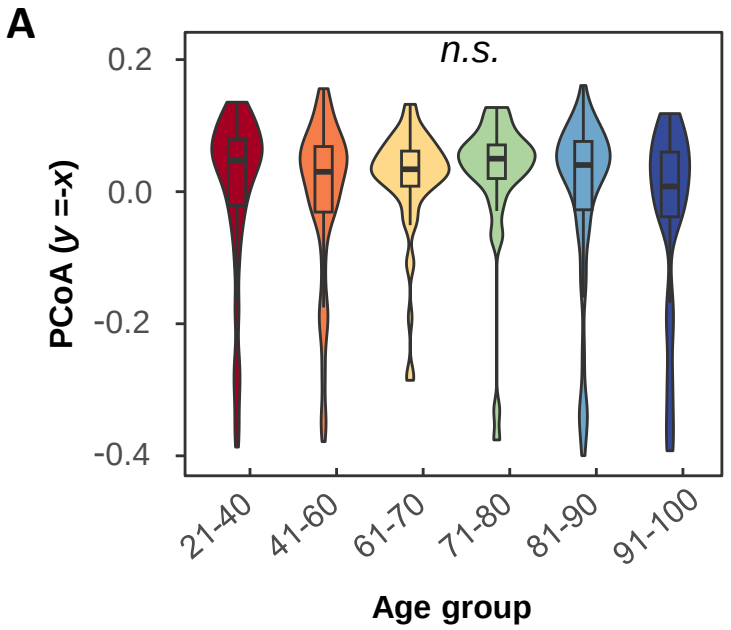
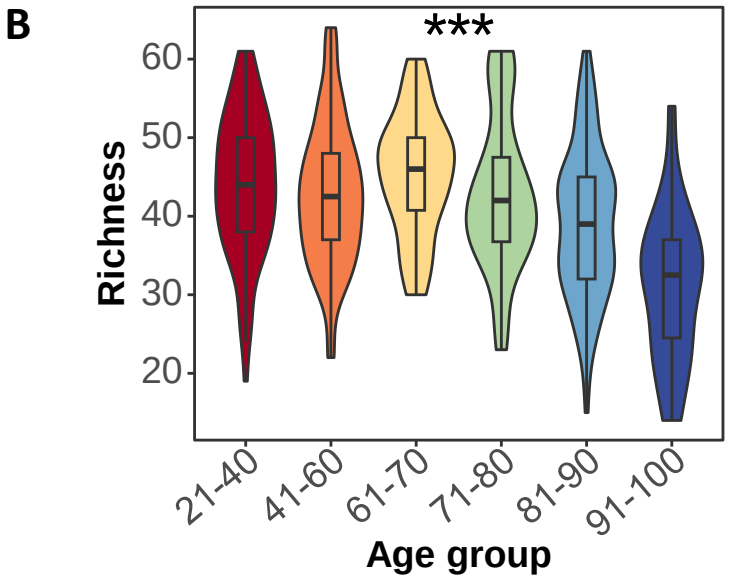
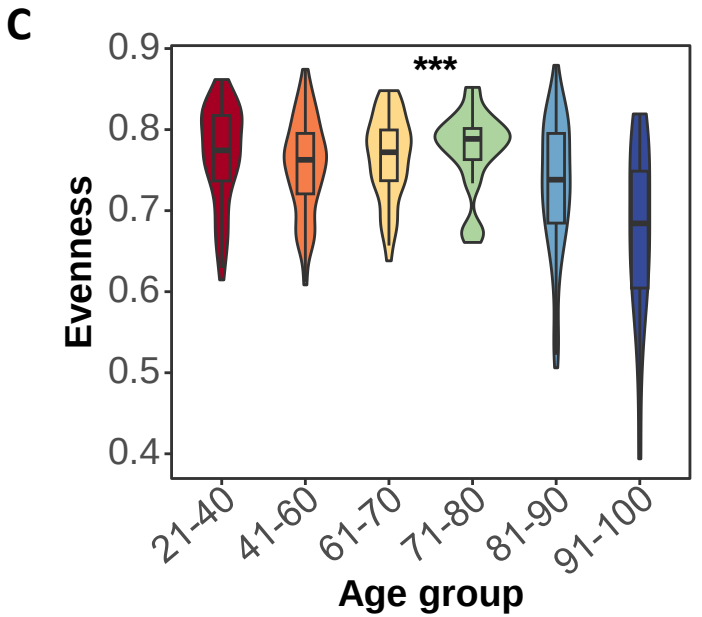
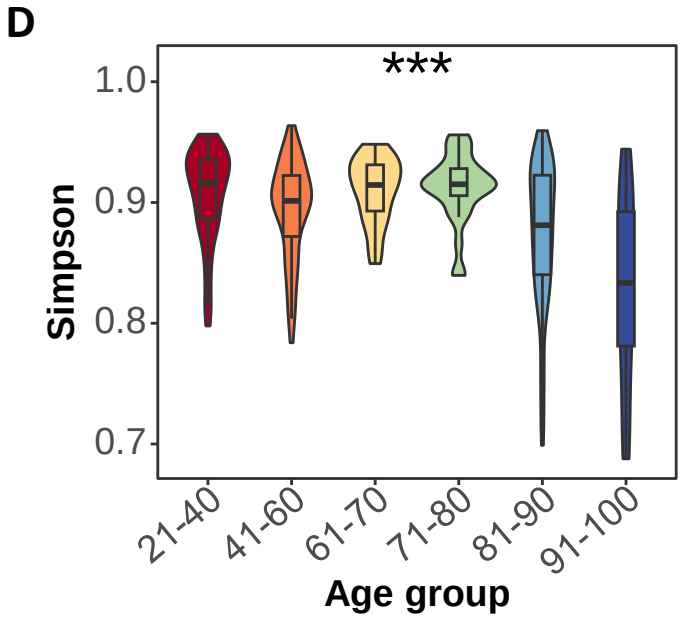
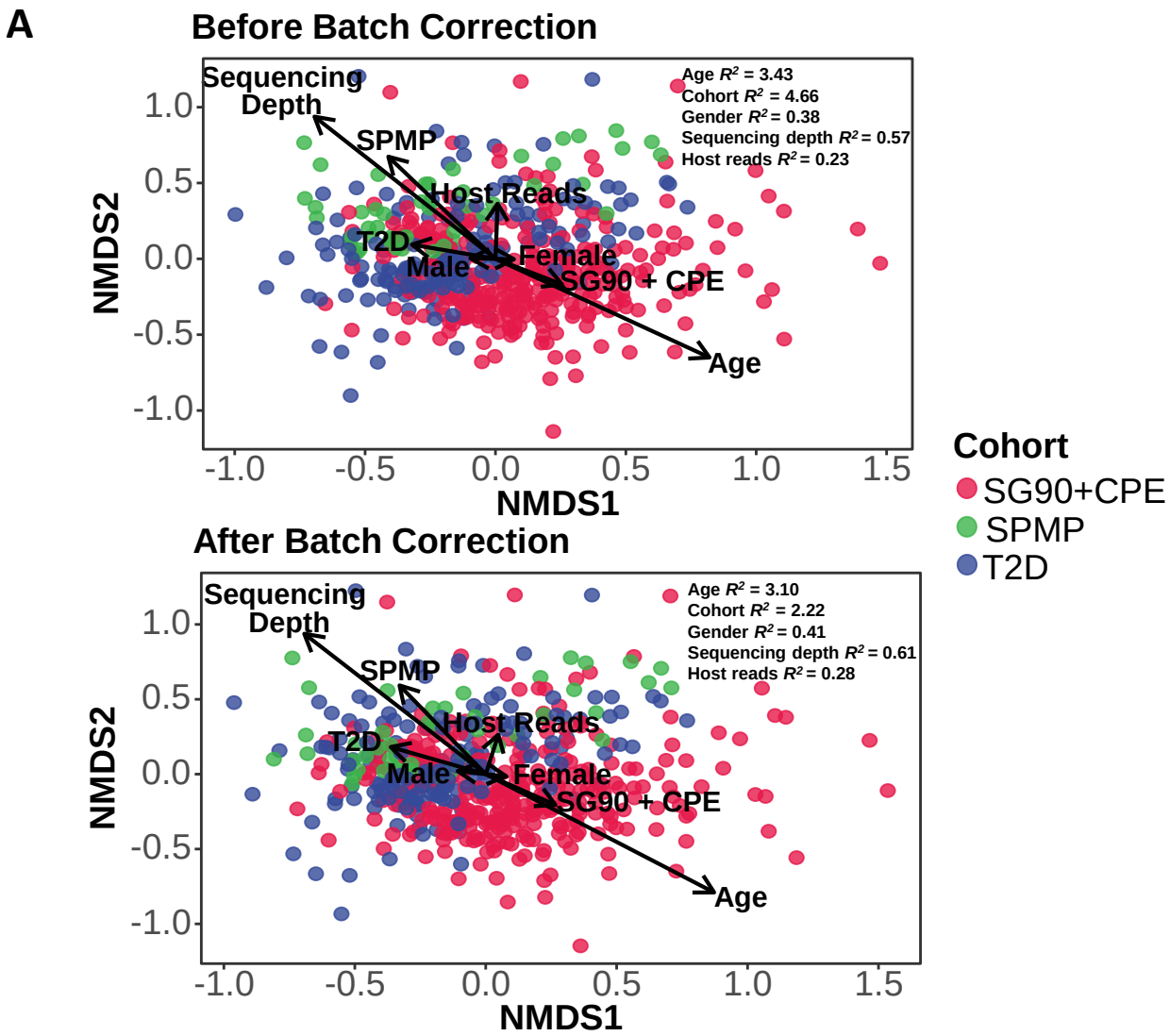


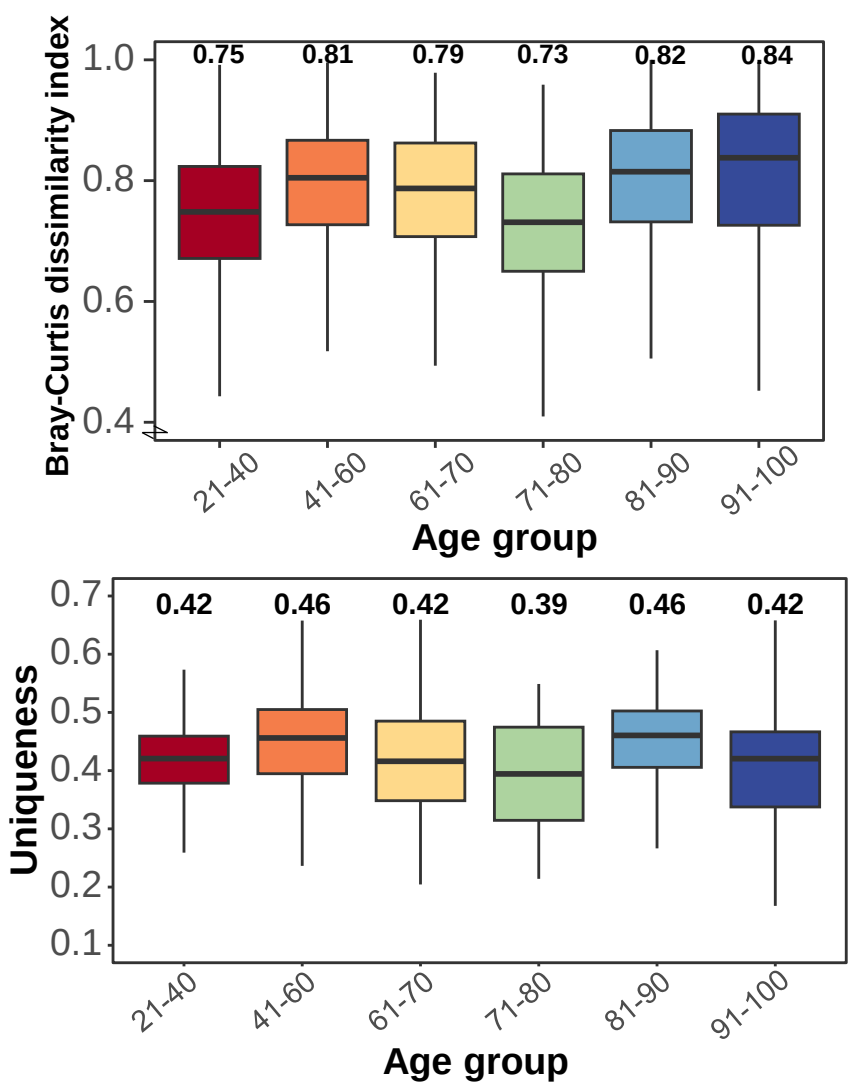
Supplementary Figure 1: **Variation in gut microbiome beta and alpha diversity metrics across age groups.** (A) Violin plot showing the variation of beta diversity on the *y*=-*x* axes corresponding to the PCoA plot in **Figure 1A**. Violin plots showing species-level (B) Richness, (C) Evenness and (D) Simpson diversity index, across age groups. All three indices were found to be significantly using two-sided test with Ordinal logistic regression across all age groups using *n*=516 independent samples ((B) Richness: 2.12×10^{-9}, (C) Evenness: 4.38×10^{-9}, (D) Simpson: 2.34×10^{-11}). Data points that are either less than Q1 – 1.5 × IQR or more than Q3 + 1.5 × IQR, where Q1, Q3 and IQR refers to the first quartile, third quartile and interquartile range, respectively, were removed for this analysis to reduce the impact of outliers. In all the inset boxplots, the center line represents the median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values. Source data are provided as a Source Data file.



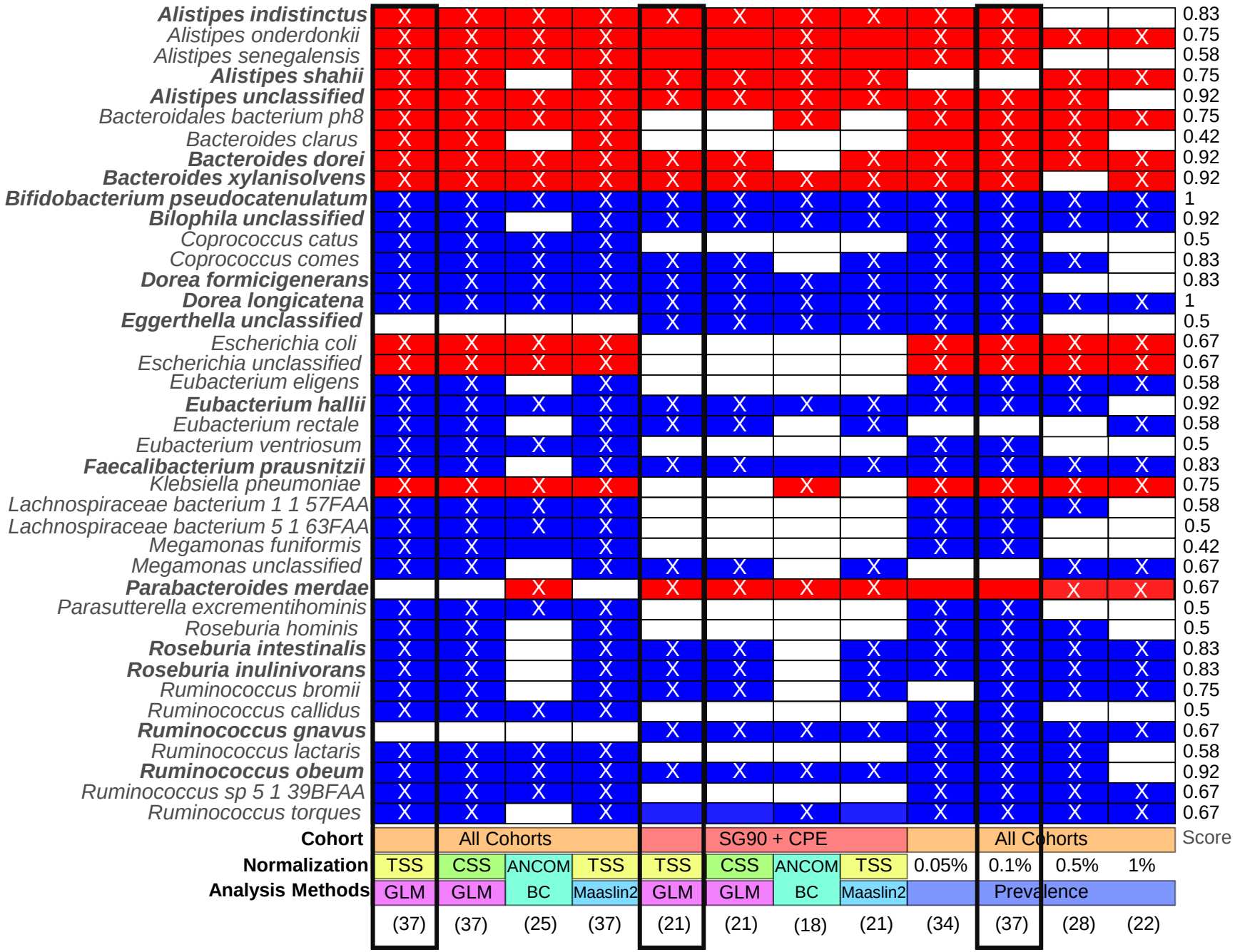








Supplementary Figure 2: Variation of beta diversity metrics. A) Nonmetric Multidimensional Scaling (NMDS) plots showing the distribution of samples for each cohort (*n*=516 independent samples) before (top) and after (bottom) batch correction. Vectors for each of the confounding variables were obtained through envfit analysis demonstrating the relationship between the ordination axes and the variables. Length of the vector indicates strength of the relationship, and the direction points to the steepest increase corresponding to the variable. Inset values for R^2 are from PERMANOVA analysis using Bray-Curtis dissimilarity. B) Box plots showing the species-level Bray-Curtis dissimilarity index (top), Uniqueness (bottom) across age groups. Uniqueness was defined as the minimum value of Bray-Curtis dissimilarity index for each individual in all boxplots, the center line represents the median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values. Median values are shown on top of the plot. Source data are provided as a Source Data file.



Direction of association

Negative

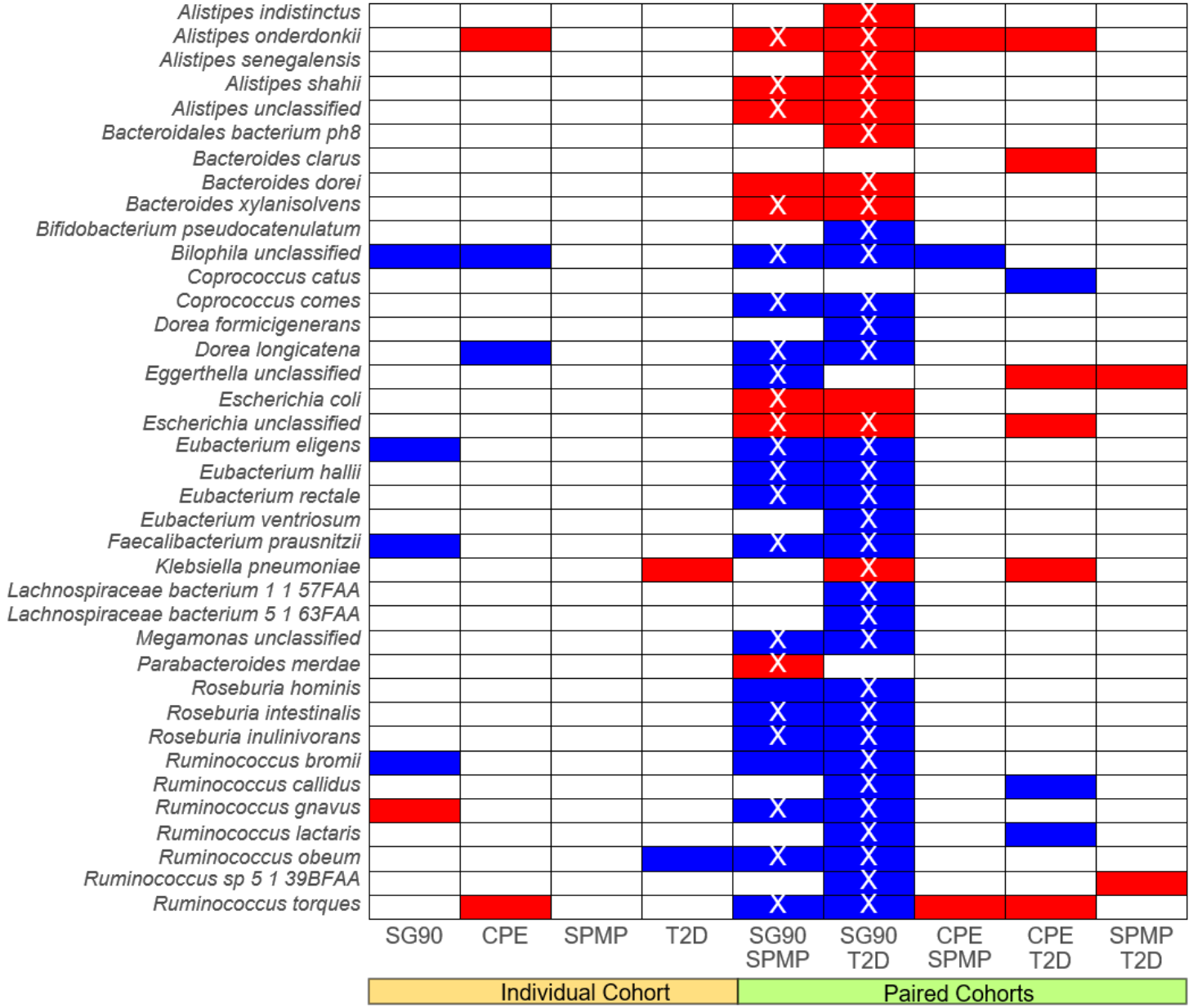
Positive

Significance

X FDR adjusted p-values

X FDR adjusted p-values

Supplementary Figure 3: Assessing the reproducibility and robustness of microbial associations with age. Heatmap showing microbial species associations with age, where (i) reproducibility was assessed through 3 distinct primary analysis approaches (black boxes) and (ii) robustness was assessed by varying their parameters, normalization and analysis settings (all columns; see **Methods)**. The robustness score (last column) captures the frequency with which a taxa was observed to be significantly associated with age across all tested methods and settings (FDR adjusted p-value<0.05 using Benjamini-Hochberg test, see Source data for exact FDR adjusted p-value). Note that all but two associations had robustness score greater than or equal to 50%. Numbers in parentheses below each column indicate the number of significant associations for each method. Red and blue colored cells represent positive and negative associations, respectively. Taxa found in the high-stringency list are in bold face. Colored cells marked with 'X' correspond to FDR adjusted p-value<0.05, while cells without an 'X' correspond to unadjusted p-value<0.05. For all cohorts, *n*=516 independent samples were used. For SG90 + CPE, *n*=295 samples were used. Source data are provided as a Source Data file.



Direction of association

Negative

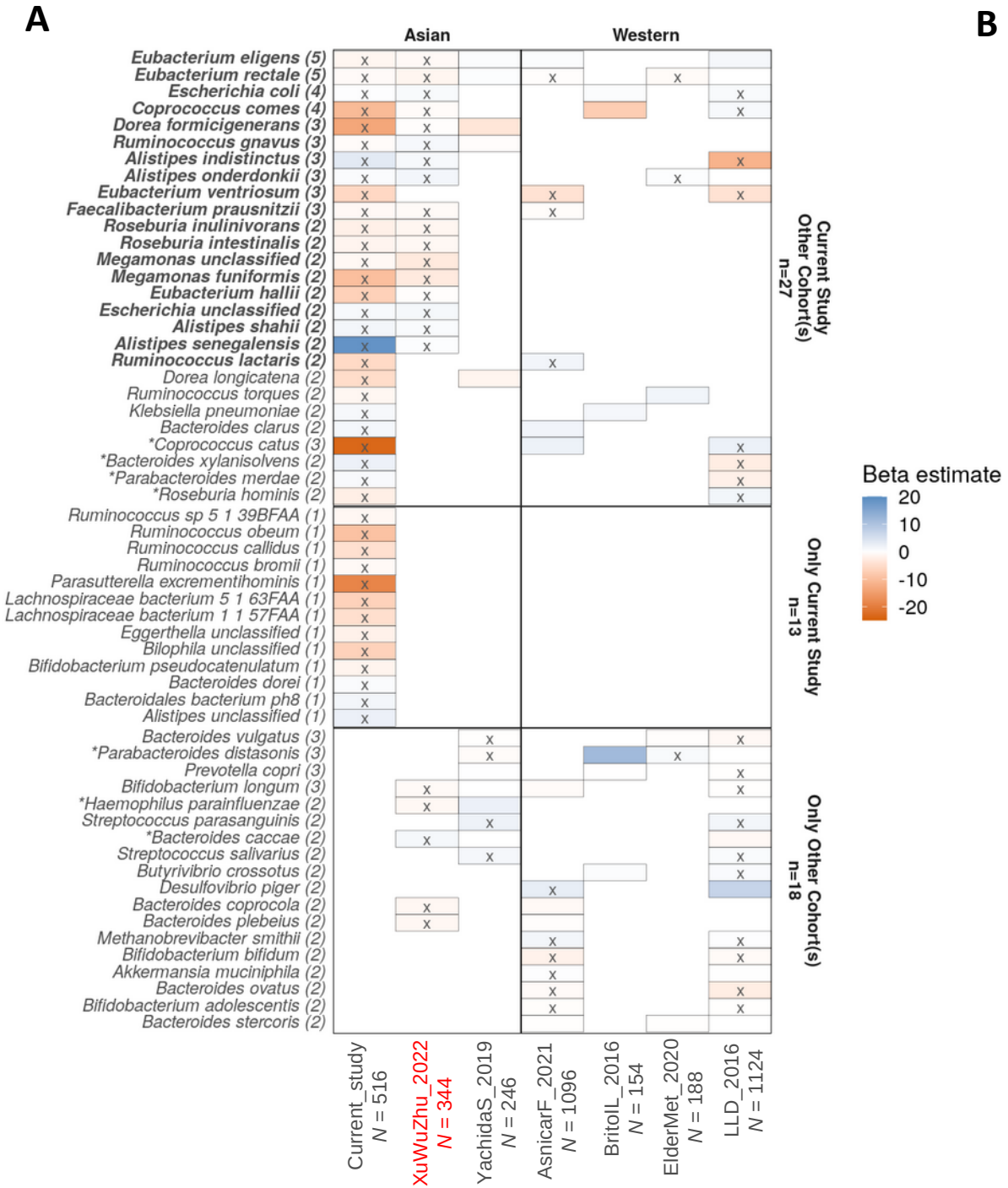
Positive

Significance

X FDR adjusted p-values

X FDR adjusted p-values

Supplementary Figure 4: Associations within individual and paired cohorts. Heatmap showing microbial species associations with age identified using independent and paired cohort analysis. Red and blue colored cells represent positive and negative associations, respectively. Colored cells marked with 'X' correspond to FDR adjusted p-value<0.05 (FDR-adjusted p-value using Benjamini-Hochberg test, see Source data for exact FDR adjusted p-value), while cells without an 'X' correspond to unadjusted p-value<0.05. As can be seen here, while some associations are identified in individual cohorts (*n*=10), joint analysis of SG90 with a cohort of younger individuals is essential to identify many more associations (*n*=38). These association are unlikely to be purely a function of batch effects as very few associations are detected when other pairs of cohorts are analyzed together (e.g. *n*=3, using CPE and SPMP), where these associations are also not significant after FDR adjustment. The number of samples used for each of these analysis are: SG90 - *n* = 213, CPE - *n* = 82, SPMP - *n* = 50, T2D - *n* = 171, SG90 + CPE - *n* = 295, SG90 + SPMP - *n* = 263, SG90 + T2D - *n* = 384, CPE + SPMP - *n* = 132, CPE + T2D - *n* = 253, T2D + SPMP - *n* = 221. Source data are provided as a Source Data file.



Supplementary Figure 5: Associations within other cohorts. A) Heatmap showing taxonomic associations with age across Asian and western cohorts. Colored cells marked with 'X' correspond to FDR adjusted p-value<0.05, while cells without an 'X' correspond to unadjusted p-value<0.05. Colors of the boxes represent whether the organism is positively (blue) or negatively (red) associated with age based on the beta estimate. The numbers in the bracket represent the number of times a particular taxa is validated in other cohorts (p-value <0.05). Taxa in bold face indicate those that are independently replicated in other cohorts (FDR adjusted p-value <0.05, FDR adjustment done using Benjamini-Hochberg test). Taxa marked with a '*' have a directionality of association that is different across cohorts. The cohort 'XuWuZhu_2022' is colored red as the association results were obtained from the original manuscript and are based on a Wilcoxon rank-sum test instead of GLM analysis. The number of independent samples analyzed are indicated in the plot. B) Barplot showing the number of subjects (indicated in every bar) for different cohorts in each age group. The number of subjects in each age group is calculated by including the lower and upper end of the range. Source data are provided as a Source Data file.

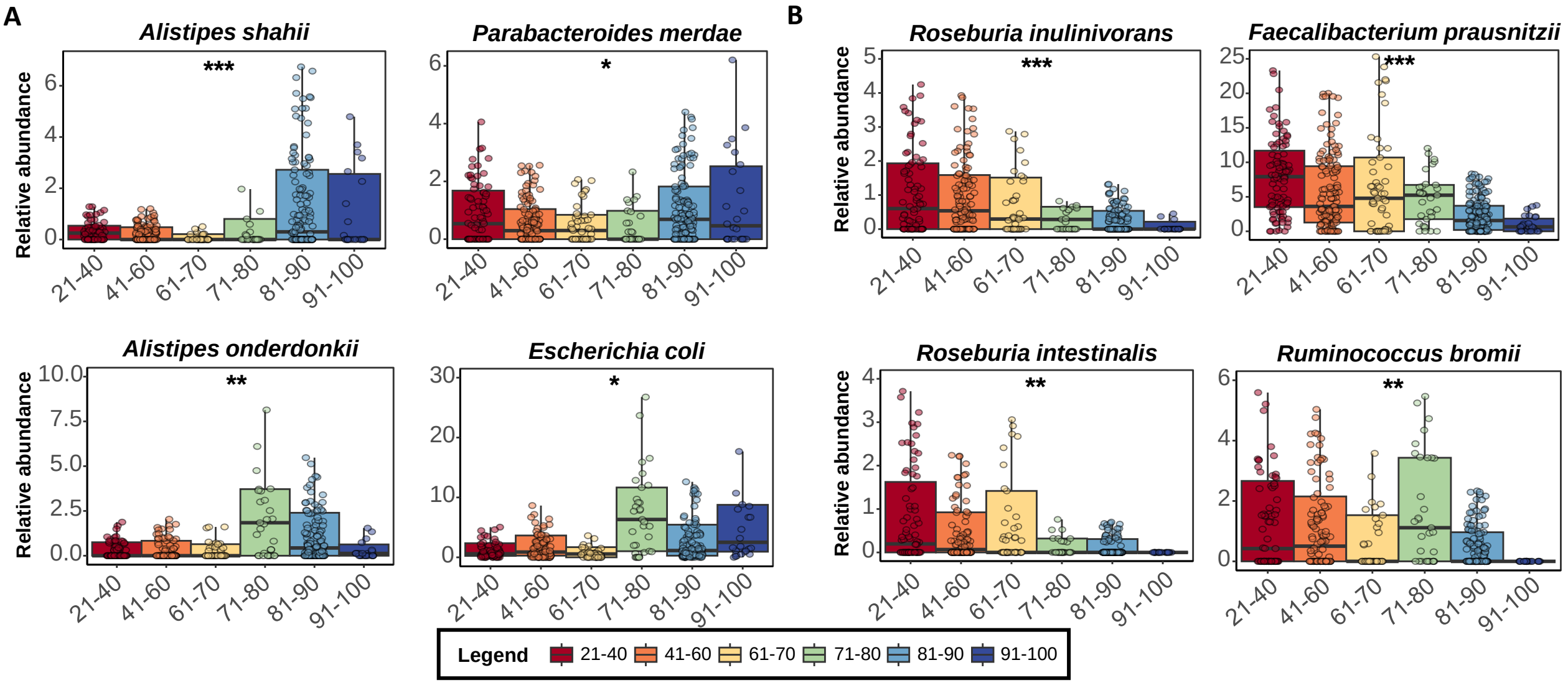
A

Enriched with age	GLM p-value	*β* for age [95% C.I.]	Remarks
Bacteroidetes	9.7E-19	0.33 [0.26, 0.39]	Enriched in Ireland elderly subjects[1] and Chinese centenarians[2]
Depleted with age			
Firmicutes	6.5E-30	-0.49 [-0.57, -0.42]	Depleted in Japanese elderly subjects[3] and Italian centenarians[4]
Actinobacteria	2.5E-05	-0.38 [-0.55, -0.22]	
Fusobacteria	0.0077	-3.92 [-6.60, -1.23]	

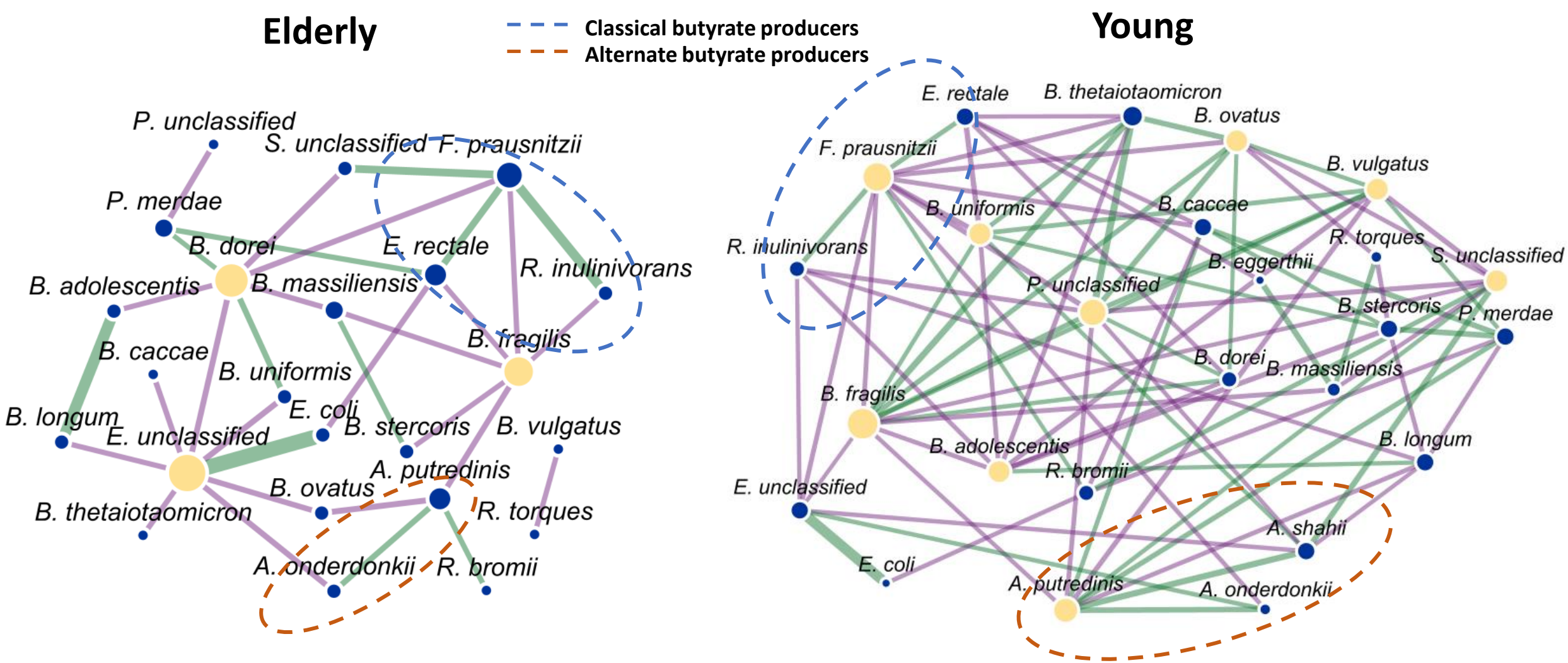
B

Enriched with age	GLM p-value	*β* for age [95% C.I.]	Remarks
Alistipes	2.64E-05	0.38 [0.22, 0.53]	Enriched in Chinese centenarians[5]
Bacteroidales noname	3.58E-05	1.37 [0.78, 1.95]	
Bacteroides	0.0064	0.13 [0.05, 0.21]	
Escherichia	0.013	0.23 [0.07, 0.38]	Enriched in Chinese[6] and Italian[7, 8] centenarians, American elderly[9]
Klebsiella	0.018	0.89 [0.25, 1.53]	
Acidaminococcaceae unclassified	0.028	6.69 [1.49, 11.88]	
Depleted with age			
Dorea	1.95E-12	-6.90 [-8.63, -5.18]	Depleted in Italian centenarians[5] and Russian elderly subjects[10]
Faecalibacterium	1.76E-12	-1.06 [-1.33, -0.79]	Depleted in Chinese[6] and Italian centenarians[11]
Roseburia	8.97E-09	-1.74 [-2.27, -1.20]	Enriched in Chinese centenarians[6] but depleted in Italian centenarians[11]
Ruminococcus	7.09E-08	-1.11 [-1.47, -0.74]	Enriched in Italian centenarians[8] and depleted in Russian elderly subjects[10]
Bilophila	2.45E-07	-7.91 [-10.64, -5.17]	Enriched in Italian centenarians[11]
Blautia	3.32E-05	-0.99 [-1.41, -0.57]	
Parasutterella	8.24E-05	-20.19 [-29.25, -11.13]	
Eubacterium	0.00035	-0.58 [-0.87, -0.29]	Depleted in Italian centenarians[8,11]
Lachnospiraceae noname	0.00035	-3.32 [-4.96, -1.69]	
Coprococcus	0.0011	-2.81 [-4.31, -1.31]	
Bifidobacterium	0.0011	-0.33 [-0.51, -0.15]	Depleted in Chinese centenarians[6] but enriched in combined analysis of Indian, Chinese, Japanese and Italians centenarians[12]
Paraprevotella	0.0013	-2.65 [-4.1, -1.2]	
Colinsella	0.0061	-1.84 [-3.00, -0.69]	
Fusobacterium	0.018	-3.60 [-6.20, -1.00]	
Coprobacillus	0.023	-10.08 [-17.67, -2.49]	
Megamonas	0.032	-0.66 [-1.19, -0.13]	

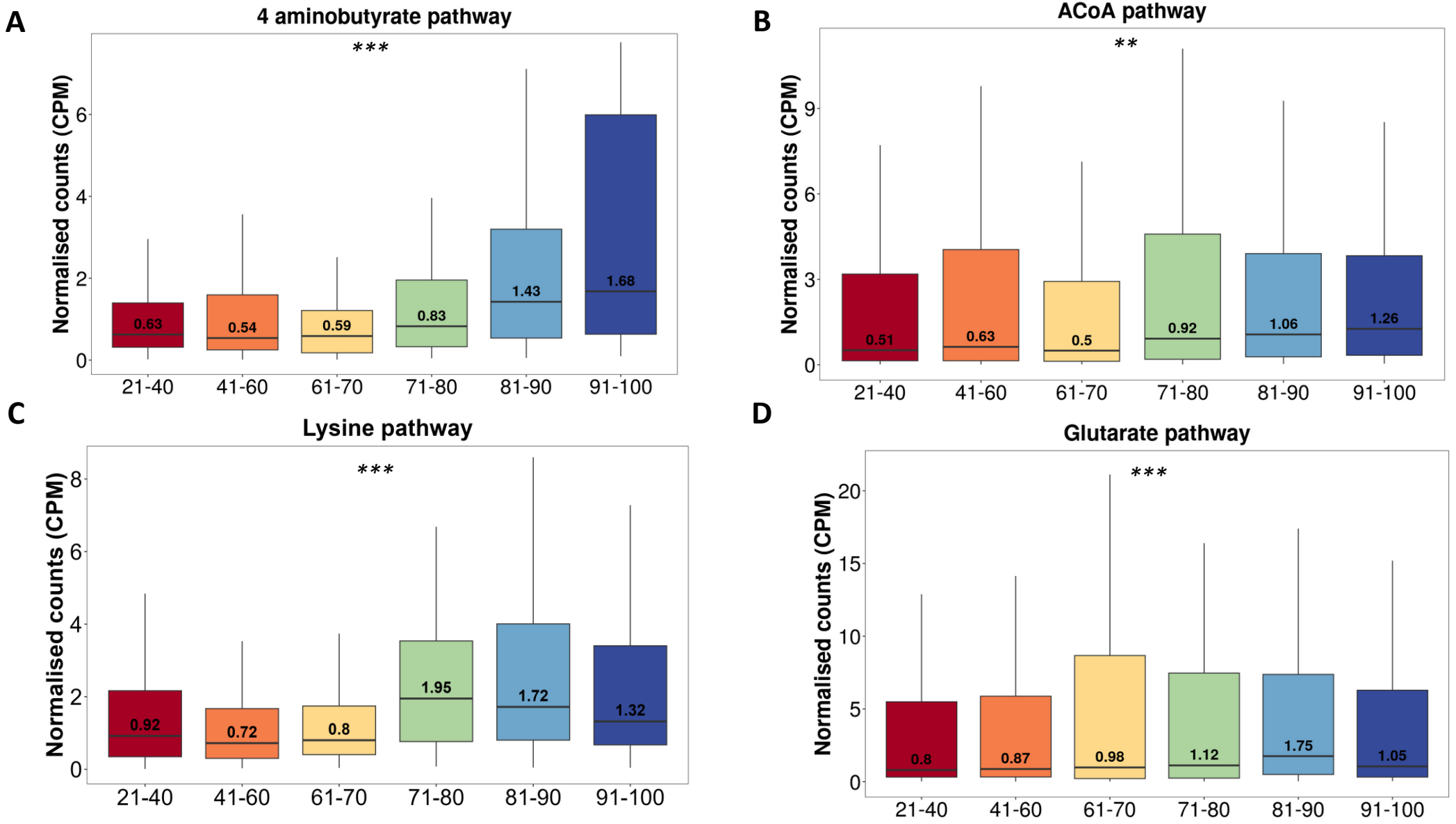
Supplementary Table 1: Higher-level taxonomic associations with age. At the (A) phylum and (B) genus level, the table reports taxa that are statistically significant with FDR-adjusted p-values≤0.05 obtained from GLM for phylum/genus association with co-variate adjustment (see **Methods**). The p-values reported here are obtained from two-sided tests from GLM and FDR-adjusted p-values are obtained using Benjamini-Hochberg test. The *β* coefficient estimates strength of association as change in age (in years) for every 1% increment of taxa relative abundance.



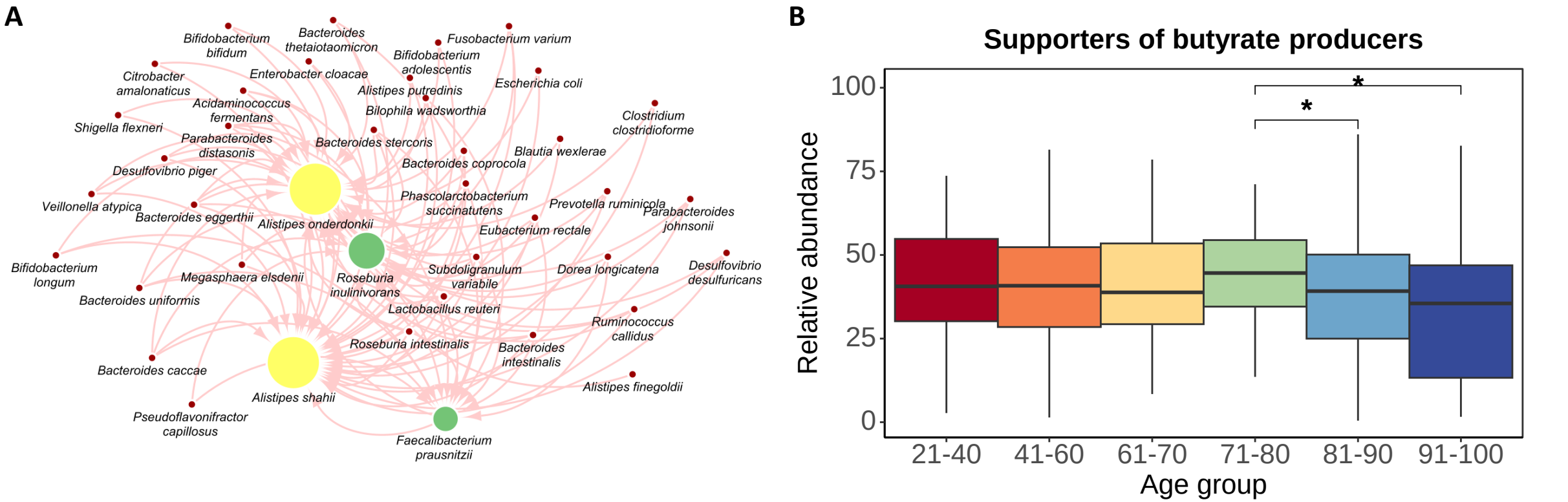
Supplementary Figure 6: Relative abundance across age groups for key gut microbial species associated with aging. Boxplots showing relative abundances across age groups for (A) species whose abundances increase with age and (B) species whose abundances decrease with age. Median FDR-adjusted p-values using Benjamini-Hochberg test from the 3 primary analysis approaches (*Alistipes shahii:* 6.81×10^{-4}, *Parabacteroides merdae:* 3.45×10^{-2}, *Alistipes onderdonkii*: 1.59×10^{-3}, *Escherchia coli*: 0.031, *Roseburia inulinivorans*: 2.3×10^{-5}, *Faecalibacterium prausnitzii*: 5.33×10^{-6}, *Roseburia intestinalis*: 5×10^{-3}, *Ruminococcus bromii*: 2.68×10^{-3}) are denoted by "*" (p-value<0.05), "**" (p-value<0.01) and "***" (p-value<0.001). In all boxplots, the center line represents the median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values (outlier points are not included in the visualization). Source data are provided as a Source Data file.



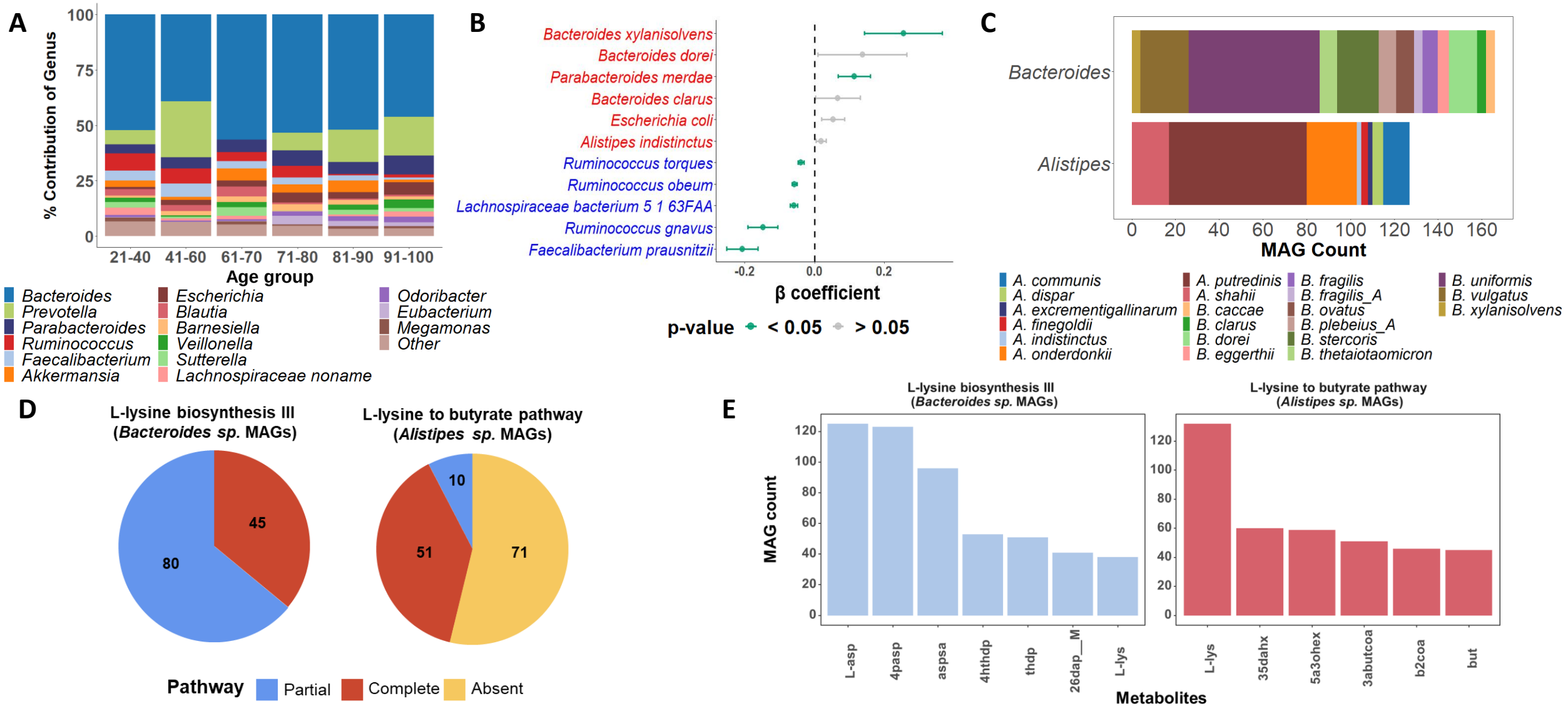
Supplementary Figure 7: Microbial co-occurrence patterns in elderly and young Asian subjects. Nodes in the network represent microbial species and edges represent significant Spearman correlations between relative abundances of species corresponding to adjacent nodes ($|r| \geq 0.1$, p-value<0.05). The network on the left (Elderly) was constructed with samples in the age range 70-100, while that on the right (Young) was constructed with samples in the age range 20-69 to ensure that sufficient number of datapoints were available. Green- and purple-colored edges represent positive and negative correlations respectively. Thickness of the edges are directly proportional to the Spearman correlation value $|r|$. Blue and orange circles identify the classical and alternate butyrate producer species, respectively. Hubs in the network are shown as yellow-colored nodes. There are 23 nodes, 32 edges, 3 hubs in the elderly network, while the young network has 25 nodes, 80 edges and 9 hubs. Source data are provided as a Source Data file.



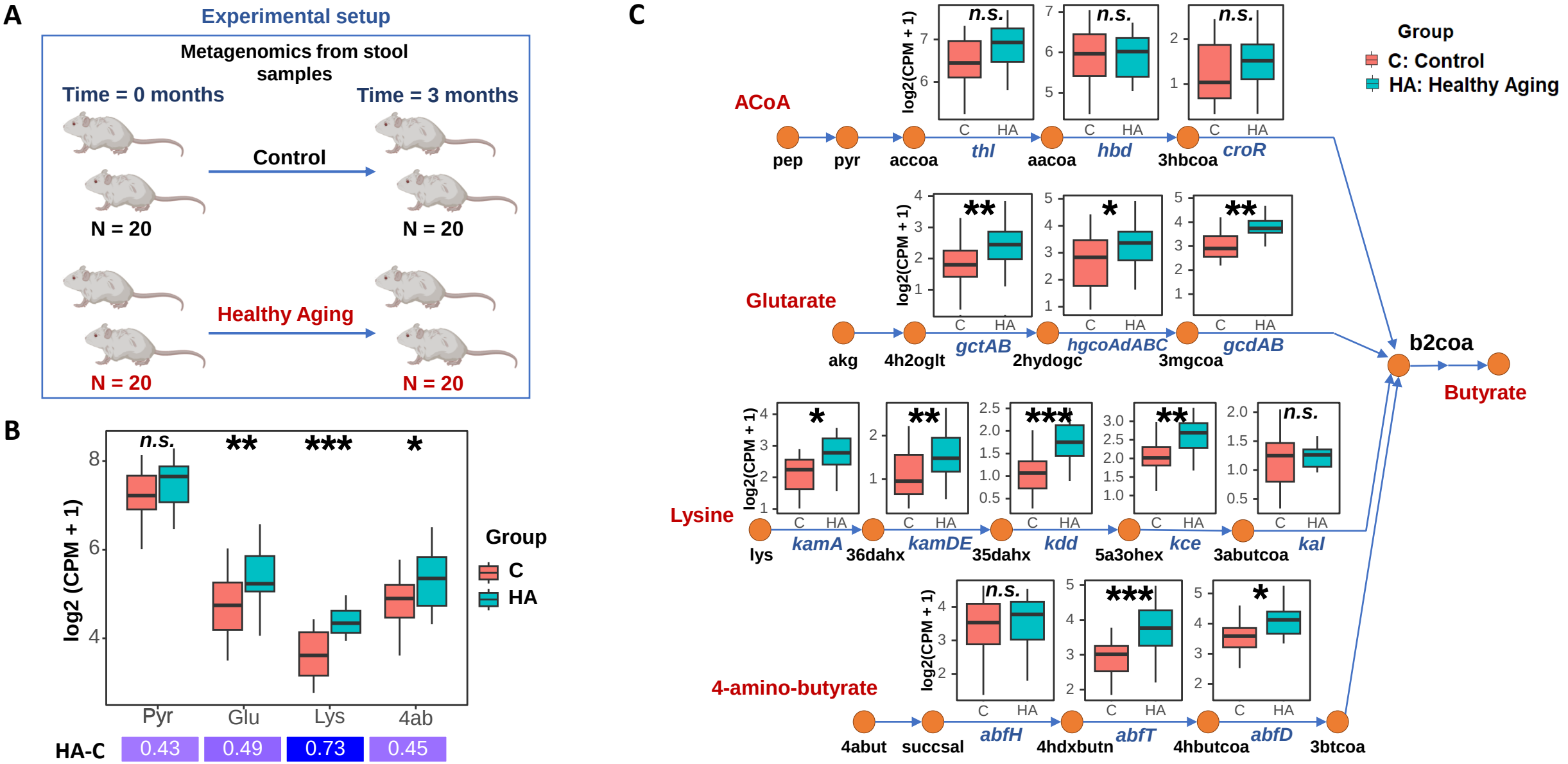
Supplementary Figure 8: Variation in butyrate production pathways across age groups. Boxplots showing the metagenomic relative abundance of pathways for butyrate production in different age groups. FDR-adjusted p-values using Benjamini-Hochberg test from a GLM test for age association (4 aminobutyrate pathway: 7.24×10^{-9}, ACoA pathway: 8.99×10^{-3}, Lysine pathway: 7.84×10^{-12}, Glutarate pathway: 1.88×10^{-17},are denoted by "**" (p-value≤0.01) and "***" (p-value<0.001). In all boxplots, the center line represents median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values (outlier points are not included in the visualization). Source data are provided as a Source Data file.



Supplementary Figure 9: Metabolic support in the gut microbiome for butyrate production. (A) Network figure depicting gut microbial species with high 'metabolic support index' for butyrate producers that receive maximum support. Directed edges go from the supporting species to the one that is being supported, and the size of the supported node is proportional to the number of incoming edges. Nodes colored in yellow and green indicate those enriched and depleted with age respectively. (B) Boxplots depicting combined relative abundances of all species that support butyrate producers (from subfigure A) in the gut microbiomes of subjects from various age groups. Wilcoxon test p-values<0.05 (p-values between groups 71-80 & 81-90 – 0.02; 81-90 & 91-100 – 0.015) are indicated with the star symbol ('*'). In all boxplots, the center line represents median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values.



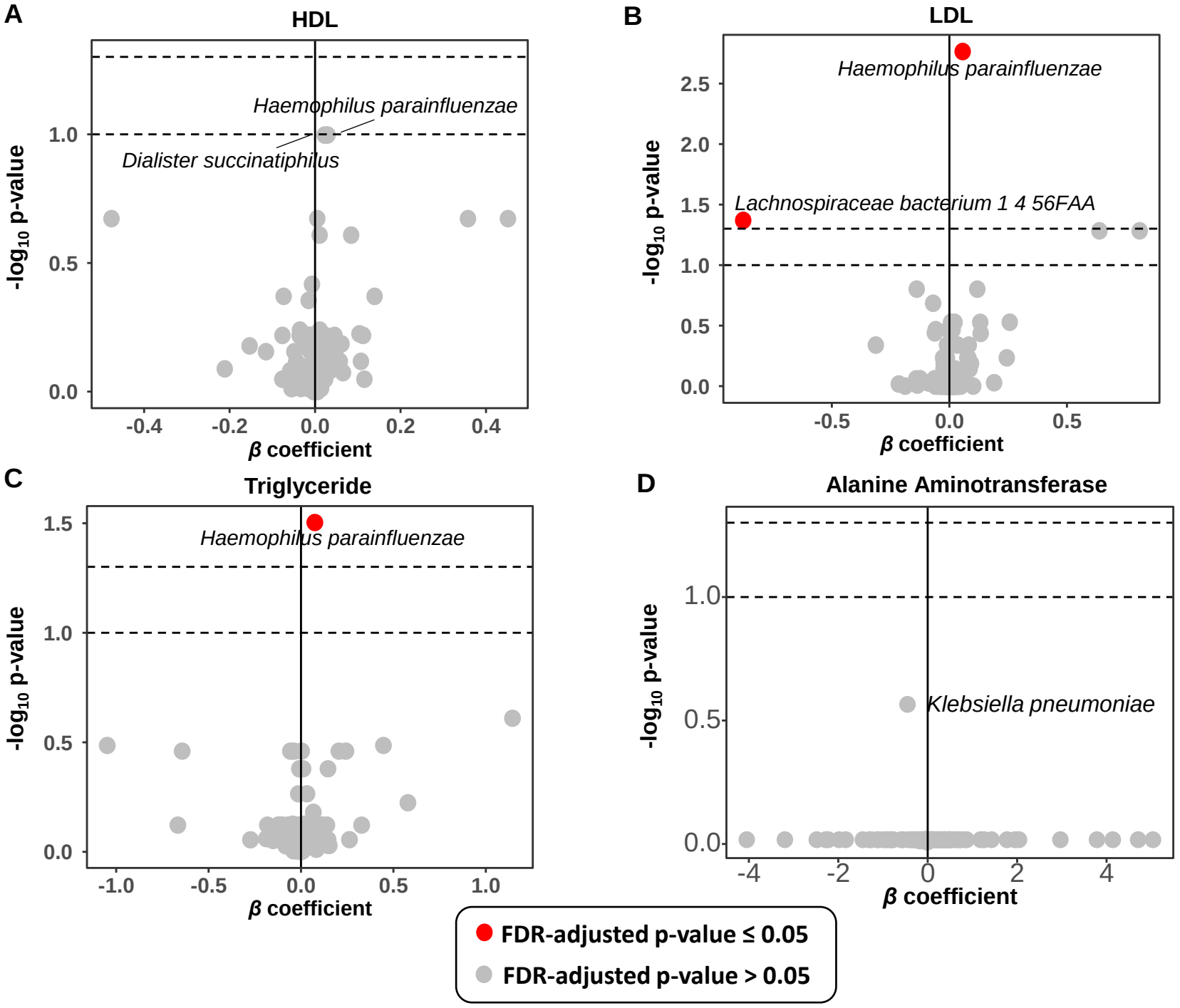
Supplementary Figure 10: Species contributions to L-lysine biosynthesis III and L-lysine to butyrate pathways (A) Stacked bar chart showing the relative contributions of genera to L-lysine biosynthesis III pathway. (B) Dot plot showing the variation of beta-coefficient of the age-associated organisms with the abundance of L-lysine biosynthesis III across age. The organisms highlighted in blue are negatively associated with age, while those in red are positively associated. Green colored points represent p-value < 0.05 (p-value obtained from two-sided tests using GLM). (C) Stacked bar chart showing the diversity of *Alistipes sp.* and *Bacteroides sp.* in the assembled MAGs. (D) Pie chart showing the number of MAGs that were annotated with L-lysine biosynthesis III and L-Lysine to butyrate pathways in *Bacteroides sp.* and *Alistipes sp*. respectively. (E) Bar chart showing the number of times a metabolite is reachable within the MAGs of *Bacteroides sp.* and *Alistipes sp*. in L-lysine biosynthesis III pathway (left) and L-lysine to butyrate pathway (right). X-axis represents the intermediate metabolites in the respective pathways. Metabolites are abbreviated based on BiGG database. Source data are provided as a Source Data file.



Supplementary Figure 11: **Butyrate production pathways in a healthy aging mouse model.** (A) Experimental setup for fecal metagenomic analysis in two groups of aged mice (18 months) with (Healthy Aging - HA) and without (Control - C) CaAKG 4% supplementation. Figure created with BioRender.com. (B) Boxplots showing relative abundance of alternative pathways for butyrate production in the two mouse groups after 3 months of supplementation. C) Boxplots for abundance of various genes in corresponding pathways where nodes represent major metabolites and edge labels represent genes involved in their conversion. In all boxplots, center line represents median, box limits represent upper and lower quartiles, and whiskers represent minimum and maximum values. The labels *"n.s."*, "*", "**" and "***" represent FDR-adjusted p-value>0.05, p-value≤0.05, p-value<0.01 and p-value<0.001 from two-sided Wilcoxon rank sum test (abfT: 0.000924, abfD: 0.0174, gctAB: 0.0021, hgcoAdABC: 0.017, gcdAB: 0.00106, kamA: 0.01742, kamDE:0.00251, kce:0.00186, kdd: 0.00092) . Source data are provided as a Source Data file.

Clinical phenotypes	Range
HDL	1.0 – 1.5 mmol/L
LDL	2.6 – 3.3 mmol/L
Triglyceride	1.7 – 2.2 mmol/L
Total cholesterol	< 5.2 mmol/L
Fasting blood glucose	< 6.1 mmol/L
High-sensitivity CRP	< 1 mg/L
Vitamin B12	160 to 950 pmol/L
Alanine Aminotransferase	6 - 66 U/L
Aspartate Aminotransferase	6 - 35 U/L

Supplementary Table 2: Reference ranges for various clinical phenotypes. Healthy ranges for various clinical phenotypes that were analyzed in this study.



Supplementary Figure 12: Microbiome associations across various metabolic markers. Volcano plots showing *β* coefficient on the x-axis and FDR-adjusted $-\log_{10}$ p-values (GLM test) on the y-axis. Red dots indicate FDR-adjusted p-value≤0.05 and grey dots indicate otherwise, with dotted lines marking p-value=0.05 and p-value=0.1 thresholds. (A), (B), and (C) Results for HDL, LDL and Triglyceride levels. (D) Results for Alanine Aminotransferase showing a weak association with *Klebsiella pneumoniae*, compared to the significant association in **Figure 4D** with AST. Source data are provided as a Source Data file.

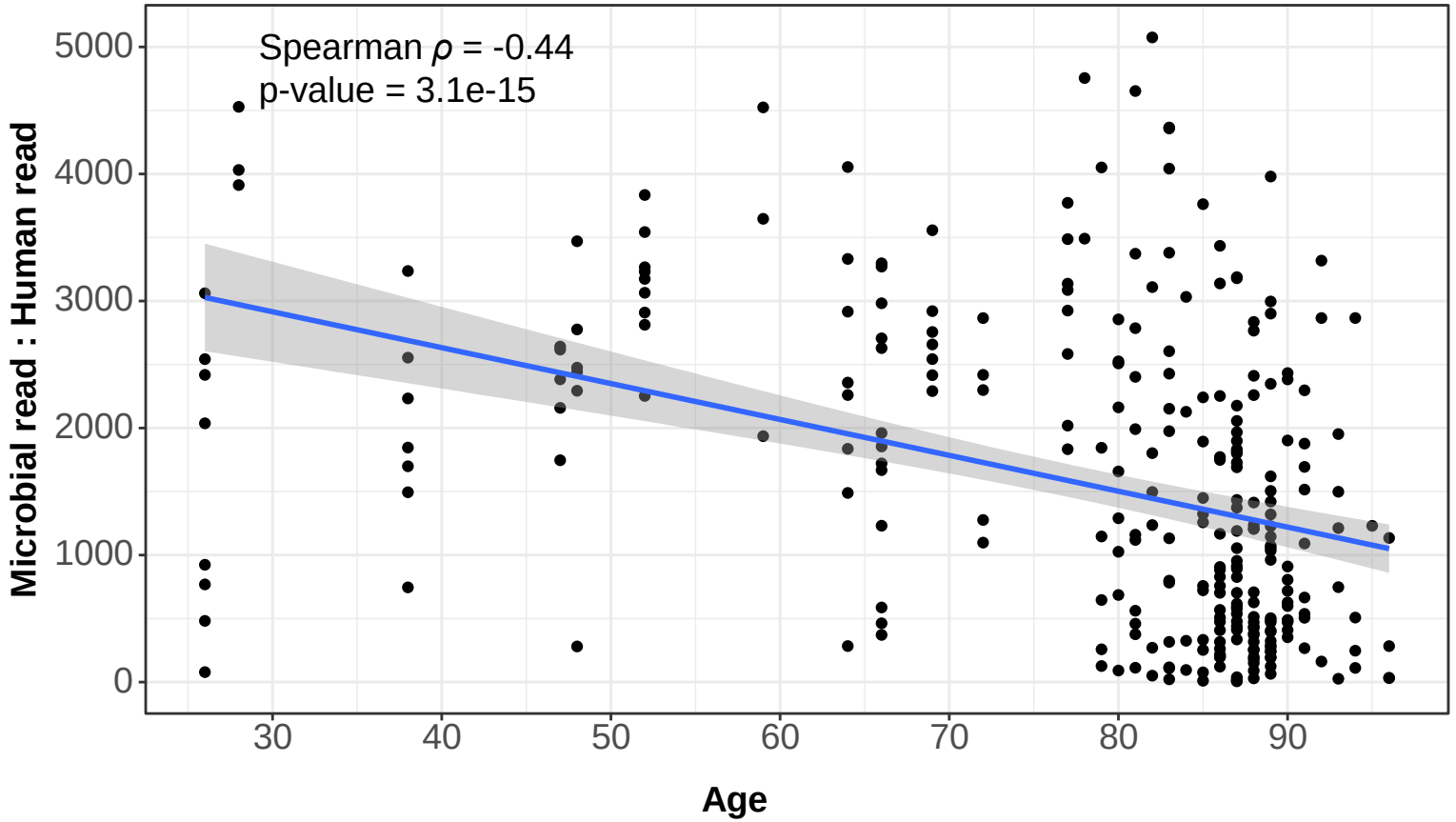
Barcode adapter, double stranded	1st strand: 5'P-GATCGGAAGAGCACACGTCT 2nd strand: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE 1.0	5’AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC* T
Index Primer	5’CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

Supplementary Table 3: Primers and adapter sequences used in this study.

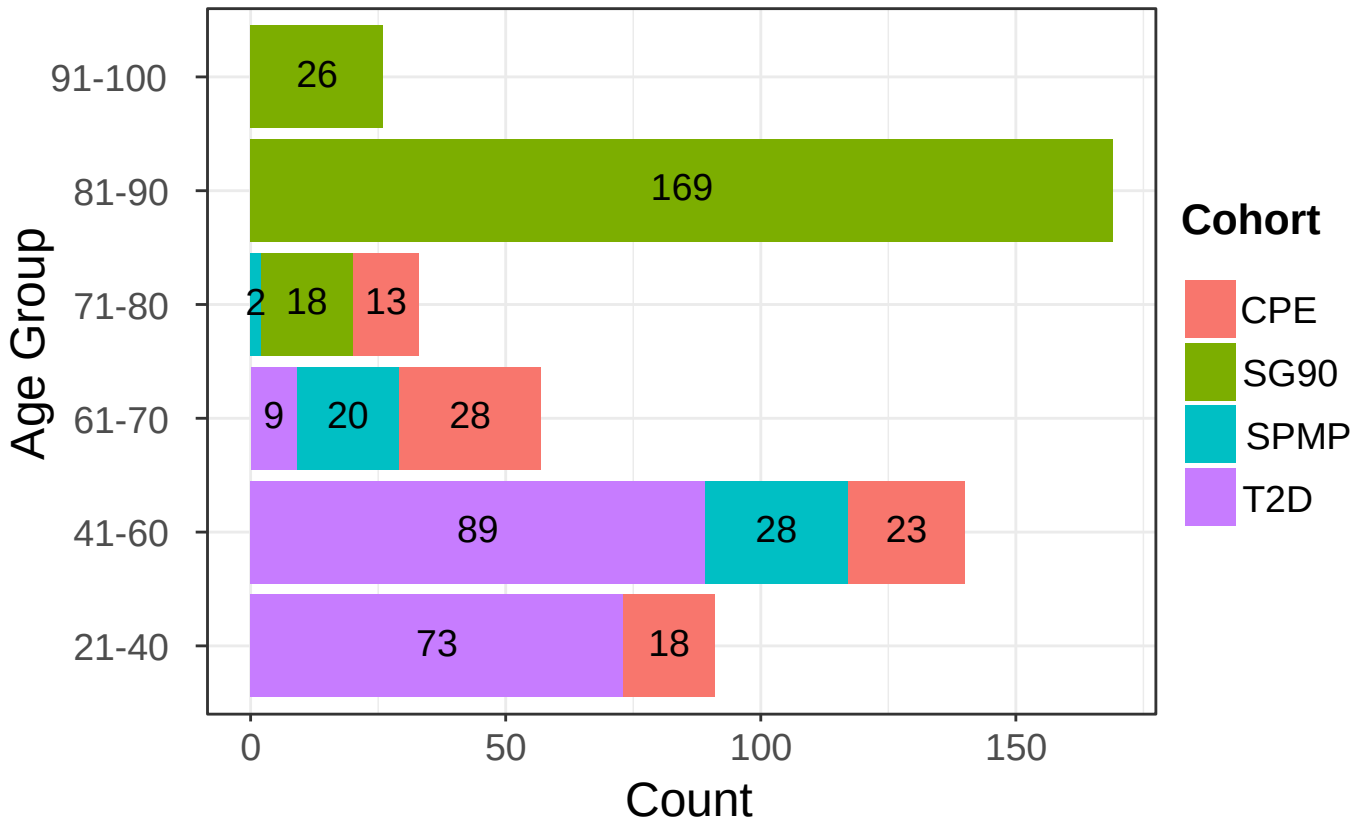
Cohort	Sample collection	DNA Extraction kit	Library Preparation kit	Sequencer	Median Sequencing Depth	Median Microbial Reads
SG90	Frozen stool samples	PowerSoil DNA Isolation Kit	Gene Read DNA Library I Core Kit	Illumina HiSeq 2500	4.3M	2.9M
				Illumina HiSeq X	9.1M	6.4M
SPMP	Frozen stool samples	QIAamp Power Fecal Pro DNA kit	NEBNext Ultra DNA Library Prep Kit	Illumina HiSeq4K	36.7M	31.2M
CPE	Frozen stool samples	PowerSoil DNA Isolation Kit	Gene Read DNA Library I Core Kit	Illumina HiSeq 2500	24M	20.6M
T2D	Frozen stool samples	Custom method[1]	Custom method[1]	Illumina GA	18.2M	NA

1. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).

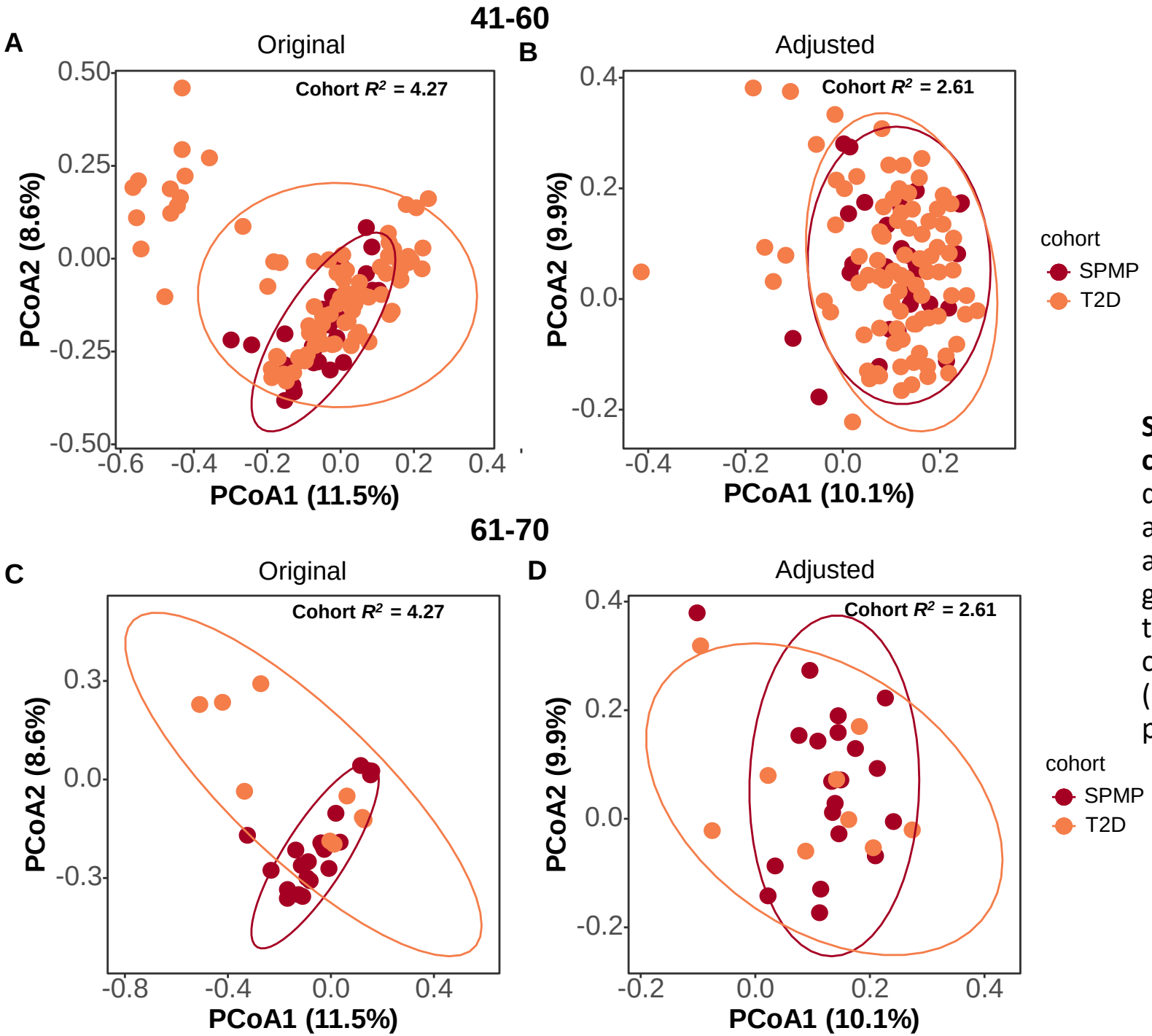
Supplementary Table 4: Experimental methods used in different cohorts for DNA extraction and sequencing.



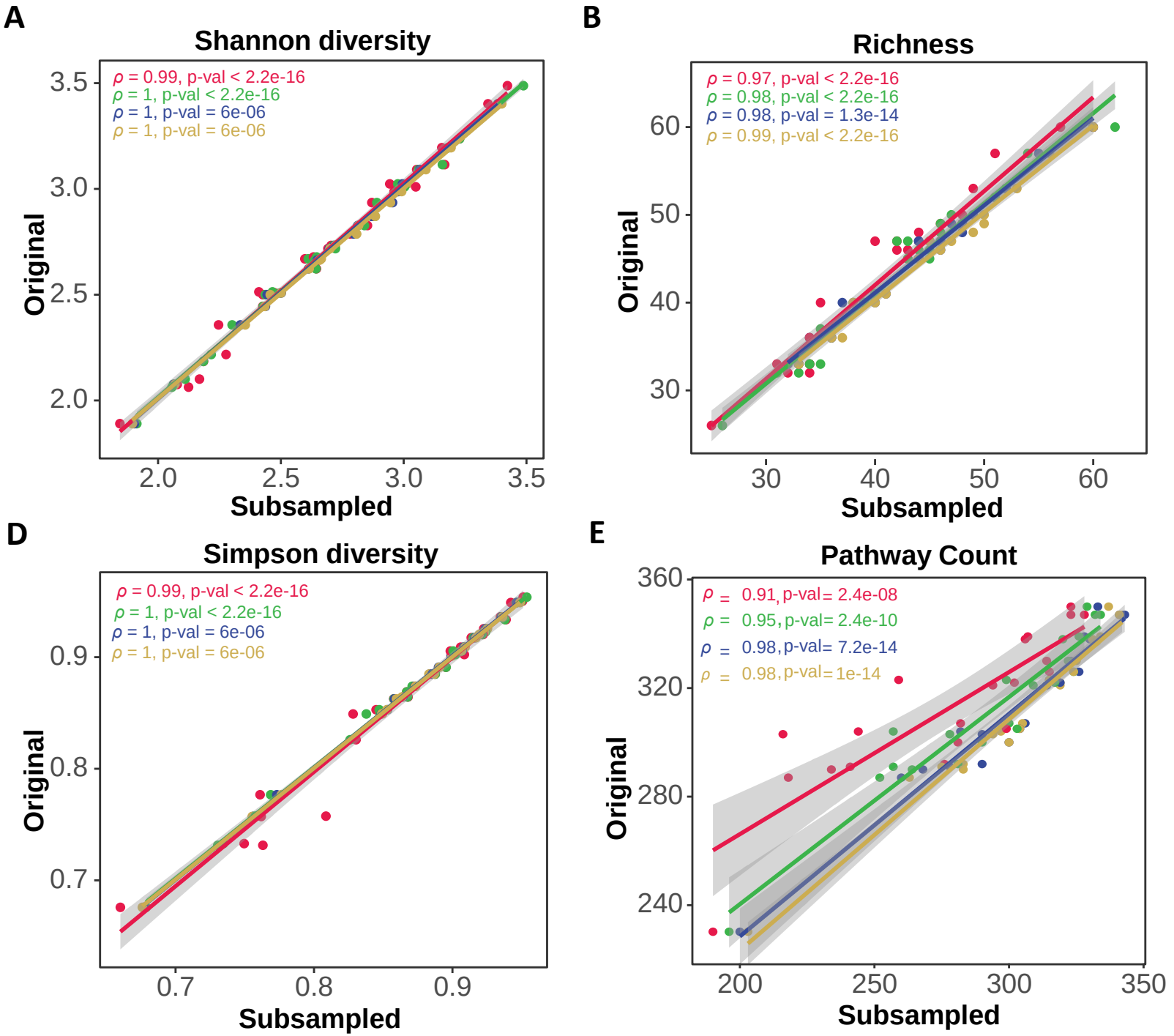
Supplementary Figure 13: Decrease in microbe-to-human read ratio with age. Scatterplot showing microbe-to-human read ratio (as a proxy for microbial biomass compared to human biomass) in relation to the age of subjects for the SG90 and CPE cohorts. Microbe-to-human read ratio was estimated as the proportion of microbial reads relative to human reads in the dataset (minimap2 mapping). Outlier points are not included in the visualization. Regression line (in blue) shows a negative correlation of microbe-to-human read ratio with age. Source data are provided as a Source Data file.



Supplementary Figure 14: Age distribution for the different cohorts. Stacked bar chart showing the number of subjects (indicated in every bar) from different cohorts in each age groups. The number of subjects in each age group is calculated by including the lower and upper end of the range. Each colour in the bar represents specific cohort as indicated in the legend. Source data are provided as a Source Data file.



Supplementary Figure 15: Effect of batch correction. PCoA plots using Bray Curtis dissimilarity for age-groups 41-60 (top panel) and 61-70 (bottom panel) before (A, C) and after (B, D) batch correction. For both age groups, batch correction appears to reduce the variation across cohorts as expected. Each cohort is represented by a colored dot (Orange – T2D, Red – SPMP). Source data are provided as a Source Data file.



Supplementary Figure 16: Correlation plots at different sequencing depths. Correlation plots between metrics for original datasets and subsampled sets at different sequencing depths for (A) Shannon diversity index, (B) Richness, (C) Evenness, (D) Simpson diversity index and (E) Pathway counts. Each colored line in each box represents regression line for different sequencing depths (Pink – 7m, Green – 14m, Blue – 21m, Yellow – 28m). Inset values indicate the Spearman ρ and p-values for the corresponding sequencing depths. The high Spearman correlation values show that all metrics are highly robust to variations in sequencing depth. Source data are provided as a Source Data file.

Supplementary References

1. O'Toole, P.W. & Jeffery, I.B. Gut microbiota and aging. Science **350**, 1214-1215 (2015).

2. Pang, S. et al. Longevity of centenarians is reflected by the gut microbiome with youth-associated signatures. Nat Aging **3**, 436-449 (2023).

3. Odamaki, T. et al. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC Microbiol **16**, 90 (2016).

4. Wu, L. et al. A Cross-Sectional Study of Compositional and Functional Profiles of Gut Microbiota in Sardinian Centenarians. mSystems **4** (2019).

5. Wang, N. et al. Enriched taxa were found among the gut microbiota of centenarians in East China. PLoS One **14**, e0222763 (2019).

6. Wang, F. et al. Gut Microbiota Community and Its Assembly Associated with Age and Diet in Chinese Centenarians. J Microbiol Biotechnol **25**, 1195-1204 (2015).

7. Biagi, E. et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One **5**, e10667 (2010).

8. Rampelli, S. et al. Functional metagenomic profiling of intestinal microbiome in extreme ageing. Aging (Albany NY) **5**, 902-912 (2013).

9. Leite, G. et al. Age and the aging process significantly alter the small bowel microbiome. Cell Rep **36**, 109765 (2021).

10. Kashtanova, D.A. et al. A Cross-Sectional Study of the Gut Microbiota Composition in Moscow Long-Livers. Microorganisms **8** (2020).

11. Biagi, E. et al. The gut microbiota of centenarians: Signatures of longevity in the gut microbiota profile. Mech Ageing Dev **165**, 180-184 (2017).

12. Tuikhar, N. et al. Comparative analysis of the gut microbiota in centenarians and young adults shows a common signature across genotypically non-related populations. Mech Ageing Dev **179**, 23-35 (2019).