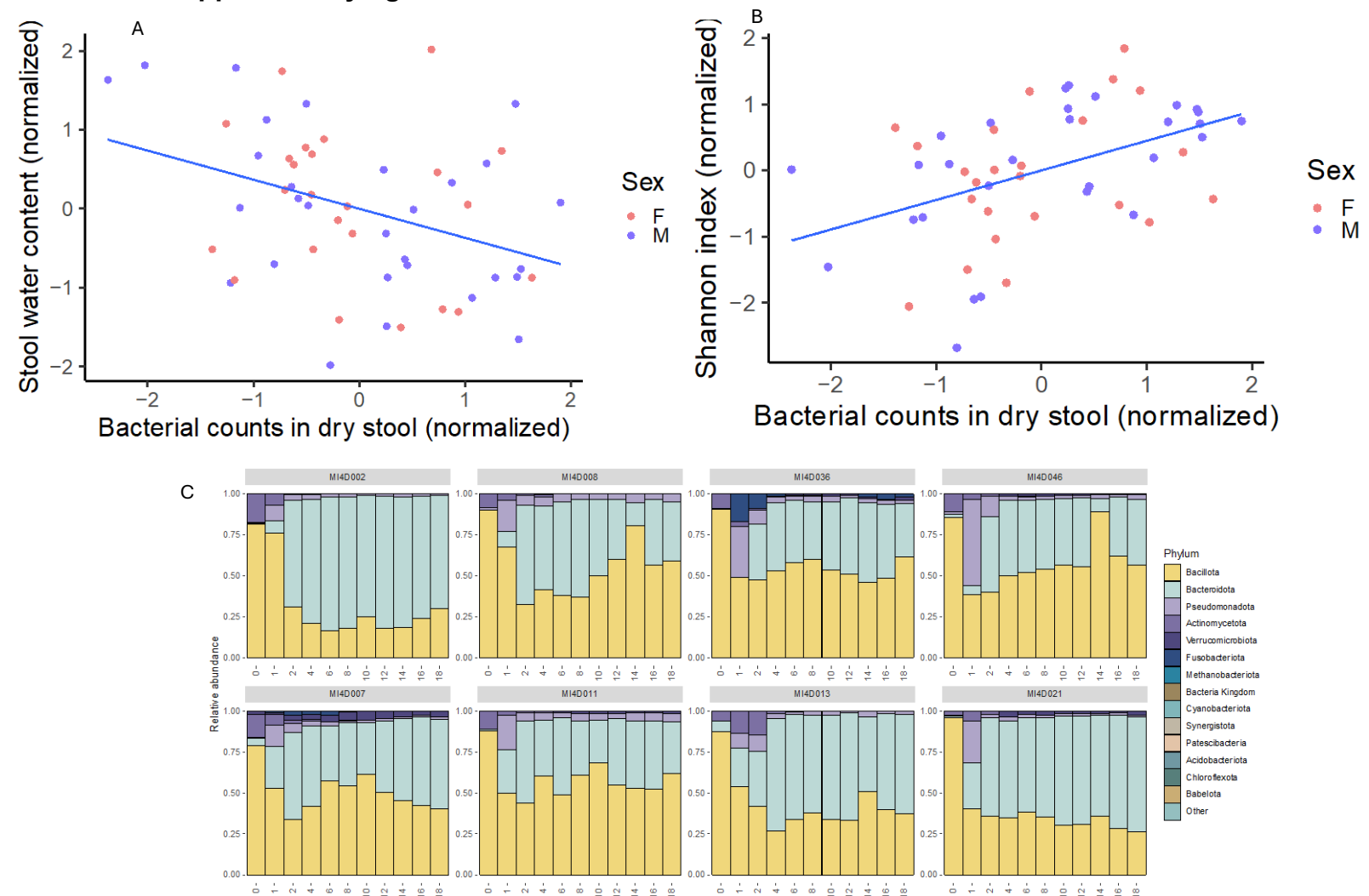


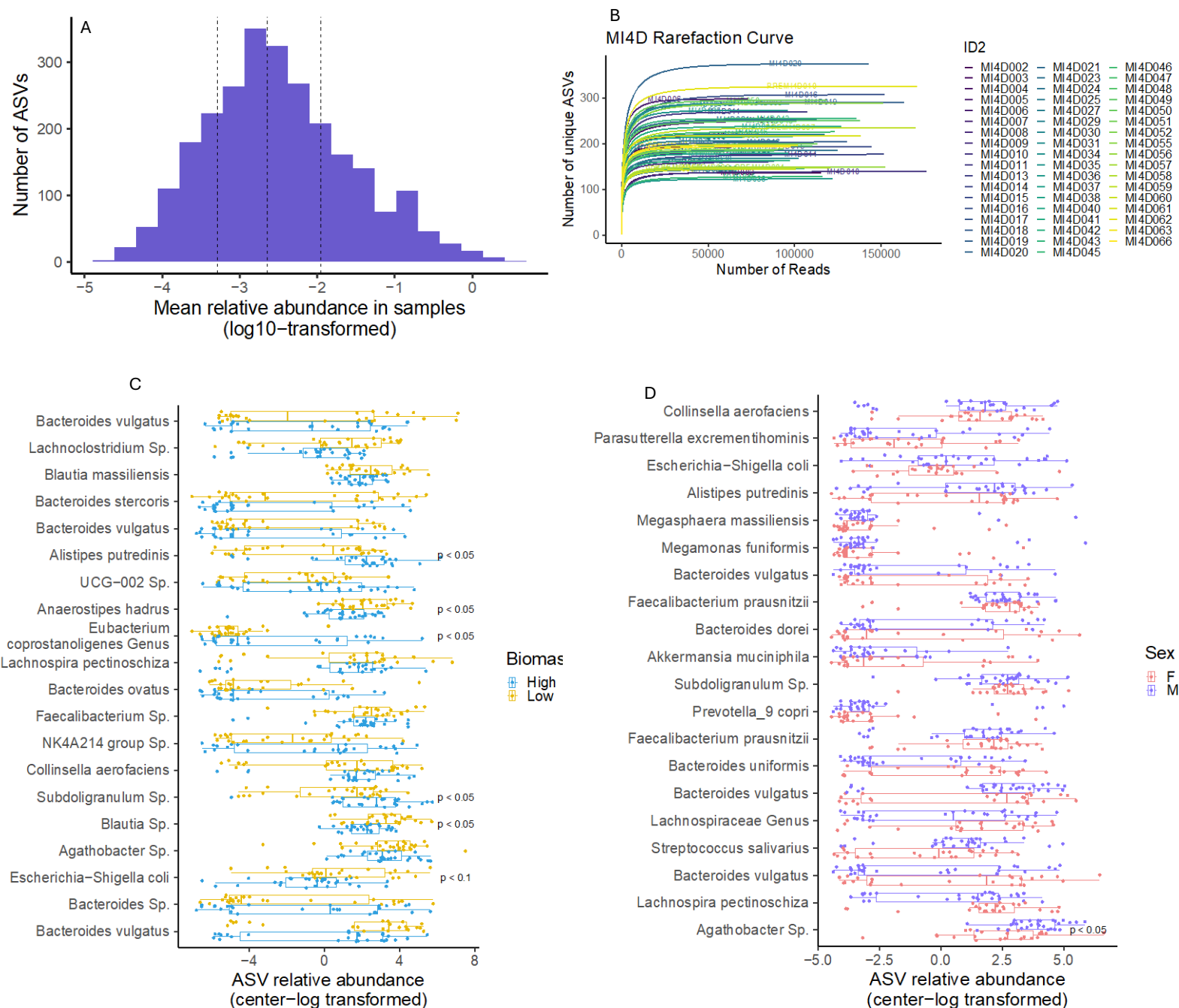
## Supplementary figures



Supplementary Fig1. (A) Association between water content and bacteria biomass in stool samples ( $p = 0.003$ )

(B) Association between Shannon index and bacteria biomass in stool samples from adolescents with obesity ( $p = 0.0008$ ). Normalization of bacterial counts in stool is completed using *caret* v6.0-85 based on center, scale, and Yeo-Johnson transformation.

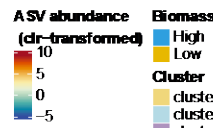
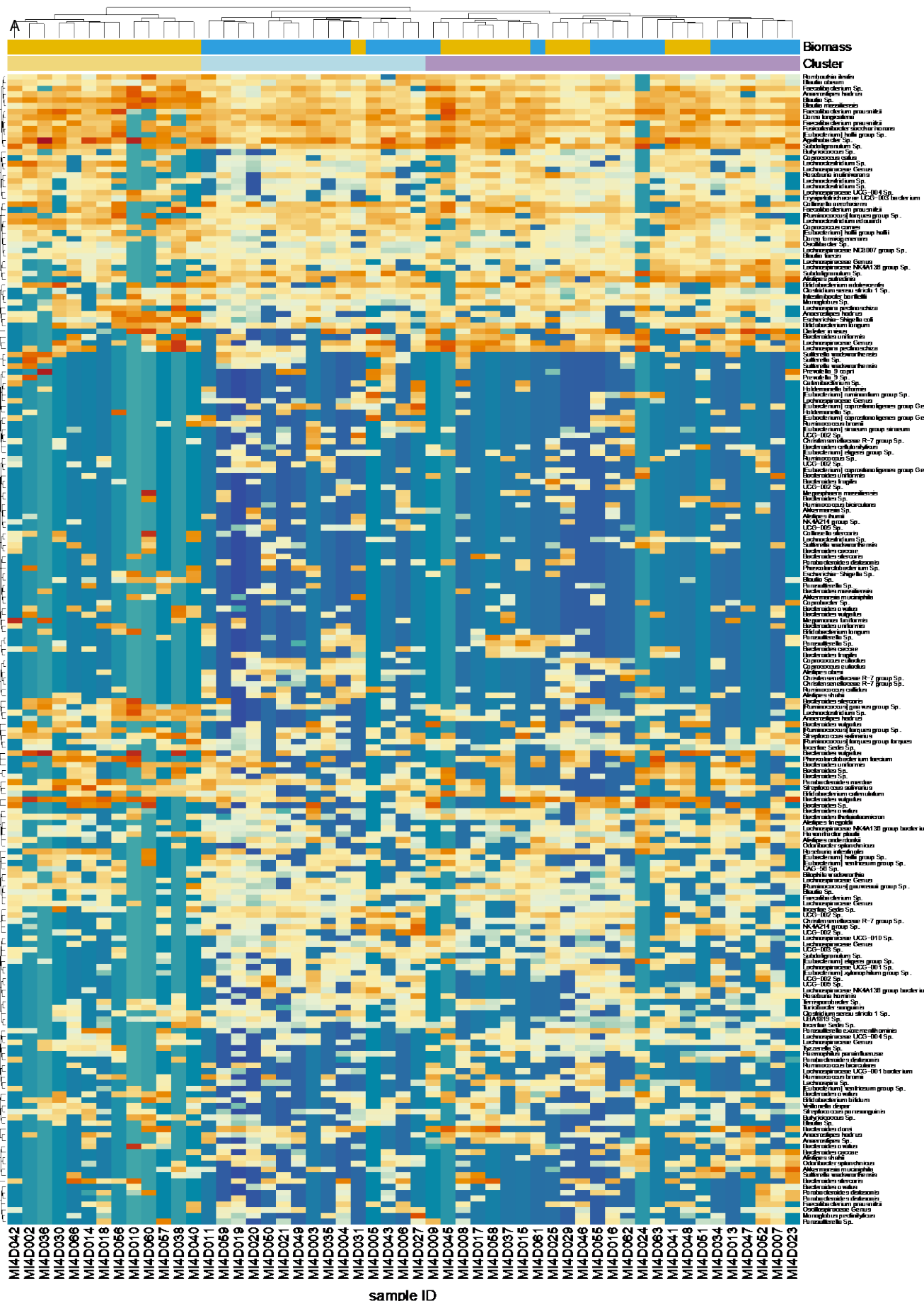
(C) Composition of the gut bacteria communities in the bioreactor culture at the phylum level. Top graphs show cultures from low fecal biomass individuals ( $n=4$ ), and the bottom graphs show cultures from high fecal biomass individuals ( $n=4$ ).



Supplementary Fig2. (A) Histogram of the mean relative abundance of all ASVs identified in fecal samples from adolescents with obesity. The dotted lines represent the first quartile, the median, and the third quartile of the distribution, respectively. (B) Rarefaction curve of 16S rRNA gene sequencing of all stool samples analyzed to

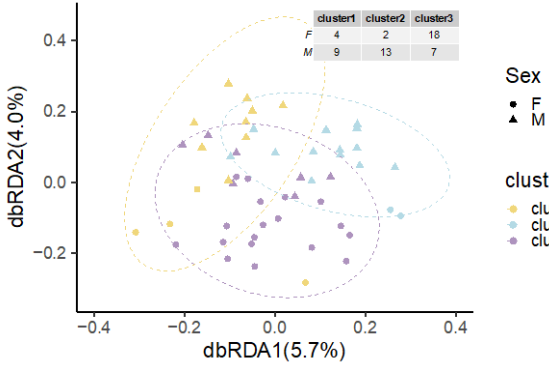
demonstrate sufficient sequencing depth for distinguishing ASVs. (C) Boxplot of the central-log-transformed reads of the top 20 ASV contributing to dbRDA axis1. (D) Boxplot of the central-log-transformed reads of the top 20 ASV contributing to dbRDA axis2. Permutation tests were used to compare differential ASV reads between the two groups. p values were adjusted for multiple testing.

ASVs

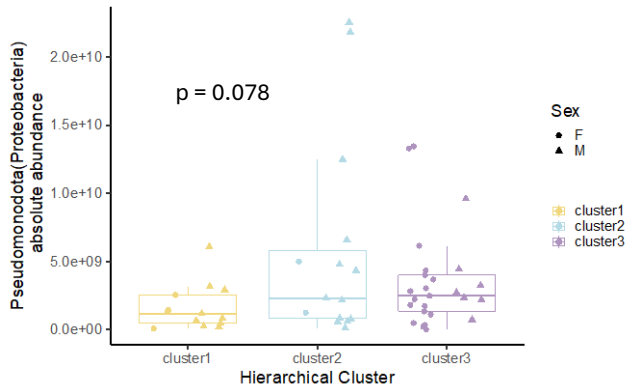


D

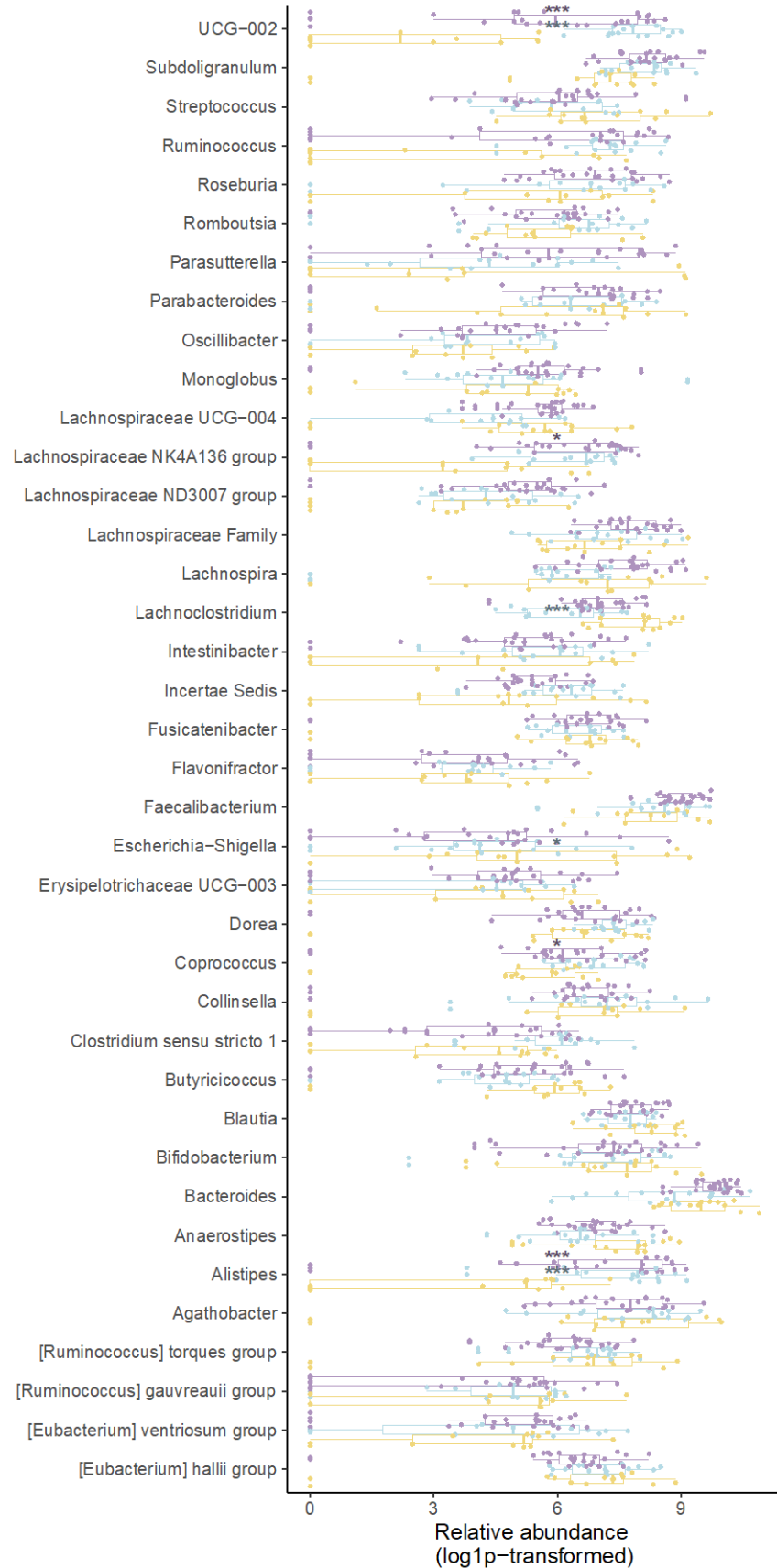
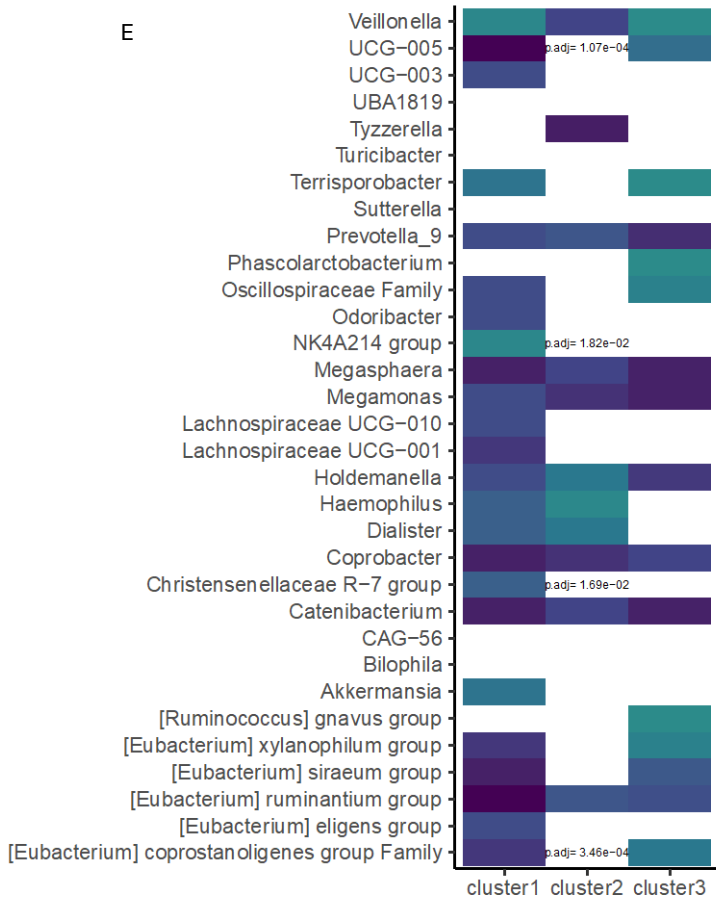
B

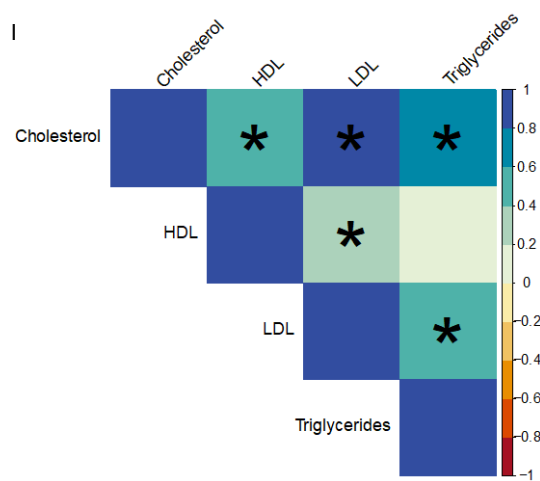
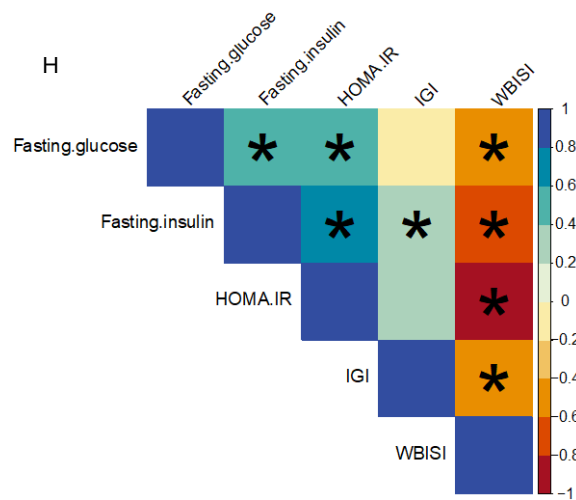
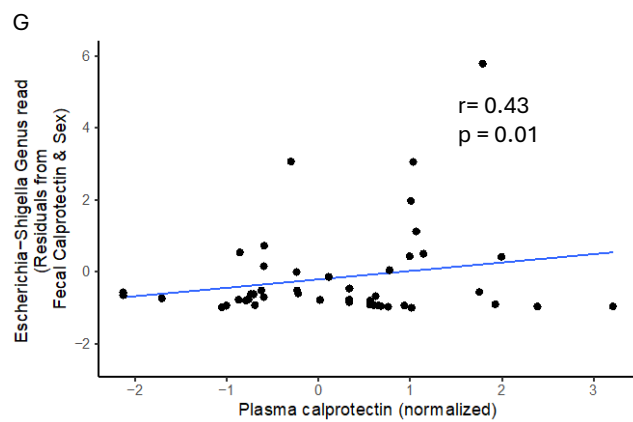
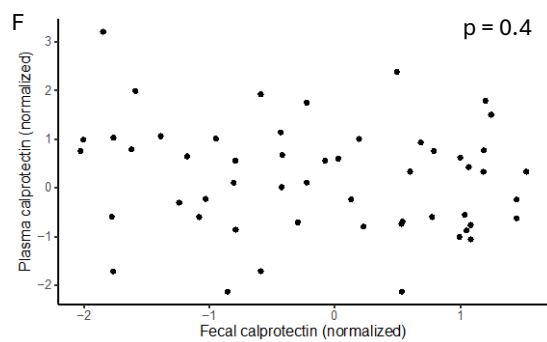


C



E





Supplementary Fig3. (A) Heatmap of central-log-transformed bacteria reads at the ASV level from the core ASVs (n = 199).

(B) Projection of clusters onto dbRDA of Bray-Curtis (BC) distance of gut microbiome, colored by cluster, shape of each sample point correspond to sex. Top-right table of participant sex in each cluster

(C) Boxplot of Pseudomonadota (Proteobacteria) absolute abundance in the three clusters (one-way ANOVA,  $p = 0.078$ )

(D) Boxplot of log1p-transformed reads of highly prevalent (>75%) genera (n = 38). Labeled are genera in clusters 2 and 3 with differential abundance from cluster 1 after multiple testing correction (Gamma generalized linear regression, adjusted for age, sex, BMI)  $p_{\text{adj}} < .001$  marked as "\*\*\*\*",  $p_{\text{adj}} < .01$  as "\*\*\*",  $p_{\text{adj}} < .05$  as "\*\*"

(E) Heatmap of genera with prevalence < 75% in entire cohort, colored by prevalence within cluster (n = 32). Labeled are genera with differential prevalence across clusters after multiple testing correction. (Chi-squared test)

(F) Scatterplot of plasma calprotectin and fecal calprotectin ( $p = 0.4$ )

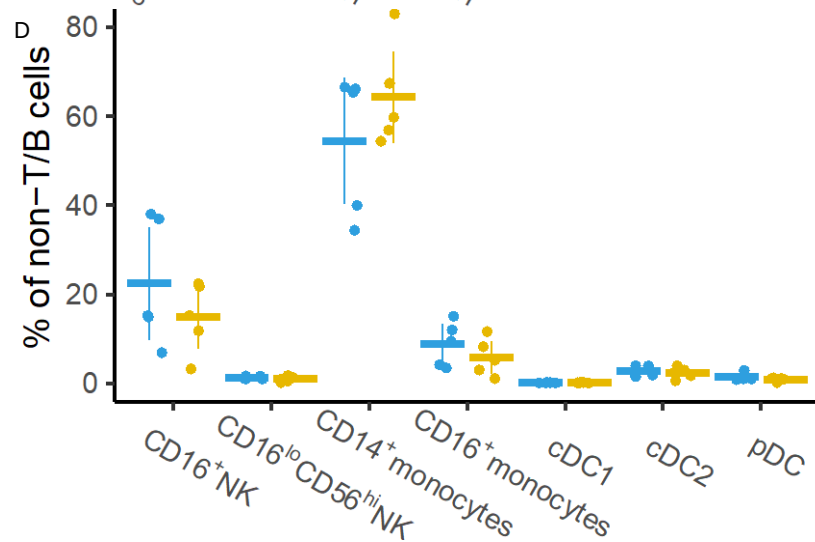
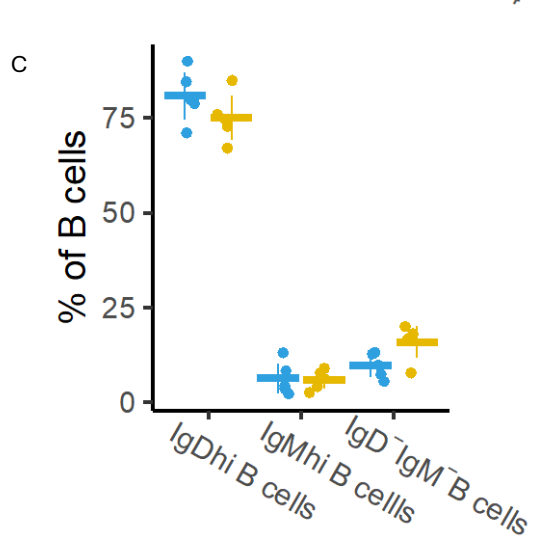
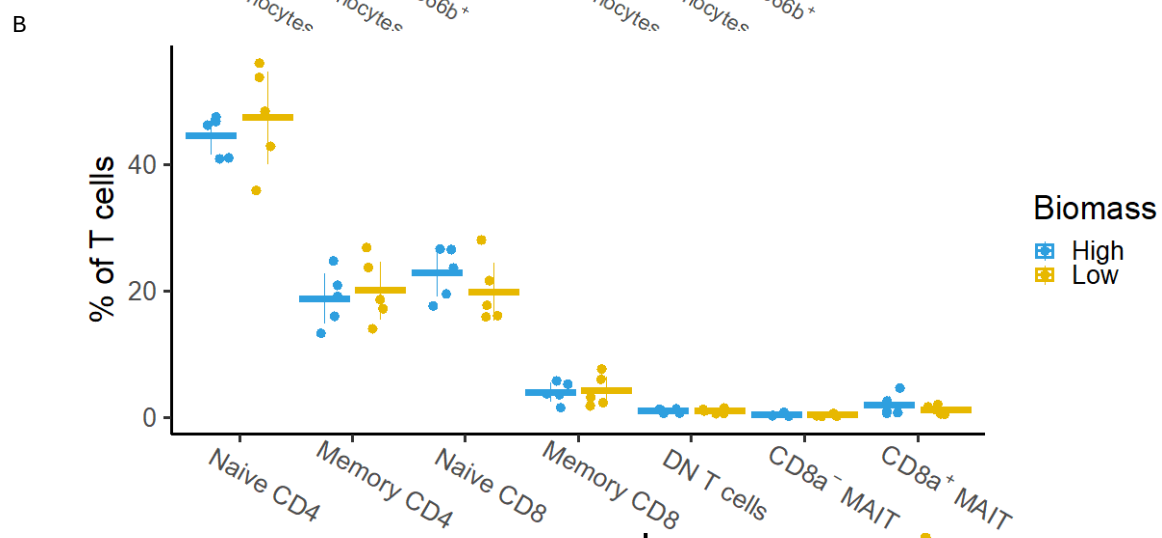
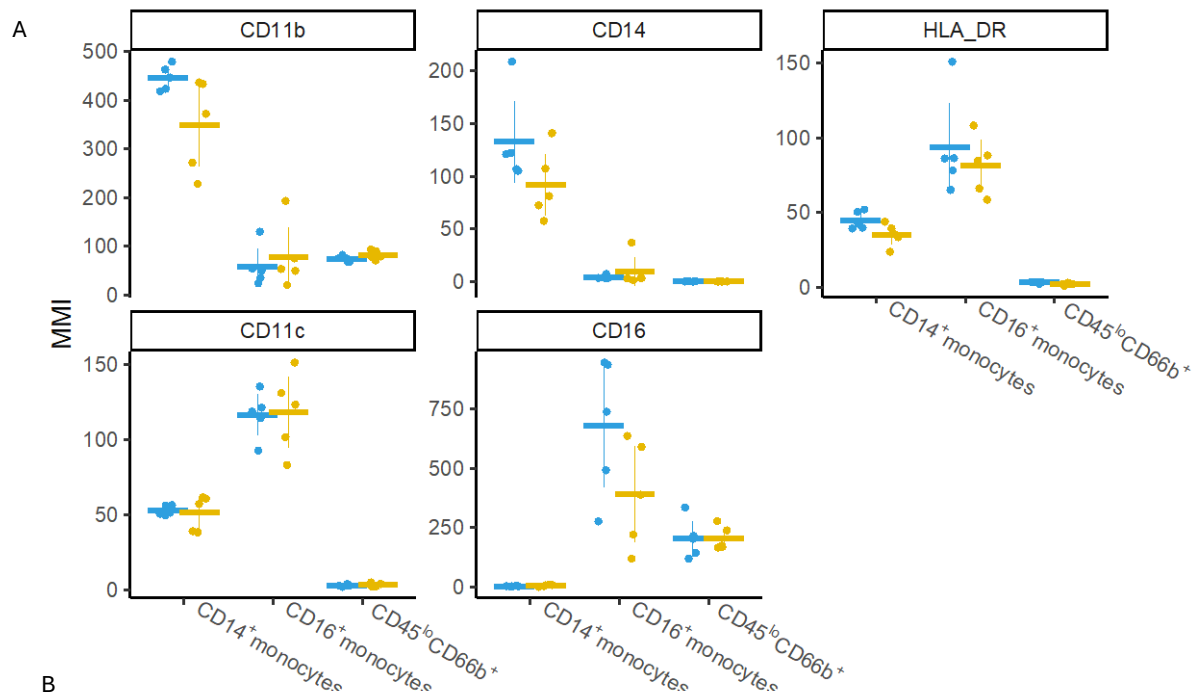
(G) Scatterplot of plasma calprotectin and residuals from Gamma generalized linear regression with fecal calprotectin and sex as plotted in Fig 3D. ( $p=0.01$ )

(H) Correlation heatmap between measures of pancreatic  $\beta$  cell function. HOMA-IR: Homeostatic model assessment for insulin resistance; IGI: insulinogenic index; WBISI: whole-body insulin sensitivity index

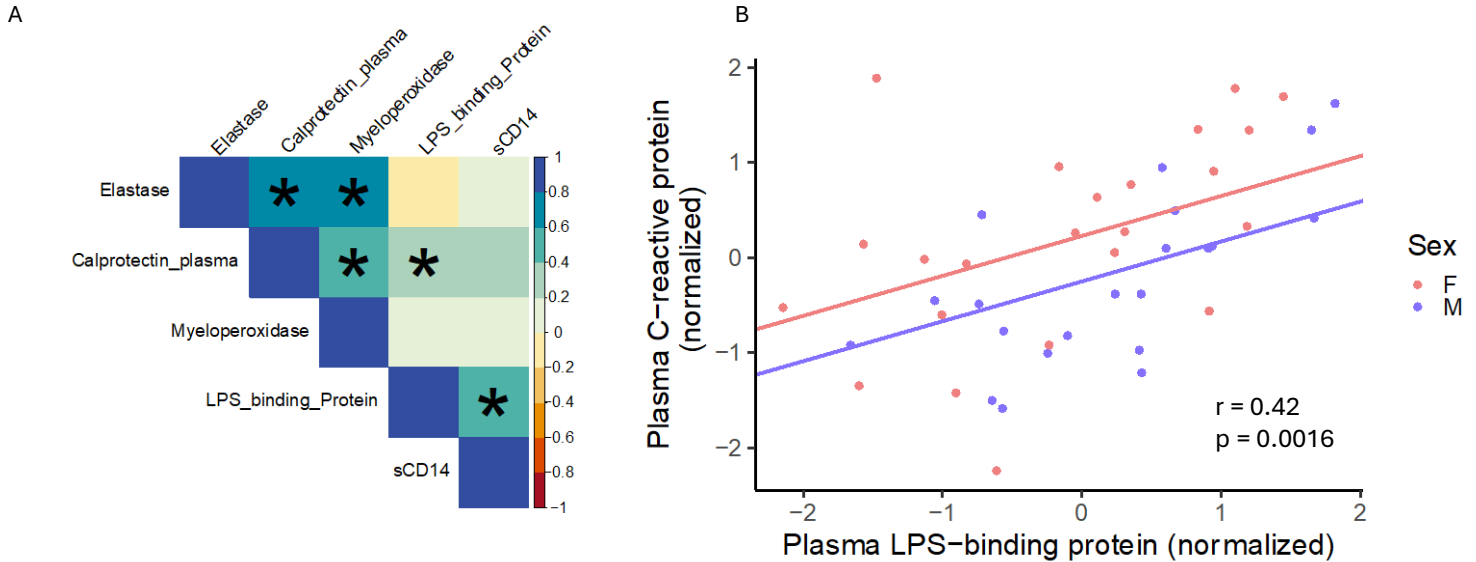
(I) Correlation heatmap between plasma lipids. HDL: High Density Lipoprotein; LDL: Low Density Lipoprotein. Correlation coefficients are colored with dark blue as perfect

positive correlation and dark red as perfect negative correlation. Significantly correlated ( $p < 0.05$ ) variables are marked with a black asterisk.





Supplementary Fig4. (A) Dot plots show median metal intensity (MMI) of markers on classical (CD14<sup>+</sup> CD16<sup>-</sup>) and non-classical (CD14<sup>-</sup> CD16<sup>+</sup>) monocytes, and LD neutrophils (CD45<sup>lo</sup> CD66b<sup>+</sup>) in PBMC. (B-D) Dot plots show frequencies of gated cell subsets in (B) TCRαβ cells, (C) B cells, and (D) non-T/B cell populations. Gating strategy is described in Methods. Crossbar represents mean frequency in each biomass group, error bar represents two standard deviations from the mean on each side. No subsets showed differential abundance between groups ( $p > 0.05$  for all cell subsets, permutation test).



Supplementary Fig5. (A) Correlation heatmap between three neutrophil-associated proteins and the two endotoxin defense proteins. sCD14: soluble CD14. Correlation coefficients are colored with dark blue as perfect positive correlation and dark red as perfect negative correlation. Significantly correlated ( $p < 0.05$ ) variables are marked with a black asterisk. (B) Association between C-reactive protein level and LPS-binding protein (LBP:  $p = 0.0016$ , Sex:  $p = 0.04$ , BMI:  $p = 0.0009$ )