Food & Function

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d6fo00158k

Association between dietary polyphenol intake and polyphenol-utilizing bacteria in healthy adults

Stephanie M. G. Wilson, [a,b,c] Andrew Oliver, [a] Zeynep Alkan, [b] Bhimanagouda S. Patil, [c,d] Mary E. Kable [a,b] and Danielle G. Lemay [*†a,b,c]

Dietary polyphenols are bioactive compounds with a bidirectional impact on the gut microbiome; they shape the microbial community and are transformed through bacterial metabolism. However, there are limited studies pairing metagenomic and dietary data to investigate the relationship between polyphenol intake and the taxonomic and functional profiles of the human gut microbiome. We examined if dietary polyphenol intake associates with microbial composition and polyphenol utilization capacity. Healthy adults participated in a cross-sectional study balanced for age, sex, and BMI. Polyphenol intake was previously estimated by mapping multiple 24 h dietary recalls to the Food Database (FooDB). We coupled intake with microbial taxonomic and functional profiles from shotgun-sequenced fecal metagenomes (n = 313). Microbial reads were mapped to dbPUP, a database with 60 experimentally characterized, gut-associated polyphenol utilization proteins (PUPs). We assessed the relationship of polyphenol intake on microbial diversity, abundance of microbes with PUP genes, PUP gene counts, and select lipopolysaccharide (LPS) producers, accounting for age, sex, BMI, fiber intake, and diet quality. Specific polyphenols associated with an increased abundance of nine PUP-containing genera. We found 117 associations between polyphenol intake and microbial PUP genes, with 85 associations involving hydrolysis PUPs. Diversity in polyphenol intake was positively associated with diversity in PUP genes but not with microbial diversity. Lastly, we detected a positive relationship between intake of olive-related polyphenol classes and abundance of order *Bacteroidales*, a producer of immunoinhibitory LPS. Dietary polyphenol intake may influence the gut microbiome's capacity for polyphenol utilization, particularly its hydrolytic activity, without impacting taxonomic diversity or composition.

Received 12th January 2026,
Accepted 21st May 2026

DOI: 10.1039/d6fo00158k

rsc.li/food-function

Introduction

Polyphenols are dietary bioactive compounds often consumed through fruits, vegetables, beverages, and spices that can also improve cardiometabolic health in the host.[1–3] A small fraction of ingested polyphenols is absorbed in the small intestine with most polyphenols passing to the colon where they can be metabolized by, and help shape, the gut microbiome.[4] Although polyphenols can disrupt microbial cell walls and membranes as well as energy pathways to inhibit bacterial growth,[5–7] it has been hypothesized that some bacteria developed pathways to break down polyphenols to combat their antimicrobial effects. Recent evidence suggests that polyphenols can also be directly utilized by select bacteria to promote growth[8–11] and promote short chain fatty acid production.[9]

Metabolism of polyphenols yields a diverse set of derivatives, with high variability between individuals. Microbial transformations of polyphenols include demethylation, decarboxylation, dihydroxylation, hydrolysis, isomerization, reduction, and ring-cleavage[12] with microbes displaying selective usage of polyphenol substrates.[13–15] Our understanding of microbial polyphenol metabolism has been largely informed by *in vitro* and *in vivo* models with purified polyphenols and prespecified microbes or microbial communities.[10] Polyphenol associated enzymes have been increasingly identified and characterized in human gut microbiomes.[10,16,17] Hadza hunter-gatherers were observed to have a greater functional capacity to metabolize polyphenols than Americans[18] with authors attributing this to dietary differences based on geographical location. Dietary data can provide insight into substrate availability for microbial populations as polyphenols are differentially absorbed and metabolized based on their physio-

[a]*Department of Nutrition, University of California, Davis, Davis, California 95616, USA. E-mail: danielle.lemay@usda.gov*
[b]*United States Department of Food and Agriculture, Agricultural Research Service Western Human Nutrition Research Center, Davis, California 95616, USA*
[c]*Texas A&M AgriLife Research, Institute for Advancing Health Through Agriculture, College Station University, Texas 77843, USA*
[d]*Vegetable and Fruit Improvement Center, Department of Horticultural Sciences, Texas A&M University, College Station, Texas 77843, USA*
† Present address: 430 West Health Sciences Drive, Davis, CA 95616, USA.

chemical properties.[13,19] Dietary patterns may influence microbial functional profiles related to polyphenols as suggested by a greater prevalence of equol producers among vegetarians and East Asian populations, who are more likely to consume soy-based foods.[20] Thus, the addition of dietary polyphenol intake to a systematic screening of polyphenol utilization genes may inform how different polyphenol intakes influence the polyphenol utilization capacity of the human gut microbiome.

Combining taxonomic and functional profiles of the gut microbiome may help explain the variability in polyphenol-associated health effects.[21] In this secondary analysis of the USDA Western Human Nutrition Research Center Nutritional Phenotyping Study, we combine our previous estimations of dietary polyphenol intake[22] with fecal shotgun metagenomes to assess relationships between polyphenol intake, the gut microbiome, and gastrointestinal permeability in healthy U.S. adults. We examine both taxonomic and functional profiling of the gut microbiome, targeting the polyphenol-associated functional capacity by mapping fecal shotgun metagenomes to a database of experimentally verified polyphenol utilization proteins (PUPs).[18] Our primary hypotheses were as follows: (1) dietary polyphenol intake is positively associated with the abundance of microbes with PUP genes and (2) diverse polyphenol intake is positively correlated with microbial diversity. Additionally, because we previously found a negative relationship between the plasma concentration of lipopolysaccharide (LPS) binding protein and the intake of polyphenols classified as prenol lipids and phenylpropanoic acids,[22] we hypothesized that intake of these two polyphenol classes correlates with the abundance of LPS producers. Lastly, we explored whether plasma LPS-binding protein, an indicator of gastrointestinal permeability, is associated with the diversity of PUP genes and the abundance of microbes carrying PUP genes.

Methods

Study design

The USDA Nutritional Phenotyping Study is an observational, cross-sectional study of healthy adults performed from May 2015 to July 2019 by the USDA Western Human Nutrition Research Center. The study protocol was approved by the Institutional Review Board at the University of California, Davis and is prospectively registered at ClinicalTrials.gov (NCT02367287). Written informed consent was obtained from all study participants. The study was performed in accordance with the Declaration of Helsinki. The work presented here is a secondary analysis of the original study which aimed to profile participants for their immunologic, physiologic, neuroendocrine, and metabolic responses to a dietary challenge and physical fitness challenge. Additional details can be found in the USDA Nutritional Phenotyping Study design paper.[23]

Participants

Generally healthy adults (without active or medicated chronic disease) were recruited from in the greater Davis, California area. Eligible adults were 18–65 years old, male or female, and had a body mass index (BMI) between 18.5 and 45 kg m^{-2}. Age, BMI, and sex categories were used to create an 18-bin sampling scheme to balance enrollment across seasons and years. Participants were excluded if pregnant, lactating, had a systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg measured on three occasions, had a diagnosed active chronic disease (including gastrointestinal disorders) treated with medication, had a recent minor (<4 weeks) or major (<16 weeks) surgery, had recent antibiotics (<4 weeks), had a recent hospitalization (<4 weeks), or use of prescription medications affecting outcomes of interest in the USDA Nutritional Phenotyping Study.[23]

In total, 393 participants enrolled into the study, 350 adults completed the two visit study with at least two dietary records that passed quality control, and 330 participants had fecal samples with sufficient DNA quality for shotgun metagenomic sequencing. A total of 313 participants had both dietary data and fecal shotgun metagenomes. The STROBE flow diagram is provided in Fig. S1.

Diet and polyphenol estimation

Dietary intake data for 24-hour recalls were collected and analyzed using the Automated Self-Administered 24-hour (ASA24) Dietary Assessment Tool, versions 2014 and 2016, developed by the National Cancer Institute, Bethesda, MD. Participants completed a training recall with study staff then were asked to complete three additional 24-hour recalls at home within 10–14 days. All recalls underwent quality control methods previously developed for the ASA24.[24] From these data, we extracted daily caloric intake, daily fiber intake (g per 1000 kcal), and the total Healthy Eating Index (HEI) score as an indicator of diet quality in compliance with the 2015–2020 Dietary Guidelines for Americans.

Reported foods were disaggregated to ingredients using a semi-automated process leveraging built-in recipe codes from the Food and Nutrient Database for Dietary Studies (FNDDS).[25] Ingredients were mapped to polyphenols in the Food Database (FooDB) using preserved metadata ids, string similarity between ingredient descriptions, and nutrient profiles. Polyphenol concentrations were standardized to mg per 100 g and then adjusted for cooking and processing using retention factors from Phenol Explorer.[26] Five additional polyphenols were added to our existing dataset as they were considered substrates for experimentally verified PUPs within the database for Polyphenol Utilization Proteins from gut microbiota (dbPUP).[18] Polyphenol estimates were averaged across recalls and standardized to average total caloric intake (mg per 1000 kcal per day). Dietary mapping and the estimation of dietary polyphenol intake are described in detail in our previous work.[22]

Blood collection and LPS-binding protein

Participants were asked to refrain from nutritional and herbal supplements for three days leading up to their fasting blood collection at the second visit. On the day before collection, participants were given a standard evening meal (30% kcal from fat, 55% kcal from carbohydrate, 15% kcal from protein) consisting of egg-stir fry, potato soup and commercially available raspberry sorbet and lemonade to consume between 18:00 and 19:00. After an overnight 12 hours fast, a blood sample was collected from each participant using venipuncture performed by a certified phlebotomist. After collection, tubes were placed on ice, then centrifuged at 1300*g* at 4 °C for 10 min. Plasma aliquots were prepared immediately and stored in Cryo-Store® vials at −80 °C until analysis. LPS-binding protein (LBP), a proxy for intestinal permeability,[27] was determined from plasma using the Abnova LBP ELISA (#KA0448) following kit instructions.

Stool collection

Participants were provided at home stool collection kits with collection instructions at their first visit and asked to self-collect a single stool sample as close to their second visit as possible. Samples were enclosed in a Ziploc bag within a lidded hard plastic commode and placed in a cooler with ice packs to be brought to researchers for same-day processing. Samples that could not be processed within 24 hours of collection were excluded from analysis (n = 30). A Stomacher paddle blender was used to homogenize samples before they were flash frozen on dry ice and separated into aliquots for freezing at −80 °C.

DNA extraction, metagenome library preparation, and sequencing

Microbial DNA was isolated using the ZymoBIOMICS DNA Miniprep kit (Zymo Research) from 100 mg of stool per manufacturer's protocol as described in detail previously.[28] The majority of DNA preparations (>95%) had $A_{260/280}$ and $A_{260/230}$ ratios greater than 1.80, and the lowest $A_{260/280}$ and $A_{260/230}$ ratios were 1.78 and 1.72, respectively. To verify DNA was intact and RNA-free prior to library construction, representative samples were resolved on an agarose gel. Evaluation of extraction controls and sequencing batch effects were previously performed with 16S sequencing using the same DNA as for metagenomic sequencing.[29] Controls included ZymoBIOMICS Microbial Mock Community standards, buffer negatives, and repeated extractions from a pooled stool sample. Sequencing batch effects were concluded to be minimal contributors to microbial composition differences.[29]

Whole-genome shotgun sequencing library preparation, library quality control, quantification, and pooling on NovaSeq6000 (Illumina) platform were performed by DNA Technologies & Expression Analysis Core Laboratory at University of California Davis Genome Center as published previously.[28]

Metagenomic sequence processing

Host DNA aligning to human genome version GRC38.p13 (NCBI GCF_000001405.39) was removed with BMTagger. Next, reads were trimmed to remove adapter and poor-quality sequences using Trimmomatic.[30] Duplicated reads were removed with FastUniq (v1.1)[31] using the default parameters. Paired-end reads were merged with FLASH (v1.2.11)[32] with an overlapping length of 10–100 bp and a mismatch ratio of 0.1. Fecal metagenomic processing for the USDA Nutritional Phenotyping Study has been described in previous work[28] and accompanying processing scripts are available at https://github.com/dglemay/ARG_metagenome.

Taxonomic and functional profiling

Taxonomic profiling was performed using MetaPhlAn (v4.0.6)[33] as described previously,[34] using the vOct22 CHOCOPhlAnSGB 202212 database.

Experimentally characterized PUPs in the human gut microbiome were recently curated from the literature into dbPUP.[18] Specifically, dbPUP contains 60 experimentally characterized PUPs from six enzymatic classes including oxidation/reduction (23 PUPs), functional group transfers (7 PUPs), hydrolysis (25 PUPs), nonhydrolytic cleaving (1 PUP), isomerization (2 PUPs), and unclassified reactions (2 PUPs).[18] Merged deduplicated reads were mapped to dbPUP by performing blastx[35] using an *e*-value cutoff of 1*e*−25 with DIAMOND (v2.1.5).[36] Gene counts were normalized to reads per kilobase per genome equivalent (RPKG) with MicrobeCensus.[37]

Taxonomy was assigned to reads that mapped to dbPUP using the branchwater (v0.9.7) implementation of sourmash.[38] Briefly, for each sample, metagenomic reads that mapped to dbPUP were extracted from files of merged reads using filterbyname.sh from the BBTools suite (v38.87).[39] Using blastn (v2.14.1+)[40] with "-max_target_seqs 1 -max_hsps 1", reads were then mapped back to assemblies, which were created with megahit (v1.2.9)[41] using default parameters. Contigs identified in the previous blast step were filtered by size (≥500 bp) using reformat.sh from BBTools. Branchwater commands "manysketch" (parameters: k = 51, scaled = 1000), "fastmultigather" (k = 51, $-t$ 500), and "tax" ($-F$ Human) were used to assign taxonomy to contigs. The taxonomic database used, genbank-2022.03-k51, was a combination of separate prepared databases for viral, archaea, protozoa, fungi, and bacteria, provided by sourmash developers. Individual databases were merged using "sourmash signature cat" before conversion to a RocksDB database with "sourmash scripts index" for use with branchwater.

Statistical analyses

All analyses and visualizations were generated using RStudio (v2024.12.1.563) running R v4.4.2.

Statistical analysis of microbial taxa. We used linear models to assess the relationship between α-diversity of the gut microbiome and energy-adjusted polyphenol intake. We examined bacterial species richness (number of unique species) and evenness (Pielou), and for diet, total polyphenol intake (mg per 1000 kcal per day) and the diversity (richness, evenness) of polyphenols in the diet. We accounted for age, BMI, and sex as they were factors that structured our study sampling scheme. As poly-

phenols can form complexes with dietary fiber[42] and co-occur with other nutrients in the food matrix, we also accounted for total fiber intake (g per 1000 kcal) and total HEI score. All covariates except for sex were treated as continuous variables.

We assessed the relationship between energy-adjusted polyphenol intake and the composition of the gut microbiome at various taxonomic levels (kingdom-strain) using the permutational analysis of variance (PERMANOVA) in the vegan package.[43] We performed this analysis within individuals grouped into the lowest and highest quartiles of total polyphenol intake (n = 157). Total polyphenol intake was treated categorically as low or high based on quartile grouping.

We modelled the relationship between energy-adjusted polyphenol intake and the abundance of LPS-producing bacteria using linear modelling, considering alternative models after review of model residuals. A logistic regression was used to model the zero-inflated *Enterobacteriaceae*, and a linear model was used to model *Gammaproteobacteria* (0.001 pseudocount added prior to log 10 transformation) and *Bacteriodales* (yeo-johnson transformed).

Statistical analysis of microbial function. We used linear models to assess whether the diversity of polyphenol intake is related to the diversity of PUP genes, accounting for sampling scheme and dietary covariates. We examined the richness and evenness (Pielou) of PUP genes, and of polyphenols in the diet.

We then examined the relationship between polyphenol intake and polyphenol utilization capacity. We assessed polyphenol intake at various resolutions (total, class, compound; mg per 1000 kcal) against (1) microbial PUP gene counts and (2) the abundance of microbial taxa with PUP genes. We utilized rank-based estimation regressions and accounted for sampling scheme and dietary covariates (total fiber intake and total HEI score). To minimize the effects of outliers and focus on more generalizable trends, we applied a 10% prevalence filter to microbial genera and a 30% prevalence filter to polyphenol compounds in the diet before modelling. Multiple comparison adjustments (Benjamini–Hochberg) were applied in both approaches ($q < 0.05$).

Statistical analysis of LPS-binding protein. We examined whether the diversity of energy-adjusted polyphenol intake and the abundance of microbes with PUP genes relate to plasma LBP. We used linear models to assess whether the diversity (richness, evenness) of microbial PUP genes is related to log-transformed LBP, adjusting for age, sex, BMI, fiber intake, and diet quality. We also used machine learning to test whether the abundance of microbial PUP genes is predictive of plasma LBP. We leveraged the hierarchical classification of PUPs (reaction class > subclass > PUP) and microbial taxonomies for the reduction of features prior to machine learning using taxaHFE-ML (https://github.com/aoliver44/taxaHFE).[44] Our taxonomically engineered feature set was combined with covariates into a random forest regression using a machine learning pipeline we developed using Tidymodels (https://github.com/aoliver44/nutrition_tools).[45] In brief, an initial 80–20 train-test split was performed followed by data preprocessing evaluation and hyperparameter tuning inside 10-fold repeated (3×) cross-validation. We repeated this process across ten random seeds to account for performance variability due to the data splits.[46] Shapley Additive exPlanations values were calculated on the best performing model and used to explain feature importance using fastshap (v0.0.07) and shapviz (v0.9.2) R packages.

Results

Participants

Overall, participants in the USDA Nutritional Phenotyping Study were 41 years old on average and considered overweight by BMI. Participants were below the Dietary Guideline for Americans recommendation for fiber intake (14 g per 1000 kcal) but above the national average for adults (57 out of 100) in overall diet quality (Table 1). On average, participants in the lowest and highest quartiles of energy-adjusted total polyphenol intake differed in their age and dietary characteristics but not by BMI. Participants within the highest quartile of polyphenol intake were generally older, had higher dietary quality and fiber intake, and were more likely to be coffee drinkers than participants within the lowest quartile. A trend was detected with sex, with higher polyphenol intake consumers more likely to be female.

Polyphenol intake and fecal microbes

We examined whether energy-adjusted total polyphenol intake associates with microbial α-diversity after accounting for age, BMI, sex, fiber intake, and diet quality. We did not detect a relationship between total polyphenol intake and microbial

Table 1 Characteristics of the USDA nutritional phenotyping cohort overall and by different polyphenol intake quartiles. A total of 313 participants had both diet and fecal shotgun metagenomes and were subsequently analyzed. Values represent mean (standard error) except for coffee consumers and sex, which displays counts. A chi-squared test was used to examine if the proportion of coffee consumers and sex differed by quartile. All other variables were tested with a two-sample t-test

		Total polyphenol intake			
	All (n = 313)	Low quartile (n = 79)	High quartile (n = 78)	Statistic	p-Value
Sex, female/male	162/151	34/45	45/33	2.81	0.094
Age (years)	41 (0.8)	36.1 (1.5)	44.1 (1.4)	3.88	<0.001
BMI (kgm^{-2})	27.2 (0.3)	28.2 (0.6)	27.2 (0.5)	−1.24	0.218
Fiber intake (g per 1000 kcal)	11.2 (0.3)	9 (0.4)	13.1 (0.6)	5.79	<0.001
Total HEI score	62.4 (0.7)	56.6 (1.2)	67.4 (1.4)	5.88	<0.001
Coffee consumer, yes/no	206/107	29/50	55/23	16.7	<0.001
Polyphenol intake (mg per 1000 kcal)	538.9 (32.7)	109.1 (6.2)	1235.5 (87.6)	12.83	<0.001

richness (p = 0.08) nor with microbial evenness (p = 0.93). Several covariate effects were detected including age with microbial richness (β = 0.94, p = 0.002), age with evenness (β = 0.0001, p = 0.002), and diet quality with microbial evenness (β = 0.0005, p = 0.05). We then assessed whether the diversity of polyphenol intake associates with microbial α-diversity. Polyphenol richness did not associate with microbial richness (p = 0.58) or evenness (p = 0.22). Similarly, polyphenol evenness was unassociated with microbial richness (p = 0.47) or evenness (p = 0.73).

We used PERMANOVA to assess whether microbial community composition differed between low and high quartiles of total polyphenol intake. We examined each taxonomic level (kingdom-strain) and accounted for age, BMI, sex, dietary quality, and fiber intake. Microbial community composition was not observed to vary with polyphenol consumption at any taxonomic level (Fig. S2).

PUP genes in fecal microbes

Of the 60 polyphenol utilization protein (PUP) genes in dbPUP, 54 were detected within the fecal microbiomes of our adult cohort. Participant fecal microbiomes had an average of 36 PUP genes (range: 25–48) across an average of 5 PUP families (range: 3–6). We detected a core group of 16 PUP genes in all 313 participants, including 12 hydrolysis PUPs, 2 oxidation reduction PUPs, and 2 functional group transfer PUPs (Table 2). A beta-glucosidase and glycoside hydrolase, both hydrolytic enzymes against isoflavones, were among the most abundant and prevalent PUP genes.

Taxonomic assignment of microbial reads mapping to dbPUP revealed that hydrolysis genes from *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* were the most prevalent across participants (Fig. S3). Genes associated with oxidation and reduction and functional group transfer reactions from Firmicutes were also detected across most participants. To further resolve the taxonomic distribution of polyphenol utilization capacity, we also summarized the number of distinct PUPs identified by bacterial family and reaction class in Fig. S4.

Polyphenol intake and the diversity of microbial PUP genes

We tested whether the diversity of polyphenol intake related to PUP gene diversity, accounting for age, sex, BMI, fiber intake, and diet quality. We found that polyphenol intake richness was positively related to the evenness of PUP genes (β = 0.00021, p = 0.03) but not the richness of PUP genes (p = 0.81). Similarly, polyphenol intake evenness was positively related to the richness of PUP genes (β = 7.18, p = 0.03) but not evenness of PUP genes (p = 0.13). No covariate effects were observed in our models on PUP alpha-diversity.

Polyphenol intake and abundance of PUP-containing microbial genera

After covariate adjustment, we identified 15 significant associations between the energy-adjusted intake of seven polyphenols (from three classes) and the abundance of 10 PUP-containing microbial genera (Fig. 1). One negative association was observed between *Anaerobutyricum* abundance and the intake of the isoflavone biochanin A. All other associations between polyphenol intake and the abundance of microbial genera with PUPs were positive. Naringin (*i.e.* naringin 7-*O*-neohesperidoside) intake positively associated with the most microbial genera including *Faecalibacillus*, *Gordonibacter*, *Lachnospira*, and *Streptococcus* (Fig. 1). The greatest effect sizes were found between hesperetin 7-neohesperidoside intake and the abundance of *Faecalibacillus*, *Lachnospira*, and *Streptococcus* (Table S1). We found *Gordonibacter* abundance positively associated with the most polyphenols, including two classes (flavones, flavonols) and three compounds (apigenin, naringin, quercetin). *Mediterraneibacter* had the second most polyphenol intake-associations, including one class (flavonols) and two compounds (kaempferol, quercetin).

Covariate effects for the abundance of PUP-containing microbial genera are summarized in Table S2. Briefly, we observed a negative effect for males on *Blautia* abundance *versus* females and a positive effect of BMI on *Dorea* abundance.

Polyphenol intake and polyphenol utilization capacity

We identified 117 associations (q < 0.05) between energy-adjusted polyphenol intake and the abundance of 25 distinct microbial PUP genes (encoded in 28 microbial genera) in our cohort after covariate adjustment (Fig. 2). Microbial PUP gene associations varied by intake resolution, with three associations found with total polyphenol intake, 16 with polyphenol classes, and 98 with individual compounds (Fig. 2). Of the 98 compound-level associations, eight associations involved polyphenols listed in dbPUP as experimentally verified substrates of the respective PUPs. The remaining 90 involved polyphenols that were not listed as experimentally verified as substrates for corresponding PUPs. Daidzein, apigenin, and diosmetin 7-rutinoside intake were the most common compounds with PUP associations, contributing to a total of 31 of the 98 PUP associations at the compound level. Flavone intake was the most common class with 5 PUP associations including three beta-glucosidase genes, an alpha-L-rhamnosidase gene, and an enoate reductase. The largest effect sizes were observed with the intake of 3,4′,5,6,7,8-hexahydroxyflavone (Table S3). Upon further inspection of its dietary sources, we found that almond consumption is the sole driver of 3,4′,5,6,7,8-hexahydroxyflavone (*i.e.* 6,8-dihydroxykaempferol) intake.

Associations between energy-adjusted polyphenol intake and PUP gene abundance were largely positive with only one negative association observed between myricetin intake and the alpha-L-rhamnosidase gene (X2CNV1) assigned to the genus *Blautia* (Table S3). Polyphenol intake was associated with 28 distinct microbial genera and 25 distinct PUP genes. PUP genes belonging to *Streptococcus* had the greatest number of dietary associations (1 total, 2 classes, 14 compounds) followed by *Roseburia* (1 total, 1 class, 12 compounds) and *Clostridium* (9 compounds).

Across all 117-polyphenol intake by PUP associations, 85 involved hydrolytic PUPs, 18 involved functional group trans-

Table 2 Abundance and prevalence of polyphenol utilization protein (PUP) genes detected in the fecal metagenomes of USDA Nutritional Phenotyping Study participants. PUPs are organized by protein reaction class and then sorted by prevalence. Several proteins listed below share the same name (functionally similar) but are derived from distinct genes

UniProt ID	Protein name	Mean gene abundance, RPKG (SD)	Prevalence, %
Functional group transfer			
A0ZZH6	Sucrose phosphorylase	0.184 (0.226)	100
Q9ZEU2	Amylosucrase	0.145 (0.053)	100
B8FRJ0	Arylsulfotransferase	0.008 (0.012)	98.7
Q9ZAR4	Dextransucrase	0.007 (0.025)	93.9
Q9RE05	Alternansucrase	0.003 (0.013)	93.3
P04830	Cyclomaltodextrin glucanotransferase	7.526×10^{-4} (0.002)	46
Q9L9F1	4-Hydroxyphenylpyruvate 3-dimethylallyltransferase	0.000 (0.000)	0
Hydrolysis			
A0A072MRT2	Beta-D-glucosideglucohydrolase	0.172 (0.180)	100
A0A072MSE8	Beta-glucosidase	0.472 (0.208)	100
A0A072MX70	Beta-glucosidase	0.368 (0.202)	100
A0A072N4Q2	Beta-glucosidase	0.291 (0.169)	100
A1S0B1	Glycoside hydrolase, family 3 domain protein	0.471 (0.121)	100
C4PG45	RhaB2	0.089 (0.046)	100
C4PG47	Alpha-L-rhamnosidase	0.075 (0.041)	100
I5AX46	Putative dehydrogenase	0.056 (0.063)	100
J9XU85	Beta-glucosidase	0.149 (0.069)	100
Q6RCI9	Alpha-L-rhamnosidase A	0.050 (0.019)	100
Q9S3L0	Alpha-L-rhamnosidase	0.146 (0.064)	100
X2CNV1	Alpha-L-rhamnosidase	0.026 (0.043)	100
Q715L4	Phloretin hydrolase	0.024 (0.023)	98.7
A0A0E3TKF2	DfgC	0.029 (0.029)	98.4
I5AX50	Sugar phosphate isomerase/epimerase	0.006 (0.013)	86.3
I5AX47	Putative dehydrogenase	0.003 (0.006)	79.6
F6IEX3	Glycoside hydrolase family protein	0.002 (0.004)	77.6
Q76LC4	Alpha-L-rhamnosidase	0.002 (0.005)	77.6
I5AX48	Sugar phosphate isomerase/epimerase	0.002 (0.003)	73.8
A0A0E3TJD6	DfgD	0.002 (0.005)	72.5
Q45VU2	Endo-1,4-beta-xylanase	0.004 (0.006)	60.1
B1MK49	DAPG_hydrolase domain-containing protein	6.233×10^{-4} (0.001)	38.3
Q6RCI8	Alpha-L-rhamnosidase B	3.380×10^{-4} (0.001)	38.3
Q51723	Beta-galactosidase	6.164×10^{-5} (3.072×10^{-4})	10.5
A0A0E3K5E4	Beta-galactosidase	0.000 (0.000)	0
Isomerization			
V9P0A9	Chalcone isomerase	0.014 (0.012)	99
I7H868	Dihydrodaizein racemase	9.371×10^{-4} (0.003)	17.9
Nonhydrolytic cleaving			
Q60FX6	Reversible 2,6-dihydroxybenzoic acid decarboxylase	9.154×10^{-4} (0.002)	47.9
Oxidation Reduction			
H3JUE4	Daidzein-to-DHD conversion enzyme	0.007 (0.009)	100
V9P074	Enoate reductase	0.052 (0.031)	100
E1CIA4	Daidzein reductase	0.003 (0.005)	93.9
A0A0U1WKA6	Daidzein and genistein reductase	8.042×10^{-4} (0.001)	72.2
Q74HL7	NADH-dependent flavin reductase subunit 1	0.004 (0.011)	72.2
E7FL40	Tetrahydrodaidzein reductase	0.002 (0.003)	59.4
M9NZ71	Daidzein reductase	3.853×10^{-4} (8.663×10^{-4})	45.7
B1XDG0	NADPH-dependent curcumin/dihydrocurcumin reductase	0.003 (0.017)	31
Q74HL8	NADH-dependent flavin reductase subunit 2	7.691×10^{-4} (0.002)	28.1
M9P0B3	Succinate dehydrogenase	3.967×10^{-4} (0.001)	22.4
E7FL41	Dihydrodaidzein reductase	7.955×10^{-4} (0.002)	21.1
H3JUE2	DHD-to-equol conversion enzyme 1	4.110×10^{-4} (0.001)	21.1
Q52028	Biphenyl dioxygenase subunit alpha	6.121×10^{-4} (0.003)	20.8
H3JUE3	DHD-to-equol conversion enzyme 2	6.473×10^{-4} (0.002)	19.2
Q2EYY8	4-Coumarate 3-hydroxylase	0.002 (0.019)	19.2
F7V1S2	Uncharacterized protein	2.592×10^{-4} (8.950×10^{-4})	15
F7V1S0	NADH:flavin oxidoreductase	7.767×10^{-5} (3.445×10^{-4})	9.6
F7V1S3	FAD_binding_2 domain-containing protein	1.582×10^{-5} (1.713×10^{-4})	1.6
M9NYU8	3-Oxoacyl-ACP reductase	2.509×10^{-5} (3.140×10^{-4})	0.6
A2VA43	Quercetinase	0.000 (0.000)	0
G2IMF2	Putative oxidoreductase	0.000 (0.000)	0
P42106	Quercetin 2,3-dioxygenase	0.000 (0.000)	0
Q2G4H9	NmrA-like protein	0.000 (0.000)	0
Unclassified			
I5AX49	Uncharacterized protein	0.005 (0.011)	73.2
Q65JC2	Glycosyl transferase family 1	0.002 (0.003)	57.2

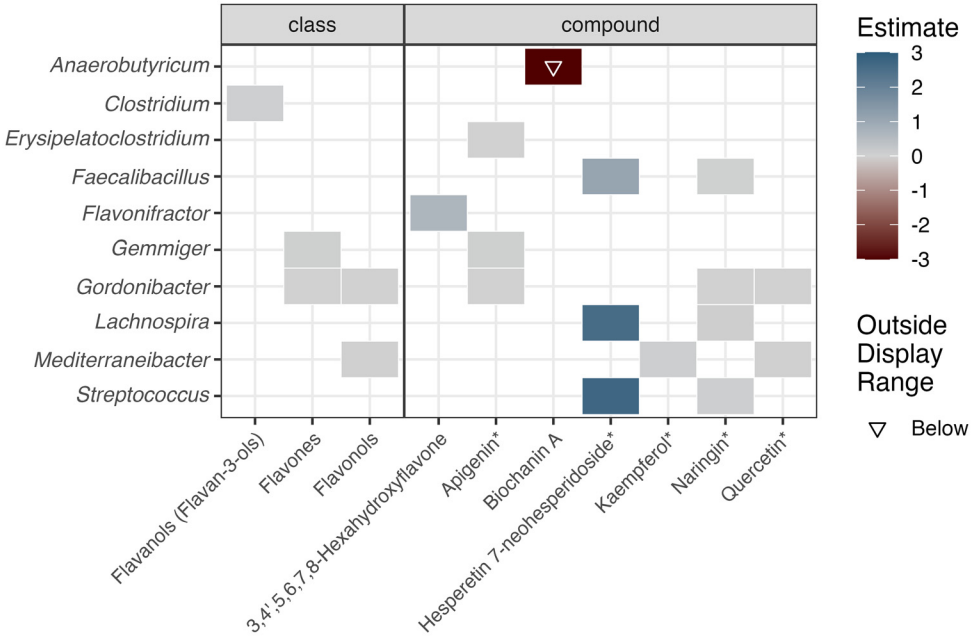

Fig. 1 Relationships (q < 0.05) between polyphenol intake and abundance of microbial genera with polyphenol utilization proteins (PUPs). Polyphenol intake was examined at the total, class, and compound level. However, no relationships were observed between total polyphenol intake and microbial abundance and are consequently not depicted on the heatmap. Beta-coefficient estimates were derived from rank-based estimation regressions that accounted for age, sex, body mass index, fiber intake, and total Healthy Eating Index (diet quality). *P*-Values were adjusted for multiple testing. *Indicates polyphenol is a substrate listed in dbPUP.

fers, 11 involved redox reactions, and 3 were with unclassified PUPs. We observed polyphenol intake associations with 16 distinct hydrolysis PUPs, 4 functional transfer PUPs, 4 redox PUPs, and 1 unclassified PUP. Alpha-L-rhamnosidase and beta-glucosidase, accounted for 56% of hydrolysis associations, with five alpha-L-rhamnosidase genes and four beta-glucosidase genes.

Covariate effects for microbial PUPs are summarized in Table S4.

Olive-associated polyphenol classes and microbial LPS producers

In our previous examination of polyphenol intake and inflammation,[22] we found a negative relationship between intake of polyphenols classified as prenol lipids and phenylpropanoic acids and LPS binding protein, a proxy for gastrointestinal permeability. Thus, we examined whether the combined intake of these two polyphenol classes (mg per 1000 kcal) relate to the abundance of microbial taxa known for production of immunoinhibitory (order *Bacteroidales*)[47] and proinflammatory LPS (phylum *Gammaproteobacteria*, family *Enterobacteriaceae*). We assessed whether the intake of polyphenols considered prenol lipids and phenylpropanoic acid was related to immunoinhibitory and inflammatory LPS-producing taxa, accounting for age, sex, BMI, fiber intake, and diet quality. With regards to inflammatory LPS producers, we did not observe a relationship between intake of prenol lipids and phenylpropanoic acids and the abundance of *Gammaproteobacteria* (Fig. 3A) nor did we find a relationship between intake and the presence of *Enterobacteriaceae* (Fig. 3B). Regarding immunoinhibitory LPS producers, we found a positive relationship between the consumption of prenol lipids and phenylpropanoic acids and the yeo-johnson transformed abundance of *Bacteroidales* (Fig. 3C). We did observe one main effect with males having 42% lower odds of fecal *Enterobacteriaceae* present than females (p = 0.02).

PUP genes and LPS-binding protein

In addition to our hypothesized relationship between LPS-producing microbes and prenol lipids, we also explored whether PUP genes were related to gastrointestinal permeability. In our linear models between PUP gene alpha-diversity and plasma LBP (log-transformed), we did not detect an LBP relationship with PUP gene richness (p = 0.56) nor PUP gene evenness (p = 0.38).

Using machine learning, we investigated if PUP gene count and the abundance of microbes with PUP genes associated with plasma LBP. Both models, including covariates and either PUP gene counts or microbial abundance, outperformed the null model (0.46 and 0.51 reduction in mean absolute error for models built with PUP gene counts and microbial abundances, respectively). BMI and sex were top features in predicting plasma LBP in both models. In our PUP-gene count model, a glycoside hydrolase gene was the most informative PUP feature predicting plasma LBP (Fig. 4A). Similarly, in our model with PUP-containing microbes, the abundance of an unclassified

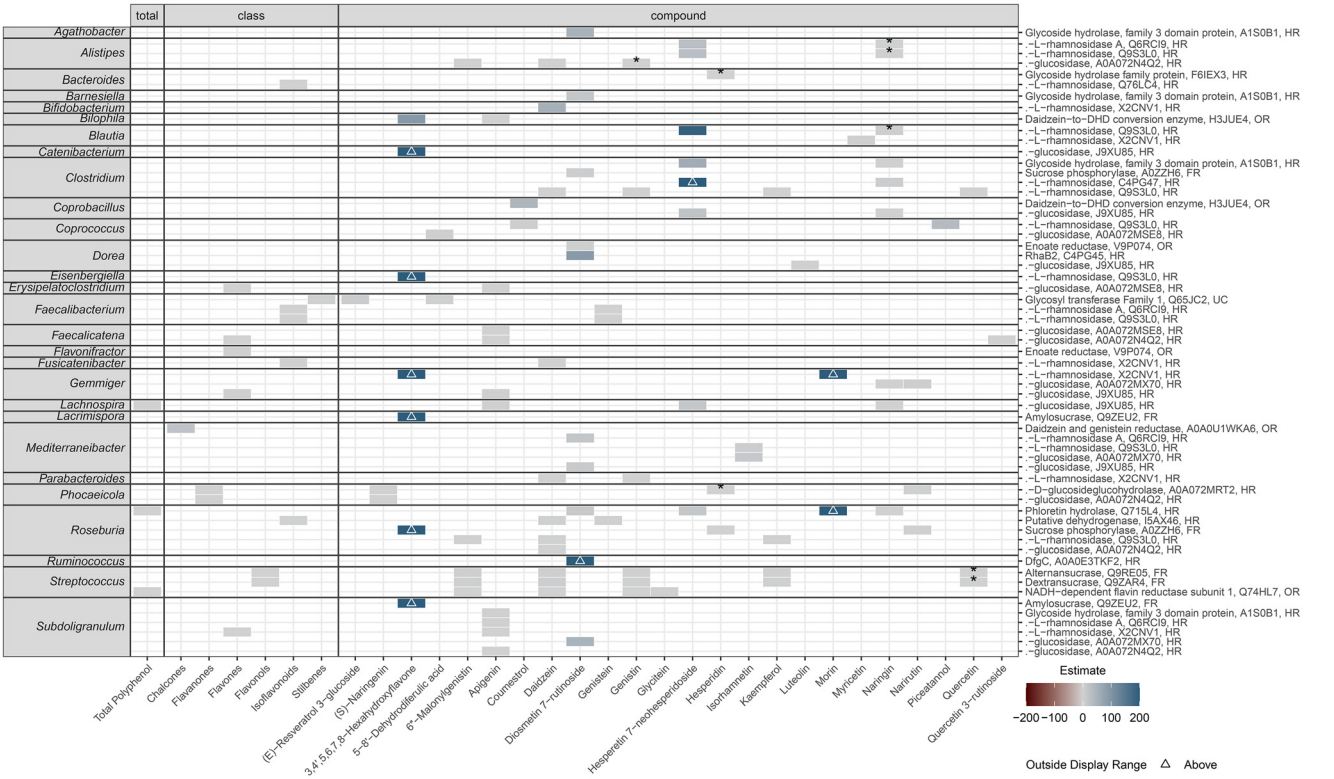

Fig. 2 Relationships ($q < 0.05$) between polyphenol intake and microbial polyphenol utilization protein (PUP) gene count. PUP labels on the right *y*-axis are formatted as PUP name, UniProt ID, and reaction classification (functional group transfer, FR; hydrolysis, HR; isomerization, IR; non-hydrolytic cleaving, NCR; oxidation/reduction, OR; unclassified; UC). Coefficients and *p*-values are from rank-based estimation regression and account for age, sex, body mass index, fiber intake, and total Healthy Eating Index (diet quality). *P*-Values were adjusted for multiple testing. *Indicates polyphenol is a substrate for specific PUP.

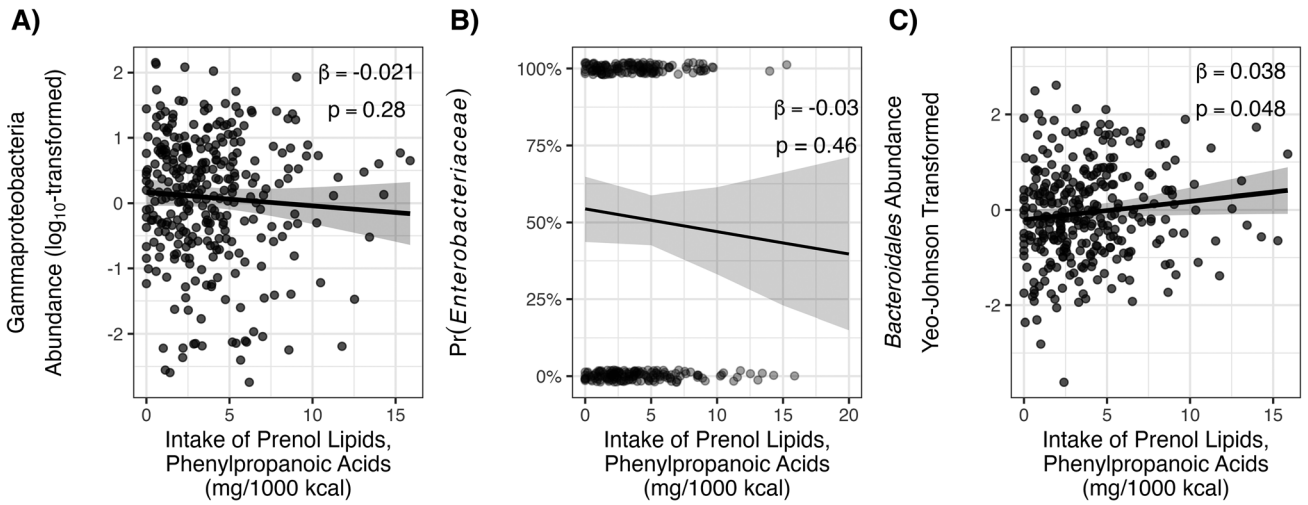

Fig. 3 Relationships between the intake of polyphenols classified as prenol lipids and phenylpropanoic acids and (A) *Gammaproteobacteria* abundance, (B) *Enterobacteriaceae* presence, and (C) *Bacteroidales* abundance. Linear models were utilized for examining the abundance of *Gammaproteobacteria* and *Bacteroidales* while logistic regression was used to model zero-inflated *Enterobacteriaceae*. All models accounted for age, sex, BMI, fiber intake, and total Healthy Eating Index (diet quality). Points represent residuals in (A) and (C) and observations in (B).

Lachnospiraceae was the most informative microbial feature (Fig. 4B). However, the raw relationships between individual model features and plasma LBP were generally weak (Fig. S5). BMI and sex showed the strongest evidence of a relationship, with higher plasma LBP observed with increasing BMI and in females compared to males.

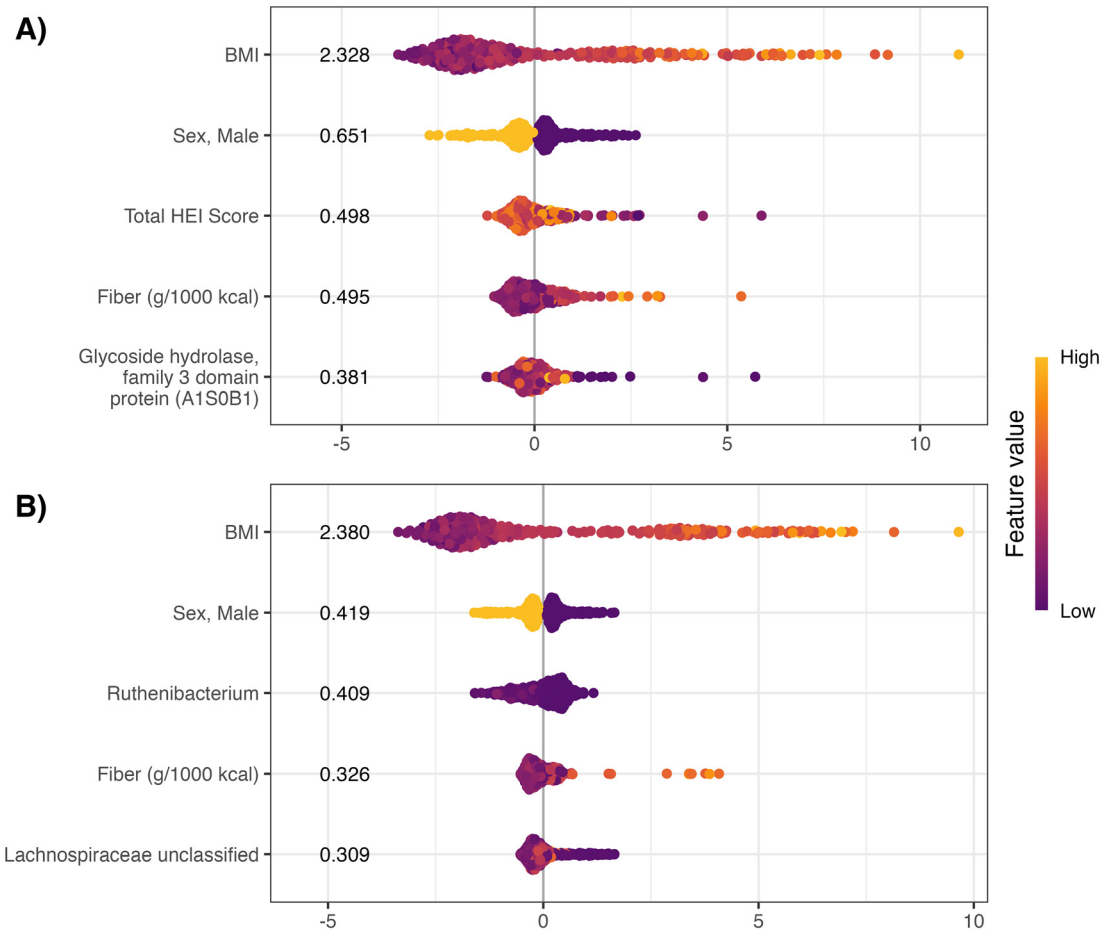

Fig. 4 Shapley Additive exPlanations (SHAP) plots from random forest regression models on LBP. The top 5 most important features for (A) microbial PUP gene counts and (B) abundance of microbes with PUP genes are shown. Features are organized by importance from top (most important) to bottom. Each point represents the individual contribution to the model output and is organized similar to a density plot. Color indicates the original value of the feature. BMI, body mass index; HEI, Healthy Eating Index.

Discussion

We analyzed the relationship between energy-adjusted polyphenol intake and the taxonomic and functional characteristics of fecal microbiomes from healthy adults. We found that intake of specific polyphenol classes and compounds was positively associated with the abundance of nine microbial genera with polyphenol utilization proteins (PUPs). Additionally, we identified 117 associations between polyphenol intake and the abundance of microbial PUP genes, with 73% of all associations involving genes for hydrolytic proteins. Although we did not find polyphenol intake to influence the gut microbial community composition or diversity, we did find that a diverse polyphenol intake was positively associated with the diversity of PUP genes. Based on prior findings with gastrointestinal permeability,[22] we examined intake of prenol lipids and phenylpropanoic acids in the context of microbial lipopolysaccharide (LPS) producers, and found that consumption of these olive-related polyphenol classes positively associated with *Bacteroidales* abundance, a producer of immunoinhibitory LPS. However, when evaluating whether PUP genes and PUP-containing microbes could improve prediction of plasma LPS-binding protein (LBP) as an indicator of gastrointestinal permeability, we found our strongest predictors were BMI and sex. Taken together, our findings support the hypothesis that polyphenol intake relates to the polyphenol utilization capacity of the gut microbiome as well as the abundance of microbes with the potential to metabolize polyphenols and exert anti-inflammatory effects.

Hydrolytic PUP genes were among the most prevalent in our cohort and associated with energy-adjusted polyphenol intake. Twelve hydrolytic PUP genes were identified in the gut microbiomes across all study participants. The widespread presence of microbial genes encoding hydrolysis enzymes may explain why nearly all microbial communities demonstrated the ability to break glycosidic linkages and ester bonds in a recent *ex vivo* incubation of human fecal samples with polyphenols.[48] Of the 117 diet-gene associations we observed, approximately 73% involved hydrolytic PUP genes. Alpha-L-rhamnosidase and beta-glucosidase were the most common

enzymes associated with polyphenol intake, even after accounting for fiber intake. Both enzymes are known for their role in fiber degradation.[49] With polyphenols, enzymatic hydrolysis removes the sugar residue from the polyphenolic backbone to release the aglycone, which can increase its bioavailability.[50] Microbial hydrolytic action can also contribute to the release of fiber-bound polyphenols, a type of non-extractable polyphenol (NEPP) that largely escapes modification in the small intestine, which also contributes to greater polyphenol bioaccessibility and bioefficacy.[51] We suspect that failing to account for dietary fiber may partially ignore the NEPP fraction within foods and obscure potential polyphenol–microbiome relationships. We repeated our analysis without covariates to understand the impact of covariate adjustment on the power to detect relationships between polyphenol intake and PUP genes and found that the significant associations were different and notably lower in number than our covariate-adjusted models (Tables S5 and S6). Our findings highlight the importance of covariate adjustment and support that polyphenol intake may relate to the hydrolytic capacity of the microbiome. Future work may clarify how NEPP and extractable polyphenols (EPP) relate to microbial polyphenol utilization capacity.

Dietary polyphenol intake associated with microbial functional potential to metabolize polyphenols. We found that greater evenness in polyphenol intake – that is, having an equal representation of various polyphenols in the diet – was positively associated with greater richness of PUP genes but not with the diversity of microbes themselves. This may suggest that consuming a diet balanced in polyphenol-rich foods may help increase the capacity of the microbiome to utilize polyphenols, compared to for example, diets high in tea but limited in fruits. In examining polyphenol intake and PUP genes, we found intake of specific polyphenol classes and compounds positively associated with the abundance of nine PUP-containing microbial genera. Notably, we observed a link between citrus polyphenols and *Lachnospira*. A prior 16S-based analysis of this cohort found non-berry fruit intake positively associated with *Lachnospira* abundance.[29] Our data suggest this may be linked to citrus-derived compounds hesperetin 7-neohesperidoside and naringin. Their poor solubility in citrus juices, which were consumed widely in our cohort,[22] may enhance colonic delivery and increase the likelihood of microbial metabolism.[52] In contrast, we observed only one negative association between the isoflavone biochanin A and abundance of *Anaerobutyricum* (formerly *Eubacterium hallii*). Although biochanin A has been shown to inhibit the growth of certain *Clostridium* and *Bifidobacterium* species *in vitro*,[53] its effect on the butyrate-producing *Anaerobutyricum* is unclear.[54] As in our functional analyses, most associations between polyphenol intake and microbial abundance were sensitive to covariate adjustment, with little overlap observed between modelling approaches. These findings suggest that polyphenol intake may support microbes with polyphenol utilizing potential and highlight the importance of accounting for demographic and dietary factors in diet-microbiome association studies.

We observed that consumption of polyphenols classified as prenol lipids and phenylpropanoic acids positively associated with the abundance of bacteria that can produce immunoinhibitory LPS. We previously found a negative relationship between plasma LBP and prenol lipid and phenylpropanoic acid intake[22] and in the current study hypothesized that consumption of these polyphenol classes was linked to the abundance of LPS producers. Prenol lipid and phenylpropanoic acid intake, driven by olive consumption, positively related with the abundance of the immunoinhibitory LPS-producer *Bacteroidales* but not with taxa known for potent proinflammatory LPS-producers such as *Enterobacteriaceae*. LPS from *Bacteroidales* was observed *in vitro* to have low-immunostimulatory capacity and to suppress LPS-induced immune activation by *E. coli*.[55] Our current findings suggest that olive-associated polyphenols may increase microbial taxa that produce immunosuppressive LPS. Future research may test the responsiveness of *Bacteroidales* to dietary modulation.

We examined whether PUP genes and PUP gene-containing microbes could predict LBP, our marker of gastrointestinal permeability. While a glycoside hydrolase and an unclassified *Lachnospiraceae* were among the most informative features in our machine learning models on plasma LBP, our post-modelling analyses revealed that individual characteristics such as BMI and sex more reliably associated with plasma LBP. This is in line with our previous findings examining the relationship between LBP and dietary variables,[56] suggesting host factors are more informative in explaining plasma LBP than singular dietary or microbial features. Given the link between intake of citrus polyphenols and greater *Lachnospira* (a genus within *Lachnospiraceae*) abundance, these findings suggest a potential pathway for reducing gastrointestinal permeability through bacterial modulation.

This study has several limitations. First, as we collected data from generally healthy adults within a targeted geographical area, these results may not be generalizable. Second, we cannot infer causality through our observational study design. A single snapshot of the gut microbial community structure does not reflect the dynamic changes that can occur after dietary exposure. *In vitro* studies examining the impact of polyphenol exposure on microbial growth and gene expression suggest that microbial genes can be altered without affecting growth.[57,58] Third, polyphenol intake estimates are subject to self-report bias from dietary recalls, limited to database constraints (*e.g.*, food and polyphenol coverage), and require further validation by measurement of polyphenols in biological samples. Lastly, though the database of polyphenol utilization proteins used in this study is the largest collection of experimentally verified polyphenol utilization proteins to date, protein coverage is relatively limited compared to the diversity of polyphenols and their microbially produced derivatives. Experimental validation of host-associated microbial polyphenol utilization has largely centered around isoflavone metabolism, common and abundant plant polyphenols, and polyphenol commercial availability.[59] Identification and validation of additional polyphenol utilization proteins, such as those with ellagic acid as a substrate, are needed to better reflect the

polyphenol utilization capacity of the gut microbiome and understand its role in the interindividual variability in responses to polyphenol intake.

In our study, we present that dietary polyphenol intake may help shape the polyphenol utilization capacity of the gut microbiome. Polyphenol intake may support both the abundance of polyphenol utilizing bacteria as well as the functional capacity of the gut microbiome, particularly by promoting PUP gene diversity and a higher count of genes involved in hydrolytic reactions. As our work leveraged a cross-sectional adult cohort, future investigation should examine whether polyphenol-rich dietary interventions affect the polyphenol utilization capacity of the gut microbiome. Our work revealed 90 potentially novel gene-substrate relationships which serve as a starting list of substrates and microbial targets for testing and experimental validation of polyphenol metabolism. These findings may inform precision nutrition efforts to tailor the gut microbiome to promote beneficial dietary responses.

Author contributions

Stephanie Wilson: investigation, software, formal analysis, visualization, methodology, writing – original draft. Andrew Oliver: investigation, software, formal analysis, writing – review & editing. Zeynep Alkan: investigation, writing – review & editing. Bhimanagouda Patil: resources, writing – review & editing. Mary Kable: resources, writing – review & editing. Danielle Lemay: conceptualized research; funding acquisition. supervision, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

Microbial metagenomes for 330 individuals are deposited in NCBI Sequence Read Archive under accession numbers SRP354271 and SRP497208. Requests for non-metagenomic data from the USDA ARS WHNRC Nutritional Phenotyping Study used in this analysis should be made *via* an email to the corresponding senior WHNRC author. Data requests will be reviewed quarterly by a committee consisting of the study investigators. Metagenomic processing scripts are available at https://github.com/dglemay/ARG_metagenome. Analysis scripts are available on Github at https://github.com/SWi1/diet_PUP_relationships.

Supplementary figures and tables are available. See DOI: https://doi.org/10.1039/d6fo00158k.

Acknowledgements

We thank Charles Stephensen for his leadership and guidance as a principal investigator of the USDA Nutritional Phenotyping Study. We also acknowledge USDA Nutritional Phenotyping Team members including Dr Ellen Bonnel for human studies management; Evelyn Holguin, Annie Kan, Dustin Burnett, Barbara Gale, Danna Juarez Rios, Justin Waller, Ashley Tovar, Christine Bowlus, Yasmine Bouzid, Yanhua "Eva" Li, and Diane Han for human studies and physiology assessment; and Joanne Arsenault for dietary quality control.

This research used resources from the Texas A&M University Institute for Advancing Health Through Agriculture (58-3091-1-018), USDA-ARS 2032-51530-026-000-D, and USDA-ARS 2032-10700-002-00D. Access to REDCAP® for the USDA Nutritional Phenotyping Study was provided by the University of California, Davis Health and supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1 TR001860. This work also utilized computational resources from the High Performance Computing Core Facility at the University of California, Davis.

References

1 T. Kiyimba, P. Yiga, M. Bamuwamye, P. Ogwok, B. Van der Schueren and C. Matthys, Efficacy of Dietary Polyphenols from Whole Foods and Purified Food Polyphenol Extracts in Optimizing Cardiometabolic Health: A Meta-Analysis of Randomized Controlled Trials, *Adv. Nutr.*, 2023, **14**, 270–282.

2 L. Hooper, P. A. Kroon, E. B. Rimm, J. S. Cohn, I. Harvey, K. A. Le Cornu, J. J. Ryder, W. L. Hall and A. Cassidy, Flavonoids, flavonoid-rich foods, and cardiovascular risk: a meta-analysis of randomized controlled trials, *Am. J. Clin. Nutr.*, 2008, **88**, 38–50.

3 H. D. Sesso, J. E. Manson, A. K. Aragaki, P. M. Rist, L. G. Johnson, G. Friedenberg, T. Copeland, A. Clar, S. Mora, M. V. Moorthy, A. Sarkissian, W. R. Carrick and G. L. Anderson, Effect of cocoa flavanol supplementation for the prevention of cardiovascular disease events: the COcoa Supplement and Multivitamin Outcomes Study (COSMOS) randomized clinical trial, *Am. J. Clin. Nutr.*, 2022, **115**, 1490–1500.

4 C. Manach, G. Williamson, C. Morand, A. Scalbert and C. Rémésy, Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies, *Am. J. Clin. Nutr.*, 2005, **81**, 230S–242S.

5 C. Cueva, M. V. Moreno-Arribas, P. J. Martín-Alvarez, G. Bills, M. F. Vicente, A. Basilio, C. L. Rivas, T. Requena, J. M. Rodríguez and B. Bartolomé, Antimicrobial activity of phenolic acids against commensal, probiotic and pathogenic bacteria, *Res. Microbiol.*, 2010, **161**, 372–382.

6 H. C. Lee, A. M. Jenner, C. S. Low and Y. K. Lee, Effect of tea phenolics and their aromatic fecal bacterial metabolites on intestinal microbiota, *Res. Microbiol.*, 2006, **157**, 876–884.

7 M. Daglia, Polyphenols as antimicrobial agents, *Curr. Opin. Biotechnol.*, 2012, **23**, 174–181.

8 L. Lavefve, L. R. Howard and F. Carbonero, Berry polyphenols metabolism and impact on human gut microbiota and health, *Food Funct.*, 2020, **11**, 45–65.
9 A. M. Alves-Santos, C. S. A. Sugizaki, G. C. Lima and M. M. V. Naves, Prebiotic effect of dietary polyphenols: A systematic review, *J. Funct. Foods*, 2020, **74**, 104169.
10 M. C. Rodríguez-Daza, E. C. Pulido-Mateos, J. Lupien-Meilleur, D. Guyonnet, Y. Desjardins and D. Roy, Polyphenol-Mediated Gut Microbiota Modulation: Toward Prebiotics and Further, *Front. Nutr.*, 2021, **8**, 689456.
11 P. Filannino, M. Gobbetti, M. De Angelis and R. D. Cagno, Hydroxycinnamic acids used as external acceptors of electrons: an energetic advantage for strictly heterofermentative lactic acid bacteria, *Appl. Environ. Microbiol.*, 2014, **80**, 7574–7582.
12 M. Marhuenda-Muñoz, E. P. Laveriano-Santos, A. Tresserra-Rimbau, R. M. Lamuela-Raventós, M. Martínez-Huélamo and A. Vallverdú-Queralt, Microbial Phenolic Metabolites: Which Molecules Actually Have an Effect on Human Health?, *Nutrients*, 2019, **11**, 2725.
13 A. L. Simons, M. Renouf, S. Hendrich and P. A. Murphy, Human Gut Microbial Degradation of Flavonoids: Structure–Function Relationships, *J. Agric. Food Chem.*, 2005, **53**, 4258–4263.
14 L. Xie, Z. Deng, J. Zhang, H. Dong, W. Wang, B. Xing and X. Liu, Comparison of Flavonoid O-Glycoside, C-Glycoside and Their Aglycones on Antioxidant Capacity and Metabolism during In Vitro Digestion and In Vivo, *Foods*, 2022, **11**, 882.
15 G. Di Pede, L. Bresciani, F. Brighenti, M. N. Clifford, A. Crozier, D. Del Rio and P. Mena, In Vitro Faecal Fermentation of Monomeric and Oligomeric Flavan-3-ols: Catabolic Pathways and Stoichiometry, *Mol. Nutr. Food Res.*, 2022, **66**, 2101090.
16 E. C. Pulido-Mateos, J. Lessard-Lord, D. Guyonnet, Y. Desjardins and D. Roy, Comprehensive analysis of the metabolic and genomic features of tannin-transforming Lactiplantibacillus plantarum strains, *Sci. Rep.*, 2022, **12**, 22406.
17 T. Goris, R. R. C. Cuadrat and A. Braune, Flavonoid-Modifying Capabilities of the Human Gut Microbiome-An In Silico Study, *Nutrients*, 2021, **13**, 2688.
18 B. Zheng, Y. He, P. Zhang, Y.-X. Huo and Y. Yin, Polyphenol Utilization Proteins in the Human Gut Microbiome, *Appl. Environ. Microbiol.*, 2022, **88**, e01851–e01821.
19 S. N. B. Selby-Pham, R. B. Miller, K. Howell, F. Dunshea and L. E. Bennett, Physicochemical properties of dietary phytochemicals can predict their passive absorption in the human small intestine, *Sci. Rep.*, 2017, **7**, 1931.
20 K. D. Setchell and C. Clerici, Equol: history, chemistry, and formation, *J. Nutr.*, 2010, **140**, 1355s–1362s.
21 D. Milenkovic, C. Morand, A. Cassidy, A. Konic-Ristic, F. Tomás-Barberán, J. M. Ordovas, P. Kroon, R. D. Caterina and A. Rodriguez-Mateos, Interindividual Variability in Biomarkers of Cardiometabolic Health after Consumption of Major Plant-Food Bioactive Compounds and the Determinants Involved, *Adv. Nutr.*, 2017, **8**, 558–570.
22 S. M. G. Wilson, A. Oliver, J. A. Larke, J. J. Naveja, Z. Alkan, J. M. Awika, C. B. Stephensen and D. G. Lemay, Fine-scale Dietary Polyphenol Intake is Associated with Systemic and Gastrointestinal Inflammation in Healthy Adults, *J. Nutr.*, 2024, 3286–3297.
23 L. M. Baldiviez, N. L. Keim, K. D. Laugero, D. H. Hwang, L. Huang, L. R. Woodhouse, D. J. Burnett, M. S. Zerofsky, E. L. Bonnel, L. H. Allen, J. W. Newman and C. B. Stephensen, Design and implementation of a cross-sectional nutritional phenotyping study in healthy US adults, *BMC Nutr.*, 2017, **3**, 79.
24 Y. Y. Bouzid, J. E. Arsenault, E. L. Bonnel, E. Cervantes, A. Kan, N. L. Keim, D. G. Lemay and C. B. Stephensen, Effect of Manual Data Cleaning on Nutrient Intakes Using the Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24), *Curr. Dev. Nutr.*, 2021, **5**, nzab005.
25 J. A. Larke, N. Bacalzo, J. J. Castillo, G. Couture, Y. Chen, Z. Xue, Z. Alkan, M. E. Kable, C. B. Lebrilla, C. B. Stephensen and D. G. Lemay, Dietary Intake of Monosaccharides from Foods is Associated with Characteristics of the Gut Microbiota and Gastrointestinal Inflammation in Healthy US Adults, *J. Nutr.*, 2023, **153**, 106–119.
26 J. A. Rothwell, J. Perez-Jimenez, V. Neveu, A. Medina-Remón, N. M'Hiri, P. García-Lobato, C. Manach, C. Knox, R. Eisner, D. S. Wishart and A. Scalbert, Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content, *Database*, 2013, **2013**, bat070.
27 B. Seethaler, M. Basrai, A. M. Neyrinck, J.-A. Nazare, J. Walter, N. M. Delzenne and S. C. Bischoff, Biomarkers for assessment of intestinal permeability in clinical practice, *Am. J. Physiol.: Gastrointest. Liver Physiol.*, 2021, **321**, 11–17.
28 A. Oliver, Z. Xue, Y. T. Villanueva, B. Durbin-Johnson, Z. Alkan, D. H. Taft, J. Liu, I. Korf, K. D. Laugero, C. B. Stephensen, D. A. Mills, M. E. Kable and D. G. Lemay, Association of Diet and Antimicrobial Resistance in Healthy U.S, Adults, *mBio*, 2022, **13**, e00101–e00122.
29 M. E. Kable, E. L. Chin, D. Storms, D. G. Lemay and C. B. Stephensen, Tree-Based Analysis of Dietary Diversity Captures Associations Between Fiber Intake and Gut Microbiota Composition in a Healthy US Adult Cohort, *J. Nutr.*, 2022, **152**, 779–788.
30 A. M. Bolger, M. Lohse and B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 2014, **30**, 2114–2120.
31 H. Xu, X. Luo, J. Qian, X. Pang, J. Song, G. Qian, J. Chen and S. Chen, FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads, *PLoS One*, 2012, **7**, e52249.
32 T. Magoc and S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, *Bioinformatics*, 2011, **27**, 2957–2963.

33 A. Blanco-Míguez, F. Beghini, F. Cumbo, L. J. McIver, K. N. Thompson, M. Zolfo, P. Manghi, L. Dubois, K. D. Huang, A. M. Thomas, W. A. Nickols, G. Piccinno, E. Piperni, M. Punčochář, M. Valles-Colomer, A. Tett, F. Giordano, R. Davies, J. Wolf, S. E. Berry, T. D. Spector, E. A. Franzosa, E. Pasolli, F. Asnicar, C. Huttenhower and N. Segata, Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4, *Nat. Biotechnol.*, 2023, **41**, 1633–1644.

34 A. Oliver, Z. Alkan, C. B. Stephensen, J. W. Newman, M. E. Kable and D. G. Lemay, Diet, Microbiome, and Inflammation Predictors of Fecal and Plasma Short-Chain Fatty Acids in Humans, *J. Nutr.*, 2024, **154**, 3298–3311.

35 C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, BLAST+: architecture and applications, *BMC Bioinf.*, 2009, **10**, 421.

36 B. Buchfink, K. Reuter and H.-G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND, *Nat. Methods*, 2021, **18**, 366–368.

37 S. Nayfach and K. S. Pollard, Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome, *Genome Biol.*, 2015, **16**, 51.

38 L. Irber, N. T. Pierce-Ward, M. Abuelanin, H. Alexander, A. Anant, K. Barve, C. Baumler, O. Botvinnik, P. Brooks, D. Dsouza, L. Gautier, M. R. Hera, H. E. Houts, L. K. Johnson, F. Klötzl, D. Koslicki, M. Lim, R. Lim, B. Nelson, I. Ogasawara, T. Reiter, C. Scott, A. Sjödin, D. Standage, S. J. Swamidass, C. Tiffany, P. Vemuri, E. Young and C. T. Brown, sourmash v4: A multitool to quickly search, compare, and analyze genomic and metagenomic data sets, *J. Open Source Softw.*, 2024, **9**, 6830.

39 B. Bushnell, *BBMap: A Fast, Accurate, Splice-Aware Aligner*, Lawrence Berkeley National Laboratory, 2014, LBNL Report #: LBNL-7065E.

40 M. Hernández-Salazar, P. Osorio-Diaz, G. Loarca-Piña, R. Reynoso-Camacho, J. Tovar and L. A. Bello-Pérez, In vitro fermentability and antioxidant capacity of the indigestible fraction of cooked black beans (Phaseolus vulgaris L.), lentils (Lens culinaris L.) and chickpeas (Cicer arietinum L.), *J. Sci. Food Agric.*, 2010, **90**, 1417–1422.

41 D. Li, C.-M. Liu, R. Luo, K. Sadakane and T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, *Bioinformatics*, 2015, **31**, 1674–1676.

42 C. Hou, Y. Chen, W. Zhang, J. Yu, M. Ji, S. Cai, W. Guo, X. Ji, L. Sun, X. Liu and Y. Wang, An insight into the full aspects of bound polyphenols in dietary fiber: Interaction, composition, function and foundation as well as alteration in food processing, *Food Chem.*, 2025, **485**, 144553.

43 J. Oksanen, G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, T. Borman, G. Carvalho, M. Chirico, M. D. Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, C. Martino, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. J. F. Ter Braak and J. Weedon, vegan: Community Ecology Package, 2026, https://vegandevs.github.io/vegan/.

44 A. Oliver, M. Kay and D. G. Lemay, TaxaHFE: a machine learning approach to collapse microbiome datasets using taxonomic structure, *Bioinform. Adv.*, 2023, **3**, vbad165.

45 M. Kuhn and H. Wickham, Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles, 2020, https://www.tidymodels.org.

46 C. Walsh, E. Stallard-Olivera and N. Fierer, Nine (not so simple) steps: a practical guide to using machine learning in microbial ecology, *mBio*, 2024, **15**, e02050–e02023.

47 R. Tume, S. E. Sherbiny, R. Bono, T. Gautier, J. P. Pais de Barros and T. Meroño, The balance between proinflammatory, "bad", and immunomodulatory, "good", lipopolysaccharide for understanding gut-derived systemic inflammation, *Front. Immunol.*, 2025, **16**, 1588129.

48 E. J. Culp, N. T. Nelson, A. A. Verdegaal and A. L. Goodman, Microbial transformation of dietary xenobiotics shapes gut microbiome composition, *Cell*, 2024, 6327–6345, DOI: 10.1016/j.cell.2024.08.038.

49 D. Ndeh and H. J. Gilbert, Biochemistry of complex glycan depolymerisation by the human gut microbiota, *FEMS Microbiol. Rev.*, 2018, **42**, 146–164.

50 Y. Sato, S. Itagaki, T. Kurokawa, J. Ogura, M. Kobayashi, T. Hirano, M. Sugawara and K. Iseki, In vitro and in vivo antioxidant properties of chlorogenic acid and caffeic acid, *Int. J. Pharm.*, 2011, **403**, 136–138.

51 J. Pérez-Jiménez, M. E. Díaz-Rubio and F. Saura-Calixto, Non-extractable polyphenols, a major dietary antioxidant: occurrence, metabolic fate and health effects, *Nutr. Res. Rev.*, 2013, **26**, 118–129.

52 A. González-Sarrías, J. C. Espín and F. A. Tomás-Barberán, Non-extractable polyphenols produce gut microbiota metabolites that persist in circulation and show anti-inflammatory and free radical-scavenging effects, *Trends Food Sci. Technol.*, 2017, **69**, 281–288.

53 O. Sklenickova, J. Flesar, L. Kokoska, E. Vlkova, K. Halamova and J. Malik, Selective Growth Inhibitory Effect of Biochanin A Against Intestinal Tract Colonizing Bacteria, *Molecules*, 2010, **15**, 1270–1279.

54 I. Attaye, J. J. Witjes, A. M. Koopen, E. W. J. van der Vossen, D. Zwirs, K. Wortelboer, D. Collard, E. M. Kemper, M. Winkelmeijer, J. J. Holst, S. L. Hazen, F. Kuipers, E. S. G. Stroes, A. K. Groen, W. M. de Vos, M. Nieuwdorp and H. Herrema, Oral *Anaerobutyricum soehngenii* augments glycemic control in type 2 diabetes, *iScience*, 2024, **27**, 110455.

55 E. d'Hennezel, S. Abubucker, L. O. Murphy and T. W. Cullen, Total Lipopolysaccharide from the Human Gut Microbiome Silences Toll-Like Receptor Signaling, *mSystems*, 2017, **2**, DOI: 10.1128/msystems.00046-17.

56 Y. Y. Bouzid, S. M. G. Wilson, Z. Alkan, C. B. Stephensen and D. G. Lemay, Lower Diet Quality Associated with

Subclinical Gastrointestinal Inflammation in Healthy U.S, Adults, *J. Nutr.*, 2024, 1449–1460.

57 J. Firrman, L. Liu, G. A. Argoty, L. Zhang, P. Tomasula, M. Wang, S. Pontious, M. Kobori and W. Xiao, Analysis of Temporal Changes in Growth and Gene Expression for Commensal Gut Microbes in Response to the Polyphenol Naringenin, *Microbiol. Insights*, 2018, **11**, DOI: 10.1177/1178636118775100.

58 J. Firrman, L. Liu, L. Zhang, G. Arango Argoty, M. Wang, P. Tomasula, M. Kobori, S. Pontious and W. Xiao, The effect of quercetin on genetic expression of the commensal gut microbes Bifidobacterium catenulatum, Enterococcus caccae and Ruminococcus gauvreauii, *Anaerobe*, 2016, **42**, 130–141.

59 A. Braune and M. Blaut, Bacterial species involved in the conversion of dietary flavonoids in the human gut, *Gut Microbes*, 2016, **7**, 216–234.