# Dietary protein source strongly alters gut microbiota composition and function

Authors: J. Alfredo Blakeley-Ruiz□*[1], Alexandria Bartlett*[1,2], Arthur S. McMillan[3], Ayesha Awan[1,3], Molly Vanhoy Walsh[1], Alissa K. Meyerhoffer[1], Simina Vintila[1], Jessie Maier[1], Tanner Richie[1], Casey M. Theriot[3], Manuel Kleiner□[1]

1. Department of Plant and Microbial Biology, College of Agricultural Sciences, North Carolina State University, Raleigh, NC, USA
2. Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA
3. Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, NC, USA

* These authors contributed equally to this manuscript
□ Please refer correspondence to J. Alfredo Blakeley-Ruiz (jablakel@ncsu.edu) or Manuel Kleiner (manuel_kleiner@ncsu.edu)

## Abstract

The source of protein in a person's diet affects their total life expectancy[1–3]. However, the mechanisms by which dietary protein sources differentially impact human health and life expectancy are poorly understood[2,4]. Dietary choices have major impacts on the composition and function of the intestinal microbiota that ultimately mediate host health[5–7]. This raises the possibility that health outcomes based on dietary protein sources might be driven by interactions between dietary protein and the gut microbiota. In this study, we determine the effects of seven different sources of dietary protein on the gut microbiota in mice. We apply an integrated metagenomics-metaproteomics approach to simultaneously investigate the effects of these dietary protein sources on the gut microbiota's composition and function. The protein abundances measured by metaproteomics can provide microbial species abundances[8], and evidence for the phenotype of microbiota members on the molecular level because measured proteins allow us to infer the metabolic and physiological processes used by a microbial community[9]. We showed that dietary protein source significantly altered the species composition and overall function of the gut microbiota. Different dietary protein sources led to changes in the abundance of microbial amino acid degrading proteins and proteins involved in the degradation of glycosylations on dietary protein. In particular, brown rice and egg white protein increased the abundance of amino acid degrading enzymes and egg white protein increased the abundance of bacteria and proteins usually associated with the degradation of the intestinal mucus barrier. These results show that dietary protein source can change the gut microbiota's metabolism, which could have major implications in the context of gut microbiota mediated diseases.

## Introduction

Source of dietary protein has a major impact on human health. People who consume high amounts of animal protein have higher mortality rates than those who consume mostly plant-based protein[1,3]. Egg protein and red meat protein in particular have been shown to lead to increased mortality rates among humans[2] and a diet high in red meat protein has been shown to increase inflammation in a model of colitis[10]. Replacing animal protein sources with plant protein sources reduces mortality rates[2]. Currently, we have a limited understanding of the underlying causes, but the gut microbiota has been implicated as potentially having a major role in the differential health impacts of different dietary protein sources[4,11]. Diet has been shown to change the gut microbiota's composition and function in ways that can be detrimental or beneficial to health[5,7,12,13]. For example, protein fermentation by the gut microbiota generates a number of toxins including ammonia, putrescine, and hydrogen sulfide[11,14], while fermentation of fiber and certain amino acids produces short-chain fatty acids, which are anti-inflammatory[15]. Previous studies demonstrate that amount of protein can have a greater impact on the gut microbiota's composition than other macronutrients[16], and that the source of dietary protein impacts the composition of the microbiota[17]. There is, however, limited data showing the mechanisms by which individual sources of dietary protein affect the gut microbiota's composition and function, which could mediate the consumption and production of compounds beneficial or detrimental to the host.

We hypothesized that dietary protein source affects the abundance of amino acid metabolizing enzymes from the gut microbiota, altering the abundance of pathways involved in

the production of toxins that could be detrimental to host health. To investigate this, we used an integrated metagenomic-metaproteomic approach. While metagenomics alone only allows us to assess the functional potential of a microbiota, its integration with metaproteomics provides a powerful tool for observing changes in microbiota function because the expressed products of genes (proteins) are quantified. We found that source of dietary protein not only alters the abundance of amino acid degrading proteins, but has an even greater impact on the abundance of glycan degrading proteins among other functions, indicating that dietary protein source can have wide-ranging effects on the gut microbiota.

## Results

### Integrated Metagenomic-Metaproteomic Analysis of Dietary Protein Effects on the Gut Microbiota

To determine how different sources of dietary protein affect the gut microbiota's function, we fed twelve mice (C57BL/6J), six females (group 1) and six males (group 2) split across four cages, a series of 9 fully defined diets (Fig. 1a, Supplementary table 1). Each diet contained a different single source of dietary protein whose mass represented either twenty or forty percent of the entire diet. In order, the diets were 20% soy protein, 20% casein protein, 20% brown rice protein, 40% soy protein, 20% yeast protein, 40% casein protein, 20% pea protein, 20% egg white protein, and 20% chicken bone protein. Mice consumed each diet for one week before switching to the next diet. We collected fecal samples from mice after 7 days of consumption of each diet. The chicken bone diet caused the mice to lose weight, so we discontinued the diet after 3 days and the mice consumed a standard chow diet for the rest of that week. No fecal samples were collected for the chicken bone diet. To control for the succession effects of a serial dietary intervention, we fed the mice the 20% soy diet or the 20% casein diet as a control at the end of the diet series. We analyzed samples from all mice and diets using an integrated metagenomic-metaproteomic approach[12,18] (Fig. 1b). We sequenced fecal samples using shotgun sequencing and used a genome-resolved metagenomics pipeline[19,20], which resulted in 454 metagenome-assembled genomes (MAGs) organized into 180 species groups. We used high-resolution mass spectrometry based metaproteomics to identify and quantify proteins in each sample using a protein sequence database derived from the metagenome and augmented with mouse and diet protein sequences[21,22]. In total, we identified 35,588 proteins (1,381-3,524 per sample), each distinguished as microbial, host, or dietary proteins(Extended Data Table 1). All taxonomic and functional data described in this study were quantified using the metaproteomic data.

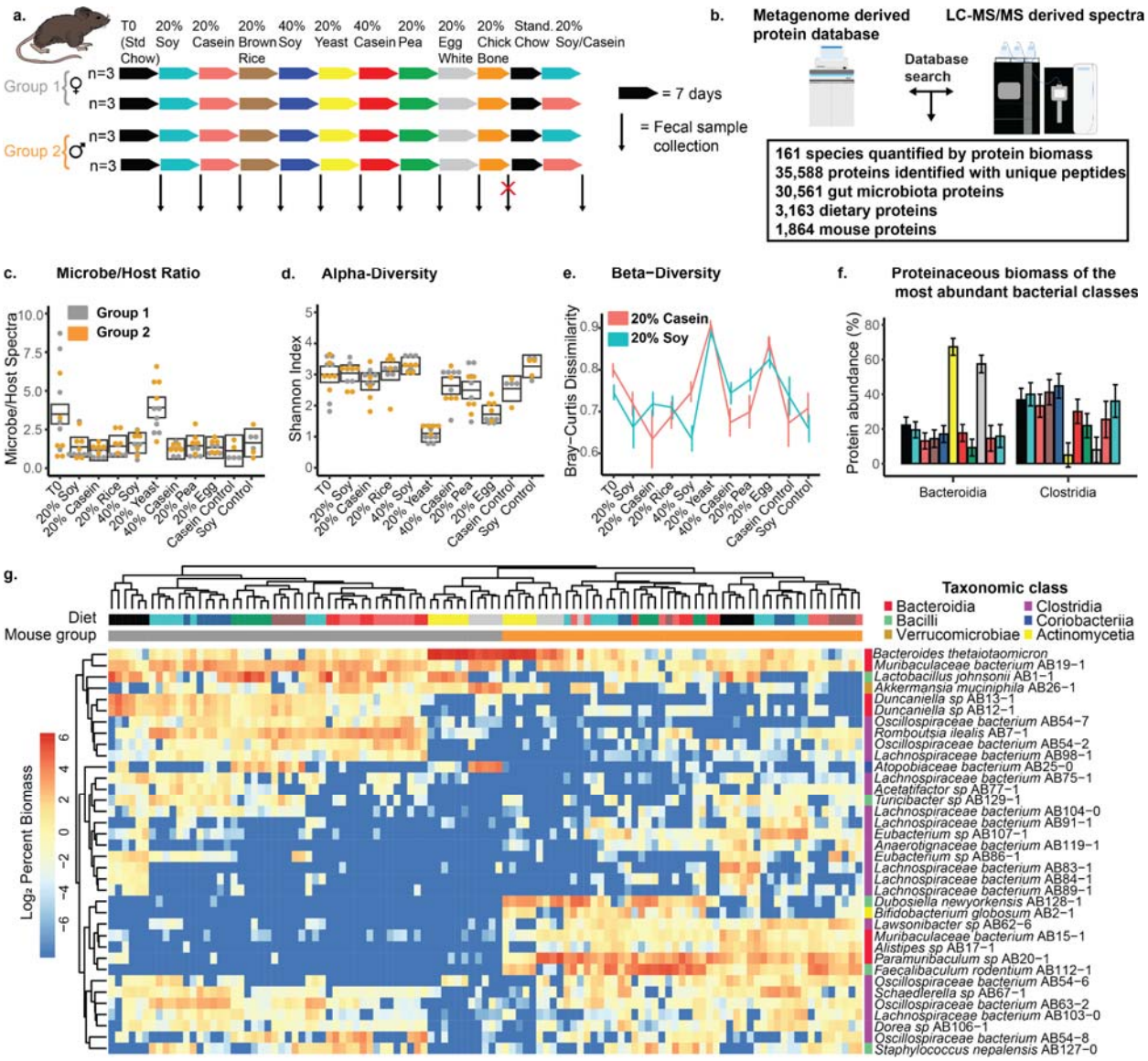


**Figure 1: Source of dietary protein alters the gut microbiota's composition.** (a) Diagram showing the experimental design, with number of cages, and order of the diets fed. Colors are used to delineate diets throughout the manuscript. Each row of arrows represents one cage (n=3 mice per cage). (b) A diagram illustrating the integrated metagenomic-metaproteomics method used to analyze the samples along with raw metrics: quantifiable species and number of proteins. (c) Ratio of spectra assigned to microbes versus the host; boxes represent 95% confidence intervals calculated on a linear mixed-effect model (Extended Data Table 2). Boxes that do not overlap denote significance (d) Shannon diversity index of the gut microbiota across all diets; boxes represent 95% confidence intervals calculated on a linear mixed-effect model (Extended Data Table 2). Boxes that do not overlap denote significance. (e) Bray-Curtis dissimilarity between the initial 20% soy diet (teal) or 20% casein diet (red) and all other diets. Error bars reflect 95% confidence intervals for all line graphs as calculated by the Rmisc package in R, but actual significance was calculated by PERMANOVA. (f) Abundances of the two most abundant bacterial classes based on summed protein abundance. Error bars are 95% confidence intervals calculated using a linear mixed effects model (Extended Data Table 2). (g) A hierarchically clustered (ward.D2 algorithm on euclidean distances) heatmap depicting the clustering by species group abundance of the 36 most abundant species in the study. Species were considered abundant if they had at least 5% of the microbial biomass in at least one sample.

**Source of dietary protein alters gut microbiota composition**

To assess the effect of dietary protein source on microbiota composition, we quantified the proteinaceous biomass for each species using metaproteomics data and obtained measurements for 161 distinct species (Extended Data Table 3) (Kleiner et al., 2017). We divided the number of spectra assigned to microbial proteins by the number of spectra assigned to host proteins to create a measure of microbial load (Extended Data Table 4). We found that the yeast protein diet significantly increased microbial load as compared to all other defined protein diets (Fig. 1c). We calculated within sample diversity (alpha diversity) and between sample compositional dissimilarity (beta diversity) based on the abundances of the quantifiable species. We found that the yeast and egg white diets significantly reduced the alpha diversity (Shannon diversity index) and richness (number of species) of the gut microbiota relative to all other diets (Fig. 1d; Supplementary Fig. 1a; Extended Data Table 4). We evaluated the effects of dietary protein on gut microbiota composition by comparing the microbiota from all diets to the initial 20% soy and 20% casein diets using the Bray-Curtis dissimilarity index. We found that the composition of the gut microbiota was most similar when the source of dietary protein was the same, regardless of the amount of protein in the diet, and that the yeast and egg white diets yielded the most dissimilar microbiota compositions (Fig. 1e). Testing with PERMANOVA ($q < 0.05$) showed that the community composition was usually significantly different when the source of dietary protein was different, but not when it was the same (Extended Data Table 5). These results show that the source of dietary protein had a greater effect on the gut microbiota than the amount of protein in the diet across three dimensions of the gut microbiota: microbial load, within sample diversity of species, and compositional dissimilarity between samples. The large differences in composition in the egg white and yeast protein diets were driven by a decrease in the abundances of species from the class Clostridia in favor of species from the class Bacteroidia (Fig. 1f). Since we observed fewer species in the class Bacteroidia overall, it makes sense that a drop in Clostridia in favor of Bacteroidia would result in lower alpha diversity in the yeast and egg white diets (Fig. 1g).

To explore what microbial taxa drive differences in microbiota composition between dietary protein sources, we focused on the most abundant individual microbial species (>5% of the microbial protein biomass in at least one sample) and hierarchically clustered them by abundance (Fig. 1g). This revealed three major clusters separating most samples by mouse group with the exception of the yeast and egg white diets which together formed a separate cluster that internally showed separation by mouse group. The T0 samples fell into the major mouse group clusters which indicates that the two mouse groups had distinct gut microbiotas at the start of the experiment. Within the mouse group clusters the microbiota clustered by source of dietary protein, which was also observed in principal component analysis (Supplementary Fig. 1). These results suggest that the source of dietary protein produces distinct gut microbiota compositions independent of starting microbiota composition (for details see Supplementary Results Section A). For example, in the yeast diet, *Bacteroides thetaiotaomicron* (*B. theta*) dominates the microbiota regardless of mouse group. *B. theta* abundance also increased in response to the egg white diet. However, there were additional species specific to each mouse group that also increased in abundance in response to the egg white diet (Fig. 1g; Supplementary Fig. 2; Supplementary. Fig. 3). In group 1, these species were *Akkermansia muciniphila* and *Atopobiaceae* bacterium AB25_9, while in group 2 these species were

*Paramuribaculum* sp. and *Dubosiella newyorkensis.* Interestingly, both *A. muciniphila* [23] and *Paramuribaculum* sp. [24] have been reported to forage on intestinal mucin and *B. theta* has been shown to switch towards mucin foraging based on diet [25]. These results show that the source of dietary protein changes the gut microbiota's composition and suggests that egg white protein could promote mucin-foraging bacteria.

**Source of dietary protein alters the gut microbiota's overall function**

To evaluate gut microbiota function, we used the normalized abundances of gut microbiota proteins as a measure of the investment of the microbiota into metabolic and physiological functions. We first used automated annotation tools to assign functions to proteins. Since the annotations from these tools were not always accurate, we manually curated the annotations of 3,959 proteins by searching the sequences against Swiss-Prot [26] and InterProScan [27]. We prioritized the most abundant proteins and used this information to extrapolate the functions for a total of 14,547 similarly annotated proteins, which in total represented between 74 and 86 percent of the total microbial protein abundance in each sample (Extended Data Table 6). Based on the annotations, we assigned broad functional categories, such as gene expression, monosaccharide metabolism, and amino acid metabolism and more detailed functional categories, such as glycolysis and ribosomal proteins to each of these proteins (Extended Data Table 6). We then used the relative protein abundances to tabulate the investment of the microbiota into each of these functions. All of the broad functional categories, except for secondary metabolism, had significant changes in abundance due to dietary protein (ANOVAs based on linear mixed effects modeling p-value $< 0.05$; Extended Data Table 7; Fig. 2, Supplementary Fig. 4, Supplementary Fig. 5), which indicates that the source of dietary protein changes the gut microbiota's function. Hierarchical clustering of all samples by abundances of broad functional categories revealed that the yeast and egg white diets clustered separately from all the other diets, similar to the results from the taxonomic clustering (Fig. 1g, Supplementary Fig. 6); however, a similar analysis at the detailed functional level revealed separate yeast, rice, and egg white clusters, with some outliers (Supplementary Fig. 7).

Among the abundant broad functional categories that significantly changed in abundance due to dietary protein were fermentation, gene expression, monosaccharide metabolism, amino acid metabolism, and glycan degradation (Supplementary Results Section B; Supplementary Fig. 4; Supplementary Fig. 5). Changes in the abundance of fermentation enzymes were driven by a decrease in short-chain fatty acid producing enzymes in the yeast and egg white diets relative to the other diets (Supplementary. Fig. 4a). Changes in the abundance of gene expression were driven by an increase in the abundance of ribosomal protein in the yeast diet relative to all other diets (Supplementary Fig. 4b). Overall changes in the abundance of monosaccharide metabolism were driven by glycolysis (Supplementary Fig. 4d); however, there were several additional changes in other pathways as well (Supplementary Results Section B). For example, glucosamine, fucose, and sialic acid metabolizing enzymes increased in abundance in the casein and egg white protein diets relative to the other diets (Supplementary. Fig 4e). The two broad functions that represented more than 1% of the total microbial proteome with the greatest effect size due to diet were amino acid metabolism and glycan degradation, with F-statistics of 28 and 92 respectively (Fig. 2; Extended Data Table 7). Glycan degradation significantly increased relative to all other diets in the yeast and egg white

diets, while amino acid metabolism increased in the rice and egg white diets relative to all other diets except the 40% casein diet (Fig 2). Significant changes in the abundance of amino acid metabolism supported our initial expectation that the response of the microbiota to different dietary protein sources would likely relate to amino acid metabolism; however, we were surprised to discover that the abundance of glycan degrading enzymes responded more strongly to the source of dietary protein than did enzymes for amino acid metabolism. This suggests that glycan degradation instead of amino acid metabolism may be the major driver of taxonomic and functional changes in the gut microbiota in response to dietary protein source. These two functions will be discussed in detail in subsequent sections.

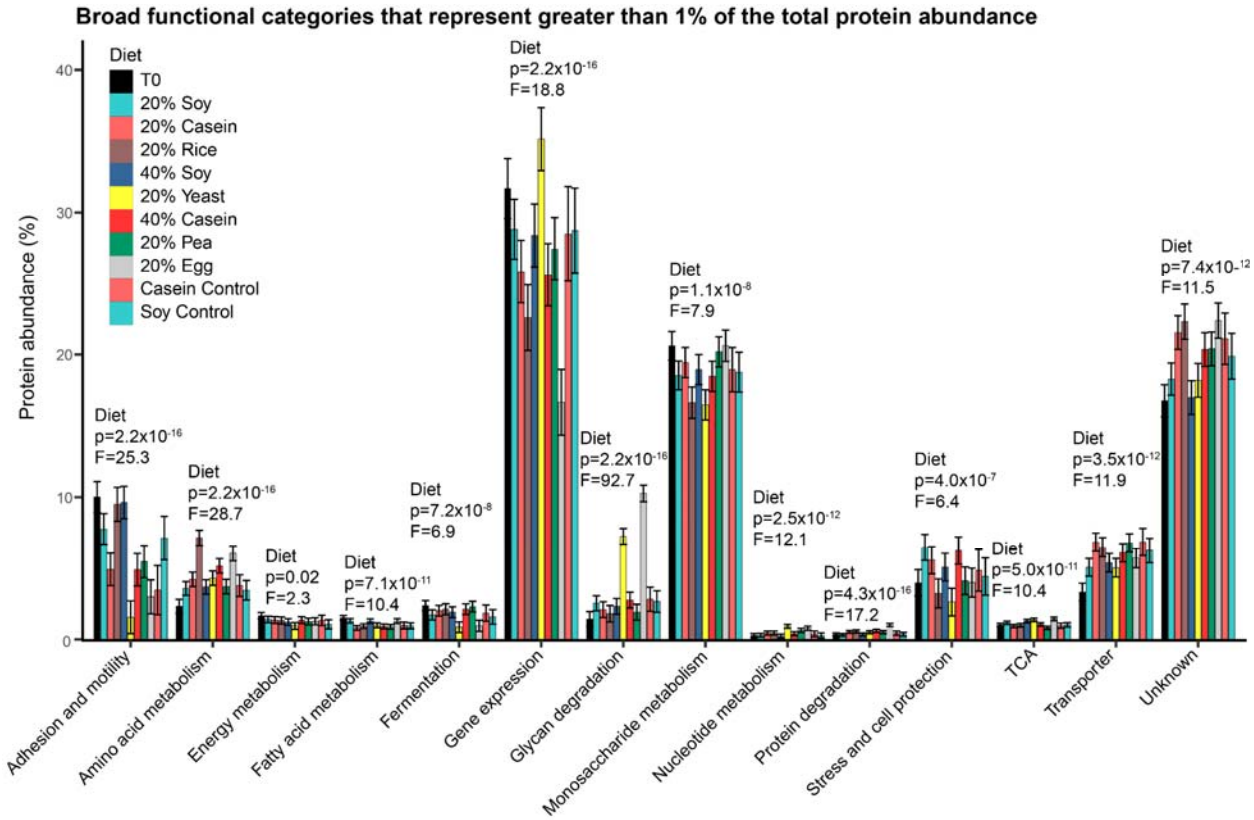


**Figure 2: Broad functional categories of microbial proteins change significantly in abundance due to the source of dietary protein.** Abundance of broad functional categories that represent at least 1% of the microbial protein abundance in at least one diet. The abundance is a modeled mean calculated from mixed effects models and the error bars represent 95% confidence intervals calculated from these models. All the categories shown here had a p-value for the diet factor below 0.05; p-value < $2.2\times10^{-16}$ is the lower limit of the method. For underlying data see Extended Data Tables 7 and 8. For higher resolution of functional categories, e.g. fermentation, see Supplementary Figs. 4-5.

## Source of dietary protein affects the abundance of amino acid degrading enzymes

To explore the effects of dietary protein source on microbiota amino acid metabolism, we manually classified 911 proteins (Extended Data Table 9) representing 68 enzyme functions (Extended Data Table 10) according to their involvement in the degradation, synthesis, interconversion, or reversible reactions of specific amino acid pathways (Supplementary Figs. 8-

19). In all diets except the yeast and standard chow diets, we observed that the microbiota was trending toward amino acid degradation instead of synthesis. We found that amino acid degrading enzymes were on average 2- to 6-fold more abundant than amino acid biosynthesis enzymes (Fig. 3a and 3d). Amino acid degrading enzymes were significantly more abundant in the rice and egg diets as compared to all the other diets (Fig. 3a), which is consistent with the observation that dietary proteins were significantly more abundant in the fecal samples of the rice and egg protein diets as compared to all other diets (Fig. 3e), suggesting that there may be a connection between the digestibility of dietary protein and amino acid degradation by the gut microbiota. Though amino acid synthesis enzymes were generally less abundant, we did observe a trend towards an increase in amino acid synthesis enzymes in the yeast protein diet relative to the other diets. This trend was not significant (Fig. 3d), but we observed several individual synthesis enzymes to be significantly increased in the yeast protein diet relative to other diets. These enzymes were involved in the synthesis of branched-chain amino acids (Supplementary Fig. 9), cysteine (Supplementary Fig. 10), lysine (Supplementary Fig. 16), proline (Supplementary Fig. 18), or tyrosine (Supplementary Fig 19).

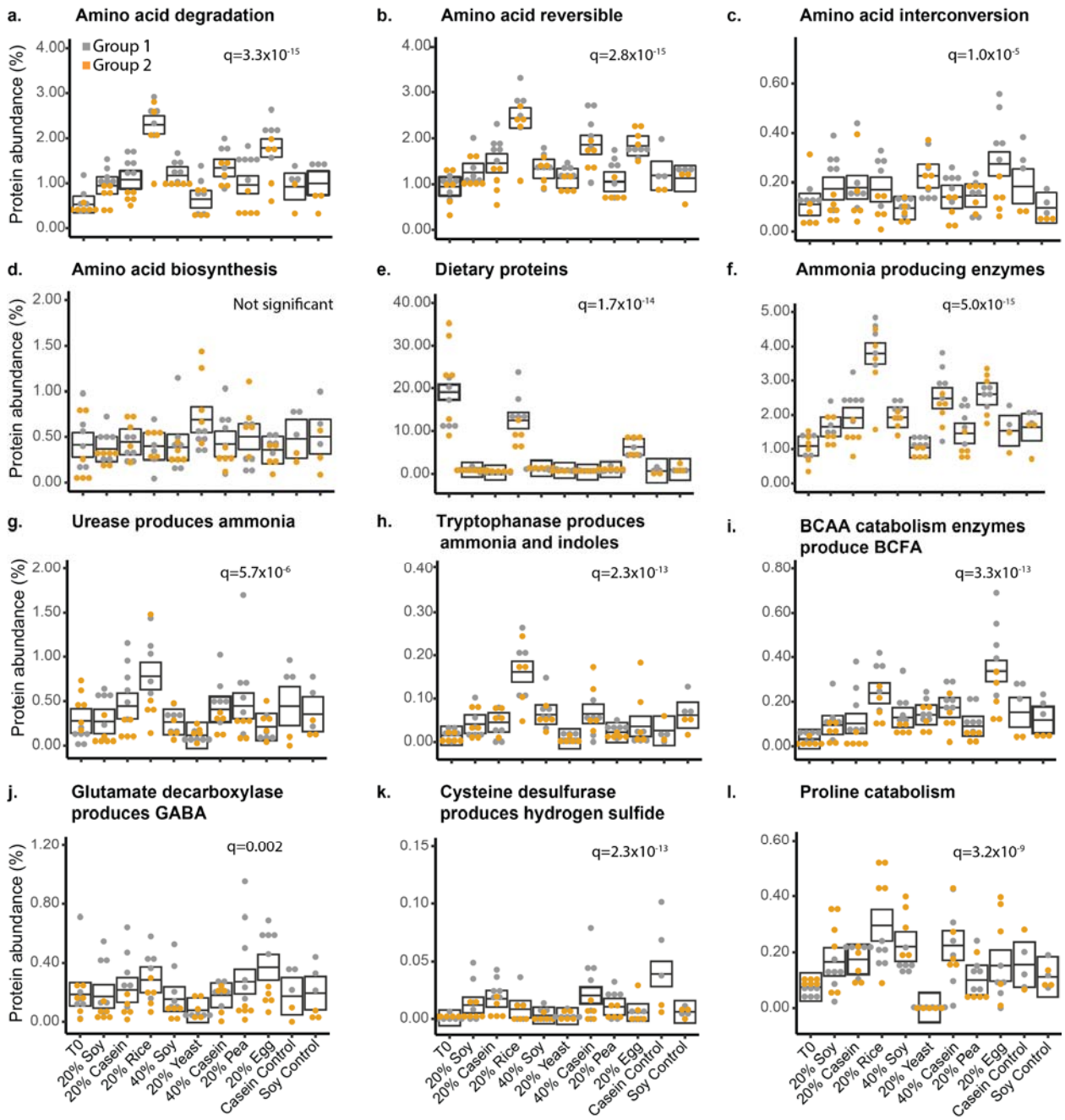


**Figure 3: Amino acid degradation increases in the rice and egg diets.** Box plots depict the percent abundance (out of the total microbial protein abundance) of different categories of microbial amino acid metabolism proteins. The exception is the dietary proteins, which are based on the percent protein abundance of the total metaproteome. The boxes represent the 95% confidence interval of the mean (line) for each diet from a complete mixed-effects model, and the q-values represent the FDR controlled p-values for the diet factor from an ANOVA on these models (q<0.05 indicates significance) (Extended Data Tables 11-12). Boxes that do not overlap indicate statistical significance. Circle dots represent actual values per sample and are colored by mouse group. Abundance of proteins classified as (a) degrading an amino acid, (b) reversible, (c) converting between two amino acids, and (d) synthesizing an amino acid. (e) Abundance of all dietary proteins detected in each condition. (f) Abundance of enzymes that are likely to produce ammonia. Enzymes classified as degrading or reversible were included as long as ammonia was one of the potential products. Summed abundance of all proteins classified as (g) urease, (h) tryptophanase, (i) involved in branched-chain amino acid (BCAA) degradation to branched-chain fatty acid (BCFA) and classified as branched-

chain amino acid aminotransferase or ketoisovalerate oxidoreductase, (j) glutamate decarboxylase, (k) cysteine desulfurase, and (l) proline racemase or D-proline reductase.

Not all amino acid degrading enzymes increased in both the brown rice and egg white protein diets; sometimes they increased in one or the other (Supplementary Figs. 7-19). For example, enzymes associated with the degradation of threonine were more abundant in the egg white protein diet (Supplementary Fig. 13), while enzymes associated with tryptophan degradation were increased in the brown rice protein diet (Supplementary Fig. 19). Rice and egg were not the only diets in which the abundances of specific amino acid degrading enzymes increased. Alanine dehydrogenase increased in the 40% soy diet relative to the pea, yeast, 20% soy, and 20% casein diets (Supplementary Fig. 10) and cysteine desulfurase increased in the 40% casein and casein control diets relative to most other diets (Supplementary Fig. 10).

Changes in amino acid degradation by the gut microbiota have potential implications for host health by directly affecting local tissues or through interactions along the gut-brain axis depending on the metabolites produced by specific pathways[11,28]. We identified six categories of amino acid degradation pathways that are relevant to host health because they produce compounds that are toxic, anti-inflammatory, neurotransmitters, or otherwise related to disease. The toxic compounds include ammonia produced by deaminating enzymes and urease[29,30], as well as hydrogen sulfide produced by cysteine desulfurase[31]. The neurotransmitters include γ-aminobutyric acid (GABA) produced by glutamate decarboxylase[32], and indoles produced by tryptophanase[33]). The anti-inflammatory metabolites are branched-chain fatty acids produced by enzymes that degrade branched-chain amino acids[34,35].

We found ammonia producing enzymes to be significantly more abundant in the brown rice protein diet as compared to all other diets, and also more abundant in the egg white protein and 40% casein diets as compared to the standard chow, 20% soy, yeast, pea, and control diets (Fig. 3f and 3g). We observed cysteine desulfurases to be significantly increased in the 40% casein and casein control diets relative to other diets (Fig. 3k). Tryptophanase significantly increased in the brown rice protein diet relative to all other diets, while glutamate decarboxylase increased in the egg white protein diet relative to all other diets except brown rice, pea, and the control diets (Fig. 3h and 3j). We observed that branched-chain amino acid degrading enzymes were significantly increased in the egg white protein diet relative to all other diets (Fig. 3i), and proline degrading enzymes were increased in the brown rice protein diet relative to other diets, except the 40% soy and 40% casein diets where we also observed proline catabolism to be significantly increased relative to the standard chow, yeast, and pea diets (Fig. 3l). These results show that the source of dietary protein can alter overall amino acid metabolism in the gut microbiome, as well as the abundance of different pathways. These changes have the potential to affect host physiology and health.

**Gut microbes express distinct glycoside hydrolases to grow on different sources of dietary protein**

Surprisingly, glycan degrading enzymes (glycoside hydrolases) showed the largest overall changes in response to dietary protein source (Fig. 2, Extended Data Table 7). Specifically, these enzymes increased significantly in abundance in the yeast and egg protein diets compared to the other diets. To further investigate the interaction of these glycan degrading enzymes with dietary protein we manually curated the functional assignments and

potential substrate specificity of the 1,059 microbial glycoside hydrolases detected in our metaproteomes (Extended Data Table 13).

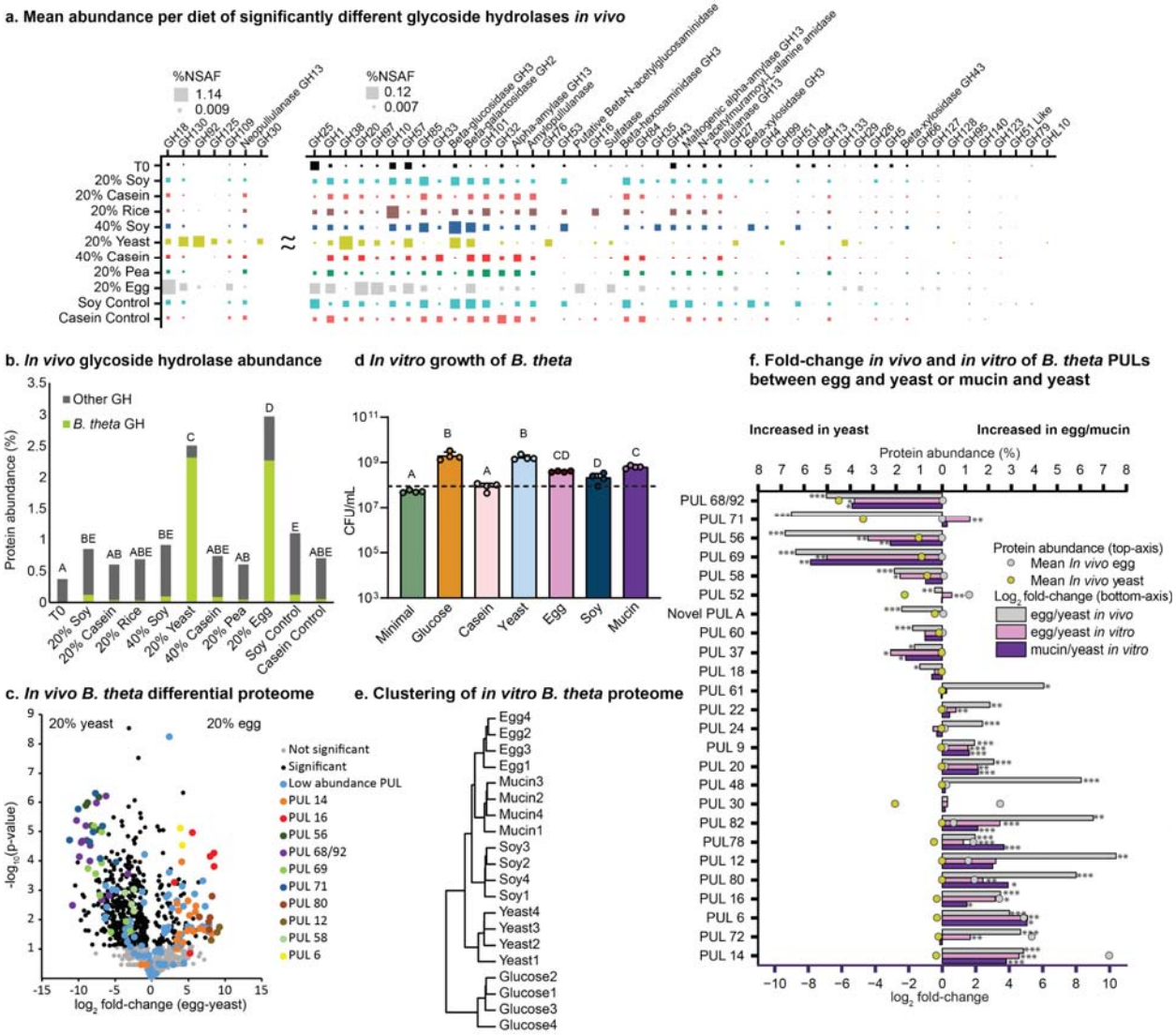


**Figure 4: Glycosylations on dietary proteins drive shifts in microbial composition.** (a) Mean summed protein abundance per diet of glycoside hydrolases with significantly different abundances between diets (q<0.05 in mixed-effects ANOVA models). (b) Mean combined protein abundance of proteins confirmed to be glycoside hydrolases. The proportion of these proteins that belong to *B. theta* is highlighted. Diets that do not have overlapping letters also have non-overlapping 95% confidence intervals for each diet calculated from a complete mixed-effects model. (c) Volcano plot of -$\log_{10}$p-values (Welch's t-test; FDR controlled at q<0.05) versus the $\log_2$ fold-change of *B. theta* proteins under the yeast and egg white protein diets *in vivo* after recalculating the protein abundance based on proteins only assigned to *B. theta*. Filled circle symbols, indicating individual proteins, were colored based on the polysaccharide utilization locus (PUL) to which the protein belongs. We only colored the proteins from PULs that had an absolute difference of 0.5% or greater between the yeast and egg diets. (d) Colony forming units per mL (CFU/mL) of *B. theta* grown in defined media with dietary proteins as the sole carbon source. The dotted line indicates T0 CFU/mL. Media that do not share letters are significantly different based on ANOVA and Tukey HSD multiple comparisons after log transformation (p-value < 0.05). (e) Hierarchical clustering (ward.D2 on Euclidean distances) of the *in vitro B. theta* proteome under different media. (f) *In vivo* and *in vitro* comparison of the summed protein abundance of PULs. The bottom axis depicts the $\log_2$ fold-change between egg white and yeast protein or mucin and yeast protein. The top axis depicts the mean protein abundance of the PULs *in vivo* in the yeast diet on

the left and in the egg diet on the right. A Welch's t-test (with FDR control) was performed to detect significant changes in PUL protein abundances (*** = q < 0.01, ** = q < 0.05, * = q < 0.1) (Extended Data Tables 18 and 20)..

We grouped the validated glycoside hydrolases into 91 families based on the CAZy database (Extended Data Table 14) (Lombard et al., 2014). Of these families, 54 significantly changed in abundance between the different dietary protein sources (ANOVA, q<0.05) (Extended Data Table 15). Different glycoside hydrolase families increased in abundance in the yeast, egg, soy, rice, and casein protein diets suggesting that distinct glycans drive their abundance changes across the different diets (Fig. 4a, Extended Data Table 16). The most abundant glycoside hydrolase families, GH18 in the case of egg and GH92 in the case of yeast, have previously been associated with the degradation of glycans conjugated onto proteins (glycosylations) as part of polysaccharide utilization loci (PULs). PULs are operons that contain all the proteins necessary to import and degrade a specific glycan structure[36]. These GH18s are endo-β-N-acetylglucosaminidases that break the bond between two acetylglucosamine residues attached to asparagine in N-linked glycoproteins. This reaction releases the glycan from the glycoprotein[37]. Meanwhile, GH92s, which are alpha-mannosidases, have been previously associated with the release of mannose residues from the glycosylations on yeast mannoproteins[37].

We found that the total abundance of glycoside hydrolases increased from <1% in the majority of diets to >2.5% in the yeast and egg diets (Fig. 4b). Additionally, we observed a general trend towards an increased abundance of glycoside hydrolases in all defined diets compared to the T0 (standard chow) diet; however, the increase was only significant for the soy, yeast and egg protein diets (Fig. 4b). The majority of the glycoside hydrolases in the yeast and egg diets came from *B. theta* (Fig. 4b). Since *B. theta* is one of the primary drivers of the changes in microbiota composition in these diets (Fig. 1g), this suggests that glycoside hydrolases are closely associated with the observed changes in microbiota composition.

To examine the specific role of *B. theta* in glycan degradation in the yeast and egg protein diets, we compared the abundances of all *B. theta* proteins in the metaproteome between the two diets. Out of 1,420 detected *B. theta* proteins, the abundances of 592 proteins significantly differed between the two diets (q < 0.05, Welch's t-test) (Fig. 4c; Extended Data Table 17). Many of the significant proteins that were the most abundant and had the greatest fold-change between the two diets came from PULs (for examples, see Supplementary Fig. 21). Between 10% and 25% of the total protein abundance of *B. theta* in the yeast and egg protein diets came from these PULs (Extended Data Table 18). Proteins within each PUL tended to be either increased in egg relative to yeast or decreased in egg diet relative to yeast (Figure 4c).

Several types of glycosylations on yeast and egg proteins have been previously characterized and many of the PULs for which we detected expression have been shown to degrade these types of glycosylations. In yeast, the cell wall is made of interconnected chitin, glucans and mannoproteins[38,39]. The second most abundant yeast protein in fecal samples from mice on the yeast diet is a homolog of the mannoprotein CWP1 from *Saccharomyces cerevisiae*. PULs 68/92 (BT3773-3792) and 69 (BT3854-3862) (Fig. 4c; Suppl. Fig. 21) have been specifically shown to degrade □-mannans attached to yeast mannoproteins in *Saccharomyces cerevisiae* [37]. Similarly, PUL 56 (BT3310-3314), another PUL increased in the yeast diet, has been shown to specifically degrade yeast β-glucans [40]. Egg white protein contains mucins called ovomucin, ovalbumin, ovotransferrin, and ovomucoid, which have been

previously shown to be N-glycosylated with acetylglucosamine and mannose containing glycans[41,42]. The majority of PULs abundant in the egg white protein diet samples, PUL14 (BT1032-1051), PUL6 (BT3017-0318), PUL16 (1280-1285), PUL 80 (BT4295-BT4299), and PUL 12 (BT0865-0867) (Supplementary Fig. 21), have been previously shown to be associated with the degradation of host intestinal mucin[43]. This suggests that similar enzymes are used to degrade egg white glycoproteins and host-derived mucin glycoproteins. An additional abundant PUL, PUL72 (BT3958-BT3965), has been previously implicated in the degradation of mannoproteins of mammalian origin[37] and our result suggests that PUL72 is also involved in the degradation of mannoproteins from non-mammalian vertebrates.

To test if *B. theta* could grow on yeast and egg white protein as predicted from the *in vivo* data, and if the expression of PULs was driven by direct responses to the dietary protein sources and not by indirect responses, such as the dietary protein source inducing increased foraging of host mucus, we characterized *B. theta* growth and gene expression on dietary protein sources *in vitro*. We used a defined culture media and supplemented purified dietary protein sources as the sole carbon source to determine if this supported *B. theta* growth. We found that *B. theta* grew in the presence of glucose (control), yeast protein, egg white protein, soy protein, and intestinal mucin (Fig. 4d). We analyzed the proteomes of *B. theta* in these five different conditions *in vitro* to determine if PULs played a role in growth (Extended Data Table 19). An overall comparison of the proteome between the media supplemented with 5 different protein sources revealed that egg white protein and mucin had the most similar proteome, and the proteome from the glucose control clustered separately from all other protein sources (Fig. 4e; Suppl. Fig. 22 and 23). We observed that 15 out of 24 PULs that were significantly different in abundance between the egg white and yeast protein diets *in vivo* were also significantly different in the same direction *in vitro* (Fig. 4f). In addition, 12 of these 15 PULs showed the same expression pattern in both the mucin and egg white protein media as compared to the yeast protein medium (Fig. 4f, Extended Data Tables 18 and 20). These results indicate that the glycosylations on yeast and egg white proteins drive the increase in abundance of *B. theta* in the yeast and egg white protein diets, and that egg white protein and intestinal mucin share similar glycosylations leading to the expression of similar PULs for their degradation.

Several glycoside hydrolases also increased in the presence of the rice and soy diets. In the soy diet, the expression of glycoside hydrolases was reproducible, increasing in abundance each of the three times the mice were fed a soy diet. Most notably β-glucosidases and β-xylosidases from the CAZy clan GH3 were increased in the soy diets, while the abundance of GH16s increased in the rice protein diet. The abundance change of GH16s in the rice protein diet was driven by a single protein from the uncultured species *Oscillospiraceae* bacterium AB63-2 that was only detected in the rice protein diet. In the presence of soy protein, but not rice protein, this bacterium reproducibly expressed different glycoside hydrolases including two β-glucosidases and one β-xylosidase from the clan GH3 (Extended Data Table 13). These results suggest that different protein sources generally affect the function and composition of the gut microbiota through the different glycan structures attached to their glycoproteins.

## Discussion

We have shown that the source of dietary protein significantly alters the gut microbiota's composition and function. We also found that changing the source of dietary protein had a

greater impact on the gut microbiota than changing the amount of protein in the diet. We further show that yeast and egg white dietary protein affect the composition of the gut microbiota substantially more than the other protein sources we tested, and surprisingly that this is more tied to their glycosylation patterns than their amino acid composition. *B. theta* proteins specifically associated with the degradation of glycosylations on yeast proteins were increased when the mice were fed the yeast protein diet[37,40], while proteins previously associated with the degradation of host mucins increased in the egg white protein diet[43]. This suggests that the gut microbiota, and *B. theta* in particular, respond to the specific glycosylations on proteins. We were able to further confirm the growth and gene expression responses to dietary protein sources for *B. theta in vitro* confirming that the responses observed *in vivo* were a direct response to the dietary protein source and not indirect responses to other diet driven host or microbiota changes. The microbiota's response to the yeast protein diet warrants further investigation in the context of potential implications for host health especially in light of literature suggesting yeast as a potential probiotic[44]. Previous studies in mice have shown that diets which promote bacteria and their enzymes that degrade the mucins in the intestinal mucus layer[6,12] can make the host more susceptible to enteric inflammation and infection[6,35]. The fact that *B. theta* shows a similar gene expression response to egg white protein and mucin *in vivo* and *in vitro* suggests that egg white protein promotes bacteria and enzymes that also degrade mucin, which could have negative implications for the host.

Further studies are needed to better understand if egg white protein driven increases in bacterial enzymes that can degrade mucin, are in fact associated with increased degradation of the mucus layer. Additionally, we need a better understanding of the effect of egg white protein on the gut microbiota in the context of different dietary compositions such as, variations in amounts and diversity of dietary fiber or diets with multiple protein sources. Finally, our work also demonstrates that the enzymes commonly used as markers for microbiota foraging on host mucins (e.g. GH18, GH109, GH20, GH29), can also be induced by specific dietary sources of protein and thus their detection can have differing implications.

Furthermore, we observed that the source of dietary protein had a significant effect on the abundance of microbial amino acid degrading enzymes. Amino acid degrading enzymes were most abundant in the brown rice and egg white protein diets, which also corresponded to the diets that had the highest proportion of dietary protein detected in the stool. This suggests a link between the digestibility of dietary protein sources and levels of amino acid catabolism by the microbiota. We also showed that among amino acid metabolizing enzymes the ones that produce ammonia were the most abundant. Additionally, we found that the microbiota increased abundances of urease proteins, which release ammonium by degrading urea, under the brown rice protein diet relative to all other diets. Urea is produced in the liver to control systemic ammonium concentrations[45]. Previous studies across multiple species have shown that increasing the amount of protein fed to animals leads to an increase in the concentration of ammonia in stool[46–48], which suggests that increased protein availability leads to increased amino acid deamination or urease activity in the gut. Here we show that, regardless of the amount of protein, the source of protein itself can lead to increases in amino acid deaminating enzymes and ureases in the intestinal microbiota. Gut microbiota urease activity and amino acid deamination have been linked to serious diseases like hepatic encephalopathy when liver function is disrupted[49]. Replacement of bacteria that produce these deaminating enzymes and

ureases with bacteria that do not has been suggested as a potential treatment[50]; our results suggest that additional adjustments in dietary protein source should be considered. Furthermore, we speculate that increased mortality rates associated with specific dietary protein sources such as eggs could at least in part be due to increased ammonia production in conjunction with the fact that liver function decreases with age[51]. Our results did not show an animal versus plant protein dichotomy in terms of changes in microbiota composition and function, but instead indicate that the effects are specific to the source of dietary protein.

Our study has at least two limitations preventing direct translation of microbiota responses to dietary protein sources into a human health context. First, we used purified dietary proteins, which differ from commonly consumed dietary proteins in that dietary protein sources also provide some amount of additional major diet components such as fats, carbohydrates, and fiber. For example, plant proteins usually come with a relevant amount of fiber, while animal proteins are often low in fiber and have relevant amounts of fats[5]. Second, we used fully defined diets and while this allowed us to track the effects of specific protein sources, we do anticipate that the dietary context of protein sources, such as co-consumption of multiple protein, fiber, fat and carbohydrate sources, will strongly influence the interactions of dietary protein sources with the microbiota. The power of our study lies in our ability to confirm that the source of dietary protein does impact gut microbiota function and should be considered when thinking about how diet impacts the microbiota and its implications for host health.

## Methods

*Animals and Housing*

In this study we included twelve C57BL/6J mice in two groups (six males, six females, Jackson Labs, Bar Harbor) aged 3-6 months. The male and female originated from different mouse rooms at the Jackson Labs and thus were expected to have different background microbiomes. Mice from both groups were housed in two separate cages (3 mice/cage) with a 12 h light/dark cycle. We autoclaved bedding, performed all cage changes in a laminar flow hood and maintained an average temperature of 70°F and 35% humidity. We conducted our animal experiments in the Laboratory Animal Facilities at the NCSU CVM campus (Association for the Assessment and Accreditation of Laboratory Animal Care accredited), which are managed by the NCSU Laboratory Animal Resources. Animals assessed as moribund were humanely euthanized via $CO_2$ asphyxiation, as approved by NC State's Institutional Animal Care and Use Committee (Protocol # 18-034-B).

*Animal Diets and Sample Collection*

We fed mice defined diets with a single source of purified protein (Suppl. Table 1). We fed each defined diet to all mice for 7 days, with the exception of the chicken bone broth diet. We observed clinical signs of disease including weigth loss in the mice after 3 days of the chicken bone broth diet and therefore replaced the diet with standard chow for the remainder of the 7 days. We fed the diets in this order: standard chow, 20% soy protein, 20% casein, 20% brown rice protein, 40% soy protein, 20% yeast protein, 40% casein, 20% pea protein, 20% egg white protein, 20% chicken bone broth protein, and lastly at the end of the experiment half of the mice in each group were fed the 20% soy protein and half the mice the 20% casein diet again as a control. Prior to the start of the defined diet, mice were fed autoclaved standard chow. All

defined diets were sterilized by -irradiation and mice were provided sterile water (Gibco). On day 7 of each defined diet, we collected fecal samples, prior to replacing food with the next diet. We collected samples in NAP preservation solution at a 1:10 sample weight to solution ratio, and roughly homogenized the sample with a disposable pestle prior to freezing at -80°C [52]. We had to sacrifice one mouse after the second diet (20% casein) so no additional samples were collected.

*Metagenomic DNA sequencing*

To create a database for metaproteomic analysis, we pooled fecal samples from each cage to create four cage specific metagenomes. We gathered one fecal sample from each cage for four different diets (20% rice, 40% soy, 20% yeast, 40% casein) for a total of 16 samples. To extract DNA, we followed the QIAamp DNA stool mini kit (Qiagen)-based protocol described by Knudsen *et al*. with modifications [53]. To remove the preservation solution from the samples, we added 5 mL of 1X Phosphate Buffered Saline solution (VWR) to the samples and centrifuged them (17,000 x g, 5 min) to pellet solids and bacterial cells in suspension. We removed the preservation solution and resuspended the fecal pellets in 1 mL of InhibitEX Buffer in Matrix E (MP Biomedicals) bead beating tubes. We beat the samples at 3.1 m/s for 3 cycles with 1 minute of ice cooling between each cycle using a Bead Ruptor Elite 24 (Omni International). We isolated DNA from the resulting lysate using the Qiagen QIAamp Fast DNA stool mini kit (cat. No. 51604). Samples were extracted individually and pooled by cage with each sample contributing a total of 200 ng of DNA.

We submitted genomic DNA (gDNA) to the North Carolina State Genomic Sciences Laboratory (Raleigh, NC, USA) for Illumina NGS library construction and sequencing. We used an Illumina TruSeq Nano Library kit with its provided protocol for library construction and performed sequencing on an Illumina NovaSeq 6000 sequencer. We obtained between 51,152,549 and 74,618,259 paired-end reads for each of the 4 samples.

*Metagenomic assembly and protein database construction*

To create a species specific database for metaproteomics, we assembled raw reads using a genome resolved metagenomics approach. We removed PhiX174 (NCBI GenBank accession CP004084.1) and mouse genome (mm10) contaminating sequences using BBSplit and removed adapters using BBDuk (BBMap, Version 38.06), parameters: mink = 6, minlength =20[54]. We assembled decontaminated reads individually using MetaSPAdes (v3.12.0) -k 33,55,99[55] and co-assembled them using MEGAHIT (v1.2.4) –kmin 31 –k-step 10[56]. We mapped reads from all four samples to all five assemblies using bbmap, and binned the contigs using MetaBAT (v2.12.1)[57]. We assessed the quality of the bins using CheckM (v1.1.3)[58] and automatically accepted medium quality bins with a completion score greater than 50% and less than 10%[59]. Since the purpose of metagenomics in our study was to generate a comprehensive protein sequence database and to assign proteins to species, we further accepted bins that were greater than >30% complete and <5% contaminated. We clustered the bins into species groups by 95% ANI using dRep (v2.6.2)[19,60] and assigned taxonomy using GTDB-Tk (v1.3.0, ref r95)[61].

We assembled the protein database by combining gene annotations from the metagenome with mouse and dietary protein databases[21]. For the metagenome, we annotated

the assemblies prior to binning and then for each bin individually using PROKKA (Version 1.14.6)[62]. If the contig was binned, we compiled the annotations from the bins. We then used CD-HIT-2D (Version 4.7), with a 90% identity cutoff, to compare the genes from the unbinned PROKKA output to the binned gene annotation[63]. If a gene was not present in a bin we added it to the database as an unbinned sequence. Once we compiled the microbial protein database we assigned each protein sequence a species code if it was species specific or an ambiguous, low-quality, or unbinned code if it was assigned to more than one species group, belonged to a low-quality bin, or was not present in a bin, respectively. In addition to the microbial sequences, we added a *Mus musculus* proteome (UP000000589, Downloaded 19Feb20), and the relevant dietary protein database for each sample: *Glycine max* (UP000008827, Downloaded 19Feb20), *Bos taurus* (UP000009136, Downloaded 19Feb20), *Cyberlindnera jadinii* (UP000094389, Downloaded 25May20), *Oryza sativa* (UP000059680, Downloaded 25May20) and *Gallus gallus* (UP000000539, Downloaded 25May20). Due to the lack of a reference proteome for the yellow pea diet, we created a custom pea reference with all available UniProtKB protein sequences for *Pisum sativum* (Taxon ID: 388 Downloaded 25Apr20) and the reference proteome of *Cajanus cajan* (UP000075243, Downloaded 25May20). For T0 samples taken when mice were fed a standard chow diet, we added proteomes from the protein sources likely to be in the diet based on the ingredient list (corn UP000007305, fish UP000000437, soy UP000008827, wheat UP000019116, Downloaded 19Feb20). We clustered the mouse and diet reference proteomes individually at a 95% identity threshold. We only searched samples against their respective dietary database.

In order to identify all sequences from the species *Bacteroides thetaiotaomicron (B. theta)* and *Lactococcus lactis* we downloaded all the sequences matching these species from UniProt [26]. We then used diamond BLASTp to identify all sequences in the protein database that matched with 95% identity or greater. The species code for these sequences was changed to BT or LAC if they were found to be *B. theta* or *L. lactis* respectively.

*Metaproteomic sample processing*

We pelleted fecal samples by centrifugation (21,000 x g, 5 min) and removed the preservation solution. We suspended dietary and fecal pellets in SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT] in Lysing Matrix E tubes (MP Biomedicals) and bead beat the samples (5 cycles of 45 s at 6.45 m/s, 1 min between cycles). After bead beating we heated the lysates to 95°C for 10 minutes. We mixed 60 µL of the resulting lysates with 400 µL of UA solution (8 M urea in 0.1 M Tris/HCl pH 8.5), loaded the sample onto a 10 kDa 500 µL filter unit (VWR International) and centrifuged at 14,000 x g for 30 minutes. We repeated this step up to three times to reach filter capacity. After loading, we added another 200 µL of UA buffer and centrifuged at 14,000 x g for another 40 minutes. We added 100 µL of IAA solution (0.05 M iodoacetamide in UA solution) to the filter and incubated at 22°C for 20 minutes. We removed IAA by centrifuging the filter at 14,000 x g for 20 minutes. We then washed the filter 3 times by adding 100 uL of UA buffer and centrifuging at 14,000 x g for 20 minutes. We then washed 3 more times by adding 100 uL of ABC buffer (50 mM Ammonium Bicarbonate) and centrifuging at 14,000 x g for 20 minutes. To digest the isolated protein, we added 0.95 µg of MS grade trypsin (Thermo Scientific Pierce, Rockford, IL, USA) mixed in 40 µL of ABC to each filter and incubated at 37°C for 16 hours. We then eluted the peptides by centrifugation at

14,000 x g for 20 minutes. We eluted again with 50 uL of 0.5 M NaCL and centrifuged at 14,000 x g for another 20 minutes. We quantified the abundance of the peptides using the Pierce Micro BCA assay (Thermo Scientific Pierce, Rockford, IL, USA) following the manufacturer's instructions.

We analyzed the samples by 1D-LC-MS/MS. Samples were run in a randomized block design. For each run, we loaded 600 ng of peptides onto a 5 mm, 300 µm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific) using an UltiMate™ 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) and desalted on the pre-column. After desalting, the pre-column was switched in line with a 75 cm x 75 µm analytical EASY-Spray column packed with PepMap RSLC C18, 2 µm material (Thermo Fisher Scientific), which was heated to 60 °C. The analytical column was connected via an Easy-Spray source to a Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer. Peptides were separated using a 140 minute reverse phase gradient [52]. We acquired spectra using the following parameters: m/z 445.12003 lock mass, normalized collision energy equal to 24, 25 s dynamic exclusion, and exclusion of ions of +1 charge state. Full MS scans were acquired for 380 to 1600 m/z at a resolution of 60,000 and a max IT time of 200 ms. Data-dependent $MS^2$ spectra for the 15 most abundant ions were acquired at a resolution of 15,000 and max IT time of 100 ms.

*Metaproteomic data processing*

We searched raw MS spectra against the diet specific protein databases using the run calibration, SEQUEST HT and percolator nodes in Proteome Discoverer 2.3 (Thermo Fisher Scientific). We used the following setting for the search: trypsin (full), 2 missed cleavages, 10 ppm precursor mass tolerance, 0.1 Da fragment mass tolerance. We included the following dynamic modifications: oxidation on M (+15.995 Da), deamidation on N,Q,R (0.984 Da) and acetyl on the protein N terminus (+42.011 Da). We also included the static modification carbamidomethyl on C (+57.021 Da). We filtered identified peptides and proteins at a false discovery rate (FDR) of 5%. Additionally, we only included proteins that had at least one protein unique peptide identified. Proteins were quantified by peptide spectral match (PSM) count (spectral counting).

*Statistical analysis and visualization*

Whenever possible in this study we tested significance of changes in abundance by applying an ANOVA on a linear effects-mixed model with the interacting fixed effects being mouse group and diet, and the random effect being the individual mouse (lme4 version 4.3.1)[64]. For multiple comparisons we calculated 95% confidence intervals for each diet using the emmeans R package (version 1.8.8)[65]. The exceptions were PERMANOVA analysis for testing significance of microbiota compositional changes (Extended Data Table 5) and Welch's t-tests to compare differences between yeast and egg white protein diets (Extended Data Table 17). For each analysis, we controlled for multiple-hypothesis testing by converting p-values to FDR-level based q-values, unless all p-values in the analysis were below 0.05 [66,67]. By definition, if all the p-values are less than 0.05 then the FDR is less than 0.05. Visualizations were produced using ggplot2 (version 3.4.3) [68], pheatmap (version 1.0.12) [69], RawGraphs [70], Microsoft Excel and Adobe Illustrator. All boxes and error bars represent 95% confidence intervals. Boxes or error bars that do not overlap denote significance. If no error bars are present then significance

is denoted by letters. In the case of beta-diversity analysis the error bars are 95% confidence intervals (Fig. 1d), but significance was tested separately by PERMANOVA.

*Compositional profiling of the microbiota*

We calculated the abundances of specific taxa in the microbiota using proteinaceous biomass[8]. Briefly, we filtered for proteins with at least 2 protein unique peptides and summed their spectra into their assigned taxonomy: microbial species, mouse, diet, ambiguous, low quality bins, unbinned bacteria (See *Metagenomic assembly and protein database construction* section for details on assignment). We calculated the microbe to host ratio by summing the spectral count assigned to microbial species, multiple microbial species, low quality bins and unbinned bacteria proteins and dividing them by the number of spectral counts assigned to mouse proteins. We considered a microbial species quantifiable if we could identify at least one protein with 2 protein unique peptides unambiguously assigned to the species. We calculated per sample species richness by simply counting the number of quantifiable species per sample. We calculated alpha (Shannon Diversity Index) and beta diversity (bray-curtis) metrics using the vegan (version 2.6-4) package in R (version 4.3.1)[71,72] on a table of the quantifiable microbial species (Statistics as described above). We also evaluated the composition of the microbiota using principal component analysis and hierarchical clustering. For principal component analysis we normalized the quantified species using centered-log ratio transformation and calculated principal components using the prcomp function in base R on all the mice and separately on each mouse group. Principal components were rendered using the ggplot2 (version 3.4.3) package in R[68]. For hierarchical clustering, we focused on the species that were most abundant, representing at least 5% of the microbial species biomass in at least one sample. We calculated the percent biomass for all the species and then extracted the species that fit the abundant species criteria. We calculated the individual significance of each abundant species using linear mixed effects models as described above. We hierarchically clustered log transformed values of these species using the R package pheatmap (version 1.0.12), using the ward.D2 algorithm and euclidean distances[69]. We compared broad taxonomic changes at the class level. For all quantifiable species we summed the abundance of the assigned class by GTDB-Tk-tk (see *Metagenomic assembly and protein database construction* for details). We then calculated confidence intervals using the linear mixed effects models and emmeans as described previously. Barcharts were rendered in ggplot2 using the estimated mean and 95% confidence intervals as error bars.

*Functional profiling of the microbiota*

For analyses of functional categories at the level of the whole microbiota we calculated the normalized spectral abundance factor (NSAF%) for each protein, which provides the relative abundance for each protein as a percentage of the summed abundance of microbiota proteins[73]. We annotated functions for all microbial proteins in our database using EggNOG-mapper[74], MANTIS[75], and Microbe Annotator[76]. We assigned glycoside hydrolase clan identifiers from the CAZy database using dbCAN2[77,78]. We manually curated these annotations by searching a subset of these proteins against the Swiss-Prot[26] and InterPro[79] databases between February 2023 and June 2023. We prioritized the proteins by abundance. If the Swiss-Prot or InterPro annotations matched the automated tool annotations we extrapolated the

assigned protein name to all proteins with the same automated annotation. Alternatively, if the annotations from the automated tools were in agreement, we consolidated the annotation into a consensus annotation. We then assigned broad functional categories, detailed functional categories, and specific names to each validated protein set. To evaluate functional changes due to diet, we summed all microbiota proteins assigned to a broad or detailed functional category, or enzyme name and applied a linear mixed-effects model to each function as described above.

*In vivo proteomic analysis of Bacteroides theta*

To analyze the *B. theta* proteome, we calculated the orgNSAF by extracting all proteins assigned to the species *B. theta* detected in the metaproteomes, and then calculating NSAF%[80]. We then compared abundances of *B. theta* proteins detected in the yeast and egg protein diets using the Welch's t-test in the Perseus software (version 1.6.14.0)[81]. To visualize polysaccharide utilization loci (PULs) we mapped the reads from one of our metagenomic samples to all the contigs that were assigned *B. theta* proteins using BBSplit (BBMap, Version 38.06). We then assembled all the mapped reads using metaSPAdes. The genes in this newly assembled genome overlapped exactly with the previous set of identified *B. theta* genes, and this *B. theta* genome was uploaded to the RAST server for further analysis[82]. PULs were detected in the metaproteome by identifying proteins labeled SusC, SusD, or TonB. The rest of the PUL was identified by visualizing the gene neighborhood in RAST. The identified genes were then cross-referenced against PULDB to assign literature described PUL numbers[83].

*In vitro growth and proteomics of B. theta*

We cultured *B. theta* in VPI-5482 in two biological replicates and at least 4 technical replicates using a defined *Bacteroides* medium similar to that described in[84]. *B. theta* cultures were grown statically at 37°C in a Coy anaerobic chamber (2.5 % $H_2$ /10 % $CO_2$ /88.5 % $N_2$) in minimal medium (100 mM $KH_2PO_4$, 8.5 mM $[NH_2]_4SO_4$, 15 mM NaCl, 5.8 µM vitamin $K_3$, 1.44 µM $FeSO_4 \cdot 7H_2O$, 1 mM $MgCl_2$, 1.9 µM hematin, 0.2 mM L-histidine, 3.69 nM vitamin $B_{12}$, 208 µM L-cysteine, and 7.2 µM $CaCl_2 \cdot 2H_2O$). The four dietary protein sources: soy protein (CA.160480), yeast protein (CA.40115), casein protein (CA.160030), and egg white protein (CA.160230), were purchased from Envigo and were the same as the protein sources used in the corresponding diets. Porcine muc2 mucin (Sigma CA. M2378) was also tested alongside controls of glucose and no carbon source. To aid in suspension in aqueous media, we pre-prepared the proteins in 200 mM NaOH water at 37°C for four days; the glucose control was also dissolved in 200 mM NaOH water. We then added the protein or glucose solution to the pre-prepared media at 0.5% (wt/v). Cultures were grown overnight in minimal media supplemented with 0.5% (wt/v) glucose before being washed and inoculated into experimental conditions at 0.01 OD and incubated at 37°C anaerobically with shaking every hour. Colony forming units (CFUs) per mL of culture were enumerated by drop plating at 0 and 24 hr post inoculation. Solid media for *B. theta* was Brain-Heart Infusion agar (Difco CA. 241830)supplemented with 10% Horse Blood (LAMPIRE CA. 7233401) (BHI-HB).

To obtain a sample for proteomics, we repeated the experiment for the glucose, yeast, egg white, mucin and soy media. After 8 hours, CFUs were enumerated to confirm growth. We

pelleted cells by centrifuging at 4,000 g for 10 minutes. We then extracted the supernatant and froze the pellets at -80°C. Protein was extracted by the same FASP protocol described above but with two differences. We lysed pellets by adding 120 uL of SDT buffer and then heating at 95°C. We used PES 10kDa filters (MilliporeSigma). We also used a similar Mass Spectrometry procedure, except the samples were run on an Exploris 480 mass spectrometer (Thermo Fisher Scientific) and 1 ug of peptide were analyzed for each sample. We searched raw MS spectra using the same Proteome Discoverer 2.3 workflow using the *B. theta* proteome downloaded from UniProt (UP000001414 downloaded January 9, 2024) as the protein sequence database. We then cross referenced PULs between the metaproteome and the *in vitro* proteome to compare them.

**Acknowledgments:**
We thank Tjorven Hinzke for advice on statistics, Heather Maughan for editing of the manuscript, Erin Baker for advice on methods, and Lawrence David and Balfour Sartor for advice and consultation on the project.
All LC-MS/MS measurements were made in the Molecular Education, Technology, and Research Innovation Center (METRIC) at North Carolina State University.
This work was supported by the National Institutes of Health through awards R35GM138362 (MK), T32DK007737 (JABR) and P30 DK034987, and by the USDA National Institute of Food and Agriculture, Hatch project 7002782. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Data availability**
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[85] partner repository with the dataset identifier PXD041586 Reviewer Username: reviewer_pxd041586@ebi.ac.uk Password: V9Jz2n4h (metaproteomic data) and PXD050296 Reviewer Username: reviewer_pxd050296@ebi.ac.uk Password:F8I9Cmcz (*B. theta in vitro* proteomics data). Metagenomic raw reads were submitted to NCBI SRA under the bioproject identifier PRJNA1026909. All metagenome assembled genomes (MAGs) with accompanying metadata were submitted to DRYAD DOI: 10.5061/dryad.x0k6djhq5. Reviewer link: https://datadryad.org/stash/share/QagcDe_b_b0GbbyQ7mPOxBapFL3QbaXt3-fhiZRvDCM

**Author Contributions**
JABR: Experimental design, data collection, data processing, data analysis, author of original manuscript, editing
AB: Conceptualization of the study, experimental design, data collection, data processing, editing
ASM: Experimental design, data collection, editing
AA: Data processing, analysis, editing
MVW: Data processing, analysis, editing
AKM: Data processing
JM: Coding, graphic design

SV: Data collection
TR: Data collection
CMT: Experimental design, data collection, editing
MK: Funding, conceptualization of the study, experimental design, data processing, data analysis, writing, editing

**References**


1. Budhathoki, S. *et al.* Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality in a Japanese Cohort. *JAMA Intern. Med.* **179**, 1509–1518 (2019).
2. Huang, J. *et al.* Association Between Plant and Animal Protein Intake and Overall and Cause-Specific Mortality. *JAMA Intern. Med.* **180**, 1173–1184 (2020).
3. Song, M. *et al.* Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality. *JAMA Intern. Med.* **176**, 1453–1463 (2016).
4. Bartlett, A. & Kleiner, M. Dietary protein and the intestinal microbiota: An understudied relationship. *iScience* **25**, 105313 (2022).
5. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
6. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353.e21 (2016).
7. Perler, B. K., Friedman, E. S. & Wu, G. D. The Role of the Gut Microbiota in the Relationship Between Diet and Human Health. *Annu. Rev. Physiol.* **85**, 449–468 (2023).
8. Kleiner, M. *et al.* Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* **8**, 1558 (2017).
9. Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci.* **109**, E1173 (2012).
10. Ahn, E., Jeong, H. & Kim, E. Differential effects of various dietary proteins on dextran sulfate sodium-induced colitis in mice. *Nutr Res Pr.* **16**, 700–715 (2022).

11. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
12. Blakeley-Ruiz, J. A. *et al.* Morphine and high-fat diet differentially alter the gut microbiota composition and metabolic function in lean versus obese mice. *ISME Commun.* **2**, 66 (2022).
13. Patnode, M. L. *et al.* Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. *Cell* **179**, 59-73.e13 (2019).
14. Yao, C. K., Muir, J. G. & Gibson, P. R. Review article: insights into colonic protein fermentation, its modulation and potential health implications. *Aliment. Pharmacol. Ther.* **43**, 181–196 (2016).
15. Sun, M. *et al.* Microbiota-derived short-chain fatty acids promote Th1 cell IL-10 production to maintain intestinal homeostasis. *Nat. Commun.* **9**, 3555 (2018).
16. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. *Science* **333**, 101–104 (2011).
17. Zhu, Y. *et al.* Beef, Chicken, and Soy Proteins in Diets Induce Different Gut Microbiota and Metabolites in Rats. *Front. Microbiol.* **8**, 1395 (2017).
18. Xiong, W., Brown, C. T., Morowitz, M. J., Banfield, J. F. & Hettich, R. L. Genome-resolved metaproteomic characterization of preterm infant gut microbiota development reveals species-specific metabolic shifts and variabilities during early life. *Microbiome* **5**, 72 (2017).
19. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
20. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
21. Blakeley-Ruiz, J. A. & Kleiner, M. Considerations for constructing a protein sequence database for metaproteomics. *Comput. Struct. Biotechnol. J.* **20**, 937–952 (2022).

22. Salvato, F., Hettich, R. L. & Kleiner, M. Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLOS Pathog.* **17**, e1009245 (2021).
23. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology* vol. 54 1469–1476 (2004).
24. Lagkouvardos, I. *et al.* Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* **7**, 28 (2019).
25. Sonnenburg, J. L. *et al.* Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science* **307**, 1955–1959 (2005).
26. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
27. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30**, 1236–1240 (2014).
28. O'Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 691–706 (2016).
29. Cagnon, L. & Braissant, O. Hyperammonemia-induced toxicity for the developing central nervous system. *Brain Res. Rev.* **56**, 183–197 (2007).
30. Rangroo Thrane, V. *et al.* Ammonia triggers neuronal disinhibition and seizures by impairing astrocyte potassium buffering. *Nat. Med.* **19**, 1643–1648 (2013).
31. Stummer, N., Feichtinger, R. G., Weghuber, D., Kofler, B. & Schneider, A. M. Role of Hydrogen Sulfide in Inflammatory Bowel Disease. *Antioxidants* **12**, (2023).
32. Luscher, B., Shen, Q. & Sahir, N. The GABAergic deficit hypothesis of major depressive disorder. *Mol. Psychiatry* **16**, 383–406 (2011).
33. Jaglin, M. *et al.* Indole, a Signaling Molecule Produced by the Gut Microbiota, Negatively

Impacts Emotional Behaviors in Rats. *Front. Neurosci.* **12**, (2018).

34. Czumaj, A., Śledziński, T. & Mika, A. Branched-Chain Fatty Acids Alter the Expression of Genes Responsible for Lipid Synthesis and Inflammation in Human Adipose Cells. *Nutrients* **14**, (2022).
35. Pereira, G. V. *et al.* Unravelling specific diet and gut microbial contributions to inflammatory bowel disease. *Res Sq Prepr.* rs.3.rs-2518251 (2023) doi:10.21203/rs.3.rs-2518251/v1.
36. Grondin, J. M., Tamura, K., Déjean, G., Wade, D. W. & Brumer, H. Polysaccharide Utilization Loci: Fueling Microbial Communities. *J. Bacteriol.* **199**, 10.1128/jb.00860-16 (2017).
37. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).
38. Kollár, R. *et al.* Architecture of the Yeast Cell Wall: β(1→6)-GLUCAN INTERCONNECTS MANNOPROTEIN, β(1→3)-GLUCAN, AND CHITIN*. *J. Biol. Chem.* **272**, 17762–17775 (1997).
39. Orlean, P. Architecture and Biosynthesis of the Saccharomyces cerevisiae Cell Wall. *Genetics* **192**, 775–818 (2012).
40. Temple, M. J. *et al.* A Bacteroidetes locus dedicated to fungal 1,6-β-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-β-glucanase. *J. Biol. Chem.* **292**, 10639–10650 (2017).
41. Hwang, H. S. *et al.* Type and branched pattern of N-glycans and their structural effect on the chicken egg allergen ovotransferrin: a comparison with ovomucoid. *Glycoconj. J.* **31**, 41–50 (2014).
42. Thaysen-Andersen, M., Mysling, S. & Højrup, P. Site-Specific Glycoprofiling of N-Linked Glycopeptides Using MALDI-TOF MS: Strong Correlation between Signal Strength and Glycoform Quantities. *Anal. Chem.* **81**, 3933–3943 (2009).
43. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and

transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).

44. Staniszewski, A. & Kordowska-Wiater, M. Probiotic and Potentially Probiotic Yeasts—Characteristics and Food Application. *Foods* **10**, (2021).
45. Adeva, M. M., Souto, G., Blanco, N. & Donapetry, C. Ammonium metabolism in humans. *Metabolism* **61**, 1495–1511 (2012).
46. Badri, D. V., Jackson, M. I. & Jewell, D. E. Dietary Protein and Carbohydrate Levels Affect the Gut Microbiota and Clinical Assessment in Healthy Adult Cats. *J. Nutr.* **151**, 3637–3650 (2021).
47. Cummings, J., Hill, M., Bone, E., Branch, W. & Jenkins, D. J. A. The effect of meat protein and dietary fiber on colonic function and metabolism II. Bacterial metabolites in feces and urine1. *Am. J. Clin. Nutr.* **32**, 2094–2101 (1979).
48. Pinna, C. *et al.* Influence of dietary protein and fructooligosaccharides on fecal fermentative end-products, fecal bacterial populations and apparent total tract digestibility in dogs. *BMC Vet. Res.* **14**, 106 (2018).
49. Rose, C. F. Ammonia-Lowering Strategies for the Treatment of Hepatic Encephalopathy. *Clin. Pharmacol. Ther.* **92**, 321–331 (2012).
50. Shen, T.-C. D. *et al.* Engineering the gut microbiota to treat hyperammonemia. *J. Clin. Invest.* **125**, 2841–2850 (2015).
51. Cieslak, K. P., Baur, O., Verheij, J., Bennink, R. J. & van Gulik, T. M. Liver function declines with increased age. *HPB* **18**, 691–696 (2016).
52. Mordant, A. & Kleiner, M. Evaluation of Sample Preservation and Storage Methods for Metaproteomics Analysis of Intestinal Microbiomes. *Microbiol. Spectr.* **9**, e0187721 (2021).
53. Knudsen, B. E. *et al.* Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. *mSystems* **1**, 10.1128/msystems.00095-16 (2016).
54. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. (2014).

55. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
56. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
57. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359–e7359 (2019).
58. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
59. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
60. Olm, M. R. *et al.* Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems* **5**, (2020).
61. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
62. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**, 2068–2069 (2014).
63. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
64. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
65. Lenth, R. V. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. (2023).
66. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300

(1995).

67. Storey, J. D., Taylor, J. E. & Siegmund, D. Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 187–205 (2004).
68. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
69. Kolde, R. *Pheatmap: Pretty Heatmaps*. (2019).
70. Mauri, M., Elli, T., Caviglia, G., Uboldi, G. & Azzi, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. in *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (Association for Computing Machinery, New York, NY, USA, 2017). doi:10.1145/3125571.3125585.
71. Oksanen, J. *et al. Vegan: Community Ecology Package*. (2022).
72. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2023).
73. Florens, L. *et al.* Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods San Diego Calif* **40**, 303–311 (2006).
74. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
75. Queirós, P., Delogu, F., Hickl, O., May, P. & Wilmes, P. Mantis: flexible and consensus-driven genome annotation. *GigaScience* **10**, giab042 (2021).
76. Ruiz-Perez, C. A., Conrad, R. E. & Konstantinidis, K. T. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics* **22**, 11 (2021).
77. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).

78. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
79. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
80. Mueller, R. S. *et al.* Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol. Syst. Biol.* **6**, 374 (2010).
81. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
82. Aziz, R. K. *et al.* The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**, 75 (2008).
83. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* **46**, D677–D683 (2018).
84. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
85. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).