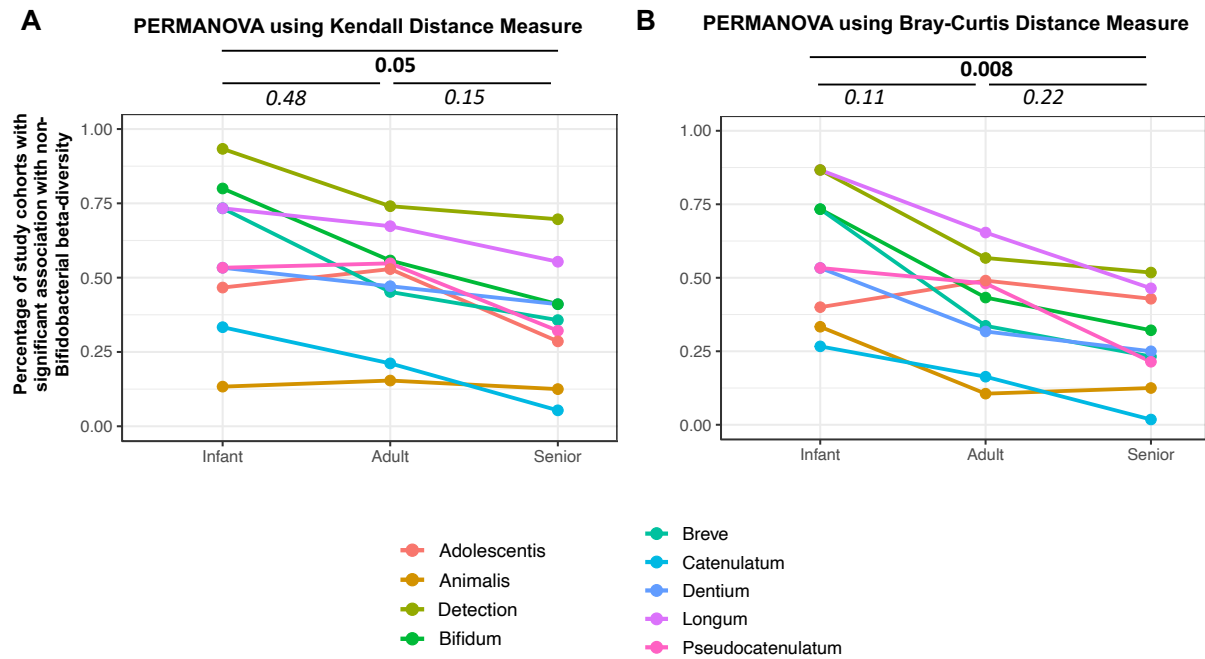
This model utilized the all the study-cohorts, where in Prevalence indicated the prevalence of the species *i* in a given study-cohort which was regressed using the SequencingType, Age-Group and CohortType of that study-cohort. This model computed the difference in the prevalence of species *i* in the infant and senior age-groups with respect to the adult age-group (after adjusting for CohortType and SequencingType); the difference in the prevalence of *i* in the RuralTribal and UrbanRuralMixed cohorts with respect to the IndustrializedUrban cohort (after adjusting for Age-Group and SequencingType) and; the prevalence difference of *i* in the 16S with respect to the WGS (after adjusting for Age-Group and CohortType). The second model:

$$\text{Model 2: } \text{lm}(\text{Prevalence} \sim \text{SequencingType} + \text{Age-Group}) \quad (7)$$

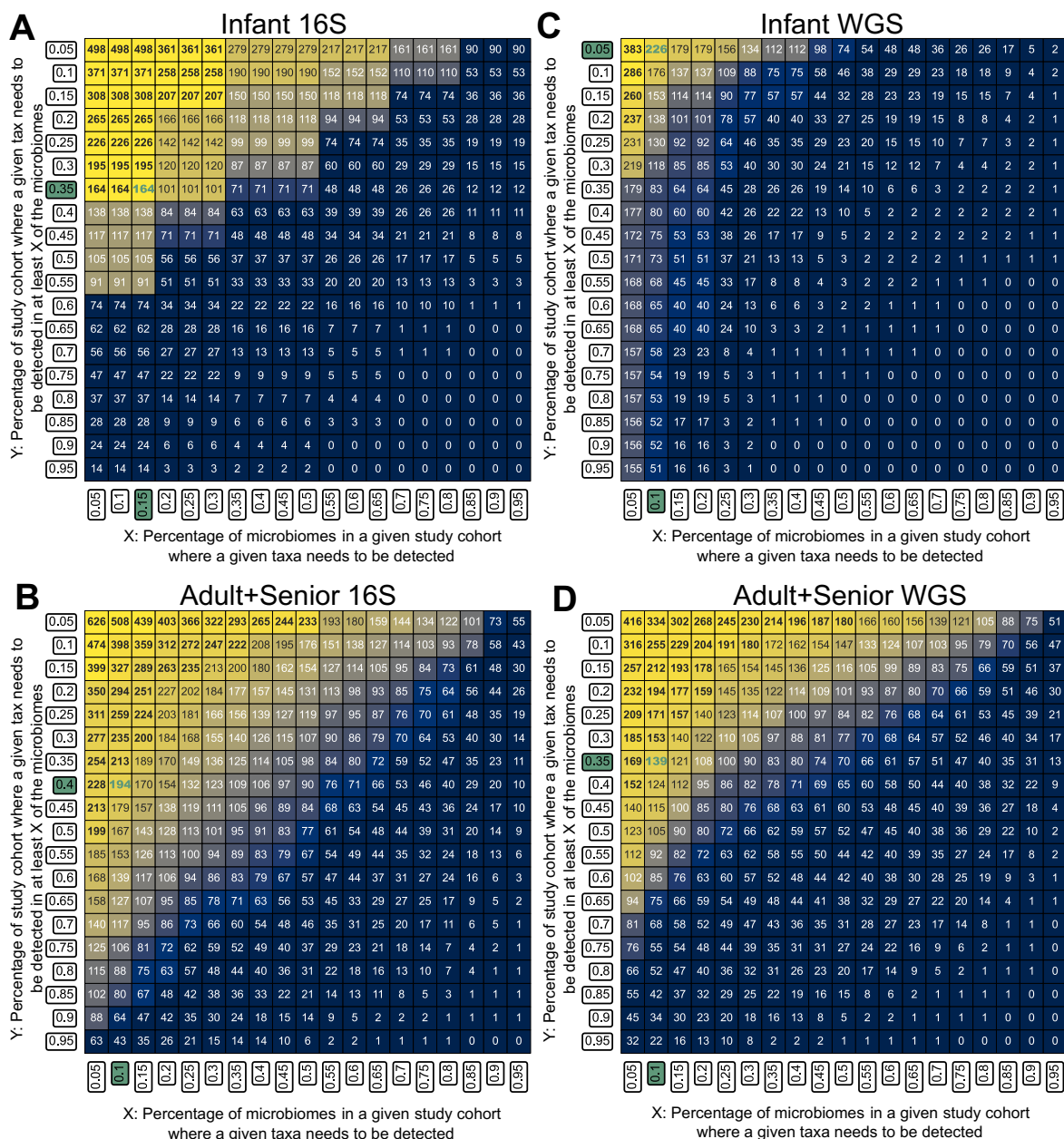
This model utilized only the infant and senior cohorts belonging to the IndustrializedUrban and computed the prevalence variation of the species *i* only between the Infant and Senior cohorts (after adjusting for the SequencingType). We did not consider RuralMixed and UrbanRuralMixed cohorts here as most cohorts from Infants and Seniors belonged to the IndustrializedUrban category.

$$\text{Model 3: } \text{lm}(\text{Prevalence} \sim \text{SequencingType} + \text{Cohort}) \quad (8)$$

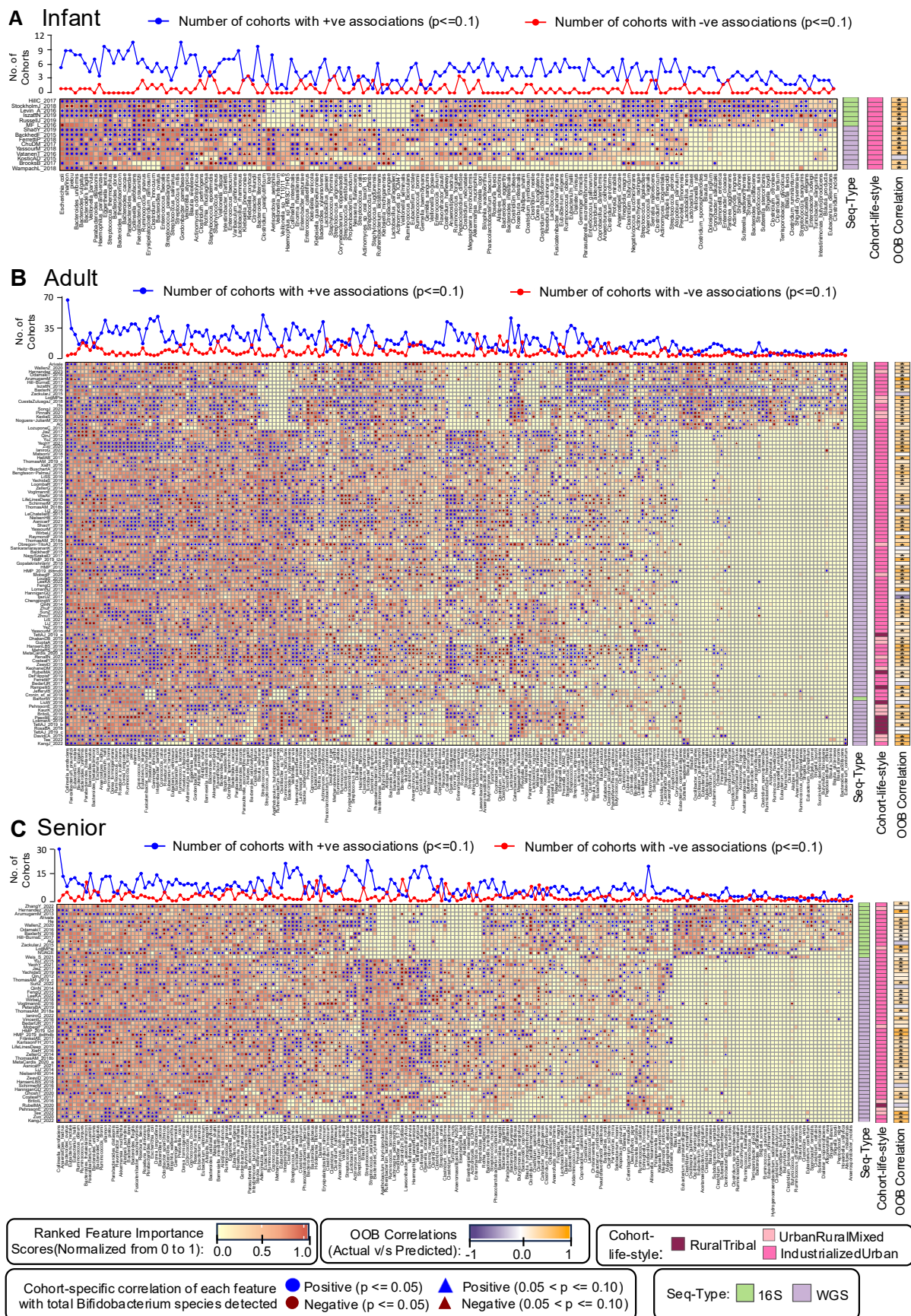
This model utilized only the RuralTribal and UrbanRuralMixed cohorts, belonging to the adult Age-Group, and profiled the difference only between the RuralTribal and UrbanRuralMixed after adjusting for the SequencingType.



Supplementary Figure 1. Line plots showing, for each of the eight major *Bifidobacterium* taxa, the percentage of study-cohorts with significant association with non-Bifidobacterial community compositions across age groups (Infant, Adult, Senior), when the community variations are processed using the **A.** Bray-Curtis and **B.** Kendall distance measures. Each line represents the eight *Bifidobacterium* taxon as indicated by the color scheme used in the figure, and points indicate the percentage of cohorts with significant associations within each age-group. The results of the Mann-Whitney tests comparing these patterns across cohorts from each age-group pair is also indicated. Differences across age-group pairs were assessed using the Mann-Whitney U test (two-sided), with $p \leq 0.05$ considered significant. No multiple testing correction was applied, as analyses were conducted independently across taxa and comparisons.



Supplementary Figure 2. Number of taxa identified using various thresholds of X and Y, where X is the percentage of microbiomes of a given study-cohort where the taxa is detected and Y is the number of study-cohorts where the taxa is detected in at least X% of the microbiomes. These values were separately obtained for **A.** Infant-16S cohorts, **B.** Adult+Senior-16S cohorts, **C.** Infant-WGS cohorts and **D.** Adult+Senior-WGS cohorts. The threshold of X and Y for each category (**A-D**) has been shown in green boxes.

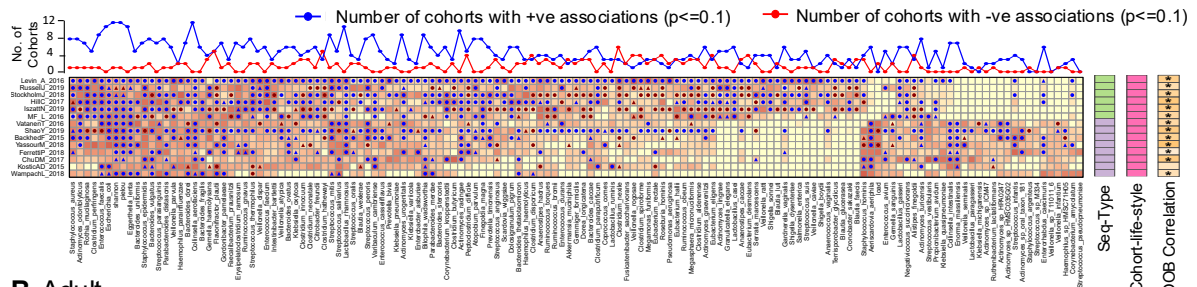


Supplementary Figure 3. Random Forest (RF) models identified distinct sets of non-Bifidobacterial taxa and microbiome properties as top predictors of the total number of

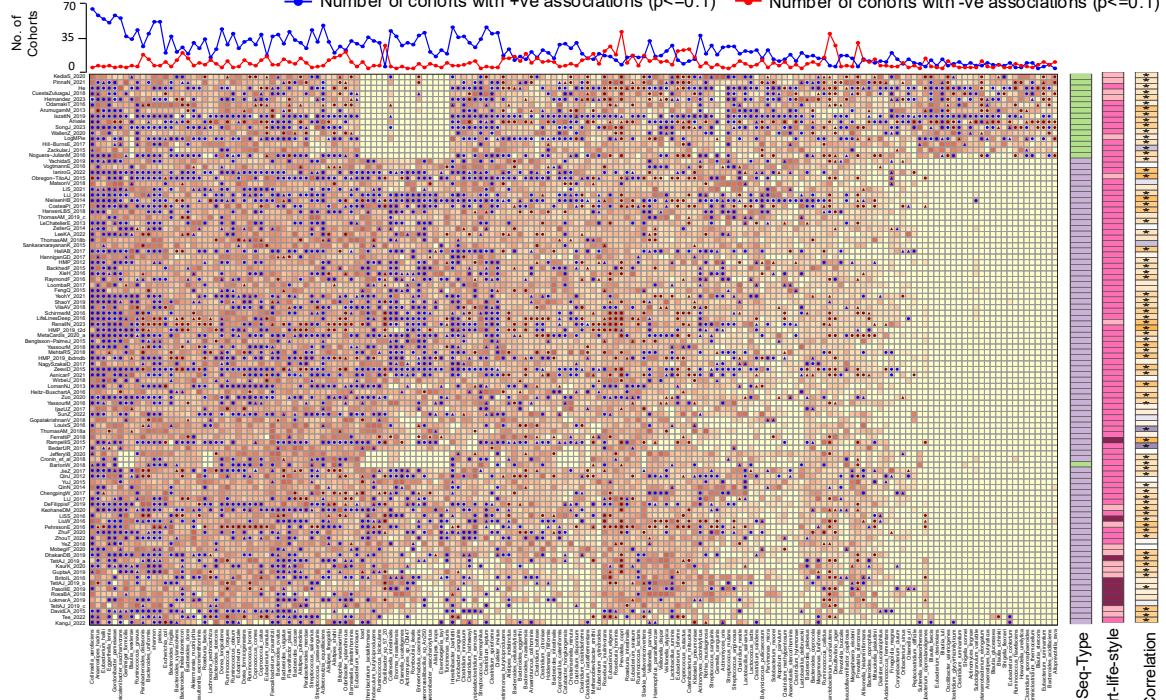
Bifidobacterial taxa detected in the gut microbiome. A-C. Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as the three microbiome properties (Shannon Index, Pielou's-Evenness Index and estimated Microbial Load) for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort-life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted total number of *Bifidobacterium* taxa for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (either non-Bifidobacterial taxon or microbiome property) in columns and Bifidobacterial detection within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with total Bifidobacterial species ($p \leq 0.1$). The extent and consistency of these associations varied by the specific non-Bifidobacterial taxon, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns; and RuralTribal cohorts displayed consistent association patterns that were distinct from those observed in IndustrializedUrban cohorts. Additionally, amongst microbiome properties, while Shannon/Pielou's-Evenness indices showed high predictive power, predictive power of estimated Microbial Load was lower. These patterns were captured in the overall Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

A Infant

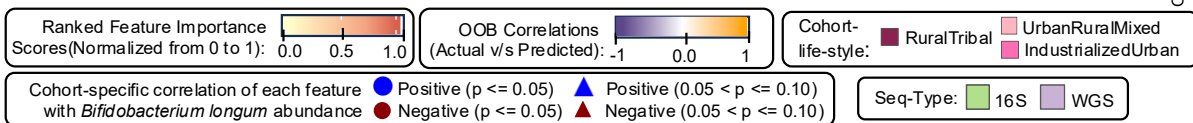
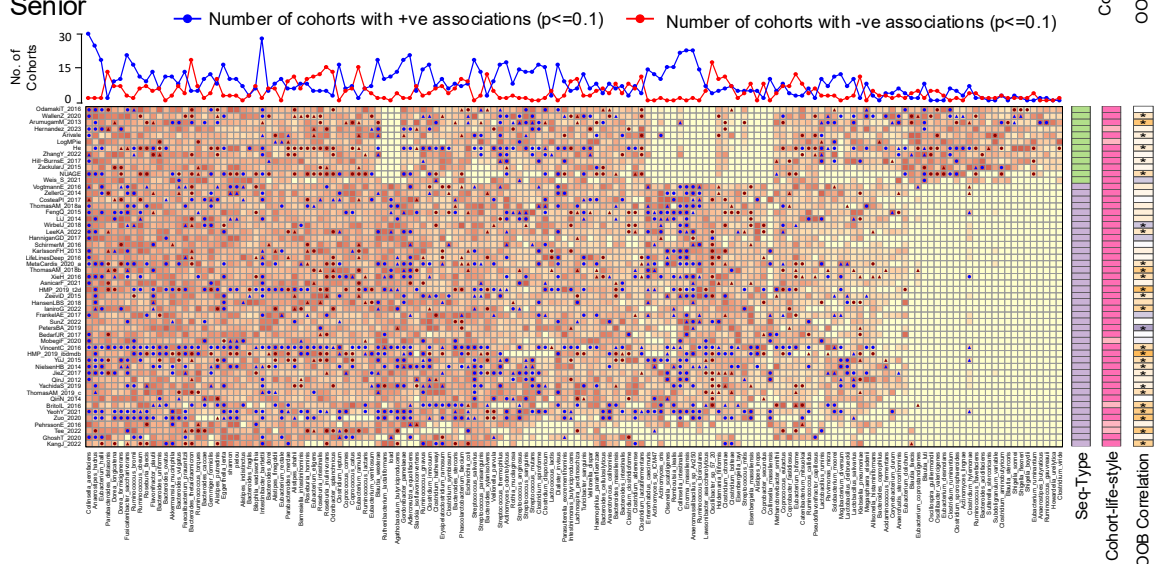
Bifidobacterium longum



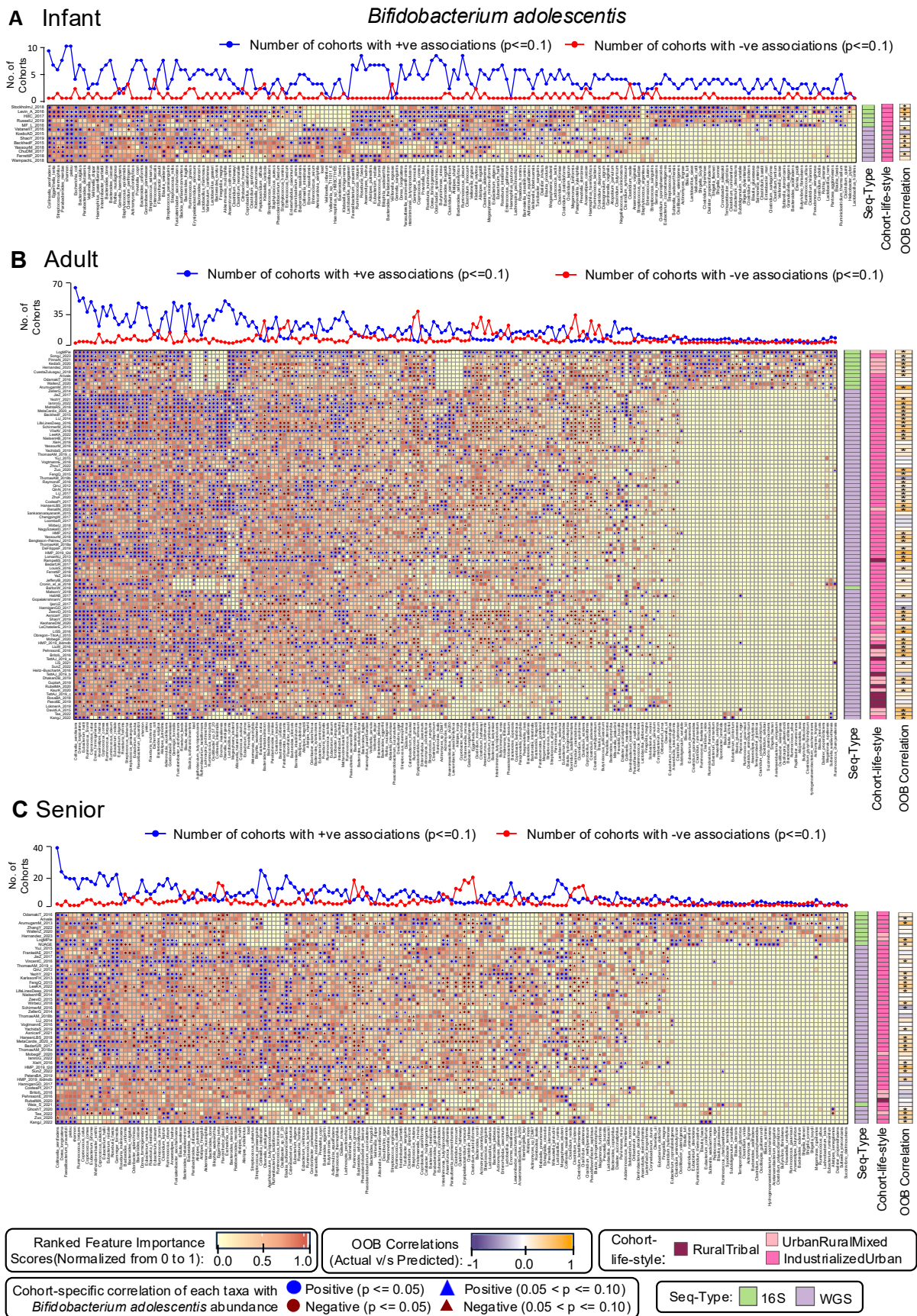
B Adult



C Senior



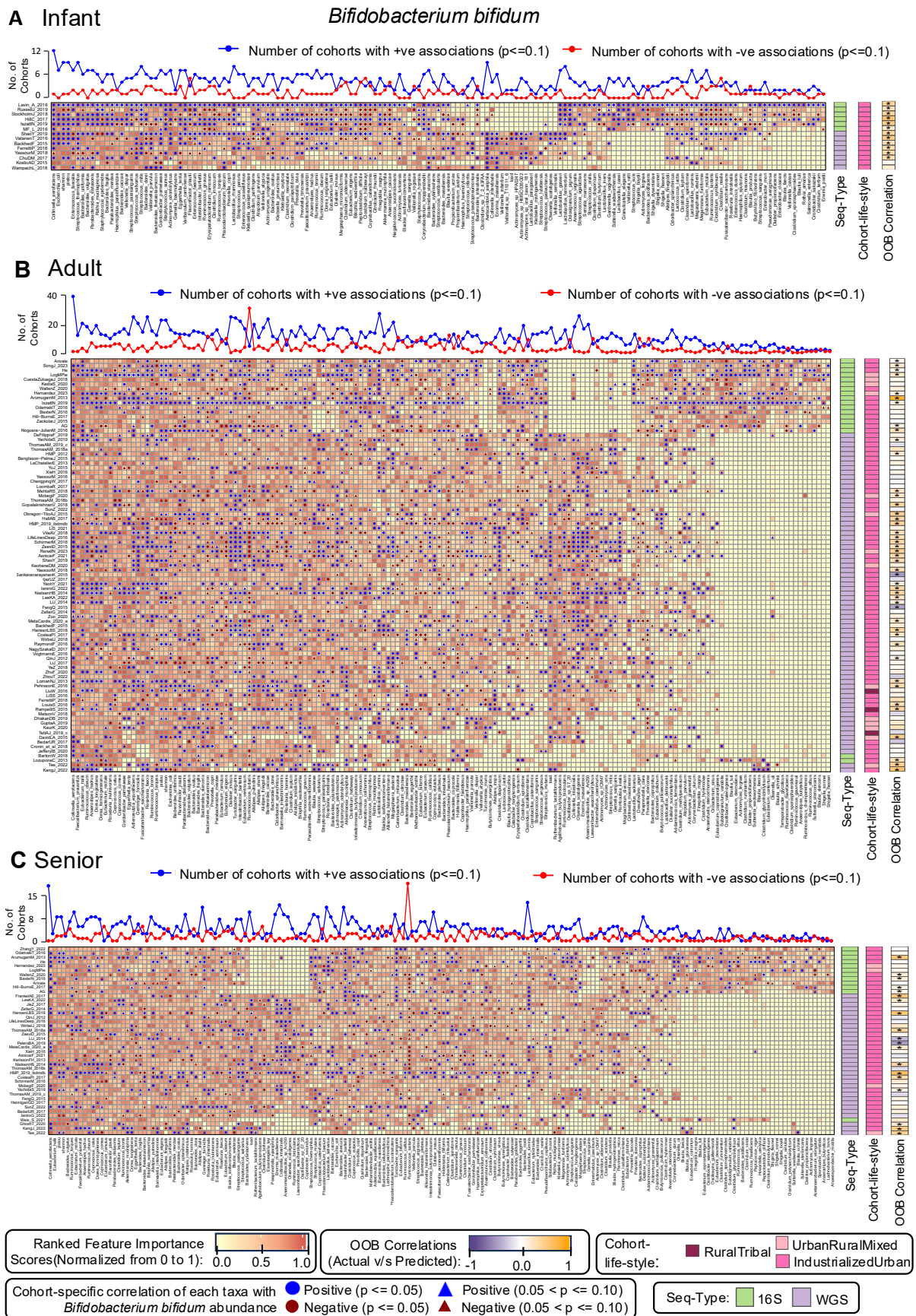
Supplementary Figure 4. Random Forest (RF) model-based identification of distinct sets of non-Bifidobacterial taxa that consistently associated with the relative abundance of *Bifidobacterium longum*, serving as top predictors of *B. longum* relative abundance in the gut microbiome. A-C. Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (A), Adult (B) and Senior (C) cohorts. The profiling strategy (Seq-Type), and the Cohort-life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. longum* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. longum* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. longum* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.



Supplementary Figure 5. Random Forest (RF) model-based identification of distinct sets of non-Bifidobacterial taxa that consistently associated with the relative abundance of

***Bifidobacterium adolescentis*, serving as top predictors of *B. adolescentis* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. adolescentis* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. adolescentis* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. adolescentis* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

***Bifidobacterium pseudocatenulatum*, serving as top predictors of *B. pseudocatenulatum* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. pseudocatenulatum* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. pseudocatenulatum* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. pseudocatenulatum* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

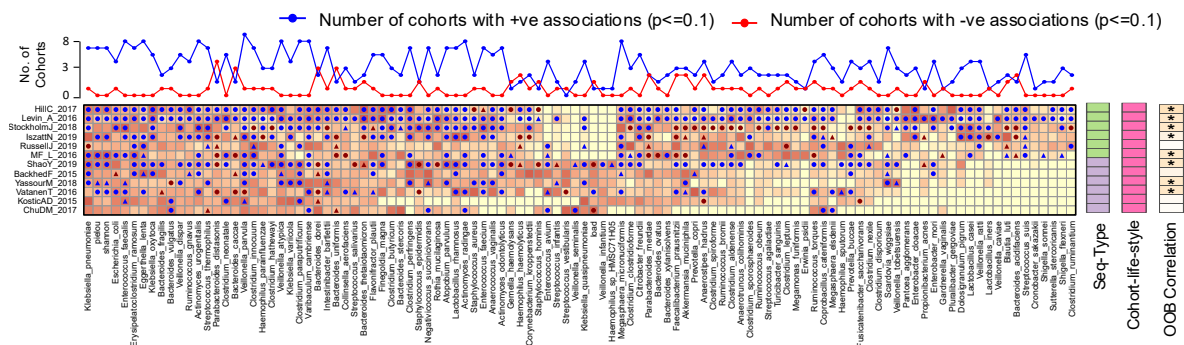


Supplementary Figure 7. Random Forest (RF) model-based identification of distinct sets of non-*Bifidobacterium* taxa that consistently associated with the relative abundance of

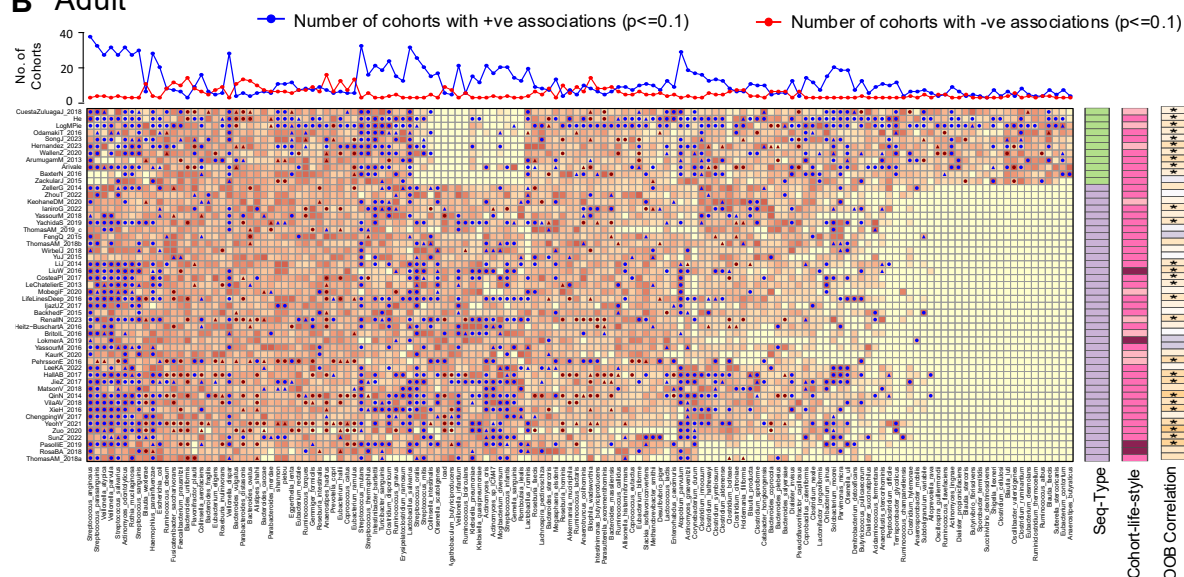
***Bifidobacterium bifidum*, serving as top predictors of *B. bifidum* relative abundance in the gut microbiome.** **A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. bifidum* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. bifidum* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. bifidum* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

A Infant

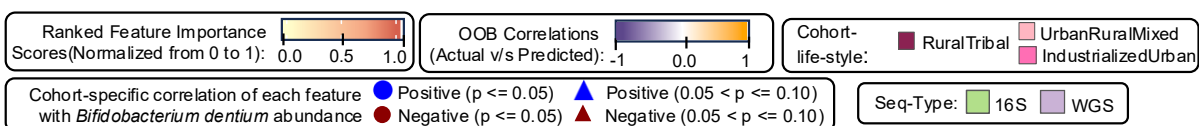
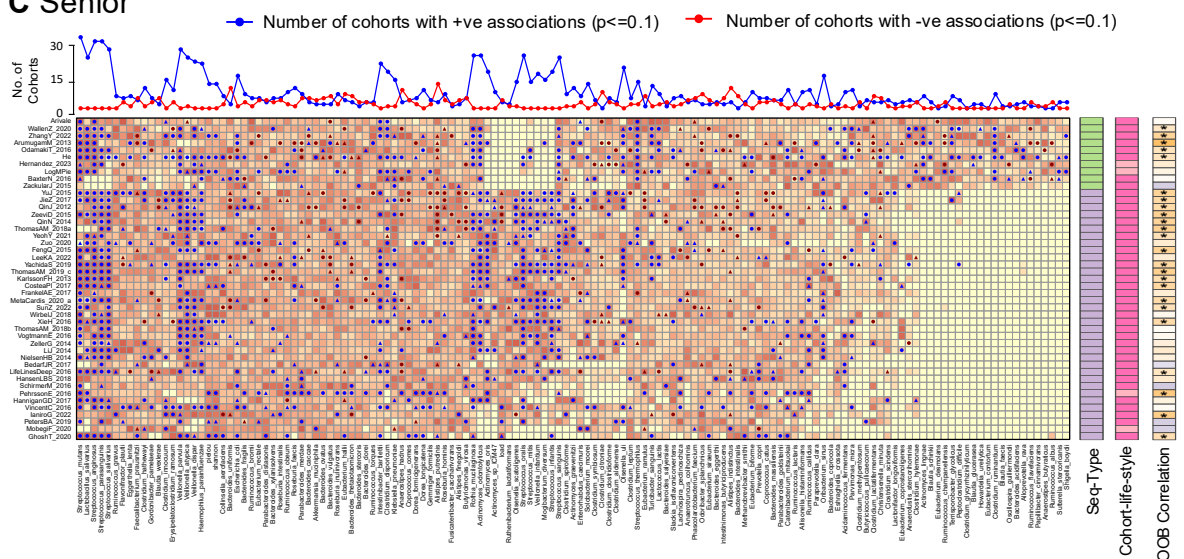
Bifidobacterium dentium



B Adult



C Senior

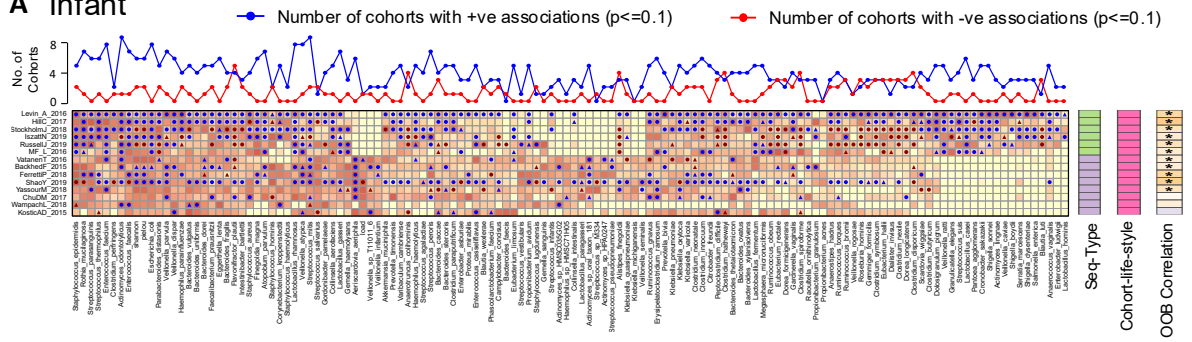


Supplementary Figure 8. Random Forest (RF) model-based identification of distinct sets of non-Bifidobacterial taxa that consistently associated with the relative abundance of

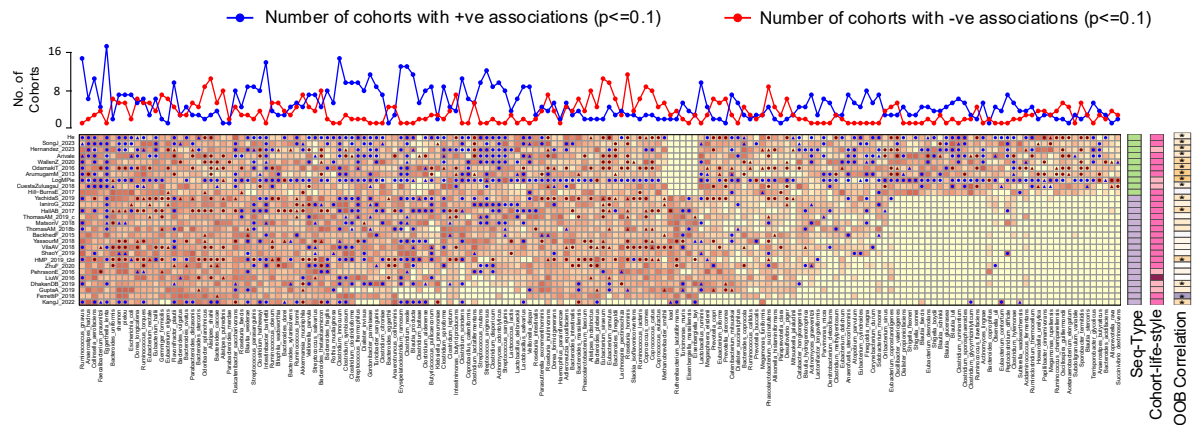
***Bifidobacterium dentium*, serving as top predictors of *B. dentium* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. dentium* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. dentium* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. dentium* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

Bifidobacterium breve

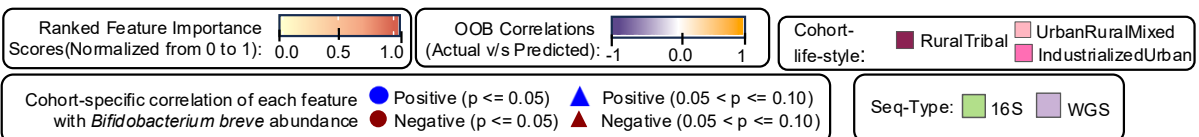
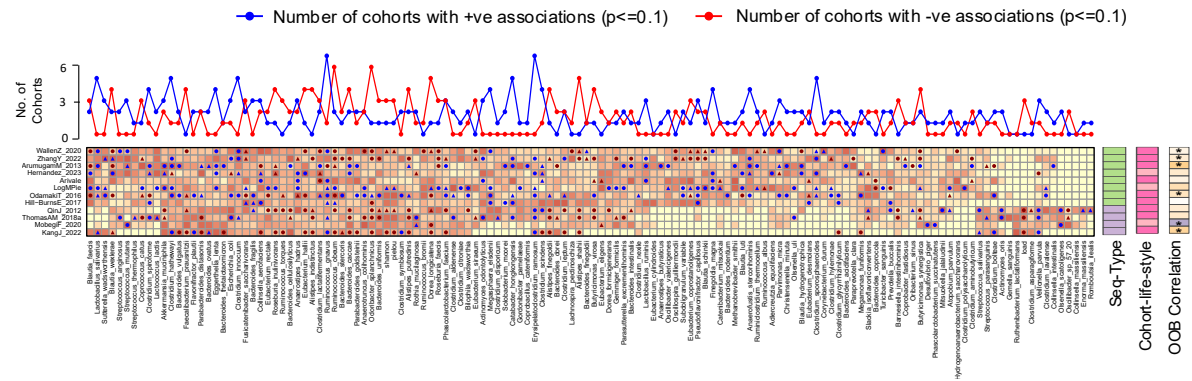
A Infant



B Adult



C Senior

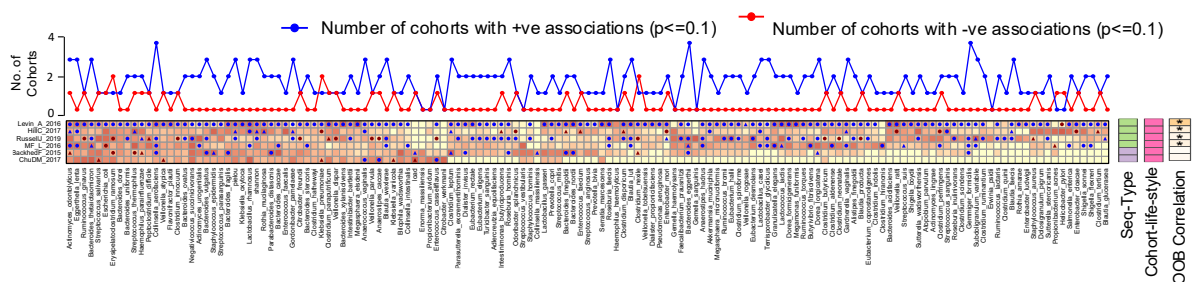


Supplementary Figure 9. Random Forest (RF) model-based identification of distinct sets of non-Bifidobacterial taxa that consistently associated with the relative abundance of

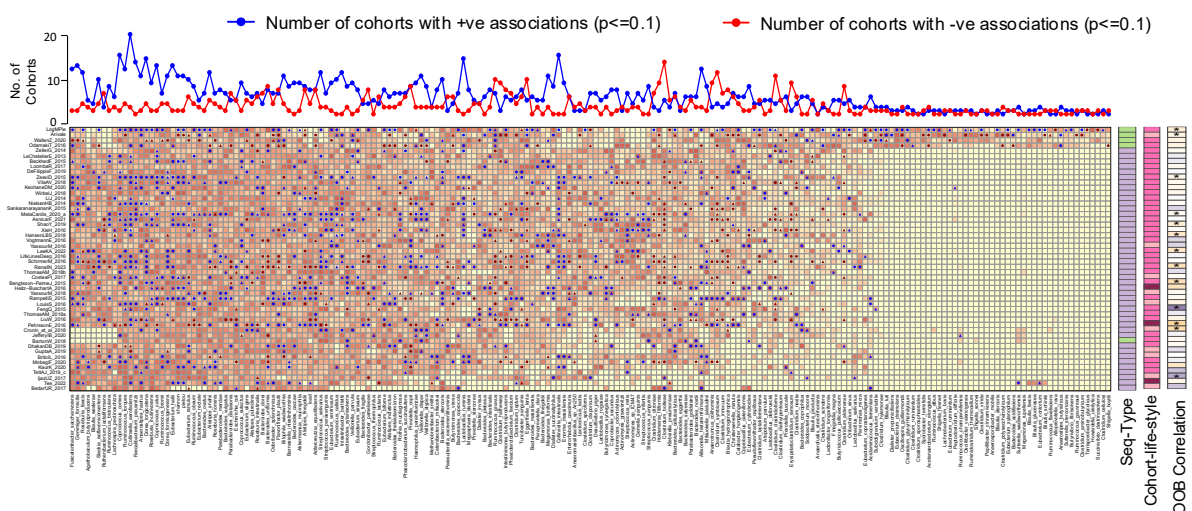
***Bifidobacterium breve*, serving as top predictors of *B. breve* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. breve* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. breve* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. breve* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

A Infant

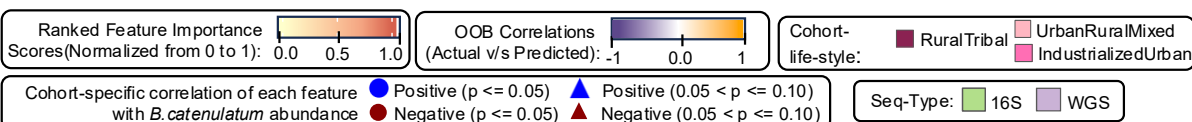
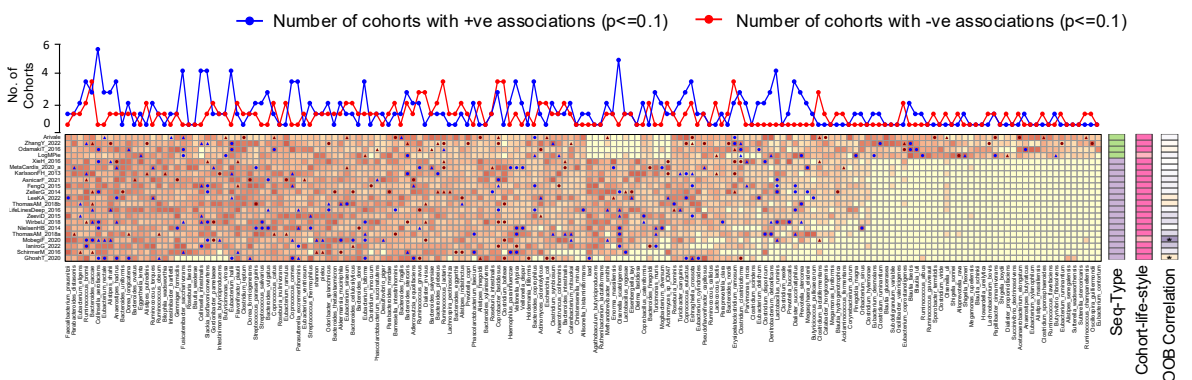
Bifidobacterium catenulatum



B Adult



C Senior

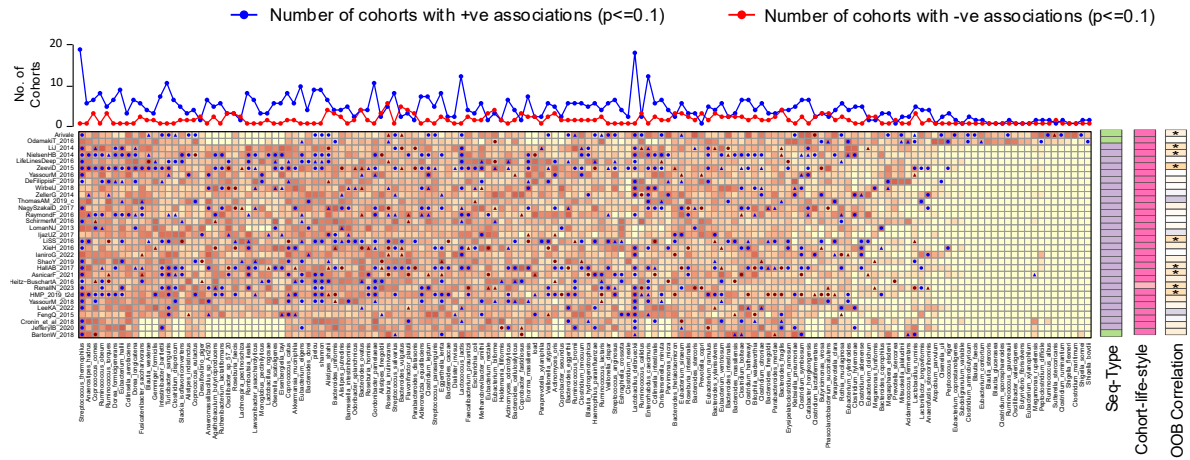


Supplementary Figure 10. Random Forest (RF) model-based identification of distinct sets of non-*Bifidobacterium* taxa that consistently associated with the relative abundance of

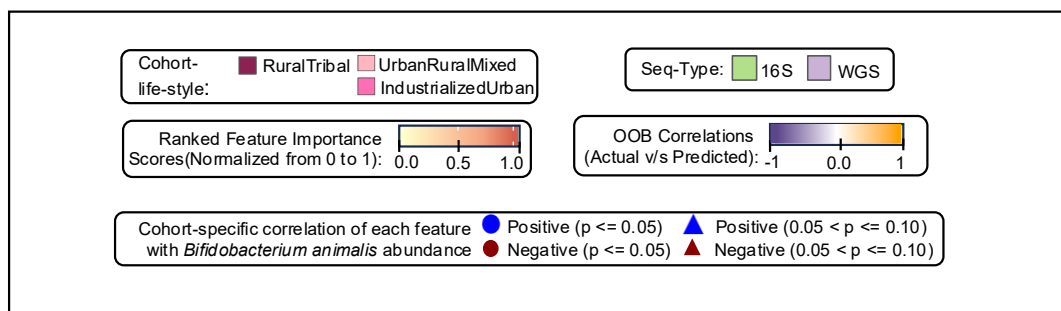
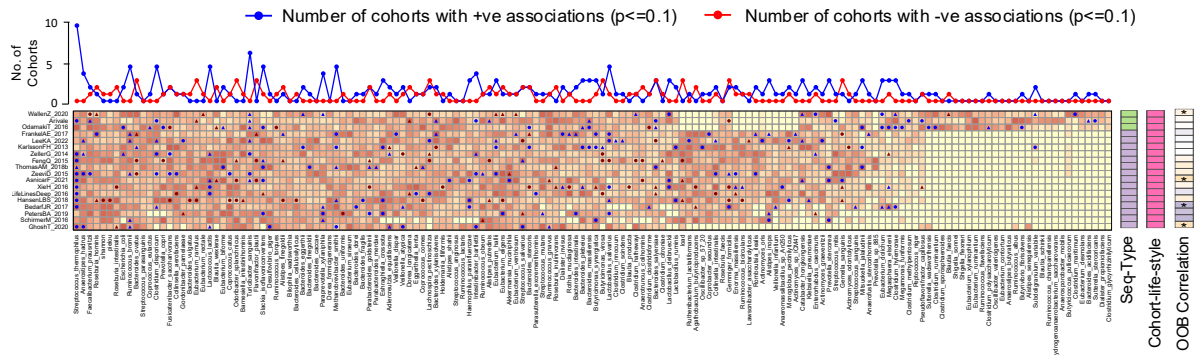
***Bifidobacterium catenulatum*, serving as top predictors of *B. catenulatum* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. catenulatum* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. catenulatum* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. catenulatum* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.

Bifidobacterium animalis

A Adult

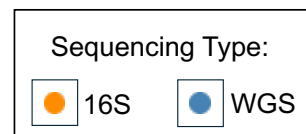
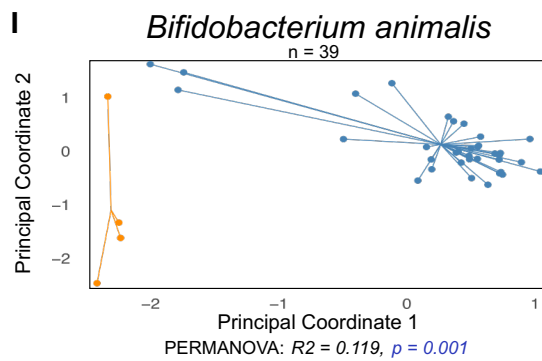
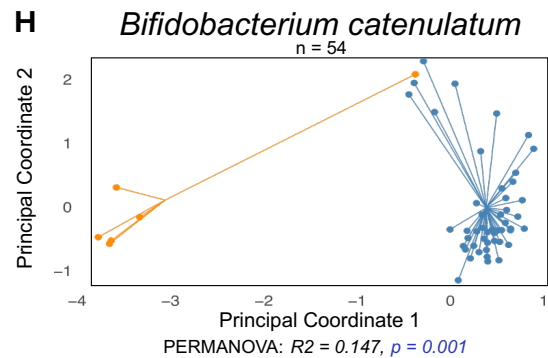
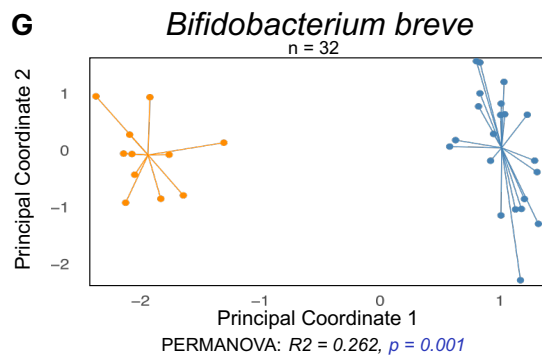
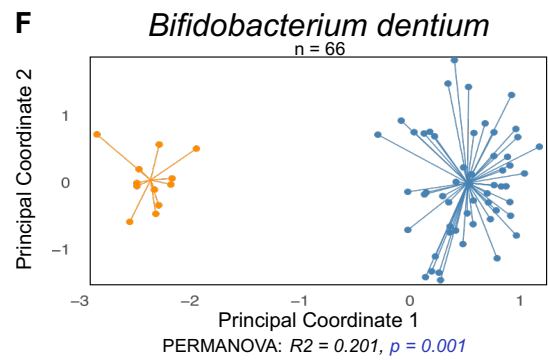
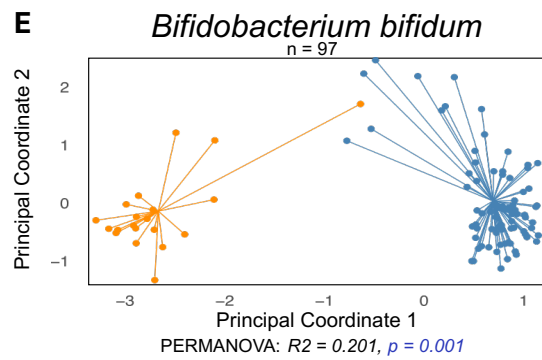
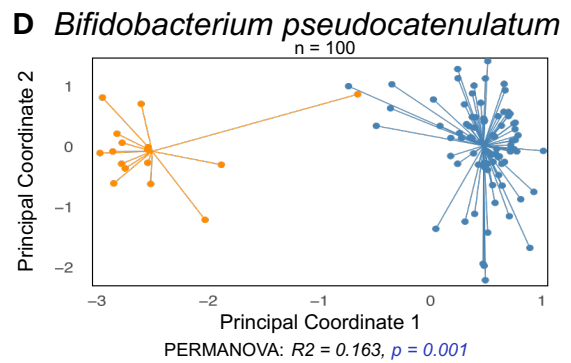
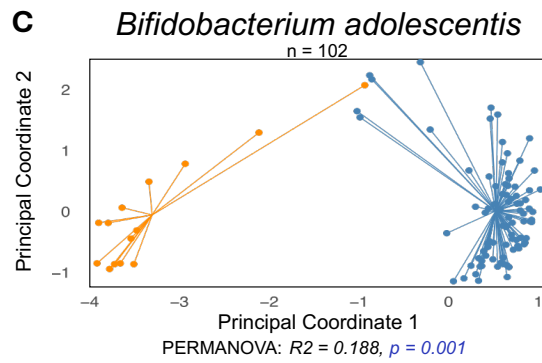
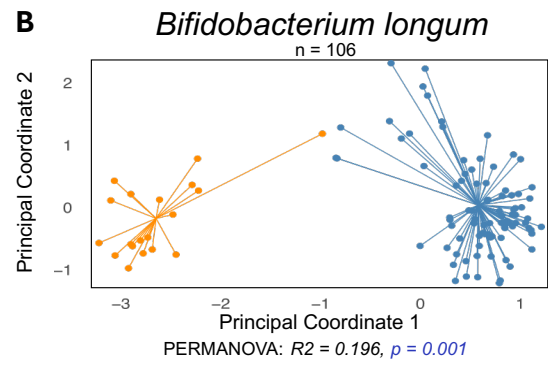
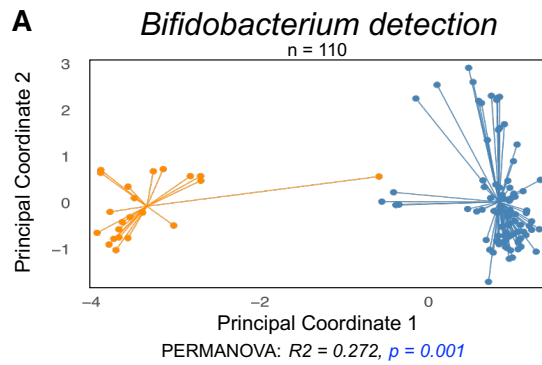


B Senior

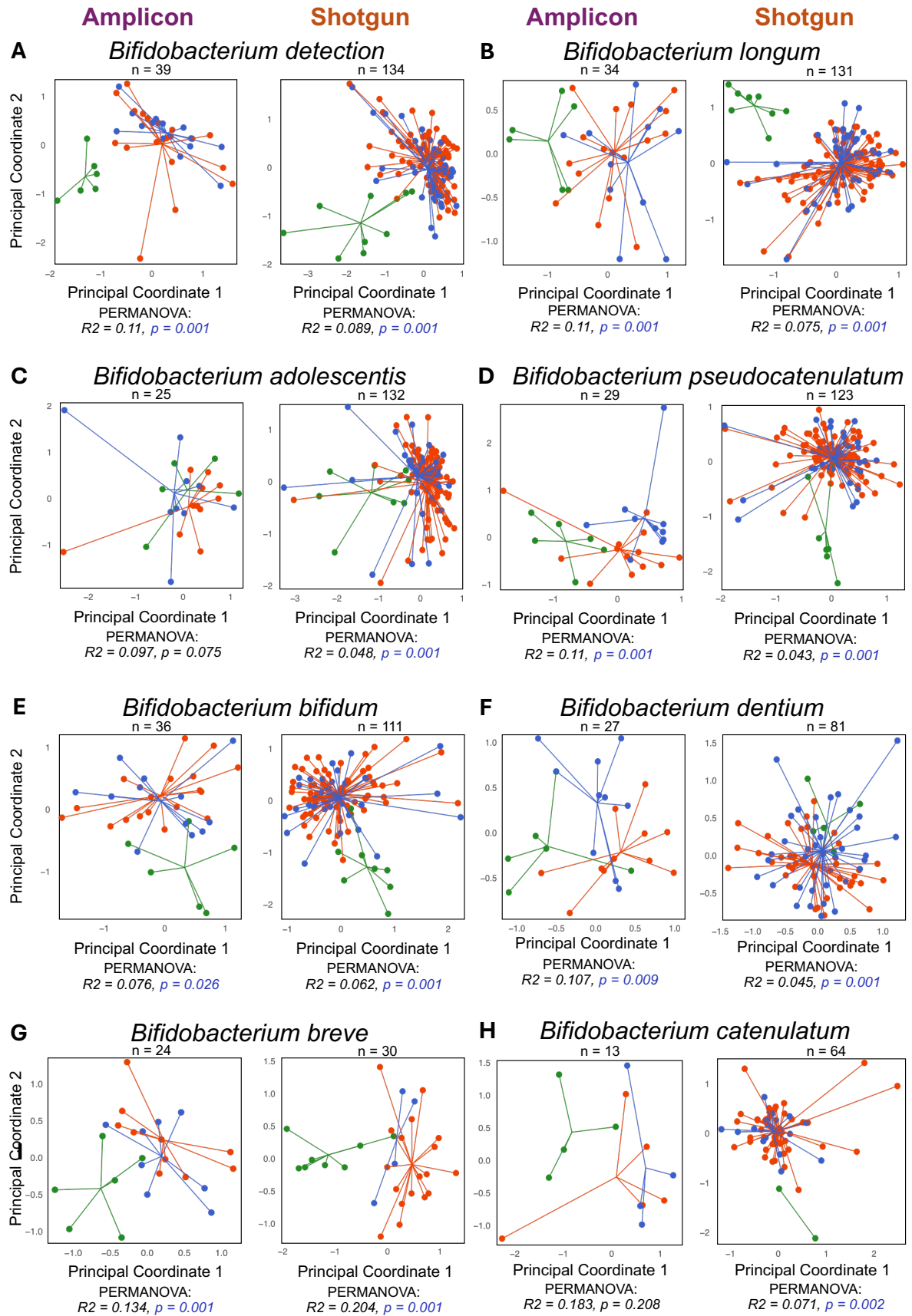


Supplementary Figure 11. Random Forest (RF) model-based identification of distinct sets of non-*Bifidobacterium* taxa that consistently associated with the relative abundance of

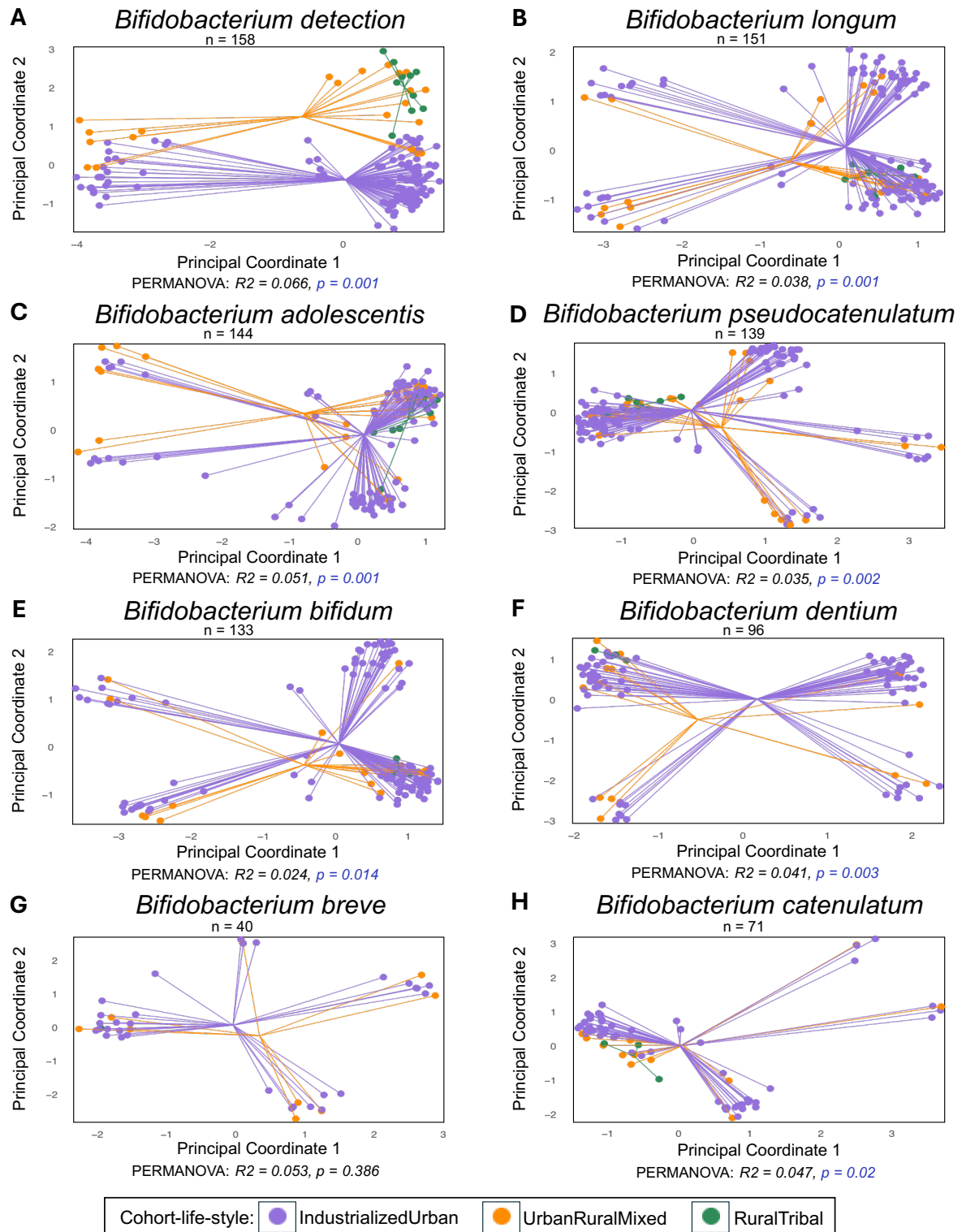
***Bifidobacterium animalis*, serving as top predictors of *B. animalis* relative abundance in the gut microbiome. A-C.** Heatmap showing the variation of the rank-normalized feature importance scores of the top predictor taxa as well as three microbiome properties- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for the Random Forest models built for Infant (**A**), Adult (**B**) and Senior (**C**) cohorts. The profiling strategy (Seq-Type), and the Cohort life-style of each cohort is indicated as strips as shown for each heatmap. Also indicated are the out-of-bag (OOB) correlations between the actual and RF-predicted *B. animalis* for each cohort. Symbols within the heatmap cells indicate both the direction (positive or negative) and strength (based on p values) of the association between each feature (non-Bifidobacterial taxon and microbiome properties as shown in columns) and *B. animalis* relative abundance within each cohort (rows). Line plots above each heatmap summarize these associations and show the number of cohorts where each feature displays significant positive or negative associations with *B. animalis* relative abundance ($p \leq 0.1$). The extent and consistency of these associations varied by specific features, with some showing notably higher predictive power and consistent directionality. Association patterns were also cohort-type-dependent, namely Infant cohorts exhibited distinct sets of associated taxa compared to other age groups (same for other age-groups); 16S and WGS cohorts showed differing patterns. We also observed variations depending upon the life-style of cohorts. These patterns were captured in the Association-Scores as described in **Figure 3**. Feature-wise associations were assessed using Spearman rank correlation (two-sided), with p values indicating strength of association. No multiple testing correction was applied, as analyses were conducted independently within each cohort.



Supplementary Figure 12. Principal Coordinate Analysis (PCoA) results showing the variation of the ranked Feature Importance Vectors for the cohort-specific Random Forest models for the different cohorts based on their profiling strategy (Seq-Type; i.e. 16S or WGS). The Feature Importance vectors contain the ranked Feature Importance scores for the different non-Bifidobacterial taxa and microbiome properties (Shannon Index, Pielou's-Evenness Index and estimated Microbial Load) obtained for the Random Forest model generated for that cohort for the prediction of **A.** Overall *Bifidobacterium* detection **B.** *B. longum* **C.** *B. adolescentis* **D.** *B. pseudocatenulatum* **E.** *B. bifidum* **F.** *B. dentium* **G.** *B. breve* **H.** *B. catenulatum* and **I.** *B. animalis*, using adult/senior studies. Each point represents a study-cohort and is colored according to sequencing-type as indicated by the color scheme used in the figure, and connected with lines to their group centroid. PERMANOVA results (shown using both the R² and p value) shows statistically significant separation for all taxa. The number of study-cohorts (n) represented in each panel is indicated within the respective plots.

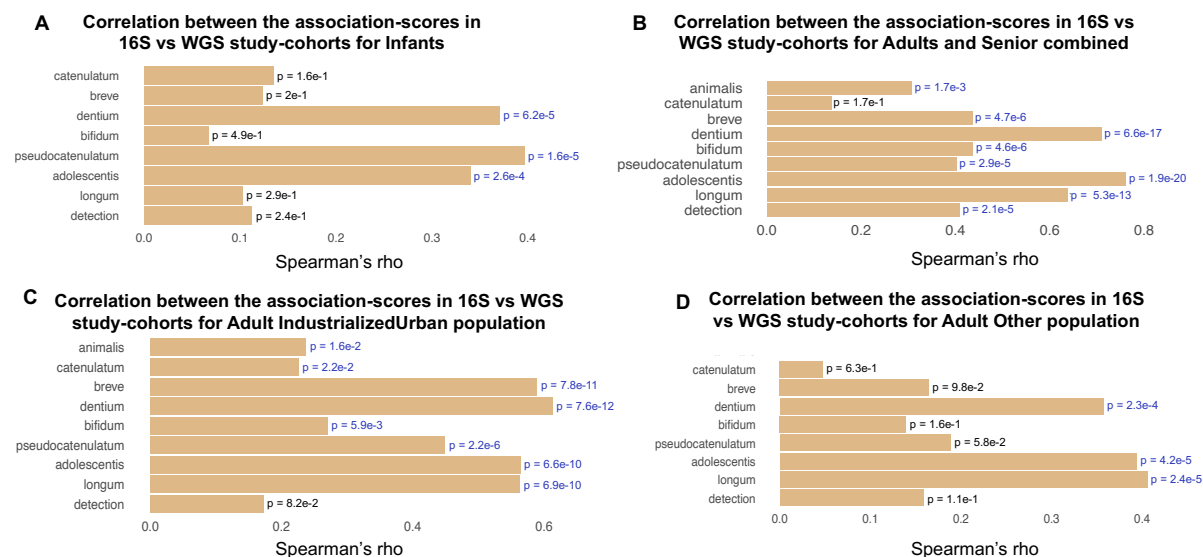


Supplementary Figure 13. Principal Coordinate Analysis (PCoA) results showing the variation of the ranked Feature Importance Vectors for the cohort-specific Random Forest models for the different cohorts based on their age-group (i.e. Infant, Adult and Senior). The Feature Importance vectors contain the ranked Feature Importance scores for the different non-Bifidobacterial taxa and microbiome properties (Shannon Index, Pielou's-Evenness Index and estimated Microbial Load) obtained for the Random Forest model generated for that cohort for the prediction of **A.** Overall *Bifidobacterium* detection **B.** *B. longum* **C.** *B. adolescentis* **D.** *B. pseudocatenulatum* **E.** *B. bifidum* **F.** *B. dentium* **G.** *B. breve* and **H.** *B. catenulatum*. Each point represents a study-cohort and is colored according to age-group as indicated by the color scheme used in the figure, and connected with lines to their group centroid. PERMANOVA results (shown using both the R² and p value) shows statistically significant separation for all taxa except *B. catenulatum*, in WGS data and *B. adolescentis*, in 16S data. The number of study-cohorts (n) represented in each panel is indicated within the respective plots.

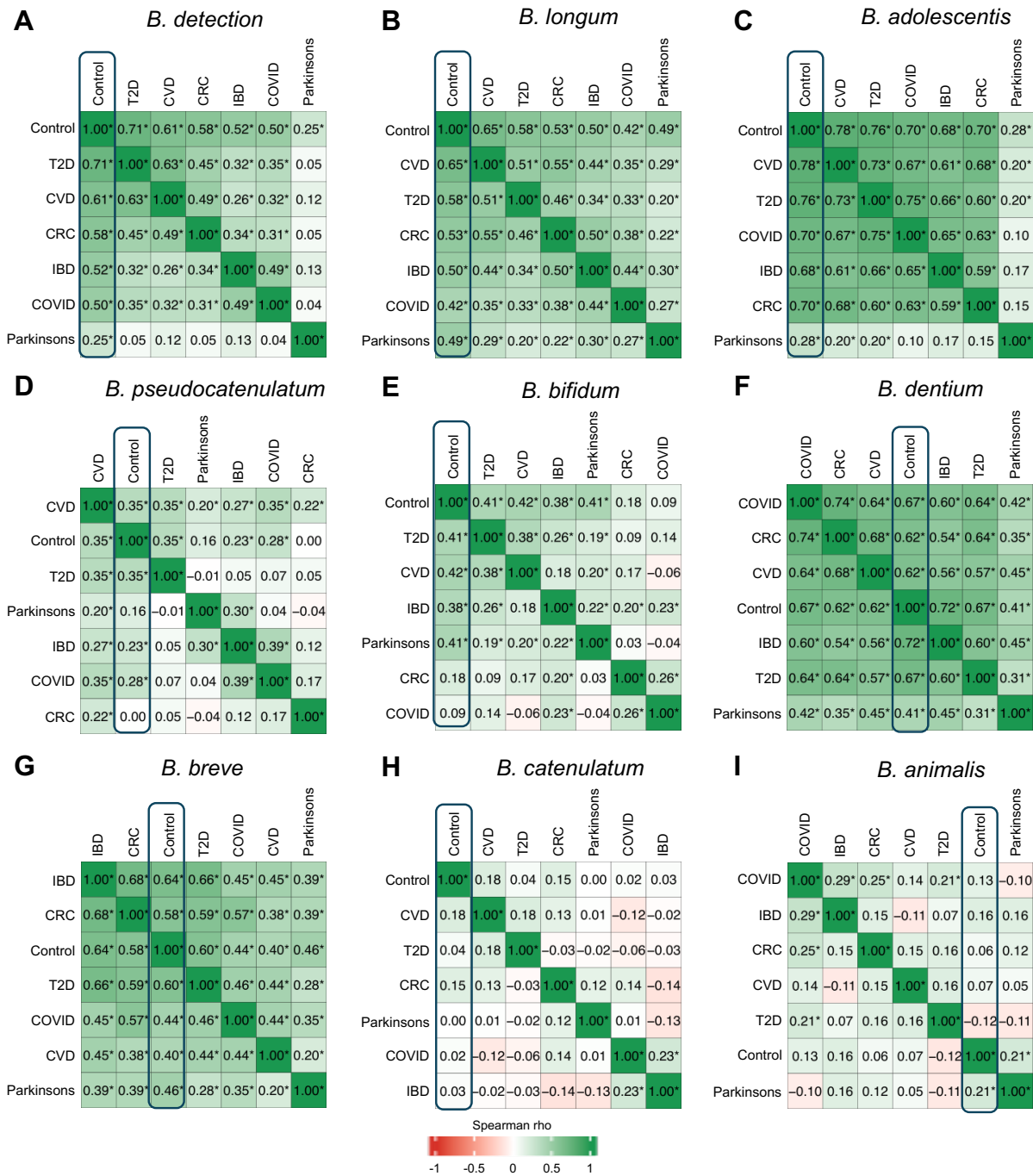


Supplementary Figure 14. Principal Coordinate Analysis (PCoA) results showing the variation of the ranked Feature Importance Vectors for the cohort-specific Random Forest models for the different cohorts based on their life-style associated Cohort-Type. This analysis was restricted to adult and senior cohorts due to the absence of Non-IndustrializedUrban cohorts (i.e., UrbanRuralMixed and RuralTribal) in the infant age-group.

Feature Importance vectors contain the ranked Feature Importance scores for the different non-Bifidobacterial taxa and microbiome properties (Shannon Index, Pielou's-Evenness Index and estimated Microbial Load) obtained for the Random Forest model generated for that cohort for the prediction of **A.** Overall *Bifidobacterium* detection **B.** *B. longum* **C.** *B. adolescentis* **D.** *B. pseudocatenulatum* **E.** *B. bifidum* **F.** *B. dentium* **G.** *B. breve* and **H.** *B. catenulatum*. Each point represents a study-cohort and is colored according to Cohort-life-style as indicated by the color scheme used in the figure, and connected with lines to their group centroid. PERMANOVA results (shown using both the R² and p value) shows statistically significant separation for all taxa except *B. breve*. The number of study-cohorts (n) represented in each panel is indicated within the respective plots.

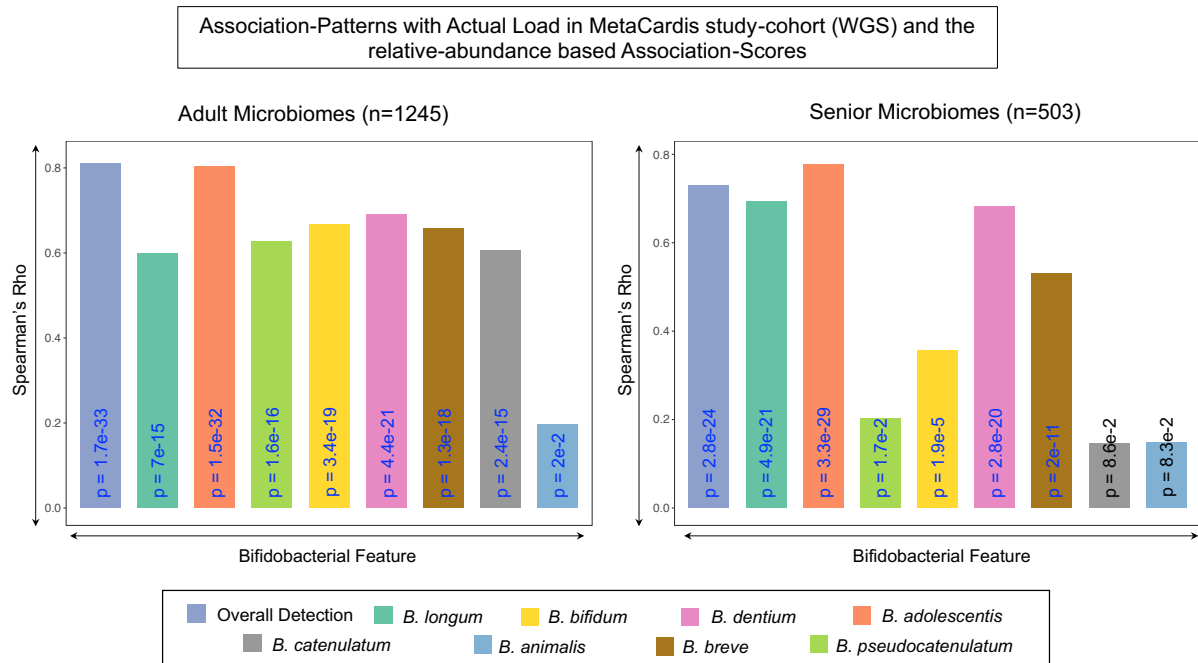


Supplementary Figure 15. Correlation between Association-Scores (obtained for different non-Bifidobacterial taxa with respect to various Bifidobacterial features) derived from 16S and WGS data in (A) Infant cohort-groups (B) Adult & Senior cohort-groups (C) IndustrializedUrban cohort groups and (D) Other non IndustrializedUrban cohort groups (UrbanRuralMixed + RuralTribal). Associations among overall *Bifidobacterium* detection and different Bifidobacterial taxa using the Association-Scores with non-Bifidobacterial taxa in 16S vs. WGS study-cohorts. Spearman's correlation coefficient has been shown in the form of bar plots for Bifidobacterial features individually between 16S and WGS profiling-strategy-based cohorts having microbiomes from infants (A), adults and senior combined (B), microbiomes from IndustrializedUrban population (C) and in other populations including UrbanRuralMixed and RuralTribal combined (D) using the Association-Scores with non-Bifidobacterial taxa. Spearman rank correlation (two-sided) was used, and all exact p values are displayed above each horizontal bar; statistically significant associations ($p \leq 0.05$) are highlighted in blue, while non-significant ones are shown in black. Some Bifidobacteria (*B. catenulatum* and *B. animalis* in infants; *B. breve* and *B. animalis* in other non IndustrializedUrban cohorts) don't show consistent association pattern with the non-Bifidobacterial taxa across either profiling strategy, hence, they have not been included.

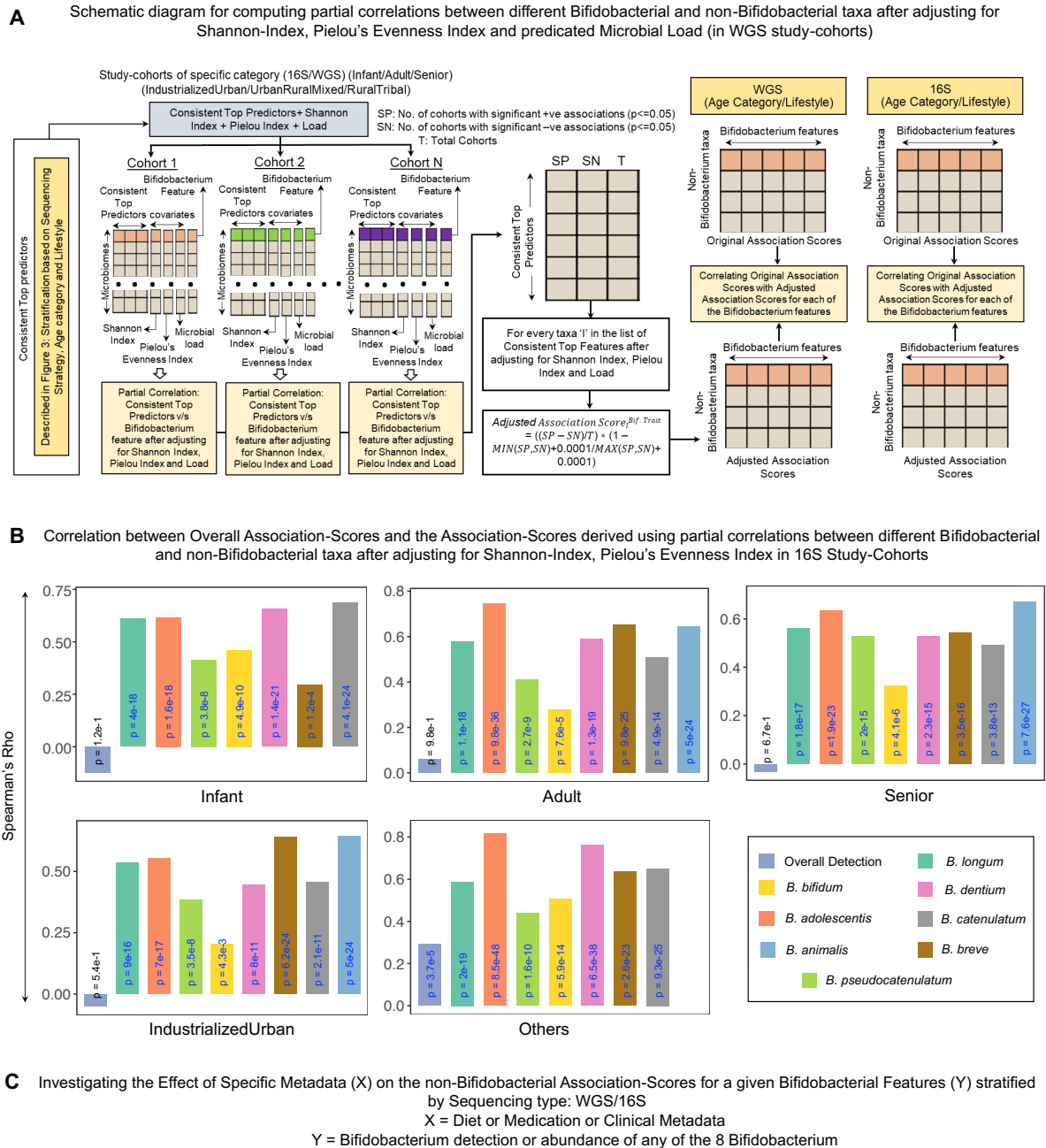


Supplementary Figure 16. Correlations between the Association-Scores obtained for the different non-Bifidobacterial taxa in the study-sub-cohorts for the six diseases and their controls with respect to the nine *Bifidobacterium* features, namely **A.** *Bifidobacterium* detection **B.** *B. longum* **C.** *B. adolescentis* **D.** *B. pseudocatenulatum* **E.** *B. bifidum* **F.** *B. dentium* **G.** *B. breve* **H.** *B. catenulatum* and **I.** *B. animalis*. Association-Scores of non-*Bifidobacterium* taxa with *Bifidobacterium* features were computed separately for apparently non-diseased sub-cohorts (disease-matched controls) and in the study-sub-cohorts containing microbiomes from six diseases. For each Bifidobacterial features, for each pair of

groups (7 conditions in total: 6 diseases + control), the spearman rank correlation (two-sided) coefficients were calculated between the Association-Scores obtained for the different non-Bifidobacterial taxa. The Heatmaps displays these pairwise correlation values, with positive correlations in green and negative correlations in red, scaled between -1 and 1 . Asterisks (*) denote statistically significant correlation ($p \leq 0.05$); the exact Spearman correlation coefficients (ρ) and corresponding p values for all comparisons are provided in **Supplementary Data 14**. The columns showing the Association-Scores between Controls and the different disease-groups are highlighted in boxes. Association-Scores for overall *Bifidobacterium* detection, and *B. longum*, *B. adolescentis*, *B. breve* and *B. dentium*, showed strong concordance across controls and disease-groups. For *B. bifidum* and *B. pseudocatenulatum*, while the Association-Scores across diseases varied, there was reasonable concordance between controls and the majority of diseases.

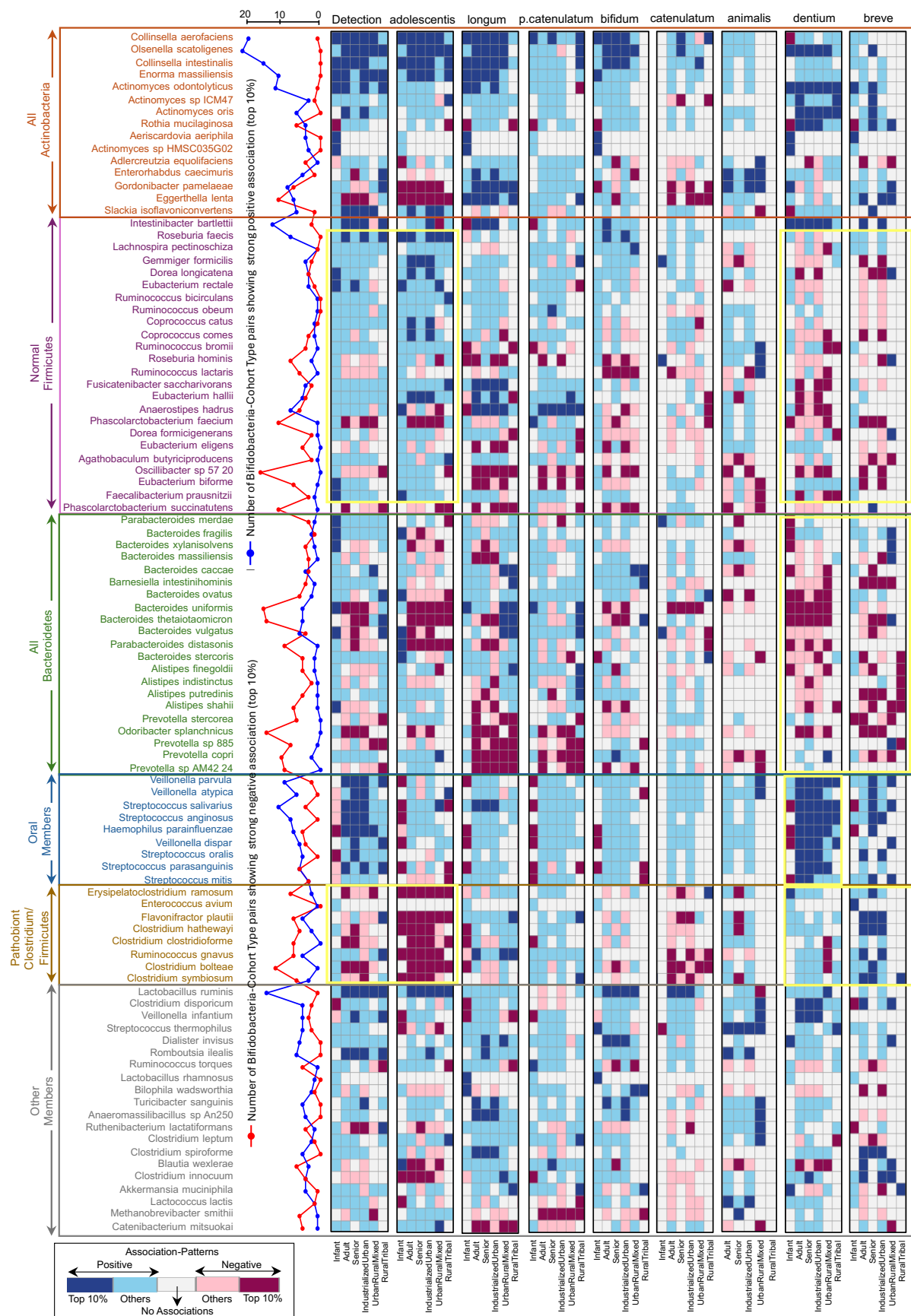


Supplementary Figure 17. Associations between quantitative-microbial-profiling (QMP) derived microbial-load-correlations between different non-Bifidobacterial taxa and the *Bifidobacteria* in the MetaCardis cohort and the relative abundance-based Association-Scores of the same taxa-pairs (computed for the WGS-sequenced Adult and Senior IndustrializedUrban cohorts). Bar plots show Spearman's rank correlation coefficients (ρ) between experimentally measured Microbial Load and relative abundance-based association-scores for overall *Bifidobacterium* detection and individual *Bifidobacterium* taxa in adult (left; $n = 1,245$) and senior (right; $n = 503$) gut microbiomes derived from whole-genome shotgun (WGS) sequencing. Each bar represents the correlation strength for a specific *Bifidobacterium* feature, with each taxon distinguished by a unique color as indicated in the figure. Spearman rank correlation (two-sided) was used, and all exact 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. There was significant positive correlation between QMP derived associations and relative abundance derived Association-Scores for all nine *Bifidobacterium* properties in adults and seven out of the nine properties in seniors.

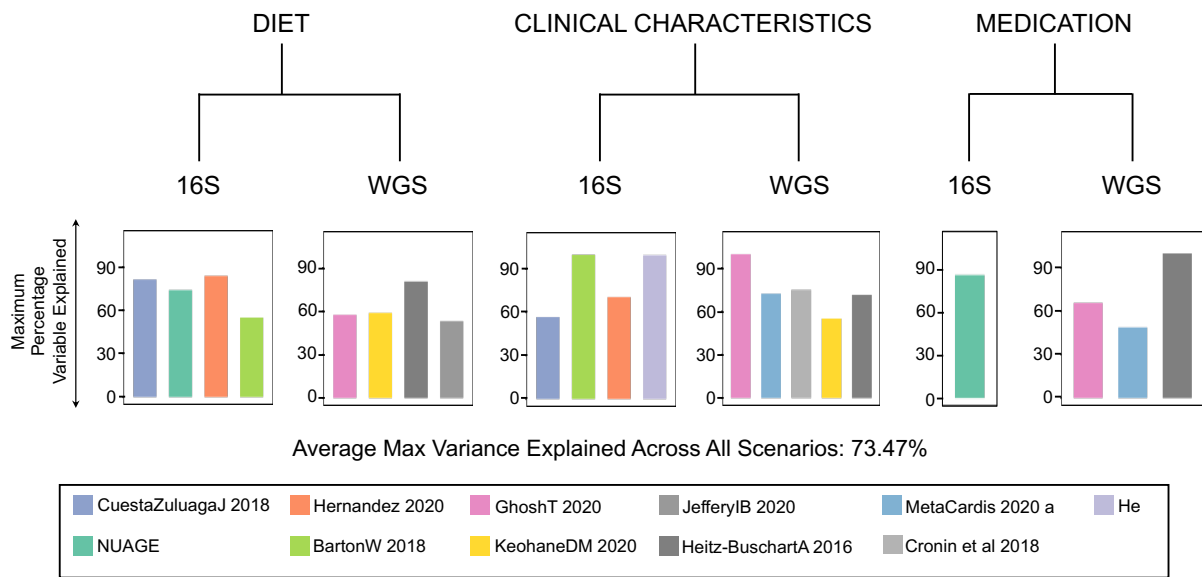


Supplementary Figure 18. Investigation of Association patterns between non-*Bifidobacterium* taxa and different *Bifidobacteria* after adjustment for microbiome/host-

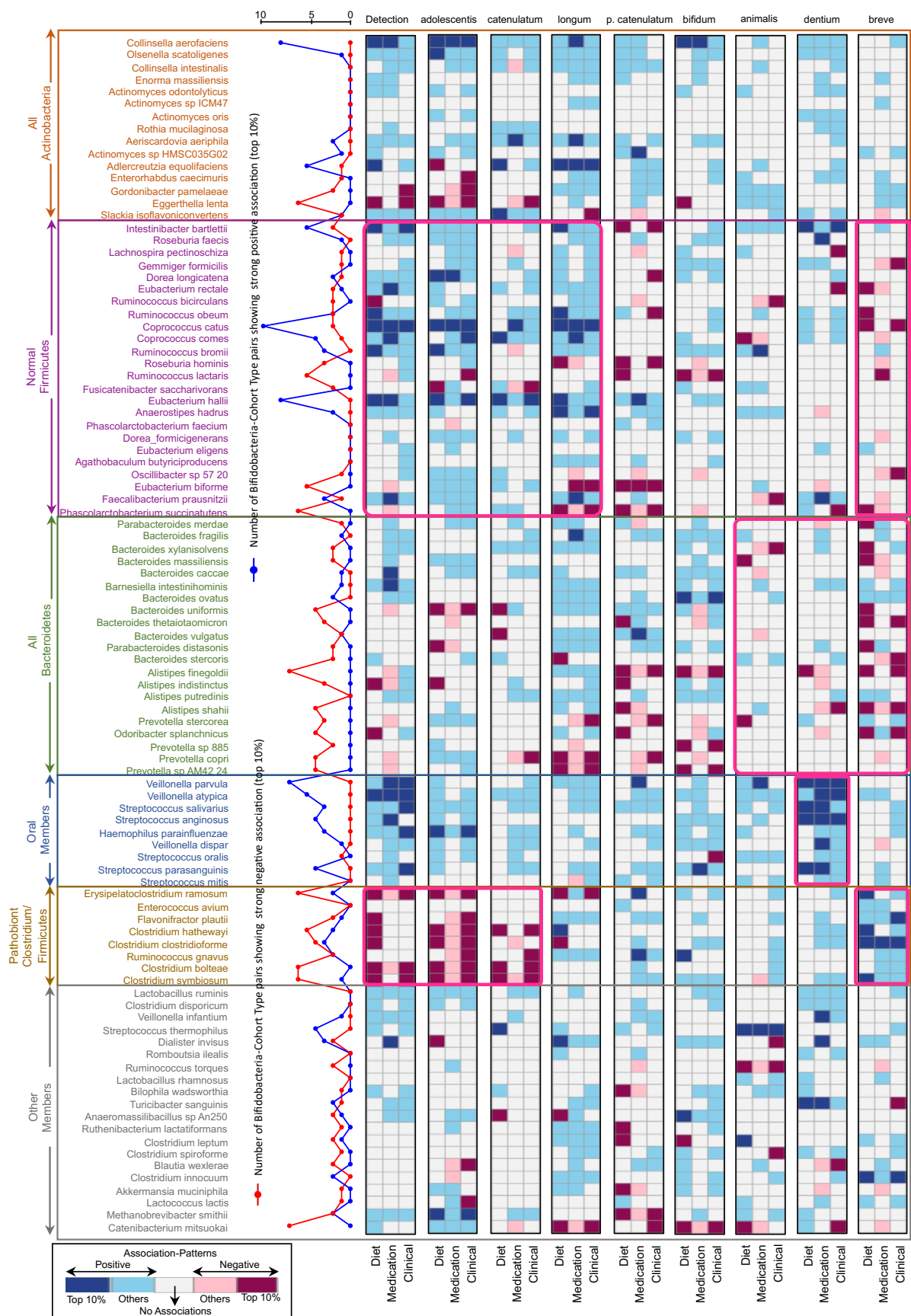
associated covariates. **A.** Schematic illustration of the workflow used to compute partial correlations between *Bifidobacterium* and non-Bifidobacterial taxa after adjustment for microbiome properties like- Shannon Index, Pielou's-Evenness Index, and estimated Microbial Load (in WGS cohorts). **B.** Correlation bar plots showing the relationship between overall Association-Scores and Association-Scores derived using partial correlations between different Bifidobacterial features and non-Bifidobacterial taxa after adjusting for the three microbiome properties described above in 16S study-cohorts. Correlations were assessed using Spearman rank correlation (two-sided). Each bar represents the correlation strength for a specific *Bifidobacterium* taxon, with each taxon distinguished by a unique color as indicated in the figure. All exact 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. No adjustment for multiple comparisons was applied, as correlations were evaluated independently for each feature. **C.** Schematic overview of the analytical workflow used to estimate partial correlations between *Bifidobacterium* and non-Bifidobacterial taxa, adjusting for diet, medication, and clinical covariates, and stratified by sequencing-type (16S and WGS).



Supplementary Figure 19. Association patterns of the non-Bifidobacterial taxa with the different Bifidobacterial features captured as adjusted Association-Scores after adjusting for three key microbiome properties namely- Shannon Index, Pielou's-Evenness Index and estimated Microbial Load in WGS study-cohorts stratified by age-group (Infant/Adult/Senior), and life-style category (IndustrializedUrban/UrbanRuralMixed/RuralTribal). Heatmap showing patterns of Association-Score computed after adjusting for Shannon Index, Pielou's-Evenness Index and estimated Microbial Load for different non-Bifidobacterial groups with various Bifidobacterial features across cohorts of different categories using WGS study-cohorts. Association patterns are shown as graded Association-Scores based on the consistent patterns (refer to **Supplementary Note 2**). In summary, the top-10% positive and negative Association-Scores are shown in deep blue and deep pink colours, respectively. Other positive and negative Association-Scores are shown in sky blue and light pink, respectively. The complete graded Association-Scores list for WGS and 16S cohorts are provided in **Supplementary Data 15**. The line plot shows the number of Bifidobacteria-cohort type pairs showing strong association (top 10% Association-Scores across a particular cohort type), where blue line indicates strong positive associations and red line indicates strong negative associations. Non-Bifidobacterial taxa are coloured depending upon the groupings based on mostly their taxonomic lineage as well as their associations with human health.



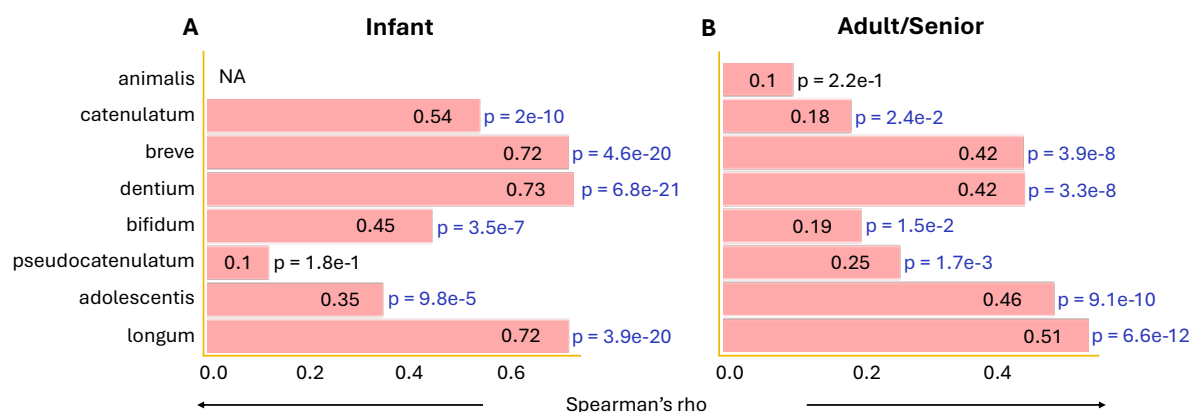
Supplementary Figure 20. Study-wise maximum variance explained across adjusted diet/clinical/medication metadata categories. Bar plots summarizing the maximum variance explained performed independently for each study across six scenarios: Diet-16S, Diet-WGS, Clinical Characteristics-16S, Clinical Characteristics-WGS, Medication-16S, Medication-WGS. Within each study, PCoA and robust PCA (rPCA) were conducted using category-specific adjusted metadata association profiles, and the maximum variance explained by the first three components was extracted. These study-wise components were subsequently used for partial correlation analysis (see **Supplementary Figure 18C**; **Supplementary Note 3**). Bars correspond to individual studies, as indicated by the color scheme used in the figure. The average maximum variance explained across all scenarios was 73.47%.



Supplementary Figure 21. Association patterns of the non-Bifidobacterial taxa with the different Bifidobacterial features captured as adjusted Association-Scores after adjusting

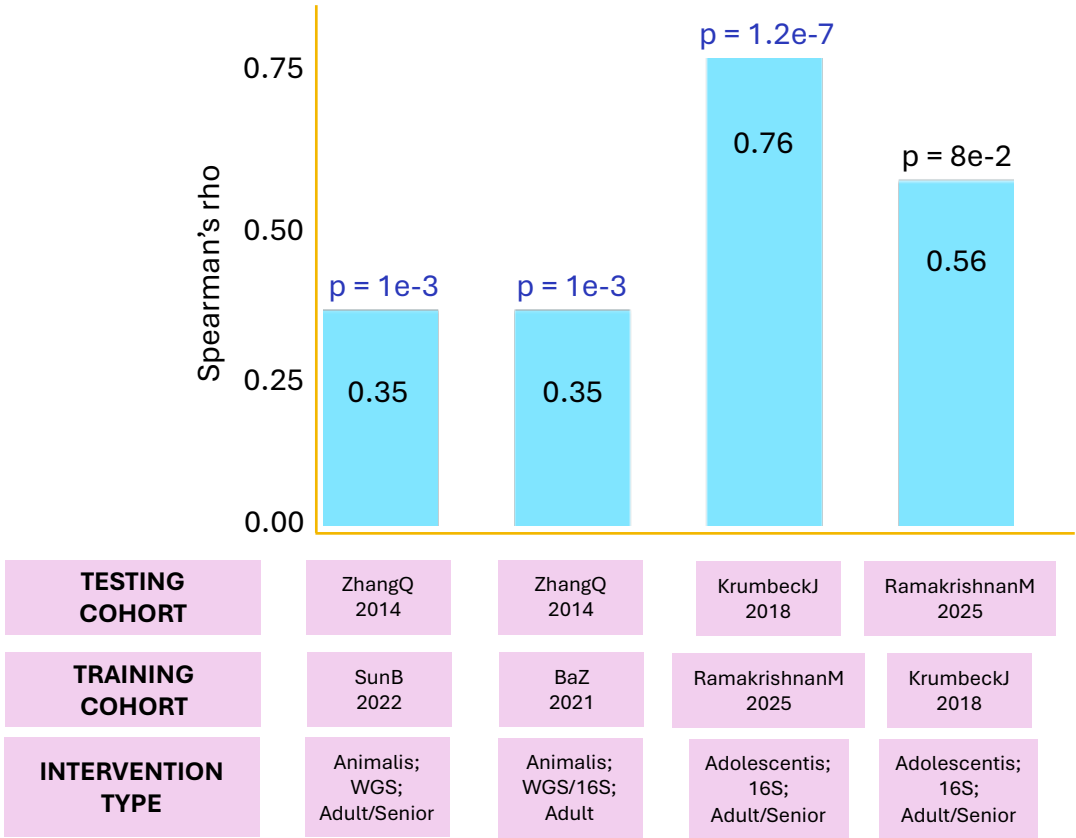
for Diet, Medication and Clinical parameters in WGS study-cohorts. Heatmap showing patterns of Association-Score computed after adjusting for Diet, Medication and Clinical parameters for different non-Bifidobacterial groups with various Bifidobacterial features across cohorts of different categories using WGS study-cohorts. Association patterns are shown as graded Association-Scores based on the consistent patterns (refer to **Supplementary Note 2**). The complete graded Association-Scores list for WGS and 16S cohorts are provided in **Supplementary Data 17**. The line plot shows the number of Bifidobacteria-cohort type pairs showing strong association (top 10% Association-Scores across a particular cohort type), where blue line indicates strong positive associations and red line indicates strong negative associations. Non-Bifidobacterial taxa are coloured depending upon the groupings based on mostly their taxonomic lineage as well as their associations with human health.

Bar plots showing Spearman correlations between individual non-*Bifidobacterium* taxa and the eight major *Bifidobacterium* species across discovery and validation 16S-profiled IndustrializedUrban cohorts



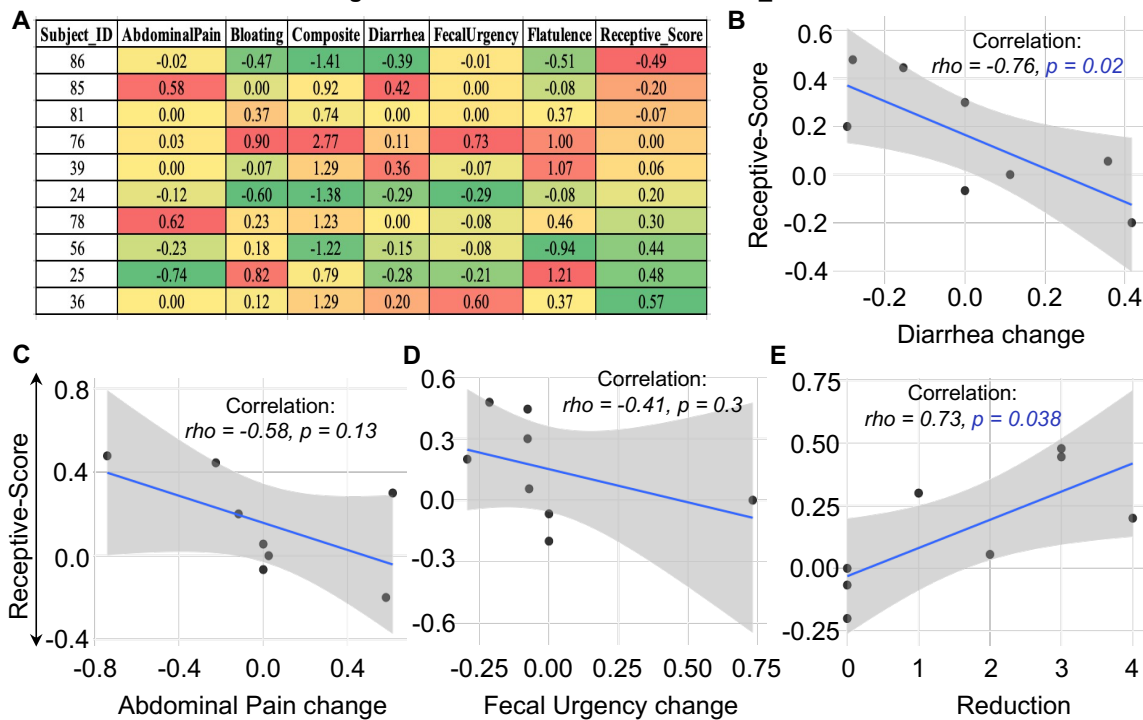
Supplementary Figure 22. Cross-cohort Spearman correlations of *Bifidobacterium* association patterns in IndustrializedUrban 16S datasets. A. Bar-plots showing the Spearman rank correlations (two-sided) between the Association-Scores of different non-*Bifidobacterium* taxa with the abundances of seven *Bifidobacteria* (in the infant-IndustrializedUrban-16S cohorts of the discovery datasets) and their abundance correlations with the corresponding *Bifidobacteria* in the infant/new-born-associated Bazanella *et al* cohort validation dataset. B. Bar-plots showing the Spearman rank correlations (two-sided) between the Association-Scores of the different non-*Bifidobacterium* taxa with the eight major *Bifidobacteria* computed for the 16S-profiled Adult/Senior cohorts with IndustrializedUrban population-type in the discovery dataset with the same Association-Scores re-computed within the cohorts of the same microbiome-type in the validation dataset. Each bar represents the correlation strength for a specific *Bifidobacterium* taxon. All exact p values are displayed above each horizontal bar; statistically significant associations ($p \leq 0.05$) are highlighted in blue, while non-significant ones are shown in black.

Cross-Cohort Validation of Receptive-Scores Across Independent Studies

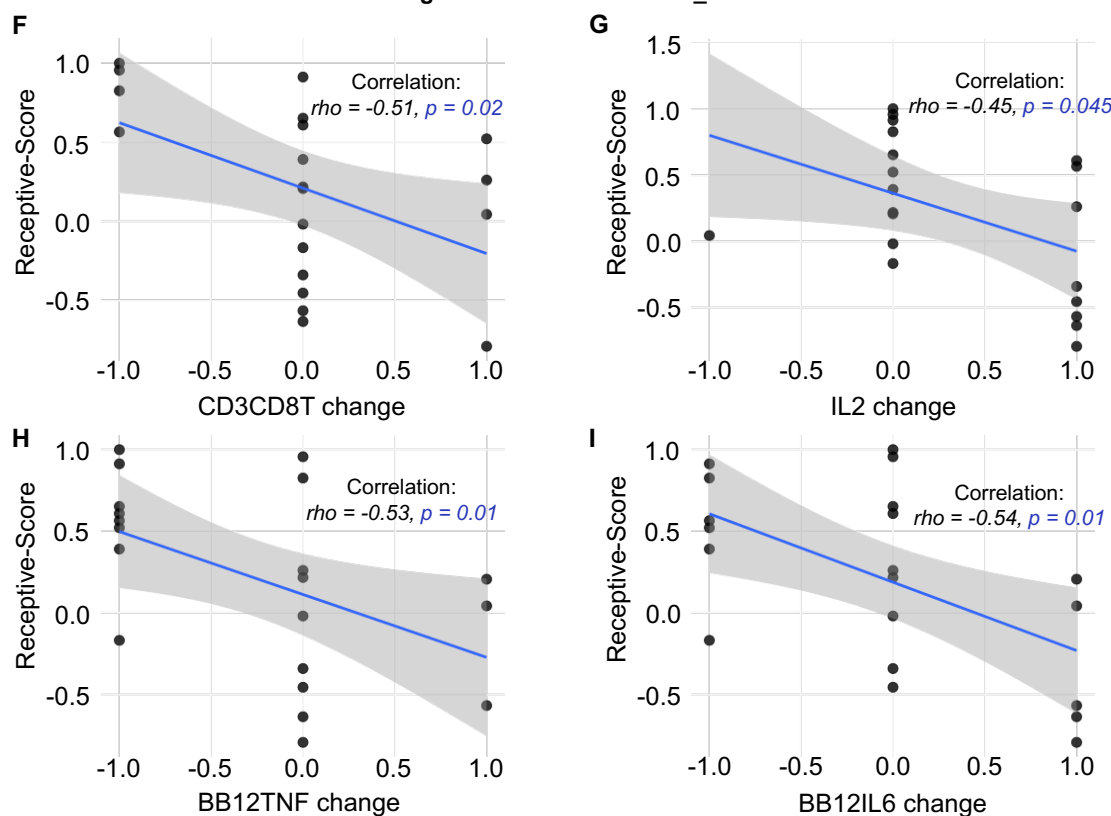


Supplementary Figure 23. Bar plots showing Cross-cohort validation of *B. animalis* and *B. adolescentis* Receptive-Scores and baseline abundance models using *ZhangQ*, *KrumbeckJ* and *RamakrishnanM* as the testing cohorts. Validation of predictive models for post-treatment increase in *B. animalis* and *B. adolescentis* using Receptive-Scores and baseline abundance derived from independent cohorts. Bar-plots showing the Spearman rank correlations (two-sided) between predicted and observed increase *B. animalis* and *B. adolescentis* across trials. Each bar represents the correlation strength and all exact p values are displayed above each bar; statistically significant associations ($p \leq 0.05$) are highlighted in blue, while non-significant ones are shown in black.

Association Between Clinical Symptom Changes (T3 -T1) and *B. adolescentis* Receptive -Score following intervention in the RamakrishnanM_2025 Trial

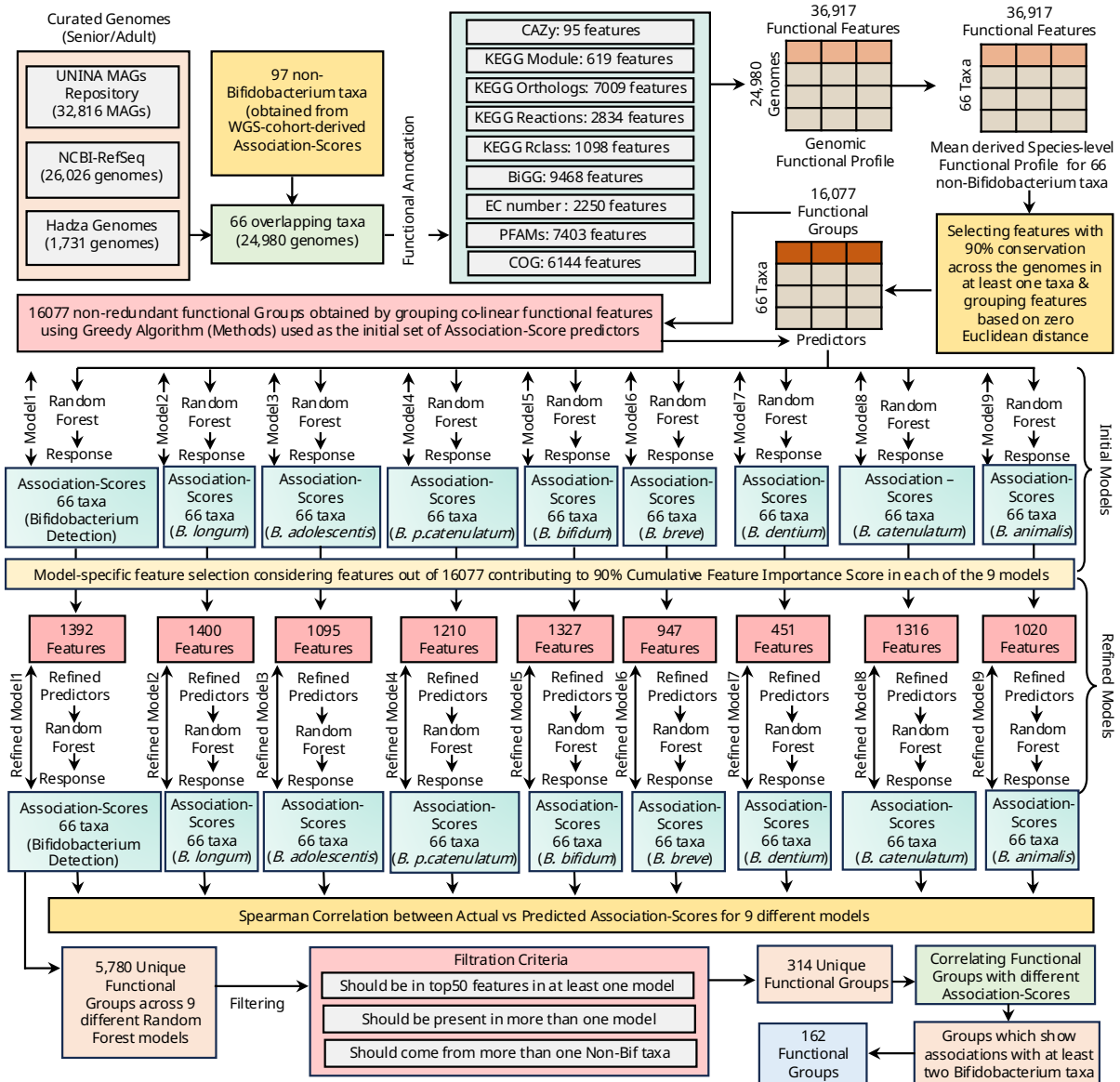


Association Between Clinical Symptom Changes (PRE -Baseline) and *B. animalis* Receptive -Score following intervention in the BaZ_2021 Trial



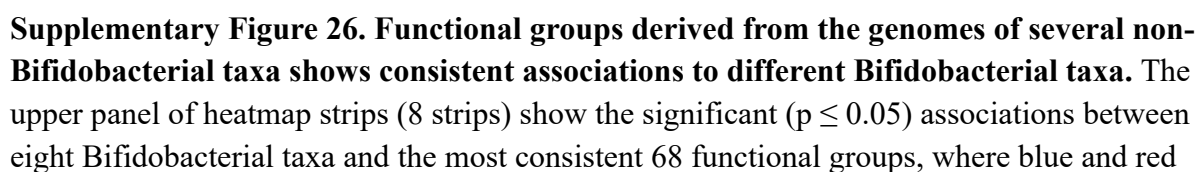
Supplementary Figure 24. Association between Receptive-Scores and clinical response following *Bifidobacterium* interventions. (A-E) Association between *B. adolescentis* Receptive-Scores and changes in gastrointestinal symptoms (T3-T1) in the

RamakrishnanM_2025 trial. Except, the subjects 86 and 36, all other subjects followed pattern where baseline Receptive-Scores were positively associated with multiple clinical outcomes post-treatment. **(A)** Heatmap summarizing the individual-level changes in clinical symptoms alongside Receptive-Score. **(B-D)** Scatterplots showing Spearman correlations between Receptive-Score and changes in: **(B)** Diarrhea, **(C)** Abdominal Pain and **(D)** Fecal Urgency. **(E)** Scatterplot showing the association between the Receptive-Score and the overall symptom reduction score, calculated as the number of gastrointestinal symptoms showing a negative (improving) change per individual. Spearman correlation coefficients (ρ) and corresponding p values are indicated in each panel. **(F-I)** Association between baseline *B. animalis* Receptive-Scores and clinical marker changes (pre-post intervention) in the BaZ_2021 trial. Scatter plots depict correlations between Receptive-Scores and changes in CD3/CD8 T-cell ratio **(F)**, Interleukin-2 or IL2 **(G)**, Tumor Necrosis Factor or TNF **(H)**, and Interleukin-6 or IL6 **(I)**, with Spearman ρ and p values shown. 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 y-axis, calculated as the fitted value $\pm 1.96 \times$ standard error (s.e.) of the linear regression derived fit based on variable on the x-axis. Spearman rank correlation (two-sided) was used; correlation coefficients (ρ) and corresponding p values are shown; statistically significant associations ($p \leq 0.05$) are highlighted in blue, while non-significant ones are shown in black. No multiple testing correction was applied, as analyses were conducted independently across clinical outcomes. Sample sizes were $n = 8$ for the RamakrishnanM_2025 trial and $n = 20$ for the BaZ_2021 trial.



Supplementary Figure 25. This panel shows the schematic workflow of how the mean functional feature abundance profile is generated using a set of 24,980 high quality genomes including MAGs generated from Senior/Adults microbiomes from different genomic repositories, then it shows the approach of how the co-linear functional features were grouped into unique functional groups using a Greedy algorithm. Furthermore, the random forest-based feature selection and identifying top features which constitute 90% cumulative feature importance scores in each Bifidobacterial features, followed by re-training of the random forest-based models with the top features to reduce noise and finally correlating the actual and predicted Association-Scores using Spearman rank correlation (two-sided). The unique top functional-groups coming from nine different models were further reduced based on three

selection criteria and then associating the most relevant features to each of the nine types of Association-Scores to capture the association patterns.



Supplementary Figure 26. Functional groups derived from the genomes of several non-Bifidobacterial taxa shows consistent associations to different Bifidobacterial taxa. The upper panel of heatmap strips (8 strips) show the significant ($p \leq 0.05$) associations between eight Bifidobacterial taxa and the most consistent 68 functional groups, where blue and red

colored cells indicate significant positive and negative associations respectively. Correlations were assessed using Spearman rank correlation (two-sided). The lower heatmap panel shows the rank-scaled abundance of functional groups across different non-Bifidobacterial taxa. Consistent associations are highlighted through boxes (in deep blue) and some of the already known links have been shown by asterisks (*). Moreover, some of the non-Bifidobacterial taxa (in rows) as well as the functions (in columns) have been labelled in red color for which the associations are described in the results and in **Supplementary Note S8**.

References

1. Goel, A. *et al.* Toward a health-associated core keystone index for the human gut microbiome. *Cell Reports* **44**, (2025).
2. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2**, e01202 (2013).
3. Alvarez-Silva, C. *et al.* Trans-ethnic gut microbiota signatures of type 2 diabetes in Denmark and India. *Genome Medicine* **13**, 37 (2021).
4. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* **6**, 320–329 (2012).
5. Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer. *PLOS ONE* **7**, e39743 (2012).
6. Willing, B. P. *et al.* A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology* **139**, 1844–1854.e1 (2010).
7. Zhang, Z. *et al.* Large-Scale Survey of Gut Microbiota Associated With MHE Via 16S rRNA-Based Pyrosequencing. *Official journal of the American College of Gastroenterology | ACG* **108**, 1601 (2013).
8. Schubert, A. M. *et al.* Microbiome Data Distinguish Patients with *Clostridium difficile* Infection and Non-*C. difficile*-Associated Diarrhea from Healthy Controls. *mBio* **5**, 10.1128/mbio.01021-14 (2014).
9. Yoshida, N. *et al.* *Bacteroides vulgatus* and *Bacteroides dorei* Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. *Circulation* **138**, 2486–2498 (2018).
10. Singh, P. *et al.* Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* **3**, 45 (2015).
11. Yoshida, N. *et al.* Effect of Resistant Starch on the Gut Microbiota and Its Metabolites in Patients with Coronary Artery Disease. *Journal of Atherosclerosis and Thrombosis* **26**, 705–719 (2019).
12. Boertien, J. M. *et al.* Fecal microbiome alterations in treatment-naïve de novo Parkinson's disease. *npj Parkinsons Dis.* **8**, 129 (2022).

13. Hopfner, F. *et al.* Gut microbiota in Parkinson disease in a northern German cohort. *Brain Research* **1667**, 41–45 (2017).
14. Aho, V. T. E. *et al.* Gut microbiota in Parkinson's disease: Temporal stability and relations to disease progression. *eBioMedicine* **44**, 691–707 (2019).
15. Flemer, B. *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* **66**, 633–643 (2017).
16. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders* **30**, 350–358 (2015).
17. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* **8**, 1784 (2017).
18. Clooney, A. G. *et al.* Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. <https://doi.org/10.1136/gutjnl-2020-321106> (2021) doi:10.1136/gutjnl-2020-321106.
19. Liu, Q. *et al.* Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome. *Gut* **71**, 544–552 (2022).
20. Zhang, F. *et al.* Prolonged Impairment of Short-Chain Fatty Acid and L-Isoleucine Biosynthesis in Gut Microbiome in Patients With COVID-19. *Gastroenterology* **162**, 548-561.e4 (2022).
21. Zhang, P. *et al.* Specific gut microbiota alterations in essential tremor and its difference from Parkinson's disease. *npj Parkinsons Dis.* **8**, 98 (2022).
22. Dong, Y. *et al.* Characterization of gut microbiota in adults with coronary atherosclerosis. *PeerJ* **11**, e15245 (2023).
23. Wong, V. W.-S. *et al.* Molecular Characterization of the Fecal Microbiota in Patients with Nonalcoholic Steatohepatitis – A Longitudinal Study. *PLOS ONE* **8**, e62885 (2013).
24. Pietrucci, D. *et al.* Dysbiosis of gut microbiota in a selected population of Parkinson's patients. *Parkinsonism & Related Disorders* **65**, 124–130 (2019).
25. Heintz-Buschart, A. *et al.* The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Movement Disorders* **33**, 88–98 (2018).
26. Qian, Y. *et al.* Alteration of the fecal microbiota in Chinese patients with Parkinson's disease. *Brain, Behavior, and Immunity* **70**, 194–202 (2018).
27. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* **2**, 17004 (2017).
28. Youngster, I. *et al.* Fecal Microbiota Transplant for Relapsing Clostridium difficile Infection Using a Frozen Inoculum From Unrelated Donors: A Randomized, Open-Label, Controlled Pilot Study. *Clin Infect Dis* **58**, 1515–1522 (2014).
29. 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.e15 (2025).

30. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol* <https://doi.org/10.1038/s41587-023-01688-w> (2023) doi:10.1038/s41587-023-01688-w.
31. 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).
32. 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).
33. Bazanella, M. *et al.* Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome†§1. *The American Journal of Clinical Nutrition* **106**, 1274–1286 (2017).
34. Allard, G., Ryan, F. J., Jeffery, I. B. & Claesson, M. J. SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* **16**, 324 (2015).
35. 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**, 10.1128/msphere.00084-21 (2021).
36. 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 Pathogens* **17**, 19 (2025).
37. Ramakrishnan, M. *et al.* Two-week supplementation of *Bifidobacterium adolescentis* iVS-1 reduces symptoms associated with lactose intolerance in lactose maldigesters. *Gut Microbes Reports* **2**, 2508199 (2025).
38. 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).
39. 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).
40. 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**, 10.1128/msphere.00084-21 (2021).
41. Ramakrishnan, M. *et al.* Two-week supplementation of *Bifidobacterium adolescentis* iVS-1 reduces symptoms associated with lactose intolerance in lactose maldigesters. *Gut Microbes Reports* **2**, 2508199 (2025).
42. 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).
43. Bazanella, M. *et al.* Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome†§1. *The American Journal of Clinical Nutrition* **106**, 1274–1286 (2017).

44. Ramakrishnan, M. *et al.* Two-week supplementation of *Bifidobacterium adolescentis* iVS-1 reduces symptoms associated with lactose intolerance in lactose maldigesters. *Gut Microbes Reports* **2**, 2508199 (2025).
45. Carter, M. M. *et al.* Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell* **186**, 3111–3124.e13 (2023).
46. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662.e20 (2019).
47. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).
48. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019).
49. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* **50**, D571–D577 (2022).
50. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457–D462 (2016).
51. Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research* **53**, D444–D456 (2025).
52. Bairoch, A. The ENZYME database in 2000. *Nucleic Acids Research* **28**, 304–305 (2000).
53. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **44**, D515–D522 (2016).
54. Böttcher, J., Sibon, O. C. M. & El Aidy, S. Coenzyme A metabolism: a key driver of gut microbiota dynamics and metabolic profiles. *FEMS Microbiol Rev* **49**, fuaf051 (2025).
55. Tanno, H. *et al.* Characterization of fructooligosaccharide metabolism and fructooligosaccharide-degrading enzymes in human commensal butyrate producers. *Gut Microbes* **13**, 1869503 (2021).
56. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology* **19**, 29–41 (2017).
57. Xu, J., Zhang, J., Guo, Y. & Zhang, W. Genetically modifying aspartate aminotransferase and aspartate ammonia-lyase affects metabolite accumulation in l-lysine producing strain derived from *Corynebacterium glutamicum* ATCC13032. *Journal of Molecular Catalysis B: Enzymatic* **113**, 82–89 (2015).
58. Jeckelmann, J.-M. *et al.* Structure and function of the glucose PTS transporter from *Escherichia coli*. *Journal of Structural Biology* **176**, 395–403 (2011).
59. Cerasi, M., Ammendola, S. & Battistoni, A. Competition for zinc binding in the host-pathogen interaction. *Front. Cell. Infect. Microbiol.* **3**, (2013).