Supplementary Figures for

Association Between Dietary Polyphenol Intake and Polyphenol-Utilizing Bacteria in Healthy Adults

Stephanie M.G. Wilson, Andrew Oliver, Zeynep Alkan, Bhimanagouda S. Patil, Mary E. Kable, Danielle G. Lemay

Corresponding author:
Dr. Danielle G. Lemay, dglemay@usda.gov

Table of Contents

Figure S1. **STROBE analytical sample size flowchart of this secondary analysis of the USDA Nutritional Phenotyping Study.**

Of the 1768 individuals interested in the study, 393 individuals enrolled into the study and 358 completed the study. Of those who completed the study, 313 individuals had both dietary and fecal shotgun metagenomic data and were used for analysis. Insufficient DNA quality was determined if DNA was low in concentration (<100 ng/µL) or quality ($A_{260/280}$ of <1.78; $A_{260/230}$ of <1.72).

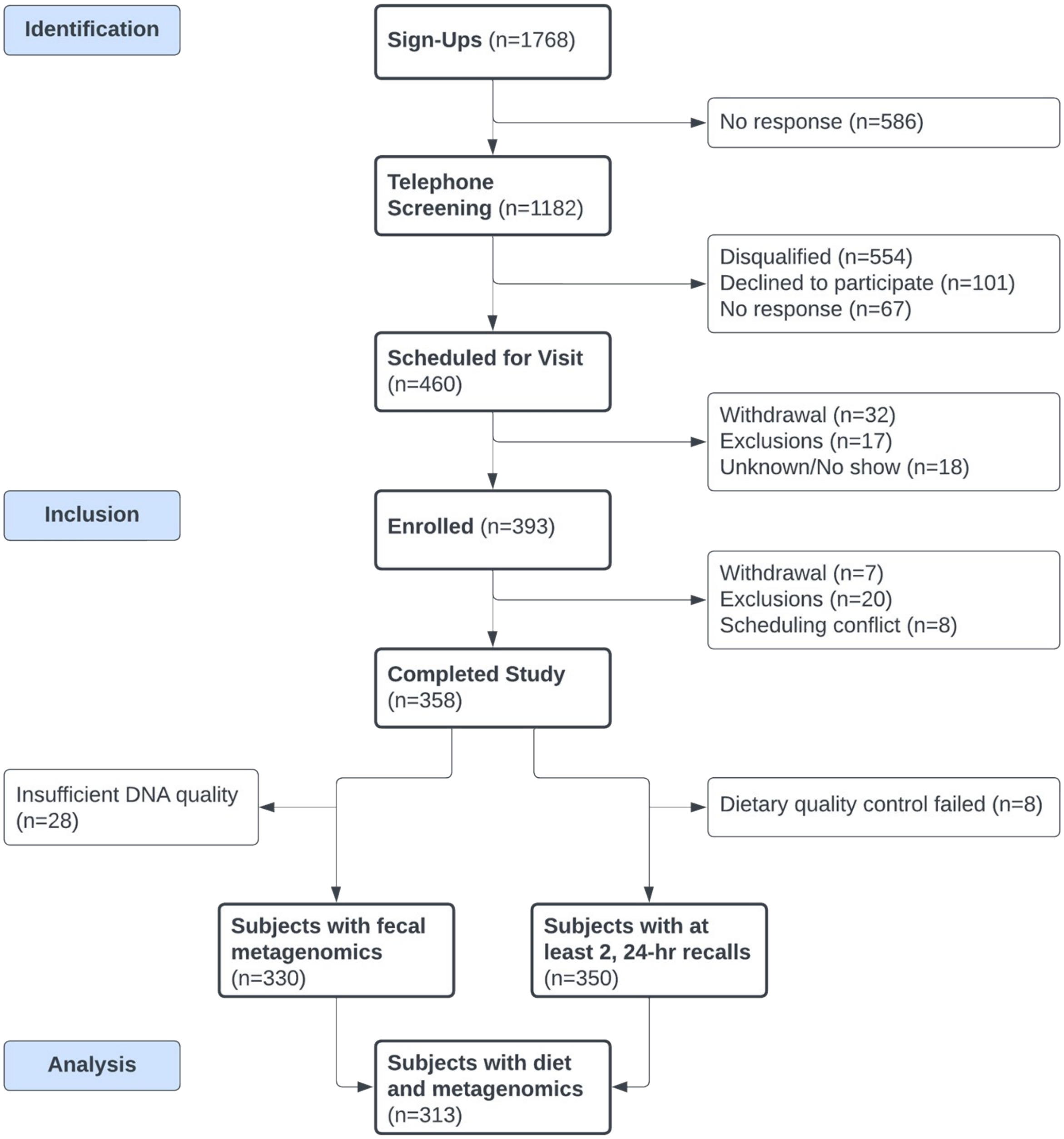

Figure S2. **Variance explained by polyphenol intake at each taxonomic level of the gut microbiome as determined by PERMANOVA.**

P-values for the polyphenol intake predictor are presented above each bar. Polyphenol intake grouping was set at the lowest and highest quartile of total polyphenol intake. PERMANOVA models accounted for age, body mass index, sex, fiber intake, and dietary quality as represented by the total Healthy Eating Index 2015.

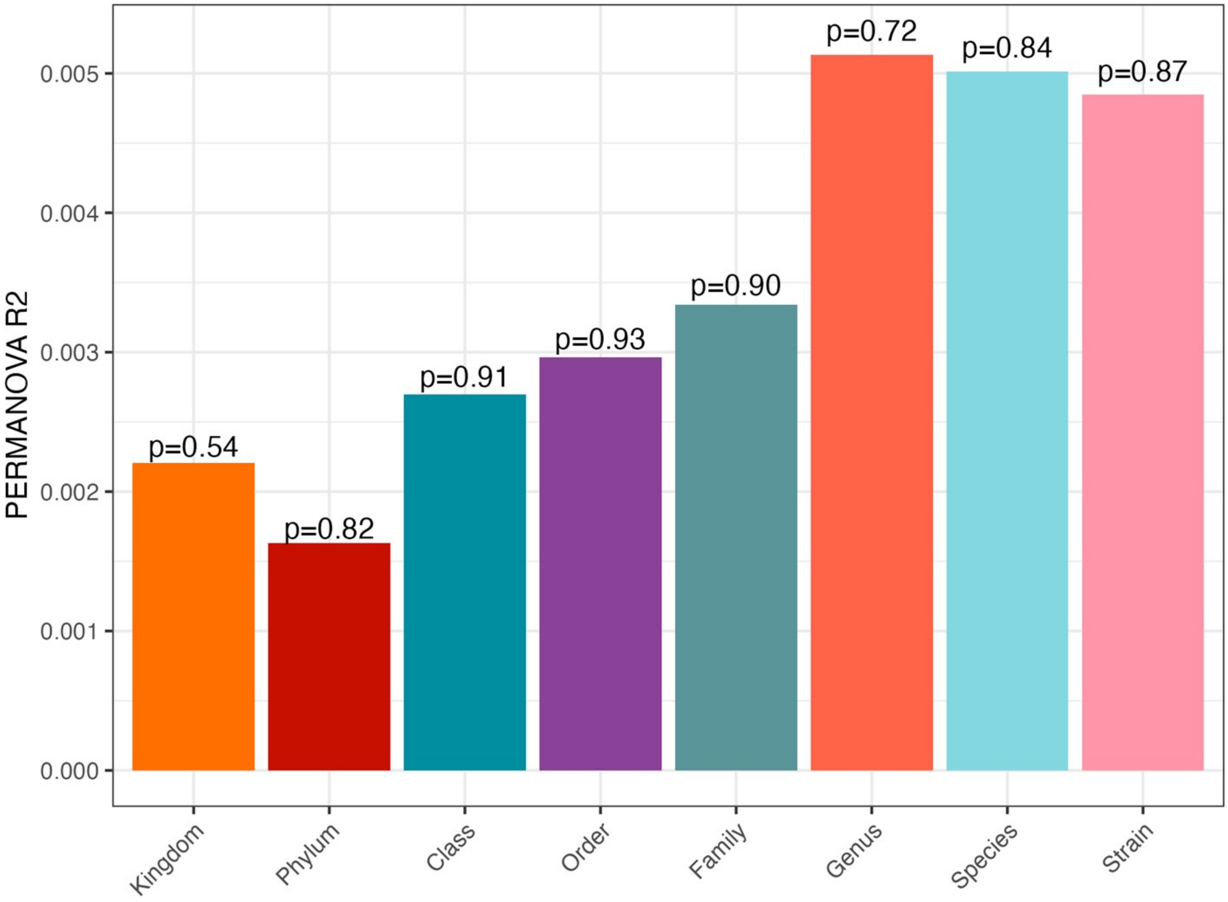

Figure S3. **Microbial phyla mapping to polyphenol utilization protein (PUP) reads**.

PUPs and their corresponding UniProt IDs are listed on the y-axis, faceted by protein family classification. Color scale represents the proportion of participants that contain the PUP mapping to the respective phylum, with darker red colors indicating greater number of participants that contained PUPs mapping to the bacterial phylum. Abbreviations: Functional group transfer, FR; Hydrolysis, HR; Isomerization, IR; Non-hydrolytic cleaving, NCR; Oxidation/Reduction, OR; Unclassified; UC

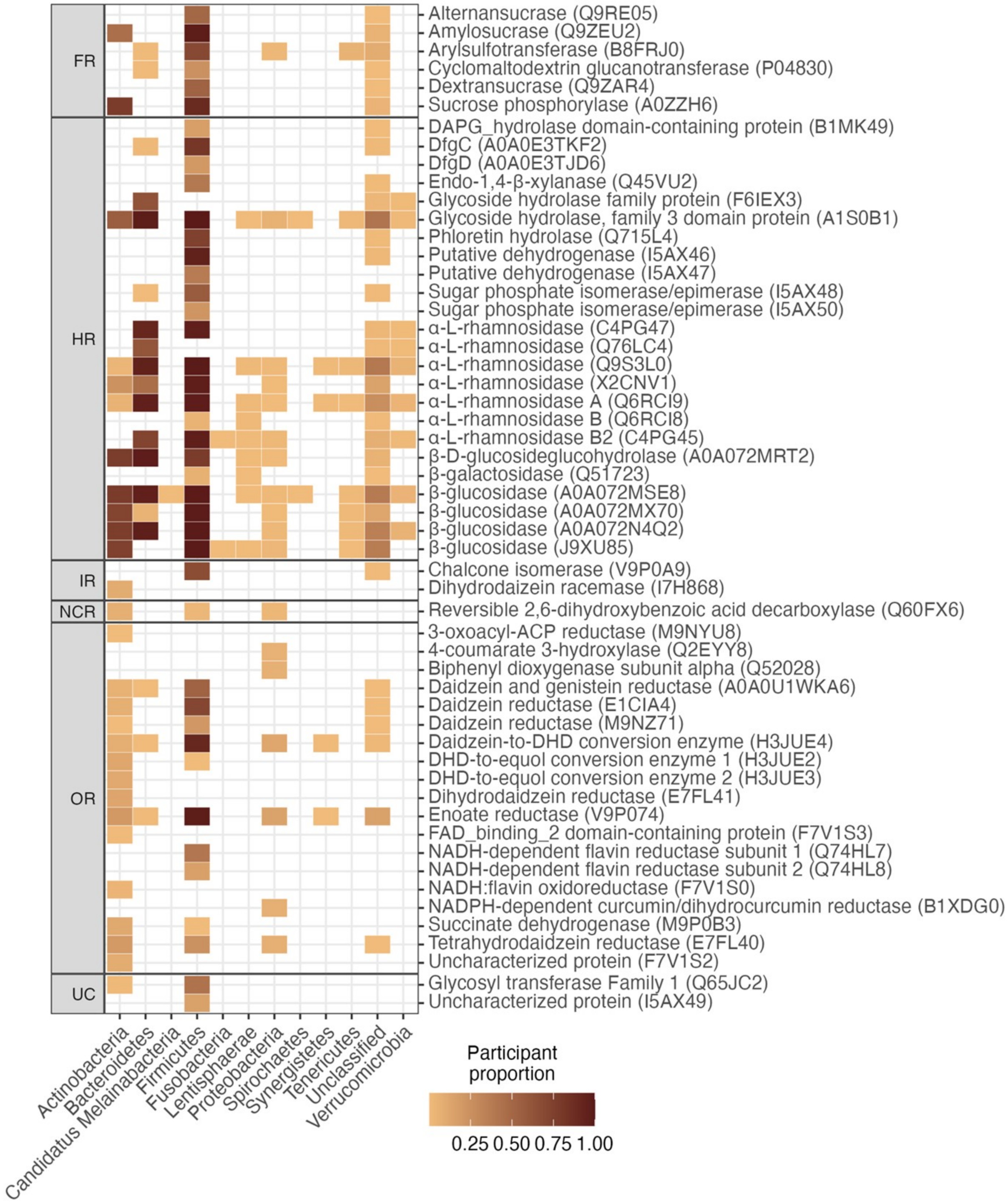

Figure S4. **Distinct polyphenol utilization proteins (PUPs) by bacterial family and reaction class**.

Each tile represents the number of distinct PUPs identified within a bacterial family (y-axis, right) associated with a specific reaction class (x-axis). Left facet indicates phylum-level taxonomy. Color intensity reflects the number of distinct PUPs observed. Only families with at least one annotated PUP are shown. Abbreviations: Functional group transfer, FR; Hydrolysis, HR; Isomerization, IR; Non-hydrolytic cleaving, NCR; Oxidation/Reduction, OR; Unclassified; UC.

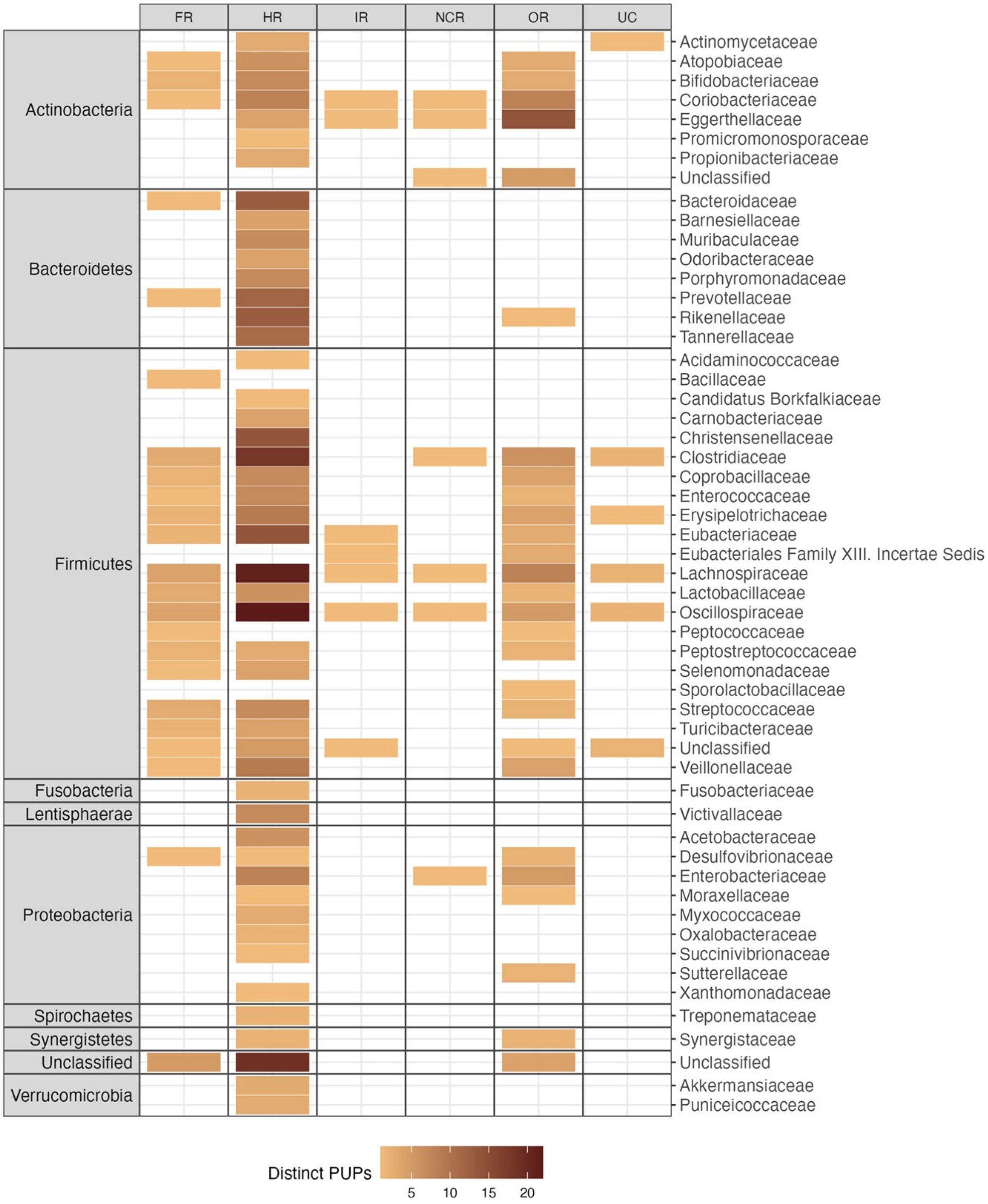

Figure S5. **Relationships between top machine learning model features and plasma LPS-binding protein**.

Top 5 features from each random forest regression model are shown, with feature importance determined by averaging absolute Shapley Additive exPlanations values. Spearman correlations were performed for numeric model features, and a Wilcoxon signed rank t-test were performed for categorical features. Red diamonds on categorical features represent the group mean. BMI, body mass index; HEI, Healthy Eating Index.