

1      **Dietary protein source strongly alters gut microbiota composition and**  
2      **function**

3  
4  
5  
6      Authors: J. Alfredo Blakeley-Ruiz† \*<sup>1</sup>, Alexandria Bartlett\*<sup>1,2</sup>, Arthur S. McMillan<sup>3</sup>, Ayesha  
7      Awan<sup>1,3</sup>, Molly Vanhoy Walsh<sup>1</sup>, Alissa K. Meyerhoffer<sup>1</sup>, Simina Vintila<sup>1</sup>, Jessie L. Maier<sup>1</sup>, Tanner  
8      Richie<sup>1</sup>, Casey M. Theriot<sup>3</sup>, Manuel Kleiner† <sup>1</sup>

- 9  
10     1. Department of Plant and Microbial Biology, College of Agricultural Sciences, North  
11     Carolina State University, Raleigh, NC, USA  
12     2. Department of Molecular Genetics and Microbiology, Duke University, Durham, NC,  
13     USA  
14     3. Department of Population Health and Pathobiology, College of Veterinary Medicine,  
15     North Carolina State University, NC, USA

16  
17  
18     \* These authors contributed equally to this manuscript

19  
20     † Please refer correspondence to J. Alfredo Blakeley-Ruiz (jablakel@ncsu.edu) or Manuel  
21     Kleiner ([manuel\\_kleiner@ncsu.edu](mailto:manuel_kleiner@ncsu.edu))

22     Orcid:

23     JABR: <https://orcid.org/0000-0001-7638-5849>

24     AB: <https://orcid.org/0000-0003-4350-0542>

25     ASM: <https://orcid.org/0000-0002-0115-0225>

26     AA: <https://orcid.org/0009-0002-3421-4798>

27     MVW: <https://orcid.org/0000-0003-2723-7297>

28     AKM: <https://orcid.org/0009-0001-7635-1275>

29     SV: <https://orcid.org/0000-0003-0018-0016>

30     JLM: <https://orcid.org/0009-0001-8575-5386>

31     TR: <https://orcid.org/0000-0002-2554-6005>

32     CMT: <https://orcid.org/0000-0002-1895-8941>

33     MK: <https://orcid.org/0000-0001-6904-0287>

38 **Abstract**

39 **BACKGROUND**

40 The source of protein in a person's diet affects their total life expectancy. However, the  
41 mechanisms by which dietary protein sources differentially impact human health and life  
42 expectancy are poorly understood. Dietary choices have major impacts on the composition and  
43 function of the intestinal microbiota that ultimately modulate host health. This raises the  
44 possibility that health outcomes based on dietary protein sources might be driven by interactions  
45 between dietary protein and the gut microbiota. In this study, we determined the effects of seven  
46 different sources of dietary protein on the gut microbiota of mice. We applied an integrated  
47 metagenomics-metaproteomics approach to simultaneously investigate the effects of these  
48 dietary protein sources on the gut microbiota's composition and function.

49 **RESULTS**

50 Different dietary protein sources significantly altered the species composition of the gut  
51 microbiota. Yeast and egg-white protein had the greatest effect on the composition of the gut  
52 microbiota driven by an increase in the abundance of *Bacteroides thetaiotaomicron*. The  
53 abundance of enzymes associated with different broad functional categories also significantly  
54 changed due to dietary protein sources. In particular, the abundance of amino acid degrading  
55 enzymes increased in the presence of brown rice and egg white protein, while glycoside  
56 hydrolases increased in the presence of yeast and egg white protein. The glycoside hydrolases  
57 increased in the yeast and egg white protein diets were mostly *B. thetaiotaomicron* enzymes  
58 previously associated with the degradation of yeast cell-wall glycoproteins in the case of the  
59 yeast diet, and the degradation of mucins in the case of the egg white diet. We validated that *B.*  
60 *thetaiotaomicron* expresses these glycoside hydrolases when grown on mucin, yeast, and egg  
61 white protein *in vitro*.

62 **CONCLUSION**

63 These results show that the source of dietary protein can alter the composition and  
64 function of the gut microbiota through the specific glycosylations present on dietary  
65 glycoproteins. Both amino acid degradation and mucin metabolism by the microbiota have been  
66 previously linked to playing a role in modulating gut health. Our study is important because it  
67 shows that dietary protein sources should be considered, in addition to fiber and fat, when  
68 designing diets for a healthy gut microbiome.

69

70 **Keywords: Gut Microbiome, Metaproteomics, Metagenomics, Dietary Intervention, Mus**  
71 **musculus, Mice**

72

### 73 **BACKGROUND**

74 Source of dietary protein has a major impact on human health. People who consume  
75 high amounts of animal protein have higher mortality rates than those who consume mostly  
76 plant-based protein [1,2]. Egg protein and red meat protein in particular have been shown to  
77 lead to increased mortality rates among humans [3] and a diet high in red meat protein has  
78 been shown to increase inflammation in a model of colitis [4]. Replacing animal protein sources  
79 with plant protein sources reduces mortality rates [3]. Currently, we have a limited  
80 understanding of the underlying causes, but the gut microbiota has been implicated as  
81 potentially having a major role in the differential health impacts of different dietary protein  
82 sources [5,6]. Diet has been shown to change the gut microbiota's composition and function in  
83 ways that can be detrimental or beneficial to health [7–10]. For example, protein fermentation by  
84 the gut microbiota generates a number of toxins including ammonia, putrescine, and hydrogen  
85 sulfide [6,11], while fermentation of fiber and certain amino acids produces anti-inflammatory  
86 short-chain fatty acids [12]. Previous studies demonstrate that the amount of protein can have a  
87 greater impact on the gut microbiota's composition than other macronutrients [13], and that  
88 source of dietary protein impacts the composition of the microbiota [14]. There is, however,

89 limited data showing the mechanisms by which individual sources of dietary protein affect the  
90 gut microbiota's composition and function, which could mediate the consumption and production  
91 of compounds beneficial or detrimental to the host.

92 Metaproteomics represents a powerful tool for characterizing the mechanisms  
93 underlying dietary effects on the gut microbiota [7,9]. Metaproteomics is defined as the large-  
94 scale characterization of the proteins present in a microbiome [15]. Protein abundances  
95 measured by metaproteomics simultaneously provide microbial species abundances [16], and  
96 evidence for the metabolic and physiological phenotype of microbiota members [17,18].

97 Metaproteomes are usually measured using a shotgun proteomics approach where proteins  
98 extracted from a sample are digested into peptides, separated by liquid chromatography, and  
99 measured on a mass spectrometer [20]. Proteins are then identified and quantified using a  
100 database search algorithm, which matches the measured peptides to a database of protein  
101 sequences [21]. Due to the heterogeneous nature of complex microbial communities it is usually  
102 best to construct the protein database using gene predictions from metagenomes measured  
103 from the same samples [19]. When metaproteomics is coupled to a genome-resolved  
104 metagenomic database it is possible to evaluate strain and species level function even if the  
105 microbes have not been previously characterized [20,21]. We call this approach integrated  
106 metagenomics-metaproteomics.

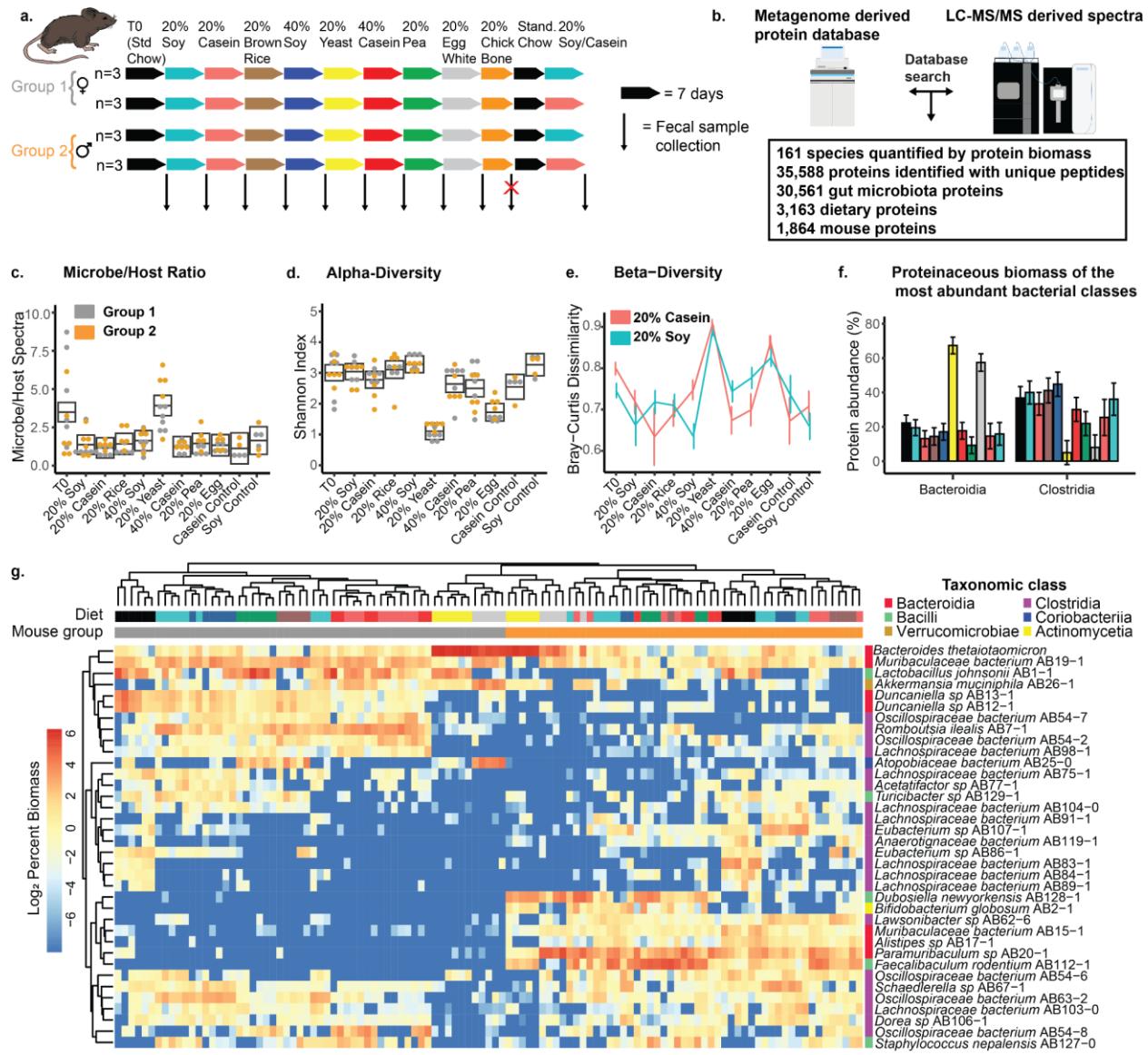
107 We used an integrated metagenomic-metaproteomic approach to investigate the effects  
108 of dietary protein source on gut microbiota's composition and function. We hypothesized that  
109 dietary protein source affects the abundance of amino acid metabolizing enzymes from the gut  
110 microbiota, altering the abundance of pathways involved in the production of toxins detrimental  
111 to host health. We found that the source of dietary protein not only alters the abundance of  
112 amino acid degrading enzymes, but has an even greater impact on the abundance of glycan  
113 degrading proteins among other functions, indicating that dietary protein sources can have wide  
114 ranging effects on the gut microbiota.

115 **RESULTS**

116 **Integrated Metagenomic-Metaproteomic Analysis of Dietary Protein Effects on the Gut**

117 **Microbiota**

118 To determine how different sources of dietary protein affect the gut microbiota, we fed  
119 mice (C57BL/6J), half female (group 1) and half male (group 2) a series of 9 fully defined diets  
120 (Fig. 1a, Supplementary table 1). Each diet contained purified protein from a different single  
121 source of dietary protein whose mass represented either twenty or forty percent of the entire  
122 diet. In order of feeding, the diets were 20% soy protein, 20% casein protein, 20% brown rice  
123 protein, 40% soy protein, 20% yeast protein, 40% casein protein, 20% pea protein, 20% egg  
124 white protein, and 20% chicken bone protein. Mice consumed each diet for one week before  
125 switching to the next diet. We collected fecal samples from mice after 7 days of consumption of  
126 each diet. The chicken bone diet caused the mice to lose weight, so we discontinued the diet  
127 after 3 days and the mice consumed a standard chow diet for the rest of that week. No fecal  
128 samples were collected for the chicken bone diet. To control for the succession effects of a  
129 serial dietary intervention, we fed the mice the 20% soy diet or the 20% casein diet as a control  
130 at the end of the diet series. We analyzed samples from all mice and diets using an integrated  
131 metagenomic-metaproteomic approach [7,21] (Fig. 1b). We sequenced fecal samples using  
132 shotgun sequencing and used a genome-resolved metagenomics pipeline [22,23], which  
133 resulted in 454 metagenome-assembled genomes (MAGs) organized into 180 species groups.  
134 We used high-resolution mass spectrometry based metaproteomics to identify and quantify  
135 proteins in each sample using a protein sequence database derived from the metagenome and  
136 augmented with mouse and diet protein sequences [19,24]. In total, we identified 35,588  
137 proteins, each distinguished as microbial, host, or dietary proteins (Extended Data Table 1). All  
138 taxonomic and functional data described in this study were quantified using the metaproteomic  
139 data [16].



140  
141  
142 **Figure 1: Source of dietary protein alters the gut microbiota's composition.** (a) Diagram showing the  
143 experimental design, with number of cages, and order of the diets fed. Colors depicting the diets are used throughout  
144 the manuscript. Each row of arrows represents one cage. We collected 10-12 samples for each experimental diet and  
145 5-6 samples for each control diet. (b) A diagram illustrating the integrated metagenomic-metaproteomics method  
146 used to analyze the samples along with raw metrics: quantifiable species and number of proteins. (c) Ratio of spectra  
147 assigned to microbes versus the host; boxes represent 95% confidence intervals calculated on a linear mixed effect  
148 model (Extended Data Table 2). (d) Shannon diversity index of the gut microbiota across all diets; boxes represent  
149 95% confidence intervals calculated on a linear mixed effect model (Extended Data Table 2). (e) Bray-Curtis  
150 dissimilarity between the initial 20% soy diet (teal) or 20% casein diet (red) and all other diets. Error bars reflect 95%  
151 confidence intervals for all line graphs as calculated by the Rmisc package in R. (f) Abundances of the two most  
152 abundant bacterial classes based on summed protein abundance. Error bars are 95% confidence intervals calculated  
153 using a linear-mixed effects model (Extended Data Table 2). (g) A hierarchically clustered (ward.D2 algorithm on  
154 euclidean distances) heatmap depicting the clustering by species group abundance of the 36 most abundant species  
155 in the study. Species were considered abundant if they had at least 5% of the microbial biomass in at least one  
156 sample.

157

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

159 To assess the effect of dietary protein source on microbiota composition, we quantified  
160 the proteinaceous biomass for each species using metaproteomics data and obtained  
161 measurements for 161 distinct species (Extended Data Table 3) [16]. We divided the number of  
162 spectra assigned to microbial proteins by the number of spectra assigned to host proteins to  
163 create a measure of microbial load (Extended Data Table 4). We found that the yeast protein  
164 diet significantly increased microbial load as compared to all other defined protein diets (Fig.  
165 1c). We calculated within sample diversity (alpha diversity) and between sample compositional  
166 dissimilarity (beta diversity) based on the abundances of the quantifiable species. We found that  
167 the yeast and egg white diets significantly reduced the alpha diversity (Shannon diversity index)  
168 and richness (number of species) of the gut microbiota relative to all other diets (Fig. 1d;  
169 Supplementary Fig. 1a; Extended Data Table 4). We evaluated the effects of dietary protein on  
170 gut microbiota composition by comparing the microbiota from all diets to the initial 20% soy and  
171 20% casein diets using the Bray-Curtis dissimilarity index. We found that the composition of the  
172 gut microbiota was most similar when the source of dietary protein was the same, regardless of  
173 the amount of protein in the diet, and that the yeast and egg white diets yielded the most  
174 dissimilar microbiota compositions (Fig. 1e). Testing with PERMANOVA ( $q < 0.05$ ) showed that  
175 the community composition was significantly different when the source of dietary protein was  
176 different (43 out of 49 comparisons), but not when it was the same (Extended Data Table 5).  
177 These results show that the source of dietary protein had a greater effect on the gut microbiota  
178 than the amount of protein in the diet across three dimensions of the gut microbiota: microbial  
179 load, within sample diversity of species, and compositional dissimilarity between samples. The  
180 large differences in microbial composition in the egg white and yeast protein diets were driven  
181 by a decrease in the abundances of species from the class Clostridia in favor of species from  
182 the class Bacteroidia (Fig. 1f). Since we observed fewer species in the class Bacteroidia overall,

183 it makes sense that a drop in Clostridia in favor of Bacteroidia would result in a lower alpha  
184 diversity in the yeast and egg white diets (Fig. 1g; Extended Data Table 3).

185 To identify which specific microbial taxa drive differences in microbiota composition  
186 between dietary protein sources, we focused on the most abundant microbial species (>5% of  
187 the microbial protein biomass in at least one sample) and hierarchically clustered them by  
188 abundance across the different dietary protein sources/groups (Fig. 1g). This revealed three  
189 major clusters separating most samples by mouse group with the exception of the yeast and  
190 egg white diets which together formed a separate cluster that internally showed separation by  
191 mouse group. The T0 samples fell into the major mouse group clusters which indicates that the  
192 two mouse groups had distinct gut microbiotas at the start of the experiment. Within the mouse  
193 group clusters the microbiota clustered by source of dietary protein, which was also observed in  
194 principal component analysis (Supplementary Fig. 1, Supplementary Results Section A). In the  
195 yeast diet, *Bacteroides thetaiotaomicron* (*B. theta*) dominates the microbiota regardless of  
196 mouse group. *B. theta* abundance also increased in response to the egg white diet. However,  
197 there were additional species specific to each mouse group that also increased in abundance in  
198 response to the egg white diet (Fig. 1g; Supplementary Fig. 2). In group 1, these species were  
199 *Akkermansia muciniphila* and *Atopobiaceae* bacterium AB25-9, while in group 2 these species  
200 were *Paramuribaculum* sp. and *Dubosiella newyorkensis*. Interestingly, both *A. muciniphila* [25]  
201 and *Paramuribaculum* sp. [26] have been reported to forage on intestinal mucin and *B. theta*  
202 has been shown to switch towards mucin foraging based on diet [27]. These results show that  
203 the source of dietary protein changes the gut microbiota's composition and suggests that an egg  
204 white diet could promote mucin-foraging bacteria.

205

## 206 **Source of dietary protein alters gut microbiota function**

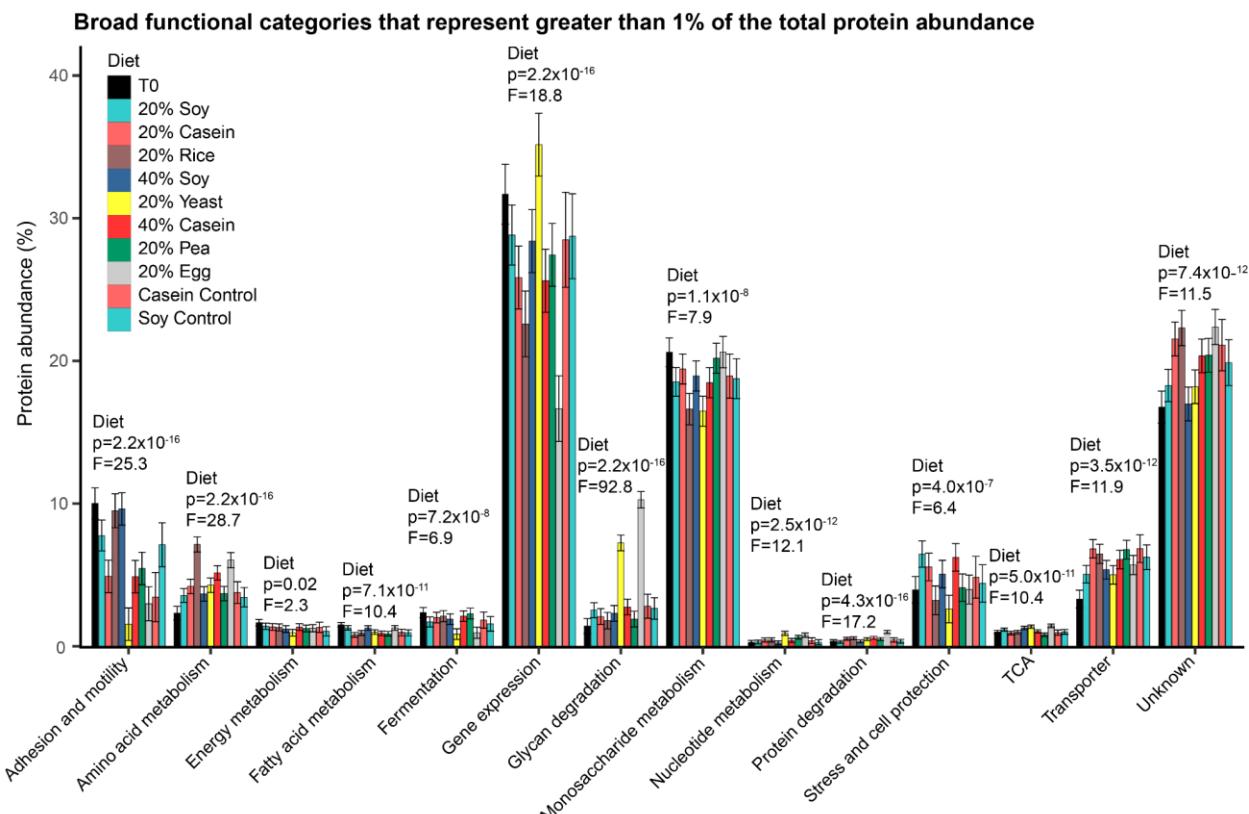
207 To evaluate gut microbiota function, we used the normalized abundances of gut  
208 microbiota proteins as a measure of the investment of the microbiota into metabolic and

209 physiological functions [17,18,28]. We first used automated annotation tools to assign functions  
210 to proteins. Since the annotations from these tools were not always accurate, we manually  
211 curated the annotations of 3,959 proteins and then extrapolated the functions to 14,547 similarly  
212 annotated proteins, which in total represented between 74 and 86 percent of the total microbial  
213 protein abundance in each sample (Extended Data Table 6). Based on the annotations, we  
214 assigned broad functional categories, such as amino acid metabolism, gene expression, glycan  
215 degradation, and monosaccharide metabolism and more detailed functional categories, such as  
216 ribosomal proteins and glycolysis to each of these proteins (Extended Data Table 6;  
217 Supplementary Fig. 3; Supplementary Results Section B). We then used the relative protein  
218 abundances to determine the investment of the microbiota into each of these functions. All of  
219 the broad functional categories, except for secondary metabolism, had significant changes in  
220 abundance due to dietary protein (ANOVA, p-value < 0.05; Extended Data Table 7; Fig. 2;  
221 Supplementary Fig. 4), which indicates that the source of dietary protein changes the gut  
222 microbiota's metabolism and physiology. Hierarchical clustering of all samples by abundances  
223 of broad functional categories revealed that the yeast and egg white diets clustered separately  
224 from all the other diets (Supplementary Fig. 5), similar to the results from the taxonomic  
225 clustering (Fig. 1g); however, a similar analysis at the detailed functional level revealed separate  
226 yeast, rice, and egg white clusters, with some outliers (Supplementary Fig. 6).

227 The two abundant broad functional categories (>1% of the total protein biomass) that  
228 had the greatest effect size due to diet were amino acid metabolism and glycan degradation,  
229 with F-statistics of 29 and 93 respectively (Fig. 2; Extended Data Table 7). Amino acid  
230 metabolism increased in the brown rice and egg white diets relative to all other diets except the  
231 40% casein diet and glycan degradation significantly increased in yeast and egg white diets  
232 relative to all other diets (Fig 2). Significant changes in the abundance of amino acid metabolism  
233 supported our initial expectation that the response of the microbiota to different dietary protein  
234 sources would likely relate to amino acid metabolism; however, we were surprised to find that

235 the abundance of glycan degrading enzymes responded more strongly to the source of dietary  
 236 protein than did enzymes for amino acid metabolism. This suggests that glycan degradation  
 237 instead of amino acid metabolism may be the major driver of taxonomic and functional changes  
 238 in the gut microbiota in response to dietary protein source. These two functions will be  
 239 discussed in detail in subsequent sections. In addition, we observed specific changes in the  
 240 abundance of enzymes associated with the gene expression, monosaccharide metabolism,  
 241 fermentation, and stress and cell protection functional categories (Supplementary Results  
 242 Section B; Supplementary Fig 3).

243



244  
 245 **Figure 2: Broad functional categories of microbial proteins change significantly in abundance due to the**  
 246 **source of dietary protein.** Abundance of broad functional categories that represent at least 1% of the microbial  
 247 protein abundance in at least one diet. The abundance is a modeled mean calculated from mixed effects models and  
 248 the error bars represent 95% confidence intervals calculated from these models. All the categories shown here had a  
 249 p-value for the diet factor below 0.05; p-value  $< 2.2 \times 10^{-16}$  is the lower limit of the method. For underlying data see  
 250 Extended Data Tables 7 and 8. For higher resolution of functional categories, e.g. fermentation, see Supplementary  
 251 Figs. 3-4.

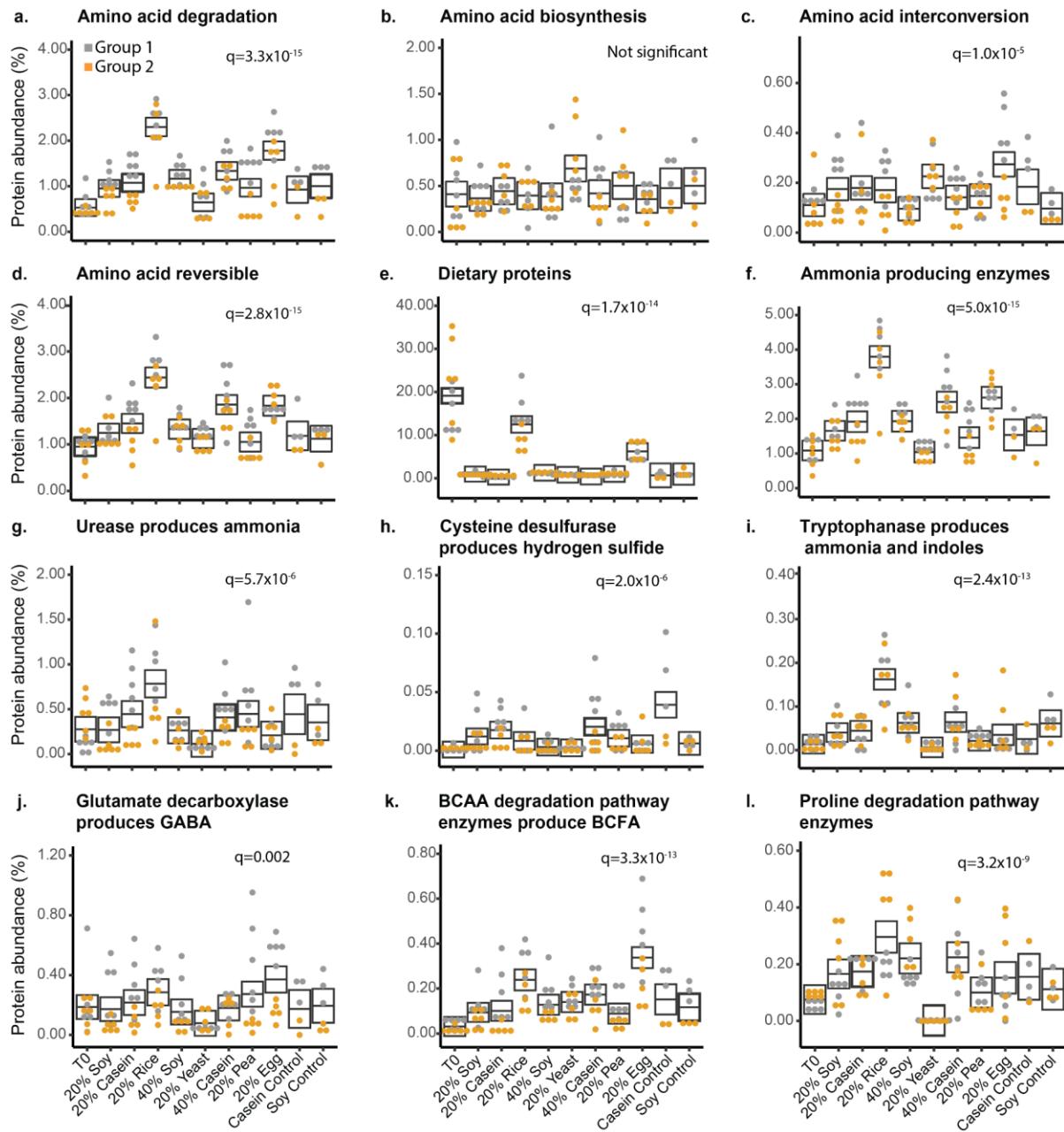
252

253

254 **Source of dietary protein alters the abundance of amino acid degrading enzymes**

255 To explore the effects of dietary protein source on microbiota amino acid metabolism, we  
256 manually classified 911 proteins (Extended Data Table 9) representing 68 enzyme functions  
257 (Extended Data Table 10) according to their involvement in the degradation (Fig. 3a), synthesis  
258 (Fig. 3b), interconversion (Fig. 3c), or reversible (Fig. 3d) reactions of specific amino acid  
259 pathways (Supplementary Figs. 7-18). In all diets except the yeast and standard chow diets, we  
260 observed that the microbiota was trending towards amino acid degradation instead of synthesis.  
261 We found that amino acid degrading enzymes were on average 2- to 6-fold more abundant than  
262 amino acid biosynthesis enzymes (Fig. 3a and 3b). Amino acid degrading enzymes were  
263 significantly more abundant in the rice and egg diets as compared to all the other diets (Fig. 3a),  
264 which is consistent with the observation that dietary proteins were significantly more abundant in  
265 the fecal samples of the brown rice and egg diets as compared to all other diets (Fig. 3e),  
266 suggesting that there may be a connection between the digestibility of dietary protein and amino  
267 acid degradation by the gut microbiota. Though amino acid synthesis enzymes were generally  
268 less abundant, we did observe a trend towards an increase in amino acid synthesis enzymes in  
269 the yeast protein diet relative to the other diets. This trend was not significant (Fig. 3b), but we  
270 observed several individual synthesis enzymes to be significantly increased in the yeast protein  
271 diet relative to other diets. These enzymes were involved in the synthesis of branched-chain  
272 amino acids (Suppl. Fig. 8), cysteine (Suppl. Fig. 9), lysine (Suppl. Fig. 15), proline (Suppl. Fig.  
273 17), or tyrosine (Suppl. Fig 18).

274



275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289

**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) synthesizing an amino acid, (c) converting between two amino acids, and (d) reversible. (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) cysteine desulfurase, (i) tryptophanase, (j) glutamate decarboxylase, (k) involved in branched-chain amino acid (BCAA) degradation to branched-chain fatty acid (BCFA) (includes branched-chain amino acid aminotransferase or ketoisovalerate oxidoreductase), and (l) involved in proline degradation (includes proline racemase or D-proline reductase).

290 Not all amino acid degrading enzymes increased in both the brown rice and egg white  
291 diets; sometimes they increased in one or the other (Supplementary Figs. 7-18). For example,  
292 enzymes associated with the degradation of threonine were more abundant in the egg white diet  
293 (Supplementary Fig. 12), while enzymes associated with tryptophan degradation were  
294 increased in the brown rice diet (Supplementary Fig. 18). Brown rice and egg were not the only  
295 diets in which the abundances of specific amino acid degrading enzymes increased. Alanine  
296 dehydrogenase increased in the 40% soy diet relative to the pea, yeast, 20% soy, and 20%  
297 casein diets (Supplementary Fig. 9) and cysteine desulfurase increased in the 40% casein and  
298 casein control diets relative to most other diets (Supplementary Fig. 9).

299 Changes in amino acid degradation by the gut microbiota have potential implications for  
300 host health by directly affecting local tissues or through interactions along the gut-brain axis  
301 depending on the metabolites produced by specific pathways [6,29]. We identified six categories  
302 of amino acid degradation pathways that are relevant to host health because they produce  
303 compounds that are toxic, anti-inflammatory, neurotransmitters, or otherwise related to disease.  
304 The toxic compounds included ammonia, produced by deaminating enzymes (Fig. 3f) and  
305 urease (Fig. 3g) [30,31] and hydrogen sulfide, produced by cysteine desulfurase (Fig. 3h) [32].  
306 The neurotransmitters included indoles produced by tryptophanase (Fig. 3i) [33] and  $\gamma$ -  
307 aminobutyric acid (GABA) produced by glutamate decarboxylase (Fig. 3j) [34]. The anti-  
308 inflammatory metabolites were branched-chain fatty acids produced by enzymes that degrade  
309 branched-chain amino acids (Fig. 3k) [35,36]. Finally, we included the enzymes in the proline  
310 degradation pathway (Fig. 3l), as an example of a specific amino acid degrading pathway  
311 affected by dietary protein source and relevant to the gut-brain axis [37] and enteric infections  
312 [38]. We found ammonia producing enzymes to be significantly more abundant in the brown rice  
313 diet as compared to all other diets, and also more abundant in the egg white and 40% casein  
314 diets as compared to the standard chow, 20% soy, yeast, pea, and control diets (Fig. 3f and 3g).  
315 We observed cysteine desulfurases to be significantly increased in the 40% casein and casein

316 control diets relative to other diets (Fig. 3h). Tryptophanase significantly increased in the brown  
317 rice diet relative to all other diets, while glutamate decarboxylase increased in the egg white diet  
318 relative to all other diets except brown rice, pea, and the control diets (Fig. 3i and 3j). We  
319 observed that branched-chain amino acid degrading enzymes were significantly increased in  
320 the egg white protein diet relative to all other diets (Fig. 3k), and proline degrading enzymes  
321 were increased in the brown rice diet relative to other diets, except the 40% soy and 40% casein  
322 diets where we also observed proline degradation to be significantly increased relative to the  
323 standard chow, yeast, and pea diets (Fig. 3l). These results show that the source of dietary  
324 protein can alter overall amino acid metabolism in the gut microbiome, as well as the  
325 abundance of different pathways. These changes have the potential to affect host physiology  
326 and health.

327

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

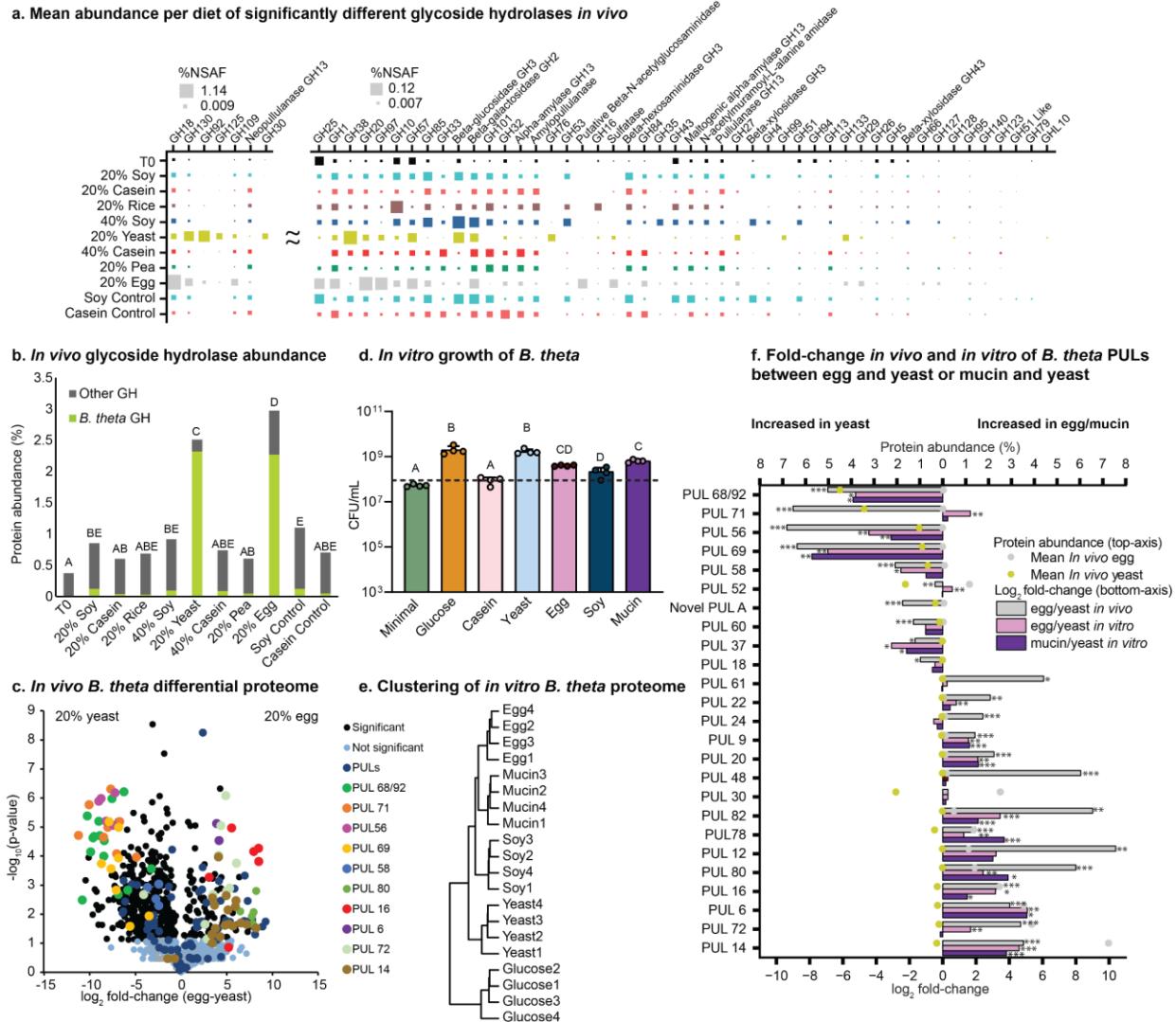
330 Surprisingly, glycan degrading enzymes (glycoside hydrolases) showed the largest  
331 overall changes in response to dietary protein source (Fig. 2, Extended Data Table 7).  
332 Specifically, these enzymes increased significantly in abundance in the yeast and egg white  
333 diets compared to the other diets. To further investigate the interaction of these glycan  
334 degrading enzymes with dietary protein we manually curated the functional assignments and  
335 potential substrate specificity of the 1,059 microbial glycoside hydrolases detected in our  
336 metaproteomes (Extended Data Table 13).

337 We grouped the validated glycoside hydrolases into 91 families based on the CAZy  
338 database (Extended Data Table 14) [39]. Of these families, 54 significantly changed in  
339 abundance between the different dietary protein sources (ANOVA,  $q < 0.05$ ) (Extended Data  
340 Table 15). Different glycoside hydrolase families increased in abundance in the soy, casein,  
341 brown rice, yeast, and egg white diets suggesting that distinct glycans drive their abundance

342 changes across the different diets (Fig. 4a, Extended Data Table 14-16, Supplementary Results  
343 Section C). The most abundant glycoside hydrolase families, GH18 in the case of egg white and  
344 GH92 in the case of yeast, have previously been associated with the degradation of glycans  
345 conjugated onto proteins (glycosylations) as part of polysaccharide utilization loci (PULs). PULs  
346 are operons that contain all the proteins necessary to import and degrade a specific glycan  
347 structure [40]. These GH18s are endo- $\beta$ -N-acetylglucosaminidases that break the bond  
348 between two acetylglucosamine residues attached to asparagine in N-linked glycoproteins. This  
349 reaction releases the glycan from the glycoprotein [41]. Meanwhile, GH92s, which are alpha-  
350 mannosidases, have been previously associated with the release of mannose residues from the  
351 glycosylations on yeast mannoproteins [41].

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

360



**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 utilizing locus (PUL) operon 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 between each comparison to detect significant changes in PUL protein abundances ( $^{***} = q < 0.01$ ,  $^{**} = q < 0.05$ ,  $^{*} = q < 0.1$ ) (Extended Data Tables 18 and 20).

382 To examine the specific role of *B. theta* in glycan degradation in the yeast and egg white  
383 diets, we compared the abundances of all *B. theta* proteins in the metaproteome between the  
384 two diets. Out of 1,420 detected *B. theta* proteins, the abundances of 592 proteins significantly  
385 differed between the two diets ( $q < 0.05$ , Welch's t-test) (Fig. 4c; Extended Data Table 17).  
386 Many of the significant proteins that were the most abundant and had the greatest fold-change  
387 between the two diets came from PULs (Fig. 4c; Supplementary Fig. 19). Between 10% and  
388 25% of the total protein abundance of *B. theta* in the yeast and egg white diets came from these  
389 PULs (Extended Data Table 18). The proteins belonging to each PUL tended to be expressed  
390 together either being significantly increased in egg white or the yeast diet (Figure 4c).

391 Several of the PULs that increased when we fed mice the yeast diet have previously  
392 been shown to specifically degrade the glycosylations on yeast cell wall proteins. PULs 68/92  
393 (BT3773-3792) and 69 (BT3854-3862) (Fig. 4c; Supplementary Fig. 19) degrade  $\alpha$ -mannans  
394 attached to yeast mannoproteins in *Saccharomyces cerevisiae* [41], while PUL 56 (BT3310-  
395 3314), degrades yeast  $\beta$ -glucans also attached to yeast cell wall mannoproteins [42].  
396 Conversely, the majority of the PULs increased when we fed mice the egg white diet had been  
397 previously linked to growth on mucin glycan conjugates: PUL14 (BT1032-1051), PUL6 (BT3017-  
398 0318), PUL16 (1280-1285), PUL 80 (BT4295-BT4299), and PUL12 (BT0865-0867)  
399 (Supplementary Fig. 19) [43]. An additional abundant PUL, PUL72 (BT3983-BT3994), has been  
400 previously implicated in the degradation of mannoproteins of mammalian origin [41] and our  
401 result suggests that PUL72 is also involved in the degradation of mannoproteins from non-  
402 mammalian vertebrates.

403 To test if *B. theta* could grow on yeast and egg white protein as predicted from the *in*  
404 *vivo* data, and if the expression of PULs was driven by direct responses to the dietary protein  
405 sources, we characterized *B. theta* growth and its proteome on dietary protein sources *in vitro*.  
406 We used a defined culture media and supplemented purified dietary protein sources as the sole  
407 carbon source to determine if this supported *B. theta* growth. We found that *B. theta* grew in the

408 presence of glucose (control), yeast protein, egg white protein, soy protein, and intestinal mucin  
409 (Fig. 4d, Tukey HSD adj P < 0.05). We analyzed the proteomes of *B. theta* in these five different  
410 conditions *in vitro* to determine if PULs played a role in growth (Extended Data Table 19). An  
411 overall comparison of the proteome between the media supplemented with 4 different protein  
412 sources revealed that egg white protein and mucin had the most similar proteome, and the  
413 proteome from the glucose control clustered separately from the protein-sources (Fig. 4e;  
414 Supplementary Fig. 20). We observed that 15 out of 24 PULs that were significantly different in  
415 abundance between the egg white and yeast diets *in vivo* were also significantly different in the  
416 same direction *in vitro* (Fig. 4f). In addition, 12 of these 15 PULs showed the same expression  
417 pattern in both the mucin and egg white protein media as compared to the yeast protein medium  
418 (Fig. 4f, Extended Data Tables 18 and 20). The relationship between mucin and egg white  
419 metabolism in microbiota species *in vivo* is further supported by the fact that five of the six  
420 species with greater than 5% abundance in an egg white sample (*B. theta*, *A. muciniphila*,  
421 *Atopobiaceae* bacterium AB25\_9 *Paramuribaculum* sp., *D. newyorkensis*) had abundant  
422 enzymes associated with the metabolism of sugars usually thought to be derived from mucin.  
423 These enzymes, of which several were among the top 100 most abundant proteins of these  
424 organisms, catalyze the metabolism of sialic acid (N-acetylneuraminate lyase, N-  
425 acylglucosamine 2-epimerase), N-acetylglucosamine (N-acetylglucosamine-6-phosphate  
426 deacetylase, glucosamine-6-phosphate deaminase, PTS system N-acetylglucosamine-specific,  
427 or fucose (fucosidase, fucose isomerase) (Extended Data Table 6). In summary, these results  
428 indicate that the glycosylations on yeast and egg white proteins drive the increase in abundance  
429 of *B. theta* in the yeast and egg white diets, and that egg white proteins and intestinal mucin  
430 share similar glycosylations leading to the expression of similar PULs for their degradation.

431

432 **DISCUSSION**

433 In this study, we sought to characterize how dietary protein source affects the gut  
434 microbiota's composition and function by measuring species-resolved proteins using integrated  
435 metagenomics-metaproteomics. We showed that source of dietary protein significantly alters the  
436 gut microbiota's composition, more so than amount of protein, and that yeast and egg white  
437 protein had the greatest effect on the composition driven by an increase in the relative  
438 abundance of *B. theta* and a decrease of bacteria from the class Clostridia. We also showed  
439 that the source of dietary protein altered the overall functional profile of the gut microbiota as  
440 reflected by changes in the abundance of microbial proteins assigned to broad functional  
441 categories. In particular, proteins involved in amino acid metabolism increased in abundance in  
442 the brown rice and egg white diets, while enzymes assigned to glycan degradation increased in  
443 the yeast and egg white diets.

444 The increase in amino acid metabolizing enzymes in the brown rice and egg white diets  
445 was driven by amino acid degrading enzymes. Previous studies across multiple species have  
446 shown that increasing the amount of protein fed to animals leads to an increase in the ammonia  
447 concentration in stool [44–46], which suggests that increased protein availability leads to  
448 increased amino acid deamination or urease activity in the gut. Here we show that, regardless  
449 of the amount of protein, the source of protein itself can lead to increases in amino acid  
450 deaminating enzymes and ureases from the intestinal microbiota. Gut microbiota urease activity  
451 and amino acid deamination have been linked to serious diseases like hepatic encephalopathy  
452 when liver function is disrupted [47]. Replacement of bacteria that produce these deaminating  
453 enzymes and ureases with bacteria that do not has been suggested as a potential treatment  
454 [48], our results suggest that adjustments in dietary protein source could be considered as well.

455 Since amino acids are the backbone of protein, we expected to observe changes in the  
456 abundance of amino acid degrading enzymes between the different sources of dietary protein;  
457 however, surprisingly the effect of dietary protein source on the abundance of glycan degrading  
458 proteins was even greater than the effect on amino acid degrading enzymes. Our results

459 suggest that the increase in glycan degrading proteins in the yeast and egg white diets is due to  
460 the glycosylations conjugated to these proteins. Yeast and egg white proteins have distinct  
461 glycan conjugate structures [49–52]. In the presence of yeast dietary protein we were able to  
462 show, *in vivo* and *in vitro*, increased expression of PULs associated with the degradation of  
463 yeast mannoprotein glycan conjugates. In the presence of egg white protein, we observed an  
464 increase in PULs previously linked to the degradation of the glycan conjugates of mucin. This  
465 combined with increases in mucin foraging bacteria *Akkermansia muciniphila* and  
466 *Paramuribaculum* sp. suggests that egg white protein promotes the abundance of mucin  
467 foraging bacteria and their proteins. The link between the foraging of mucin and egg white  
468 protein in retrospect makes sense, as egg white protein contains mucins called ovomucin and  
469 other proteins: ovalbumin, ovotransferrin, and ovomucoid, which have been previously shown to  
470 be N-glycosylated with acetylglucosamine and mannose containing glycans [49,52]. Previous  
471 studies in mice have shown that diets, which promote bacteria and their enzymes that degrade  
472 mucins, can make the host more susceptible to enteric inflammation and infection [36,53]. Since  
473 egg white protein also promotes these functions, these results suggest that diets high in egg  
474 protein may be detrimental to gastrointestinal health, which could explain the prior results from  
475 population level studies that eggs lead to increased mortality rates among humans [3].

476 Our study has at least two limitations preventing direct translation of microbiota  
477 responses to dietary protein sources into a human health context. First, we used purified dietary  
478 proteins, which differ from commonly consumed dietary proteins in that regular dietary protein  
479 sources also provide some amount of additional major dietary components such as fats,  
480 carbohydrates, and fiber. For example, plant proteins usually come with a relevant amount of  
481 fiber, while animal proteins are often low in fiber and have higher content fats [8]. Second, we  
482 used fully defined diets and while this allowed us to track effects to specific protein sources, we  
483 do anticipate that the dietary context of protein sources such as co-consumption of multiple  
484 protein, fiber, fat and carbohydrate sources will strongly influence the interactions of dietary

485 protein sources with the microbiota. The power of our study lies in our ability to confirm that the  
486 source of dietary protein does impact gut microbiota function and should be considered when  
487 thinking about how diet impacts the microbiota and its implications for host health. Future  
488 studies that determine how the effect of dietary protein source on the gut microbiota impacts  
489 gastrointestinal diseases are needed.

490

491

## 492 **MATERIALS AND METHODS**

### 493 *Animals and Housing*

494 In this study we included twelve C57BL/6J mice in two groups (six males, six females,  
495 Jackson Labs, Bar Harbor) aged 3-6 months. The males and females originated from different  
496 mouse rooms at the Jackson Labs and thus were expected to have different background  
497 microbiomes. Mice from both groups were housed in two separate cages (3 mice/cage) with a  
498 12 h light/dark cycle. We autoclaved bedding, performed all cage changes in a laminar flow  
499 hood and maintained an average temperature of 70°F and 35% humidity. We conducted our  
500 animal experiments in the Laboratory Animal Facilities at the NCSU CVM campus (Association  
501 for the Assessment and Accreditation of Laboratory Animal Care accredited), which are  
502 managed by the NCSU Laboratory Animal Resources. Animals assessed as moribund were  
503 humanely euthanized via CO<sub>2</sub> asphyxiation, as approved by NC State's Institutional Animal  
504 Care and Use Committee (Protocol # 18-034-B).

505

### 506 *Animal Diets and Sample Collection*

507 We fed mice defined diets with a single source of purified protein (Supplementary Table  
508 1). We fed each defined diet to all mice for 7 days, with the exception of the chicken bone broth  
509 diet. We observed clinical signs of disease including weight loss in the mice after 3 days of the

510 chicken bone broth diet and therefore replaced the diet with standard chow for the remainder of  
511 the 7 days. We fed the diets in this order: standard chow, 20% soy protein, 20% casein, 20%  
512 brown rice protein, 40% soy protein, 20% yeast protein, 40% casein, 20% pea protein, 20% egg  
513 white protein, 20% chicken bone broth protein, and lastly at the end of the experiment half of the  
514 mice in each group were fed the 20% soy protein and half the mice the 20% casein diet again  
515 as a control. Prior to the start of the defined diet, mice were fed autoclaved standard chow. All  
516 defined diets were sterilized by  $\gamma$ -irradiation and mice were provided sterile water (Gibco). On  
517 day 7 of each defined diet, we collected fecal samples, prior to replacing food with the next diet.  
518 We collected samples in NAP preservation solution at a 1:10 sample weight to solution ratio,  
519 and roughly homogenized the sample with a disposable pestle prior to freezing at -80°C [54].  
520 We had to sacrifice one mouse during the second diet (20% casein) so no additional samples  
521 were collected. We also were unable to collect a sample from one of the mice during the brown  
522 rice and egg white diets so only 10 samples were collected for those diets.

523

#### 524 *Metagenomic DNA sequencing*

525 To create a database for metaproteomic analysis, we pooled fecal samples from each  
526 cage to create four cage specific metagenomes. We gathered one fecal sample from each cage  
527 for four different diets (20% rice, 40% soy, 20% yeast, 40% casein) for a total of 16 samples. To  
528 extract DNA, we followed the QIAamp DNA stool mini kit (Qiagen)-based protocol described by  
529 Knudsen *et al.* with modifications [55]. To remove the preservation solution from the samples,  
530 we added 5 mL of 1X Phosphate Buffered Saline solution (VWR) to the samples and  
531 centrifuged them (17,000 x g, 5 min) to pellet solids and bacterial cells in suspension. We  
532 removed the preservation solution and resuspended the fecal pellets in 1 mL of InhibitEX Buffer  
533 in Matrix E (MP Biomedicals) bead beating tubes. We beat the samples at 3.1 m/s for 3 cycles  
534 with 1 minute of ice cooling between each cycle using a Bead Ruptor Elite 24 (Omni  
535 International). We isolated DNA from the resulting lysate using the Qiagen QIAamp Fast DNA

536 stool mini kit (cat. No. 51604). Samples were extracted individually and pooled by cage with  
537 each sample contributing a total of 200 ng of DNA.

538 We submitted genomic DNA (gDNA) to the North Carolina State Genomic Sciences  
539 Laboratory (Raleigh, NC, USA) for Illumina NGS library construction and sequencing. We used  
540 an Illumina TruSeq Nano Library kit with its provided protocol for library construction and  
541 performed sequencing on an Illumina NovaSeq 6000 sequencer. We obtained between  
542 51,152,549 and 74,618,259 paired-end reads for each of the 4 samples.

543

544 *Metagenomic assembly and protein database construction*

545 To create a species specific database for metaproteomics, we assembled raw reads  
546 using a genome resolved metagenomics approach. We removed PhiX174 (NCBI GenBank  
547 accession CP004084.1) and mouse genome (mm10) contaminating sequences using BBSSplit  
548 and removed adapters using BBDuk (BBMap, Version 38.06), parameters: mink = 6, minlength  
549 =20[56]. We assembled decontaminated reads individually using MetaSPAdes (v3.12.0) -k  
550 33,55,99[57] and co-assembled them using MEGAHIT (v1.2.4) –kmin 31 –k-step 10 [58]. We  
551 mapped reads from all four samples to all five assemblies using bbmap, and binned the contigs  
552 using MetaBAT (v2.12.1) [59]. We assessed the quality of the bins using CheckM (v1.1.3)[60]  
553 and automatically accepted medium quality bins with a completion score greater than 50% and  
554 less than 10% [61]. Since the purpose of metagenomics in our study was to generate a  
555 comprehensive protein sequence database and to assign proteins to species, we further  
556 accepted bins that were greater than >30% complete and <5% contaminated. We clustered the  
557 bins into species groups by 95% ANI using dRep (v2.6.2) [22,62] and assigned taxonomy using  
558 GTDB-Tk (v1.3.0, ref r95) [63].

559 We assembled the protein database by combining gene annotations from the  
560 metagenome with mouse and dietary protein databases [19]. For the metagenome, we  
561 annotated the assemblies prior to binning and then for each bin individually using PROKKA

562 (Version 1.14.6) [64]. If the contig was binned, we compiled the annotations from the bins. We  
563 then used CD-HIT-2D (Version 4.7), with a 90% identity cutoff, to compare the genes from the  
564 unbinned PROKKA output to the binned gene annotation [65]. If a gene was not present in a bin  
565 we added it to the database as an unbinned sequence. Once we compiled the microbial protein  
566 database we assigned each protein sequence a species code if it was species specific or an  
567 ambiguous, low-quality, or unbinned code if it was assigned to more than one species group,  
568 belonged to a low-quality bin, or was not present in a bin, respectively. In addition to the  
569 microbial sequences, we added a *Mus musculus* proteome (UP000000589, Downloaded  
570 19Feb20), and the relevant dietary protein database for each sample: *Glycine max*  
571 (UP000008827, Downloaded 19Feb20), *Bos taurus* (UP000009136, Downloaded 19Feb20),  
572 *Cyberlindnera jadinii* (UP000094389, Downloaded 25May20), *Oryza sativa* (UP00059680,  
573 Downloaded 25May20) and *Gallus gallus* (UP000000539, Downloaded 25May20). Due to the  
574 lack of a reference proteome for the yellow pea diet, we created a custom pea reference with all  
575 available UniProtKB protein sequences for *Pisum sativum* (Taxon ID: 388 Downloaded  
576 25Apr20) and the reference proteome of *Cajanus cajan* (UP000075243, Downloaded  
577 25May20). For T0 samples taken when mice were fed a standard chow diet, we added  
578 proteomes from the protein sources likely to be in the diet based on the ingredient list (corn  
579 UP000007305, fish UP000000437, soy UP000008827, wheat UP00019116, Downloaded  
580 19Feb20). We clustered the mouse and diet reference proteomes individually at a 95% identity  
581 threshold. We only searched samples against their respective dietary database.

582 In order to identify all sequences from the species *Bacteroides thetaiotaomicron* (*B.*  
583 *theta*) and *Lactococcus lactis* we downloaded all the sequences matching these species from  
584 UniProt [66]. We then used diamond BLASTp to identify all sequences in the protein database  
585 that matched with 95% identity or greater. The species code for these sequences was changed  
586 to BT or LAC if they were found to be *B. theta* or *L. lactis* respectively.

587

588 *Metaproteomic sample processing*

589 We extracted protein using a modified FASP protocol [67]. We pelleted fecal samples by  
590 centrifugation (21,000 x g, 5 min) and removed the preservation solution. We suspended dietary  
591 and fecal pellets in SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT] in  
592 Lysing Matrix E tubes (MP Biomedicals) and bead beat the samples (5 cycles of 45 s at 6.45  
593 m/s, 1 min between cycles). After bead beating we heated the lysates to 95°C for 10 minutes.  
594 We mixed 60 µL of the resulting lysates with 400 µL of UA solution (8 M urea in 0.1 M Tris/HCl  
595 pH 8.5), loaded the sample onto a 10 kDa 500 µL filter unit (VWR International) and centrifuged  
596 at 14,000 x g for 30 minutes. We repeated this step up to three times to reach filter capacity.  
597 After loading, we added another 200 µL of UA buffer and centrifuged at 14,000 x g for another  
598 40 minutes. We added 100 µL of IAA solution (0.05 M iodoacetamide in UA solution) to the filter  
599 and incubated at 22°C for 20 minutes. We removed IAA by centrifuging the filter at 14,000 x g  
600 for 20 minutes. We then washed the filter 3 times by adding 100 uL of UA buffer and  
601 centrifuging at 14,000 x g for 20 minutes. We then washed 3 more times by adding 100 uL of  
602 ABC buffer (50 mM Ammonium Bicarbonate) and centrifuging at 14,000 x g for 20 minutes. To  
603 digest the isolated protein, we added 0.95 µg of MS grade trypsin (Thermo Scientific Pierce,  
604 Rockford, IL, USA) mixed in 40 µL of ABC to each filter and incubated at 37°C for 16 hours. We  
605 then eluted the peptides by centrifugation at 14,000 x g for 20 minutes. We eluted again with 50  
606 uL of 0.5 M NaCL and centrifuged at 14,000 x g for another 20 minutes. We quantified the  
607 abundance of the peptides using the Pierce Micro BCA assay (Thermo Scientific Pierce,  
608 Rockford, IL, USA) following the manufacturer's instructions.

609 We analyzed the samples by 1D-LC-MS/MS. Samples were run in randomized block  
610 design. For each run, we loaded 600 ng of peptides onto a 5 mm, 300 µm ID C18 Acclaim®  
611 PepMap100 pre-column (Thermo Fisher Scientific) using an UltiMate™ 3000 RSLCnano Liquid  
612 Chromatograph (Thermo Fisher Scientific) and desalting on the pre-column. After desalting, the  
613 pre-column was switched in line with a 75 cm x 75 µm analytical EASY-Spray column packed

614 with PepMap RSLC C18, 2  $\mu$ m material (Thermo Fisher Scientific), which was heated to 60 °C.  
615 The analytical column was connected via an Easy-Spray source to a Q Exactive HF Hybrid  
616 Quadrupole-Orbitrap mass spectrometer. Peptides were separated using a 140 minute reverse  
617 phase gradient [54]. We acquired spectra using the following parameters: m/z 445.12003 lock  
618 mass, normalized collision energy equal to 24, 25 s dynamic exclusion, and exclusion of ions of  
619 +1 charge state. Full MS scans were acquired for 380 to 1600 m/z at a resolution of 60,000 and  
620 a max IT time of 200 ms. Data-dependent MS<sup>2</sup> spectra for the 15 most abundant ions were  
621 acquired at a resolution of 15,000 and max IT time of 100 ms.

622

623 *Metaproteomic data processing*

624 We searched raw MS spectra against the diet specific protein databases using the run  
625 calibration, SEQUEST HT and percolator nodes in Proteome Discoverer 2.3 (Thermo Fisher  
626 Scientific). We used the following setting for search: trypsin (full), 2 missed cleavages, 10 ppm  
627 precursor mass tolerance, 0.1 Da fragment mass tolerance. We included the following dynamic  
628 modifications: oxidation on M (+15.995 Da), deamidation on N,Q,R (0.984 Da) and acetyl on the  
629 protein N terminus (+42.011 Da). We also included the static modification carbamidomethyl on  
630 C (+57.021 Da). We filtered identified peptides and proteins at a false discovery rate (FDR) of  
631 5%. Additionally, we only included proteins that had at least one protein unique peptide  
632 identified. Proteins were quantified by peptide spectral match (PSM) count (spectral counting).

633

634 *Statistical analysis and visualization*

635 Whenever possible in this study we tested significance of changes in abundance by  
636 applying an ANOVA on a linear mixed effects model with the interacting fixed effects being  
637 mouse group and diet, and the random effect being the individual mouse (lme4 version 4.3.1)  
638 [68]. For multiple comparisons we calculated 95% confidence intervals for each diet using the  
639 emmeans R package (version 1.8.8) [69]. The exceptions were PERMANOVA analysis for

640 testing significance of microbiota compositional changes (Extended Data Table 5) and Welch's  
641 t-tests to compare differences between yeast and egg white protein diets (Extended Data Table  
642 17, 18 and 20). For each analysis, we controlled for multiple-hypothesis testing by converting p-  
643 values to FDR-level based q-values, unless all p-values in the analysis were below 0.05 [70,71].  
644 By definition, if all the p-values are less than 0.05 than the FDR is less than 0.05. Visualizations  
645 were produced using ggplot2 (version 3.4.3) [72], pheatmap (version 1.0.12) [73], RawGraphs  
646 [74], Microsoft Excel and Adobe Illustrator. All boxes and error bars represent 95% confidence  
647 intervals. Boxes or error bars that do not overlap denote significance. If no error bars are  
648 present then significance is denoted by letters or asterisk. In the case of beta-diversity analysis  
649 the error bars are 95% confidence intervals (Fig. 1d), but significance was tested separately by  
650 PERMANOVA.

651

652 *Compositional profiling of the microbiota*

653 We calculated the abundances of specific taxa in the microbiota using proteinaceous  
654 biomass [16]. Briefly, we filtered for proteins with at least 2 protein unique peptides and summed  
655 their spectra into their assigned taxonomy: microbial species, mouse, diet, ambiguous, low  
656 quality bins, unbinned bacteria (See *Metagenomic assembly and protein database construction*  
657 section for details on assignment). We calculated the microbe to host ratio by summing the  
658 spectral count assigned to microbial species, multiple microbial species, low quality bins and  
659 unbinned bacteria proteins and dividing them by the number of spectral counts assigned to  
660 mouse proteins. We considered a microbial species quantifiable if we could identify at least one  
661 protein with 2 protein unique peptides unambiguously assigned to the species. We calculated  
662 per sample species richness by simply counting the number of quantifiable species per sample.  
663 We calculated alpha (Shannon Diversity Index) and beta diversity (Bray-Curtis) metrics using  
664 the vegan (version 2.6-4) package in R (version 4.3.1)[75,76] on a table of the quantifiable  
665 microbial species (statistics as described above). We also evaluated the composition of the

666 microbiota using principal component analysis and hierarchical clustering. For principal  
667 component analysis we normalized the quantified species using centered-log ratio  
668 transformation and calculated principal components using the prcomp function in base R on all  
669 the mice and separately on each mouse group. Principal components were rendered using the  
670 ggplot2 (version 3.4.3) package in R [72]. For hierarchical clustering, we focused on the species  
671 that were most abundant, representing at least 5% of the microbial species biomass in at least  
672 one sample. We calculated the percent biomass for all the species and then extracted the  
673 species that fit the abundant species criteria. We calculated the individual significance of each  
674 abundant species using linear mixed effects models as described above. We hierarchically  
675 clustered log transformed values of these species using the R package pheatmap (version  
676 1.0.12), using the ward.D2 algorithm and euclidean distances [73]. We compared broad  
677 taxonomic changes at the class level. For all quantifiable species we summed the abundance of  
678 the assigned class by GTDB-Tk (see *Metagenomic assembly and protein database construction*  
679 for details). We then calculated confidence intervals using the linear mixed effects models and  
680 emmeans as described previously. Barcharts were rendered in ggplot2 using the estimated  
681 mean and 95% confidence intervals as error bars.

682

#### 683 *Functional profiling of the microbiota*

684 For analyses of functional categories at the level of the whole microbiota we calculated  
685 the normalized spectral abundance factor (NSAF%) for each protein, which provides the relative  
686 abundance for each protein as a percentage of the summed abundance of microbiota proteins  
687 [77]. We annotated functions for all microbial proteins in our database using EggNOG-mapper  
688 [78], MANTIS [79], and Microbe Annotator [80]. We assigned glycoside hydrolase protein family  
689 identifiers from the CAZy database using dbCAN2 [39,81]. We manually curated these  
690 annotations by searching a subset of these proteins against the Swiss-Prot [66] and InterPro  
691 [82] databases between February 2023 and June 2023 (See results for exact numbers). If the

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

700

701 *In vivo proteomic analysis of B. theta*

702 To analyze the *B. theta* proteome, we calculated the orgNSAF by extracting all proteins  
703 assigned to the species *B. theta* detected in the metaproteomes, and then calculating NSAF%  
704 [18]. We then compared abundances of *B. theta* proteins detected in the yeast and egg protein  
705 diets using the Welch's t-test in the Perseus software (version 1.6.14.0) [83]. To visualize  
706 polysaccharide utilization loci (PULs), we mapped the reads from one of our metagenomic  
707 samples to all the contigs that were assigned *B. theta* proteins using BBSSplit (BBMap, Version  
708 38.06). We then assembled all the mapped reads using metaSPAdes. The genes in this newly  
709 assembled genome overlapped exactly with the previous set of identified *B. theta* genes, and  
710 this *B. theta* genome was uploaded to the RAST server for further analysis [84]. PULs were  
711 detected in the metaproteome by identifying proteins labeled SusC, SusD, or TonB. The rest of  
712 the PUL was identified by visualizing the gene neighborhood in RAST. The identified genes  
713 were then cross referenced against PULDB to assign literature described PUL numbers [85].

714

715 *In vitro growth and proteomics of B. theta*

716 We cultured *B. theta* VPI-5482 in two biological replicates and at least 4 technical  
717 replicates using a defined *Bacteroides* medium similar to that described in [86]. *B. theta* cultures

718 were grown statically at 37°C in a Coy anaerobic chamber (2.5 % H<sub>2</sub>/10 % CO<sub>2</sub>/88.5 % N<sub>2</sub>) in  
719 minimal medium (100 mM KH<sub>2</sub>PO<sub>4</sub>, 8.5 mM [NH<sub>2</sub>]<sub>4</sub>SO<sub>4</sub>, 15 mM NaCl, 5.8 µM vitamin K<sub>3</sub>, 1.44  
720 µM FeSO<sub>4</sub>· 7H<sub>2</sub>O, 1 mM MgCl<sub>2</sub>, 1.9 µM hematin, 0.2 mM L-histidine, 3.69 nM vitamin B<sub>12</sub>, 208  
721 µM L-cysteine, and 7.2 µM CaCl<sub>2</sub>· 2H<sub>2</sub>O). The four dietary protein sources: soy protein  
722 (CA.160480), yeast protein (CA.40115), casein protein (CA.160030), and egg white protein  
723 (CA.160230), were purchased from Envigo and were the same as the protein sources used in  
724 the corresponding diets. Porcine muc2 mucin (Sigma CA. M2378) was also tested alongside  
725 controls of glucose and no carbon source. To aid in suspension in aqueous media, we pre-  
726 prepared the proteins in 200 mM NaOH water at 37°C for four days; the glucose control was  
727 also dissolved in 200 mM NaOH water. We then added the protein or glucose solution to the  
728 pre-prepared media at 0.5% (wt/v). Cultures were grown overnight in minimal media  
729 supplemented with 0.5% (wt/v) glucose before being washed and inoculated into experimental  
730 conditions at 0.01 OD and incubated at 37°C anaerobically with shaking every hour. Colony  
731 forming units (CFUs) per mL of culture were enumerated by drip plating at 0 and 24 hr post  
732 inoculation. Solid media for *B. theta* was Brain-Heart Infusion agar (Difco CA. 241830)  
733 supplemented with 10% Horse Blood (LAMPIRE CA. 7233401) (BHI-HB).

734 To obtain samples for proteomics, we repeated the experiment for the glucose, yeast,  
735 egg white, mucin and soy media. After 8 hours, CFUs were enumerated to confirm growth. We  
736 pelleted cells by centrifuging at 4,000 g for 10 minutes. We then extracted the supernatant and  
737 froze the pellets at -80°C. Protein was extracted by the same FASP protocol described above  
738 but with two differences. We lysed pellets by adding 120 µL of SDT buffer and then heating at  
739 95°C. We used PES 10kDa filters (MilliporeSigma). We also used a similar Mass Spectrometry  
740 procedure, except the samples were run on an Exploris 480 mass spectrometer (Thermo Fisher  
741 Scientific) and 1 µg of peptide were analyzed for each sample. We searched raw MS spectra  
742 using the same Proteome Discoverer 2.3 workflow using the *B. theta* proteome downloaded  
743 from UniProt (UP000001414 downloaded January 9, 2024) as the protein sequence database.

744 We then cross referenced PULs between the metaproteome and the *in vitro* proteome to  
745 compare them.

746

## 747 **DECLARATIONS**

### 748 **Ethics approval and consent to participate**

749 We conducted our animal experiments in the Laboratory Animal Facilities at the NCSU CVM  
750 campus (Association for the Assessment and Accreditation of Laboratory Animal Care  
751 accredited), which are managed by the NCSU Laboratory Animal Resources. Animals assessed  
752 as moribund were humanely euthanized via CO<sub>2</sub> asphyxiation, as approved by NC State's  
753 Institutional Animal Care and Use Committee (Protocol # 18-034-B).

754

### 755 **Data availability**

756 The mass spectrometry proteomics data have been deposited to the ProteomeXchange  
757 Consortium via the PRIDE [87] partner repository with the dataset identifier PXD041586  
758 [Reviewer Access at <https://www.ebi.ac.uk/pride/login> with Reviewer Username:  
759 reviewer\_pxd041586@ebi.ac.uk Password: V9Jz2n4h] (metaproteomic data) and PXD050296  
760 [Reviewer Username: reviewer\_pxd050296@ebi.ac.uk Password:F8I9Cmcz] (*B. theta* *in vitro*  
761 proteomics data). Metagenomic raw reads were submitted to NCBI SRA under the bioproject  
762 identifier PRJNA1026909. All metagenome assembled genomes (MAGs) with accompanying  
763 metadata were submitted to DRYAD DOI: 10.5061/dryad.x0k6djhq5. [Reviewer link:  
764 [https://datadryad.org/stash/share/QagcDe\\_b\\_b0GbbyQ7mPOxBapFL3QbaXt3-fhiZRvDCM](https://datadryad.org/stash/share/QagcDe_b_b0GbbyQ7mPOxBapFL3QbaXt3-fhiZRvDCM)]

765

### 766 **Competing interests**

767 The authors declare that there are no competing interests.

768

### 769 **Funding**

770 This work was supported by the National Institutes of Health through awards R35GM138362  
771 (MK), T32DK007737 (JABR) and P30 DK034987, and by the USDA National Institute of Food  
772 and Agriculture, Hatch project 7002782. The content is solely the responsibility of the authors  
773 and does not necessarily represent the official views of the National Institutes of Health.

774

775 **Acknowledgments**

776 We thank Brandon C. Iker for inspiring this research and initial concepts, Tjorven Hinzke for  
777 advice on statistics, Heather Maughan for editing of the manuscript, Erin Baker for advice on  
778 methods, and Lawrence David and Balfour Sartor for advice and consultation on the project.  
779 All LC-MS/MS measurements were made in the Molecular Education, Technology, and  
780 Research Innovation Center (METRIC) at North Carolina State University.

781

782 **Author Contributions**

783 JABR: Experimental design, data collection, data processing, data analysis, author of original  
784 manuscript, editing

785 AB: Conceptualization of the study, experimental design, data collection, data processing,  
786 editing

787 ASM: Experimental design, data collection, editing

788 AA: Data processing, analysis, editing

789 MVW: Data processing, analysis, editing

790 AKM: Data processing

791 JM: Coding, graphic design

792 SV: Data collection

793 TR: Data collection

794 CMT: Experimental design, data collection, editing

795 MK: Funding, conceptualization of the study, experimental design, data collection, data  
796 processing, data analysis, writing, editing

797

798 **References**

799

800 1. Budhathoki S, Sawada N, Iwasaki M, Yamaji T, Goto A, Kotemori A, et al. Association of  
801 Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality in a Japanese  
802 Cohort. *JAMA Intern Med.* 2019;179:1509–18.

803 2. Song M, Fung TT, Hu FB, Willett WC, Longo VD, Chan AT, et al. Association of Animal and  
804 Plant Protein Intake With All-Cause and Cause-Specific Mortality. *JAMA Intern Med.*  
805 2016;176:1453–63.

806 3. Huang J, Liao LM, Weinstein SJ, Sinha R, Graubard BI, Albanes D. Association Between  
807 Plant and Animal Protein Intake and Overall and Cause-Specific Mortality. *JAMA Intern Med.*  
808 2020;180:1173–84.

809 4. Ahn E, Jeong H, Kim E. Differential effects of various dietary proteins on dextran sulfate  
810 sodium-induced colitis in mice. *Nutr Res Pr.* 2022;16:700–15.

811 5. Bartlett A, Kleiner M. Dietary protein and the intestinal microbiota: An understudied  
812 relationship. *iScience.* 2022;25:105313.

813 6. Oliphant K, Allen-Vercoe E. Macronutrient metabolism by the human gut microbiome: major  
814 fermentation by-products and their impact on host health. *Microbiome.* 2019;7:91.

815 7. Blakeley-Ruiz JA, McClintock CS, Shrestha HK, Poudel S, Yang ZK, Giannone RJ, et al.  
816 Morphine and high-fat diet differentially alter the gut microbiota composition and metabolic  
817 function in lean versus obese mice. *ISME Commun.* 2022;2:66.

818 8. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly  
819 and reproducibly alters the human gut microbiome. *Nature.* 2013/12/11 ed. 2014;505:559–63.

820 9. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies  
821 Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans.  
822 *Cell.* 2019;179:59-73.e13.

823 10. Perler BK, Friedman ES, Wu GD. The Role of the Gut Microbiota in the Relationship  
824 Between Diet and Human Health. *Annu Rev Physiol.* 2023;85:449–68.

825 11. Yao CK, Muir JG, Gibson PR. Review article: insights into colonic protein fermentation, its  
826 modulation and potential health implications. *Aliment Pharmacol Ther.* 2016;43:181–96.

827 12. Sun M, Wu W, Chen L, Yang W, Huang X, Ma C, et al. Microbiota-derived short-chain fatty  
828 acids promote Th1 cell IL-10 production to maintain intestinal homeostasis. *Nat Commun.*  
829 2018;9:3555.

830 13. Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a Human Gut Microbiota's Response  
831 to Diet in Gnotobiotic Mice. *Science.* 2011;333:101–4.

- 832 14. Zhu Y, Shi X, Lin X, Ye K, Xu X, Li C, et al. Beef, Chicken, and Soy Proteins in Diets Induce  
833 Different Gut Microbiota and Metabolites in Rats. *Front Microbiol.* 2017;8:1395.
- 834 15. Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel electrophoresis  
835 and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ*  
836 *Microbiol.* 2004;6:911–20.
- 837 16. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass  
838 contributions in microbial communities via metaproteomics. *Nat Commun.* 2017;8:1558.
- 839 17. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a  
840 gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon  
841 and energy use. *Proc Natl Acad Sci.* 2012;109:E1173.
- 842 18. Mueller RS, Denef VJ, Kalnejais LH, Suttle KB, Thomas BC, Wilmes P, et al. Ecological  
843 distribution and population physiology defined by proteomics in a natural microbial community.  
844 *Mol Syst Biol.* 2010;6:374.
- 845 19. Blakeley-Ruiz JA, Kleiner M. Considerations for constructing a protein sequence database  
846 for metaproteomics. *Comput Struct Biotechnol J.* 2022;20:937–52.
- 847 20. Brooks B, Mueller R, Young J, Morowitz M, Hettich R, Banfield J. Strain-resolved microbial  
848 community proteomics reveals simultaneous aerobic and anaerobic function during  
849 gastrointestinal tract colonization of a preterm infant. *Front Microbiol.* 2015;6:654.
- 850 21. Xiong W, Brown CT, Morowitz MJ, Banfield JF, Hettich RL. Genome-resolved  
851 metaproteomic characterization of preterm infant gut microbiota development reveals species-  
852 specific metabolic shifts and variabilities during early life. *Microbiome.* 2017;5:72.
- 853 22. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic  
854 comparisons that enables improved genome recovery from metagenomes through de-  
855 replication. *ISME J.* 2017/07/25 ed. 2017;11:2864–8.
- 856 23. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery  
857 of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat*  
858 *Microbiol.* 2017;2:1533–42.
- 859 24. Salvato F, Hettich RL, Kleiner M. Five key aspects of metaproteomics as a tool to  
860 understand functional interactions in host-associated microbiomes. *PLOS Pathog.*  
861 2021;17:e1009245.
- 862 25. Derrien M, Vaughan EE, Plugge CM, de Vos WM. *Akkermansia muciniphila* gen. nov., sp.  
863 nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol. Microbiology*  
864 Society; 2004. p. 1469–76.
- 865 26. Lagkouvardos I, Lesker TR, Hitch TCA, Gálvez EJC, Smit N, Neuhaus K, et al. Sequence  
866 and cultivation study of Muribaculaceae reveals novel species, host preference, and functional  
867 potential of this yet undescribed family. *Microbiome.* 2019;7:28.
- 868 27. Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, et al. Glycan  
869 Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science.* 2005;307:1955–9.

- 870 28. Kleiner M. Metaproteomics: Much More than Measuring Gene Expression in Microbial  
871 Communities. *mSystems*. 2019;4:e00115-19.
- 872 29. O'Keefe SJD. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev  
873 Gastroenterol Hepatol*. 2016;13:691–706.
- 874 30. Cagnon L, Braissant O. Hyperammonemia-induced toxicity for the developing central  
875 nervous system. *Brain Res Rev*. 2007;56:183–97.
- 876 31. Rangroo Thrane V, Thrane AS, Wang F, Cotrina ML, Smith NA, Chen M, et al. Ammonia  
877 triggers neuronal disinhibition and seizures by impairing astrocyte potassium buffering. *Nat  
878 Med*. 2013;19:1643–8.
- 879 32. Stummer N, Feichtinger RG, Weghuber D, Kofler B, Schneider AM. Role of Hydrogen  
880 Sulfide in Inflammatory Bowel Disease. *Antioxidants*. 2023;12.
- 881 33. Jaglin M, Rhimi M, Philippe C, Pons N, Bruneau A, Goustard B, et al. Indole, a Signaling  
882 Molecule Produced by the Gut Microbiota, Negatively Impacts Emotional Behaviors in Rats.  
883 *Front Neurosci* [Internet]. 2018;12. Available from:  
884 <https://www.frontiersin.org/articles/10.3389/fnins.2018.00216>
- 885 34. Luscher B, Shen Q, Sahir N. The GABAergic deficit hypothesis of major depressive  
886 disorder. *Mol Psychiatry*. 2011;16:383–406.
- 887 35. Czumaj A, Śledziński T, Mika A. Branched-Chain Fatty Acids Alter the Expression of Genes  
888 Responsible for Lipid Synthesis and Inflammation in Human Adipose Cells. *Nutrients*. 2022;14.
- 889 36. Pereira GV, Boudaud M, Wolter M, Alexander C, De Sciscio A, Grant ET, et al. Opposing  
890 diet, microbiome, and metabolite mechanisms regulate inflammatory bowel disease in a  
891 genetically susceptible host. *Cell Host Microbe* [Internet]. 2024; Available from:  
892 <https://www.sciencedirect.com/science/article/pii/S193131282400060X>
- 893 37. Mayneris-Perxachs J, Castells-Nobau A, Arnoriaga-Rodríguez M, Martin M, de la Vega-  
894 Correa L, Zapata C, et al. Microbiota alterations in proline metabolism impact depression. *Cell  
895 Metab*. 2022;34:681-701.e10.
- 896 38. Fletcher JR, Pike CM, Parsons RJ, Rivera AJ, Foley MH, McLaren MR, et al. *Clostridioides  
897 difficile* exploits toxin-mediated inflammation to alter the host nutritional landscape and exclude  
898 competitors from the gut microbiota. *Nat Commun*. 2021;12:462.
- 899 39. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-  
900 active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42:D490–5.
- 901 40. Grondin JM, Tamura K, Déjean G, Wade DW, Brumer H. Polysaccharide Utilization Loci:  
902 Fueling Microbial Communities. *J Bacteriol*. 2017;199:10.1128/jb.00860-16.
- 903 41. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron E, Pudlo NA, et al. Human gut  
904 Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature*. 2015;517:165–9.

- 905 42. Temple MJ, Cuskin F, Baslé A, Hickey N, Speciale G, Williams SJ, et al. A Bacteroidetes  
906 locus dedicated to fungal 1,6- $\beta$ -glucan degradation: Unique substrate conformation drives  
907 specificity of the key endo-1,6- $\beta$ -glucanase. *J Biol Chem.* 2017;292:10639–50.
- 908 43. Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and  
909 transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe.* 2008;4:447–  
910 57.
- 911 44. Badri DV, Jackson MI, Jewell DE. Dietary Protein and Carbohydrate Levels Affect the Gut  
912 Microbiota and Clinical Assessment in Healthy Adult Cats. *J Nutr.* 2021;151:3637–50.
- 913 45. Cummings J, Hill M, Bone E, Branch W, Jenkins DJA. The effect of meat protein and dietary  
914 fiber on colonic function and metabolism II. Bacterial metabolites in feces and urine1. *Am J Clin  
915 Nutr.* 1979;32:2094–101.
- 916 46. Pinna C, Vecchiato CG, Bolduan C, Grandi M, Stefanelli C, Windisch W, et al. Influence of  
917 dietary protein and fructooligosaccharides on fecal fermentative end-products, fecal bacterial  
918 populations and apparent total tract digestibility in dogs. *BMC Vet Res.* 2018;14:106.
- 919 47. Rose CF. Ammonia-Lowering Strategies for the Treatment of Hepatic Encephalopathy. *Clin  
920 Pharmacol Ther.* 2012;92:321–31.
- 921 48. Shen T-CD, Albenberg L, Bittinger K, Chehoud C, Chen Y-Y, Judge CA, et al. Engineering  
922 the gut microbiota to treat hyperammonemia. *J Clin Invest.* 2015;125:2841–50.
- 923 49. Hwang HS, Kim BS, Park H, Park H-Y, Choi H-D, Kim HH. Type and branched pattern of N-  
924 glycans and their structural effect on the chicken egg allergen ovotransferrin: a comparison with  
925 ovomucoid. *Glycoconj J.* 2014;31:41–50.
- 926 50. Kollár R, Reinhold BB, Petráková E, Yeh HJC, Ashwell G, Drgonová J, et al. Architecture of  
927 the Yeast Cell Wall:  $\beta$ (1 $\rightarrow$ 6)-GLUCAN INTERCONNECTS MANNOPROTEIN,  $\beta$ (1 $\rightarrow$ 3)-  
928 GLUCAN, AND CHITIN\*. *J Biol Chem.* 1997;272:17762–75.
- 929 51. Orlean P. Architecture and Biosynthesis of the *Saccharomyces cerevisiae* Cell Wall.  
930 *Genetics.* 2012;192:775–818.
- 931 52. Thaysen-Andersen M, Mysling S, Højrup P. Site-Specific Glycoprofiling of N-Linked  
932 Glycopeptides Using MALDI-TOF MS: Strong Correlation between Signal Strength and  
933 Glycoform Quantities. *Anal Chem.* 2009;81:3933–43.
- 934 53. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary  
935 Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen  
936 Susceptibility. *Cell.* 2016;167:1339–1353.e21.
- 937 54. Mordant A, Kleiner M. Evaluation of Sample Preservation and Storage Methods for  
938 Metaproteomics Analysis of Intestinal Microbiomes. *Microbiol Spectr.* 2021;9:e0187721.
- 939 55. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Priemé A, Aarestrup FM, et al. Impact of  
940 Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition.  
941 *mSystems.* 2016;1:10.1128/msystems.00095-16.

- 942 56. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National  
943 Lab.(LBNL), Berkeley, CA (United States); 2014.
- 944 57. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile  
945 metagenomic assembler. *Genome Res.* 2017;27:824–34.
- 946 58. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution  
947 for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*.  
948 2015;31:1674–6.
- 949 59. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning  
950 algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*.  
951 2019;7:e7359–e7359.
- 952 60. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the  
953 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*  
954 *Res.* 2015;25:1043–55.
- 955 61. Bowers RM, Kyprides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al.  
956 Minimum information about a single amplified genome (MISAG) and a metagenome-assembled  
957 genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35:725–31.
- 958 62. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF.  
959 Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries.  
960 *mSystems*. 2020;5.
- 961 63. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes  
962 with the Genome Taxonomy Database. *Bioinformatics*. 2020;36:1925–7.
- 963 64. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinforma Oxf Engl*.  
964 2014;30:2068–9.
- 965 65. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or  
966 nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
- 967 66. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic*  
968 *Acids Res.* 2023;51:D523–31.
- 969 67. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for  
970 proteome analysis. *Nat Methods*. 2009;6:359–62.
- 971 68. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J*  
972 *Stat Softw.* 2015;67:1–48.
- 973 69. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means [Internet].  
974 2023. Available from: <https://CRAN.R-project.org/package=emmeans>
- 975 70. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful  
976 Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57:289–300.

- 977 71. Storey JD, Taylor JE, Siegmund D. Strong Control, Conservative Point Estimation and  
978 Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach. *J R*  
979 *Stat Soc Ser B Stat Methodol.* 2004;66:187–205.
- 980 72. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* [Internet]. Springer-Verlag New  
981 York; 2016. Available from: <https://ggplot2.tidyverse.org>
- 982 73. Kolde R. *pheatmap: Pretty Heatmaps* [Internet]. 2019. Available from: <https://CRAN.R-project.org/package=pheatmap>
- 984 74. Mauri M, Elli T, Caviglia G, Ubaldi G, Azzi M. RAWGraphs: A Visualisation Platform to  
985 Create Open Outputs. *Proc 12th Biennu Conf Ital SIGCHI Chapter* [Internet]. New York, NY,  
986 USA: Association for Computing Machinery; 2017. Available from:  
987 <https://doi.org/10.1145/3125571.3125585>
- 988 75. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. *vegan: Community Ecology Package* [Internet]. 2022. Available from: <https://CRAN.R-project.org/package=vegan>
- 991 76. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. Vienna,  
992 Austria: R Foundation for Statistical Computing; 2023. Available from: <https://www.R-project.org/>
- 993 77. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, et al. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral  
994 abundance factors. *Methods San Diego Calif.* 2006;40:303–11.
- 996 78. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast  
997 Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol*  
998 *Biol Evol.* 2017;34:2115–22.
- 999 79. Queirós P, Delogu F, Hickl O, May P, Wilmes P. Mantis: flexible and consensus-driven  
1000 genome annotation. *GigaScience.* 2021;10:giab042.
- 1001 80. Ruiz-Perez CA, Conrad RE, Konstantinidis KT. MicrobeAnnotator: a user-friendly,  
1002 comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics.*  
1003 2021;22:11.
- 1004 81. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for  
1005 automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018;46:W95–101.
- 1006 82. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, et al. The  
1007 InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.*  
1008 2021;49:D344–54.
- 1009 83. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus  
1010 computational platform for comprehensive analysis of (prote)omics data. *Nat Methods.*  
1011 2016;13:731–40.
- 1012 84. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server:  
1013 Rapid Annotations using Subsystems Technology. *BMC Genomics.* 2008;9:75.

- 1014 85. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, et al. PULDB: the  
1015 expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 2018;46:D677–83.
- 1016 86. Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe EC, et al. Glycan  
1017 complexity dictates microbial resource allocation in the large intestine. *Nat Commun.*  
1018 2015;6:7481.
- 1019 87. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan  
1020 S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based  
1021 proteomics evidences. *Nucleic Acids Res.* 2022;50:D543–52.

1022 **Supplementary Information**

1023

1024

1025 **Dietary protein source strongly alters gut microbiota composition and**  
1026 **function**

1027

1028

1029

1030 Authors: J. Alfredo Blakeley-Ruiz† \*<sup>1</sup>, Alexandria Bartlett† <sup>1,2</sup>, Arthur S. McMillan<sup>3</sup>, Ayesha  
1031 Awan<sup>1,3</sup>, Molly Vanhoy Walsh<sup>1</sup>, Alissa K. Meyerhoffer<sup>1</sup>, Simina Vintila<sup>1</sup>, Jessie L. Maier<sup>1</sup>, Tanner  
1032 Richie<sup>1</sup>, Casey M. Theriot<sup>3</sup>, Manuel Kleiner\*<sup>1</sup>

- 1033
- 1034 1. Department of Plant and Microbial Biology, College of Agricultural Sciences, North  
1035 Carolina State University, Raleigh, NC, USA
- 1036 2. Department of Molecular Genetics and Microbiology, Duke University, Durham, NC,  
1037 USA
- 1038 3. Department of Population Health and Pathobiology, College of Veterinary Medicine,  
1039 North Carolina State University, NC, USA

1040

1041

1042

1043 † These authors contributed equally to this manuscript

1044 \* Please refer correspondence to J. Alfredo Blakeley-Ruiz (jablakel@ncsu.edu) or Manuel  
1045 Kleiner ([manuel\\_kleiner@ncsu.edu](mailto:manuel_kleiner@ncsu.edu))

1048 **Supplementary Results**

1049 **Section A: Effect of dietary protein on gut microbiota composition**

1050 Hierarchical clustering of the proteinaceous biomass of species revealed distinct starting  
1051 microbiota compositions for the group 1 and group 2 mice. This led us to surmise that the  
1052 source of dietary protein altered the gut microbiota regardless of starting microbiota  
1053 composition. To confirm this, we analyzed the group 1 and group 2 mice separately. We  
1054 observed that they had similar alpha diversity and richness responses to dietary protein source  
1055 regardless of mouse group and that the Bray-Curtis dissimilarity patterns were also similar (Fig.  
1056 1d and e; Supplementary Fig. 1a, b and c). Principal component analysis showed the same  
1057 separation as the hierarchical clustering between mouse group 1 and 2 along the first  
1058 component (Supplementary Fig. 1d). Further analysis along the second and third components  
1059 showed distinct dietary clusters between 1) soy, 2) T0, 3) yeast and egg white, and 4) casein,  
1060 brown rice and pea (Supplementary Fig. 1e). Principal component analysis on the separate  
1061 mouse groups revealed the same cluster groups (Supplementary Fig. 1f and g), suggesting that  
1062 the source of dietary protein alters the gut microbiota's composition regardless of starting  
1063 microbiota composition.

1064

1065 We observed dynamic species abundance responses to different sources of dietary  
1066 protein; these responses were in some cases consistent across mouse groups, while in other  
1067 cases they were mouse group specific. For example, *Bacteroides thetaiotaomicron* (*B. theta*)  
1068 increased in abundance across both mouse groups in the yeast and egg white diets (Fig. 1g;  
1069 Supplementary Fig. 2), while the abundances of *Schaederella* sp. AB67-1 and *Lachnospiraceae*  
1070 bacterium AB103-0 repeatedly increased in the presence of soy (Supplementary Fig. 2). Other  
1071 changes in species abundance due to diet were mouse group specific. For example,  
1072 *Lactobacillus johnsonii* increased in abundance in the pea and casein diets in the group 1 mice,  
1073 while *Faecalibaculum rodentium* showed a similar pattern in the group 2 mice. *Oscillospiraceae*  
1074 bacterium AB63-2 increased in abundance in the presence of the soy diets in the group 1 mice,  
1075 while *Oscillospiraceae* bacterium AB54-6 followed a similar pattern in the group 2 mice. All of  
1076 the abundant species were significantly different in abundance between at least two diets  
1077 (Supplementary Fig. 2). Supplementary Fig. 2 contains the details for the specific dynamics of  
1078 the 36 most abundant species we detected. All species had a significantly different abundance  
1079 between diets (linear mixed effect model ANOVA,  $q < 0.05$ ).  
1080

1081 **Section B: Effect of dietary protein on microbiota function**

1082 The two most abundant broad functional categories of detected peptides were gene  
1083 expression, which includes ribosomes, chaperones, and transcription related enzymes, and  
1084 monosaccharide metabolism, which includes glycolysis and the metabolism of simple sugars  
1085 other than glucose (Fig. 2). The microbial investment in gene expression enzymes increased in  
1086 the yeast diet relative to all other diets (except standard chow) and decreased in the egg white  
1087 diet relative to all other diets. This was driven by an overall increase in the abundance of  
1088 ribosomal proteins in the yeast diet (Supplemental Fig. 3b). The abundance of ribosomal

1089 proteins has been suggested to be directly correlated with bacterial growth rates<sup>1</sup> suggesting  
1090 that overall bacterial growth rate is higher when mice were fed the yeast diet. This is further  
1091 supported by the overall higher bacterial load in the yeast diet (Fig. 1c). In contrast, we  
1092 observed gene expression proteins that assist with the synthesis and folding of proteins, e.g.,  
1093 elongation factors and chaperones, to be increased in the soy diets relative to some of the other  
1094 diets (Supplemental Fig. 3b). Curiously, we also observed an increase in stress proteins,  
1095 including oxidative stress proteins, in the soy and casein diets relative to brown rice, egg white,  
1096 and yeast (Supplementary Fig. 3c). Oxidative stress interferes with the proper elongation and  
1097 folding of proteins, which could explain why chaperones and elongation factors are increased in  
1098 the soy diet<sup>2</sup>.

1099 We observed a significant decrease in the abundance of monosaccharide metabolizing  
1100 enzymes in the yeast and brown rice diets (Fig. 2). Most abundant in this category were the  
1101 enzymes belonging to the energy pay-off phase of glycolysis (Supplementary Fig. 3d). However,  
1102 many of the other functions within monosaccharide metabolism had different abundance  
1103 patterns. For example, galactose and mannose metabolism enzymes were increased in the  
1104 yeast diet (Supplementary Fig 3e), while along with galactose metabolism, we also observed a  
1105 general increase in the abundance of fucose, glucosamine, and sialic acid metabolism in the  
1106 egg white diet relative to other diets. Glucosamine and sialic acid metabolism were also  
1107 increased in the casein and pea diets relative to other diets (Supplementary Fig 3e). Fucose,  
1108 galactose, sialic acid, and acetylglucosamine are all components of mucin<sup>3</sup>. In summary, these  
1109 results suggest that the source of dietary protein impacts sugar metabolism in the gut  
1110 microbiota.

1111 Two other broad functional categories that significantly changed in abundance due to  
1112 dietary protein source were adhesion and motility proteins and fermentation proteins. The  
1113 microbiota invested significantly less in proteins categorized as adhesion and motility proteins in  
1114 the yeast and egg diets. Flagellar proteins drove this result, which can be explained by the  
1115 replacement of species from the class Clostridia with species from the class Bacteroidia  
1116 because microbes in the phylum Bacteroidota usually do not have flagella<sup>4</sup>. The microbial  
1117 investment in fermentation enzymes also decreased in the yeast and egg white diets. We  
1118 divided fermentation enzymes into three categories, ethanol producing, short-chain fatty acid  
1119 (SCFA) producing, and lactic acid producing (Supplementary Fig. 3a). This categorization  
1120 revealed that fermentation enzymes leading to SCFA metabolites were the primary drivers of  
1121 the decrease in fermentation-related proteins in the yeast and egg white diets. Production of  
1122 SCFAs has been previously linked to anti-inflammatory responses, which could suggest that the  
1123 changes in microbiota composition observed in the yeast and egg white diets may be  
1124 detrimental to host health<sup>5,6</sup>.

1125

## 1126 **Section C: Effect of brown rice and soy dietary protein on glycoside hydrolase 1127 abundance**

1128 Several glycoside hydrolases increased in the presence of the brown rice and soy diets.  
1129 In the soy diet, the expression of glycoside hydrolases was reproducible, increasing in  
1130 abundance each of the three times the mice were fed a soy diet. Most notably  $\beta$ -glucosidases  
1131 and  $\beta$ -xylosidases from the CAZy protein family GH3 were increased in the soy diets, while the

1132 abundance of GH16s increased in the brown rice diet. The abundance change of GH16s in the  
1133 brown rice diet was driven by a single protein from the uncultured species *Oscillospiraceae*  
1134 bacterium AB63-2 that was only detected in the brown rice diet. In the presence of soy protein,  
1135 but not brown rice protein, this bacterium reproducibly expressed different glycoside hydrolases  
1136 including two  $\beta$ -glucosidases and one  $\beta$ -xylosidase from the protein family GH3 (Extended Data  
1137 Table 13). These results suggest that different protein sources generally affect the function and  
1138 composition of the gut microbiota through the different glycan structures attached to their  
1139 glycoproteins.

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

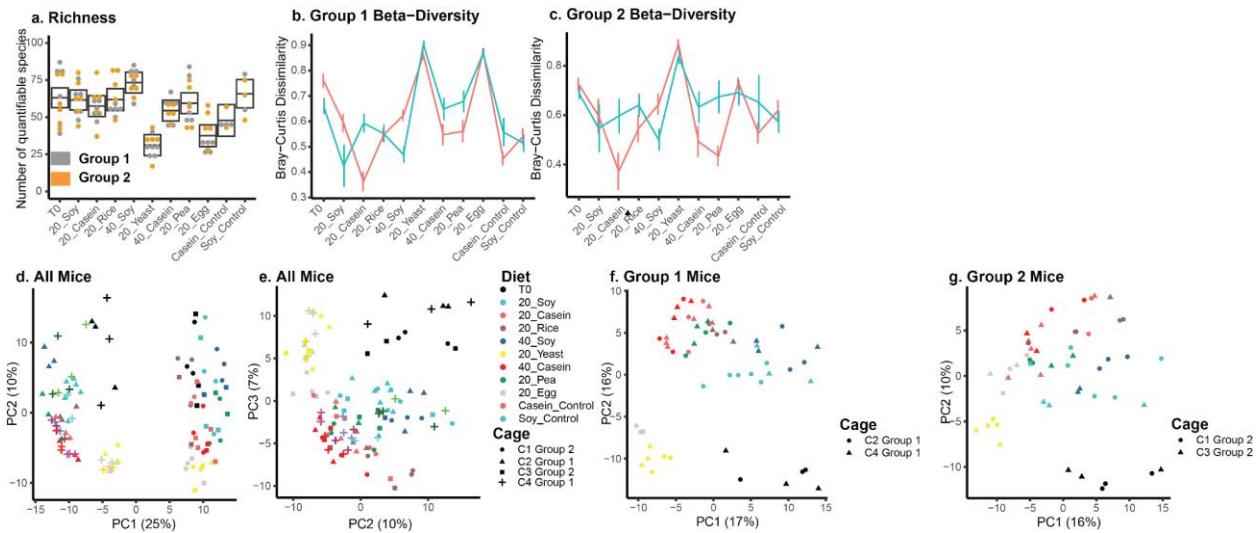
1157

1158

1159

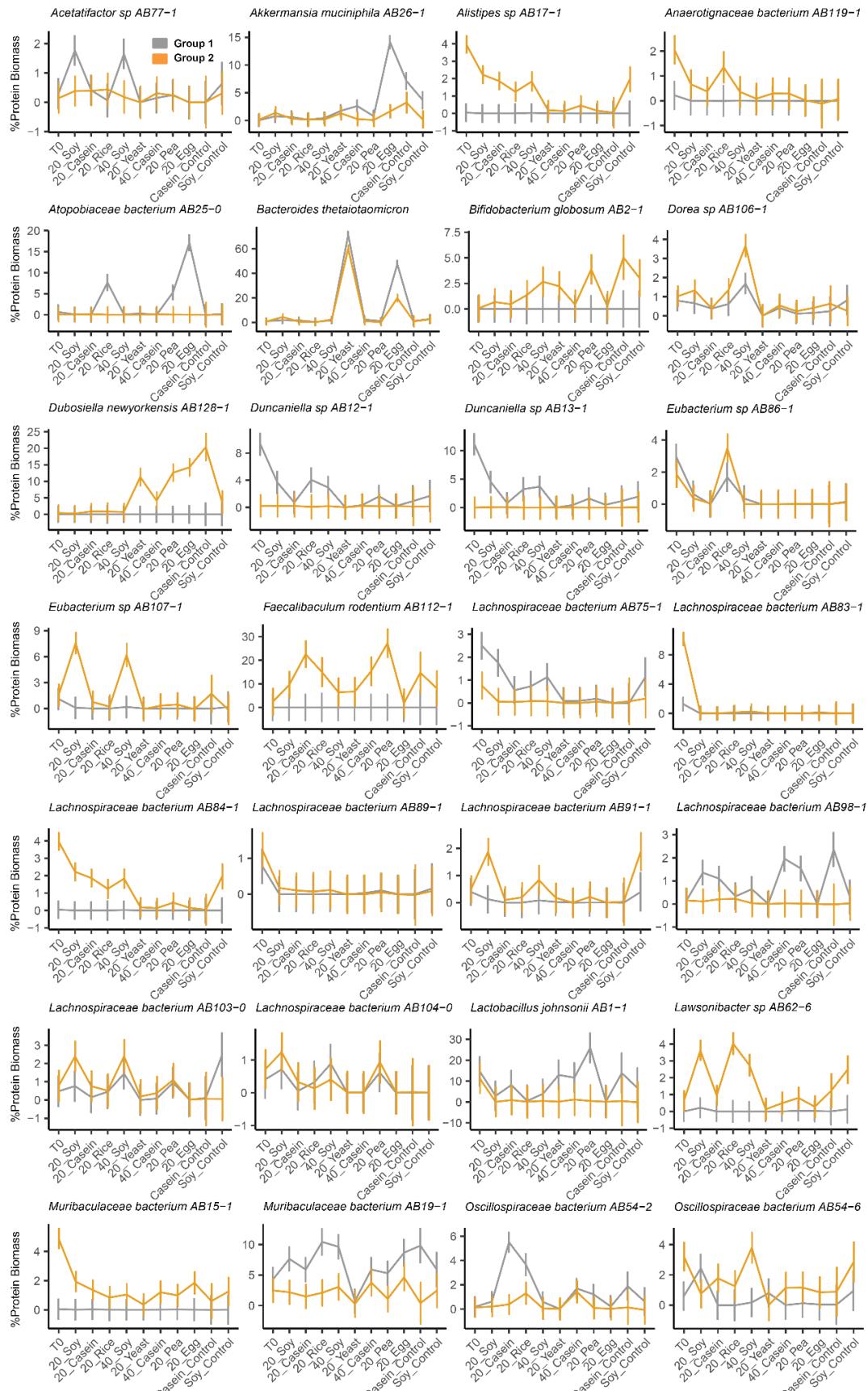
## 1160 Supplementary Table 1: Composition of nine fully defined diets

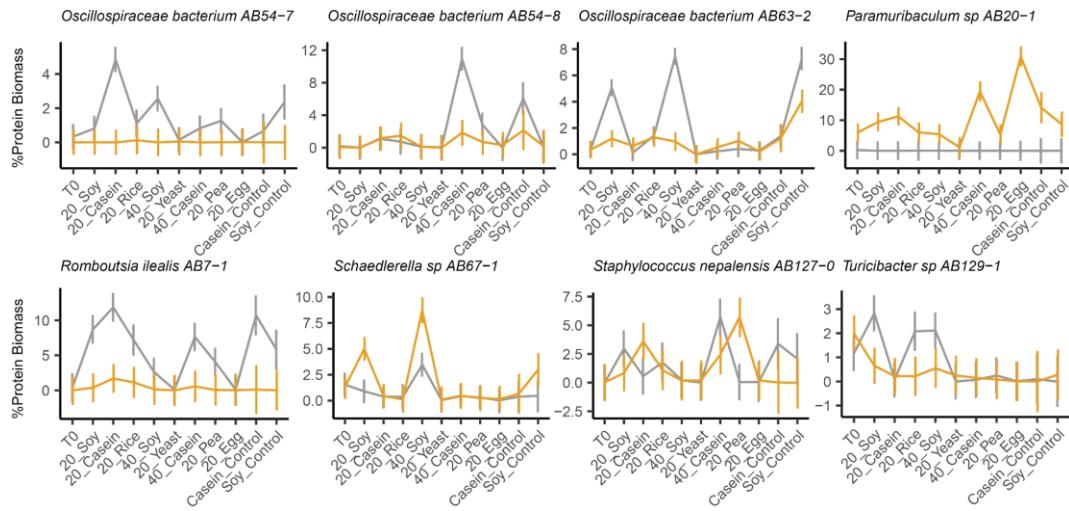
Diet	20% Soy	20% Casein	20% Brown Rice	40% Soy	20% Yeast	40% Casein	20% Pea	20% Egg White Solids, spray-dried	20% Chicken bone broth
Teklad Catalog Number	TD.190249	TD.190254	TD.190252	TD.190250	TD.190253	TD.190255	TD.190251	TD.190256	TD.190257
Protein supplier	Teklad	Teklad	Swanson	Teklad	Teklad	Teklad	NAKED	Teklad	Ancient Nutrition
Protein (g/Kg)	230	230	260	460	400	460	222.22	248.45	220
Sucrose (g/Kg)	436.1	434.7	400.43	211.785	279.06	208.96	438	412.667	450.33
Corn Starch (g/Kg)	200	200	200	200	200	200	200	200	200
Corn Oil (g/Kg)	52.3	52.3	54.6	50	42.6	50	50.82	54.6	44.7
Cellulose (g/Kg)	37.86	37.86	37.86	37.86	37.86	37.86	37.86	37.86	37.86
Vitamin Mix, Teklad (40060) (g/Kg)	10	10	10	10	10	10	10	10	10
Ethoxyquin, antioxidant	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Mineral Mix, Ca-P Deficient 79055 (g/Kg)	13.37	13.37	13.37	13.37	13.27	13.37	13.37	13.37	13.37
Calcium Phosphate dibasic (g/Kg)	15.26	16.66	23.72	6.8	0	9.6	19.72	22.54	23.73
Calcium Carbonate (g/Kg)	5.1	5.1	0	10.175	17.1	10.2	8	0.5	0



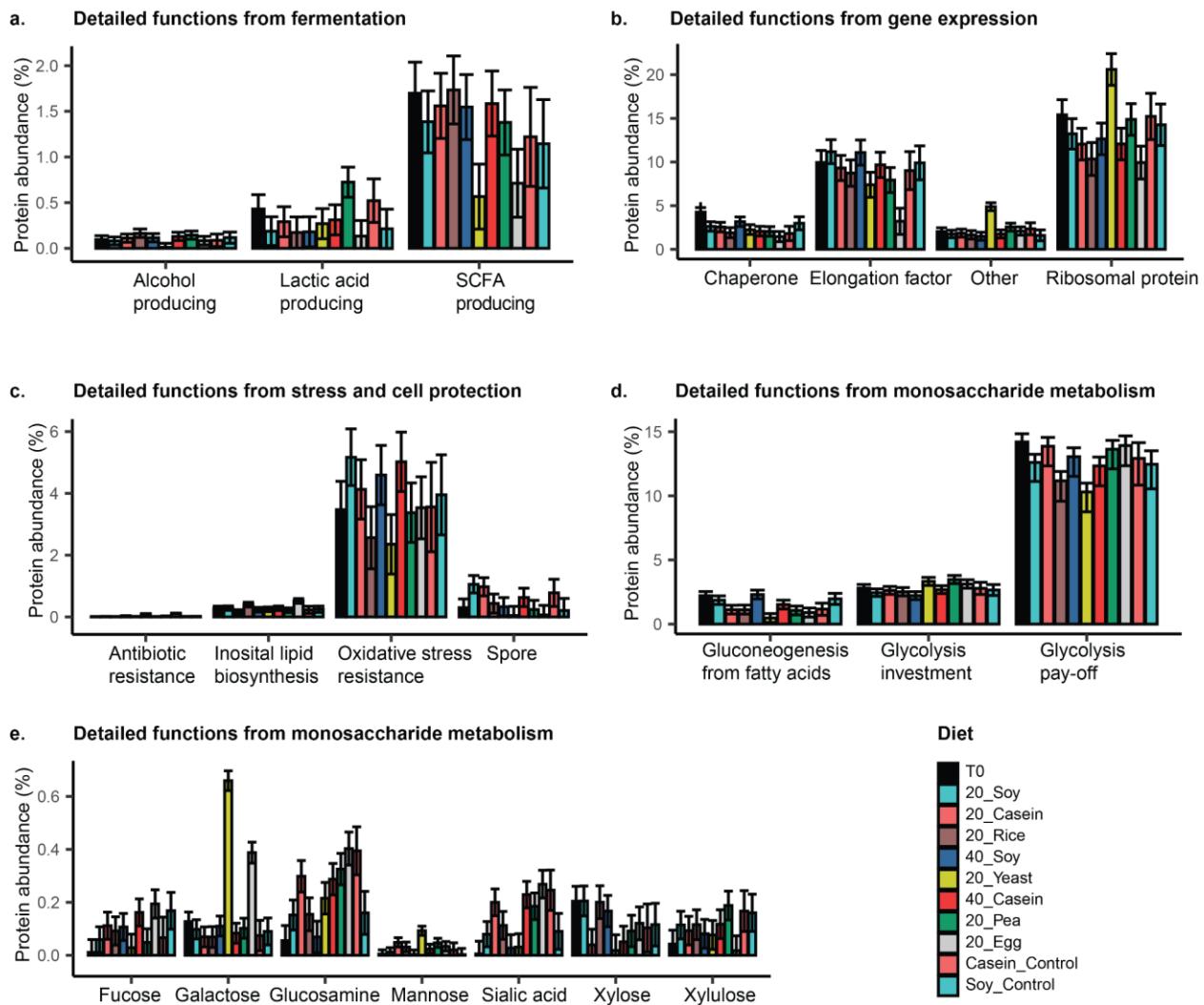
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169

**Supplementary Figure 1: Source of dietary protein alters the gut microbiota's composition.** (a) depicts the average number of quantifiable species per diet; boxes represent the 95% confidence interval based on linear mixed effects models (Extended Data Table 2). (b) and (c) depict the Bray-Curtis dissimilarity between 20% soy diets (teal) or 20% casein diets (red) and all other diets for group 1 and group 2, respectively. (d) and (e) first, second, and third principal components of microbiota composition based on species level metaproteomic proteinaceous biomass. (f) and (g) first and second components of microbiota composition based on species level metaproteomics proteinaceous biomass for the group 1 and group 2 mice, respectively.





1171  
 1172  
 1173  
 1174 **Supplementary Figure 2: Abundances of the most abundant species across diets and mouse groups.** Line  
 1175 plots depicting the average abundance of bacterial species in the group 1 mice (gray) and group 2 mice (orange)  
 1176 across all diets. Abundances were determined from metaproteomic data using a biomass assessment method<sup>7</sup>. Error  
 1177 bars represent the 95% confidence interval of the mean using mixed effects modeling. Species were defined as  
 1178 abundant if they represented at least 5% of the microbial protein mass in one sample.  
 1179  
 1180



1181  
1182  
1183 **Supplementary Figure 3: Abundance of detailed functional categories associated with fermentation, gene**  
1184 **expression, stress and cell protection, and monosaccharide metabolism.** The mean protein abundance (% of

1185 total microbial proteins) per sample of each detailed function based on a complete linear mixed effects model. Error

1186 bars represent 95% confidence intervals and error bars that do not overlap indicate significant abundance

1187 differences. (a) Detailed functions that make up the fermentation broad functional category. (b) Detailed functions that

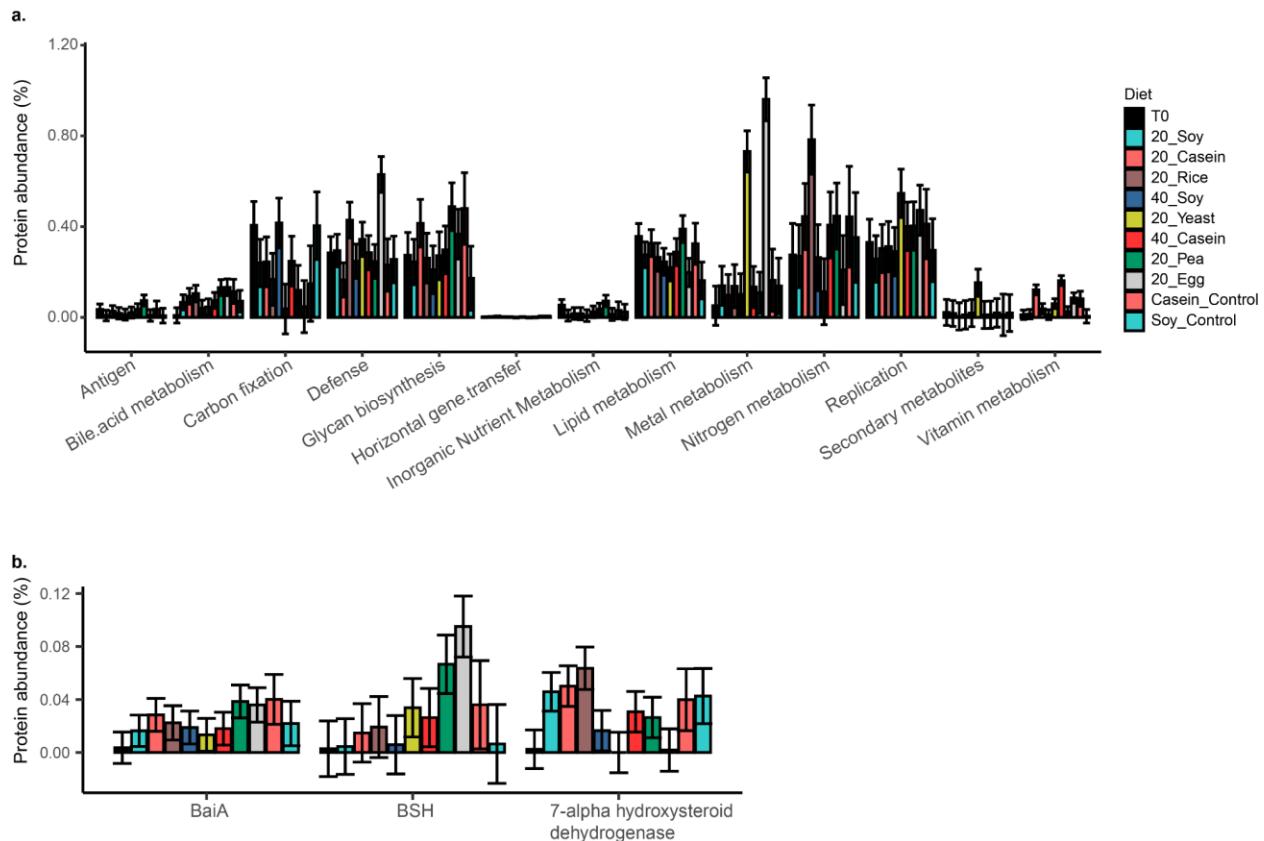
1188 make up the gene expression functional category. (c) Detailed functions that make up the stress and cell protection

1189 functional category. (d) The most abundant detailed functions that make up the monosaccharide metabolism

1190 functional category. (e) Select detailed functions that make up part of the monosaccharide metabolism category

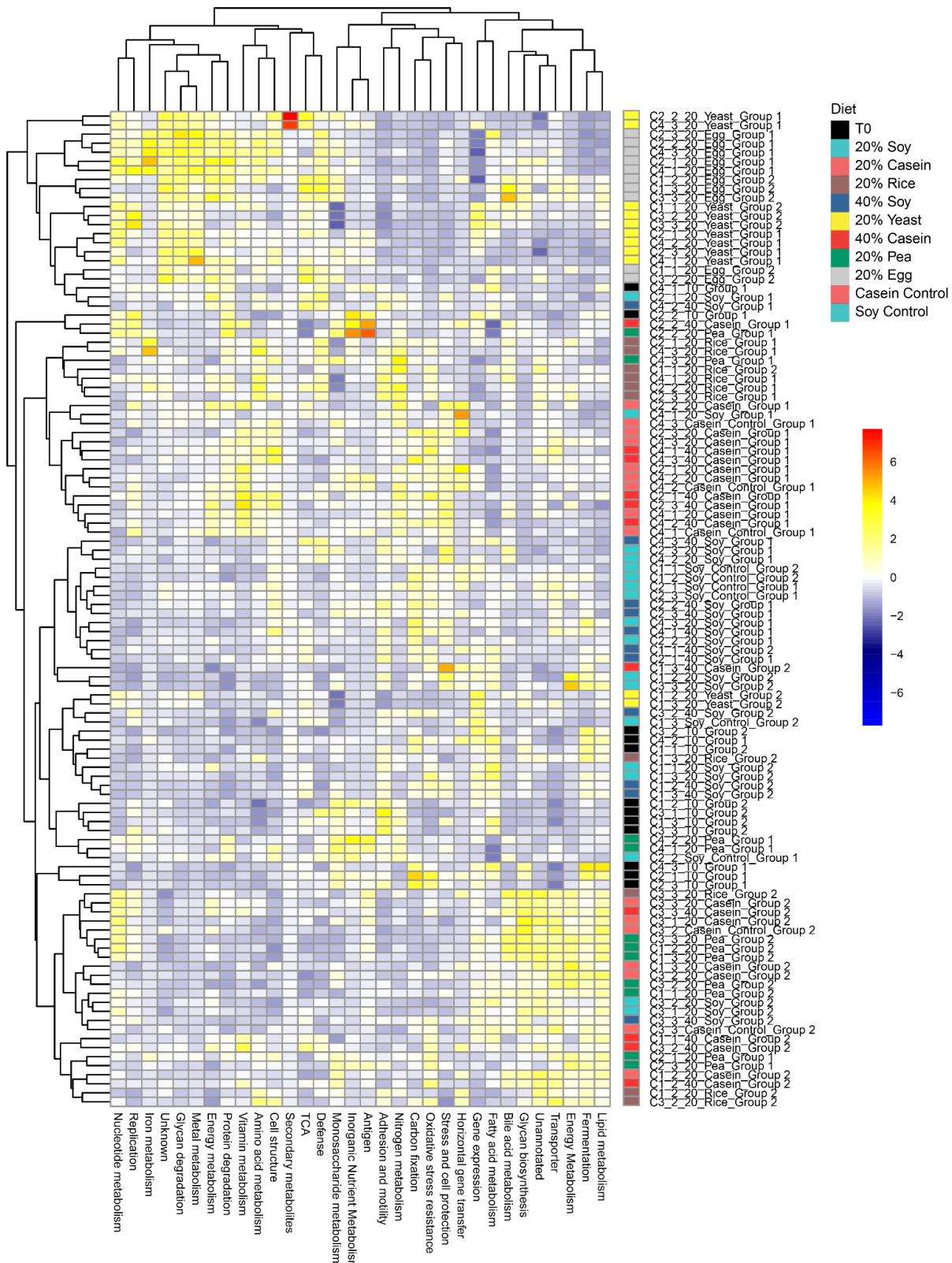
1191 (Extended Data Tables 6, 7 and 8).

1192  
1193  
1194  
1195  
1196



1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204

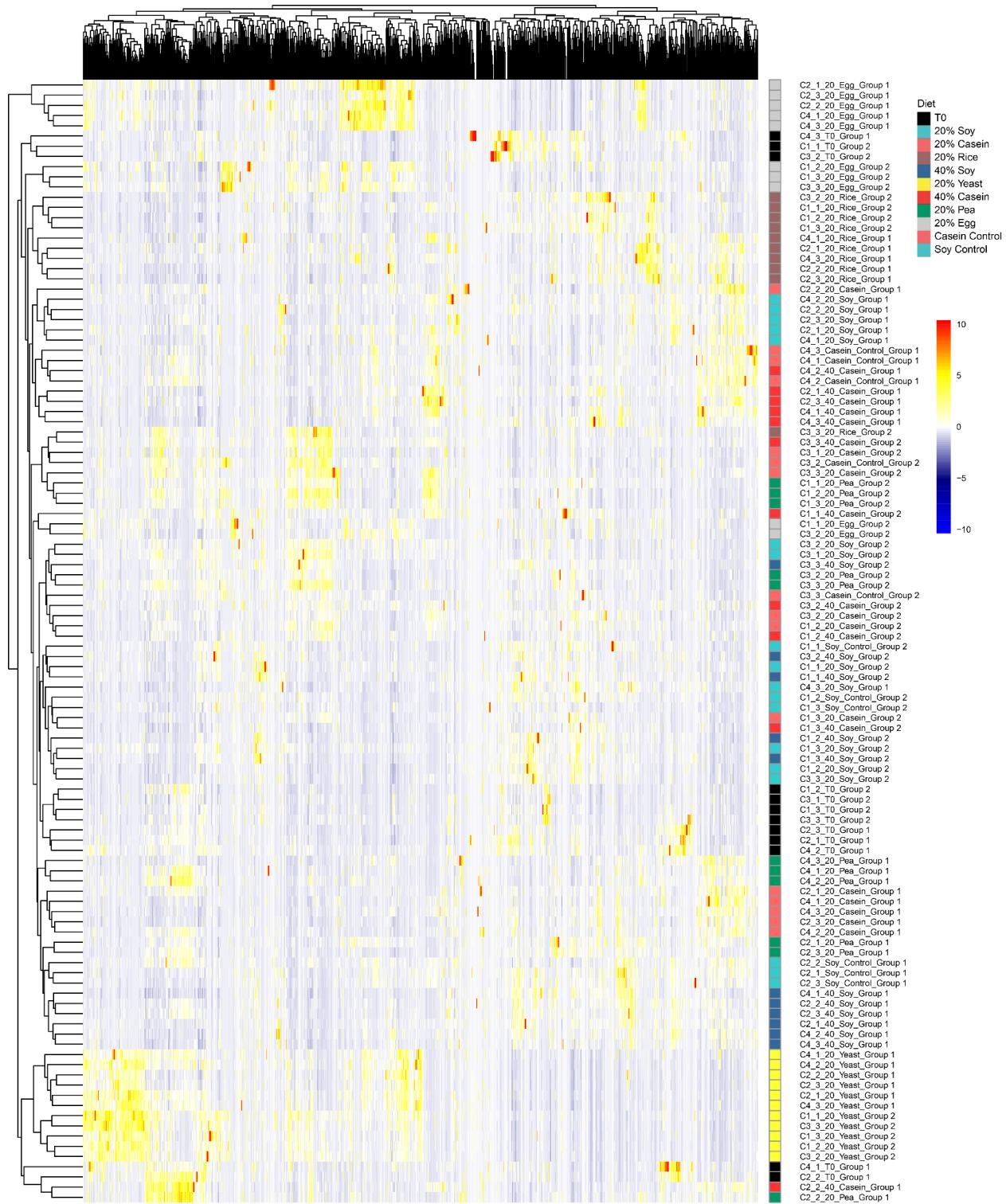
**Supplementary Figure 4: Effect of dietary protein source on lower abundance protein functions.** The mean protein abundance (% of total microbial proteins) per sample of each detailed function based on a complete linear mixed effects model. Error bars represent 95% confidence intervals and error bars that do not overlap indicate significant abundance differences. (a) Broad functional categories that represented less than 1% of the microbial protein abundance. (b) Consensus names of specific bile acid modifying enzymes (Extended Data Tables 6, 7 and 8).



1205  
1206  
1207  
1208

**Supplementary Figure 5: Hierarchical clustering of broad functional categories.** Tabulated abundances of broad functional categories (Extended Data Table 6) were z-scored by feature and then clustered using the ward.D2 method using the pheatmap package in R.

1209



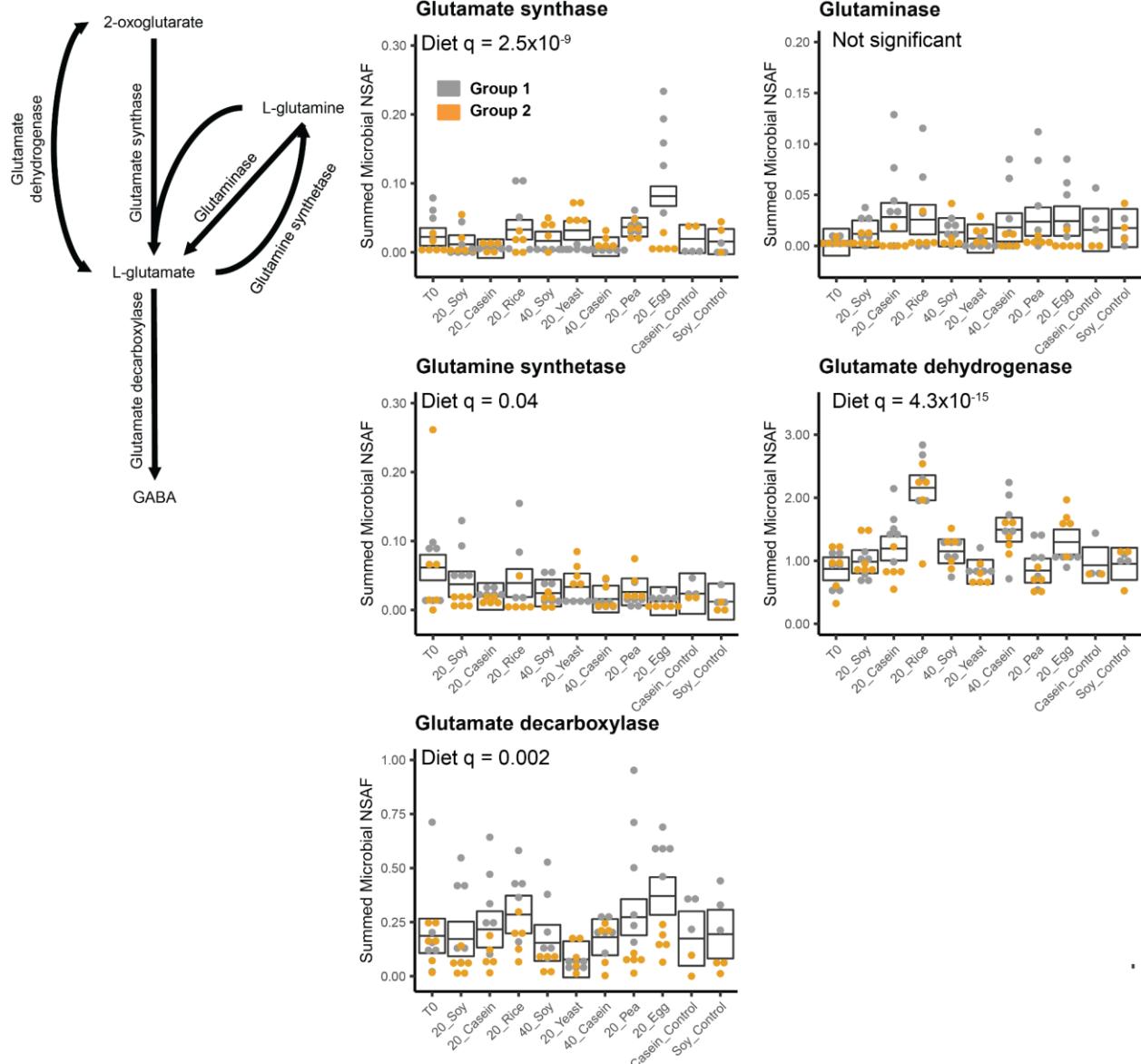
1210

1211

1212

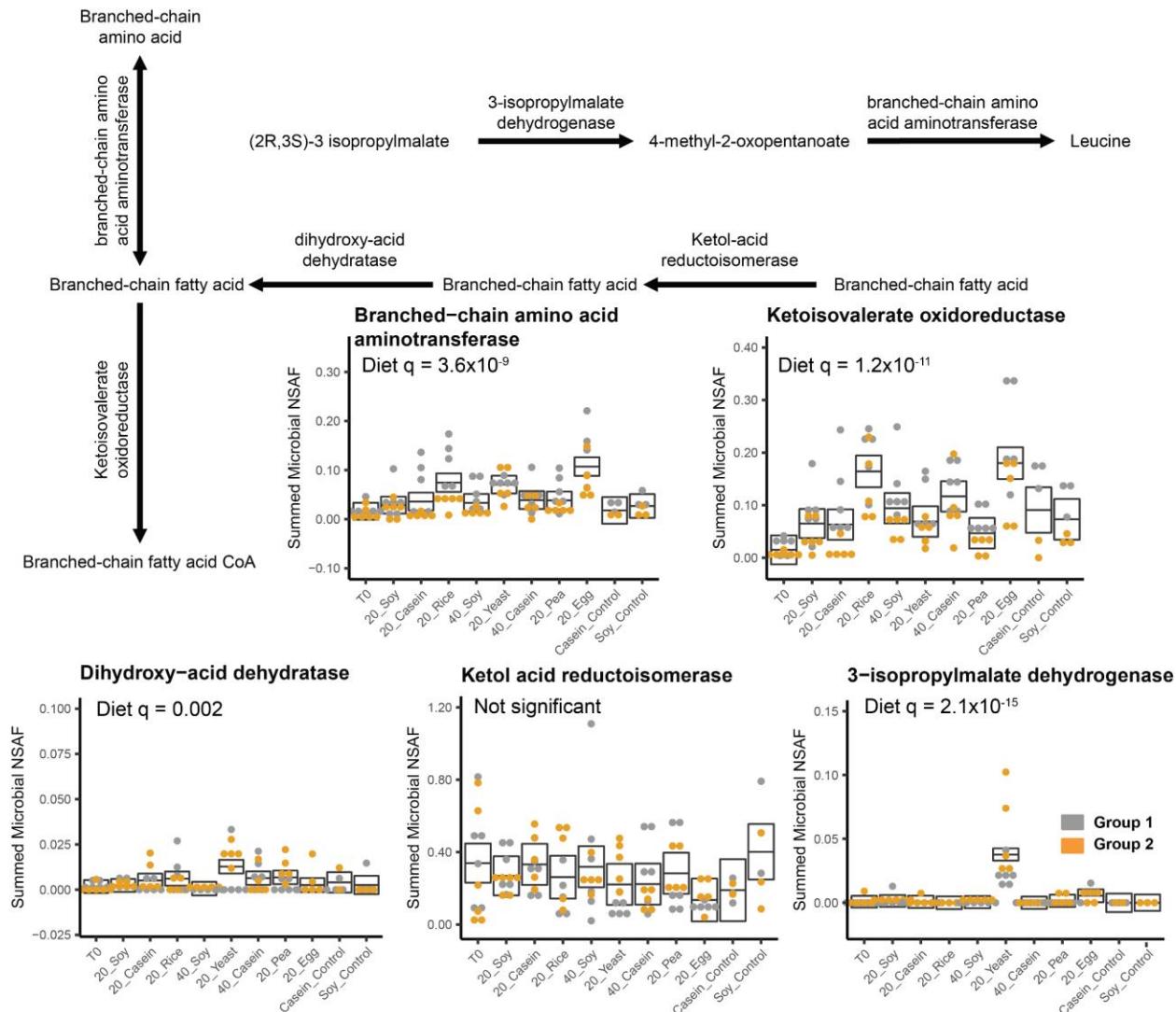
1213

**Supplementary Figure 6: Hierarchical clustering of consensus enzyme names.** Tabulated abundances of consensus enzyme names (Extended Data Table 6) were z-scored by feature and then clustered using the ward.D2 method using the pheatmap package in R.



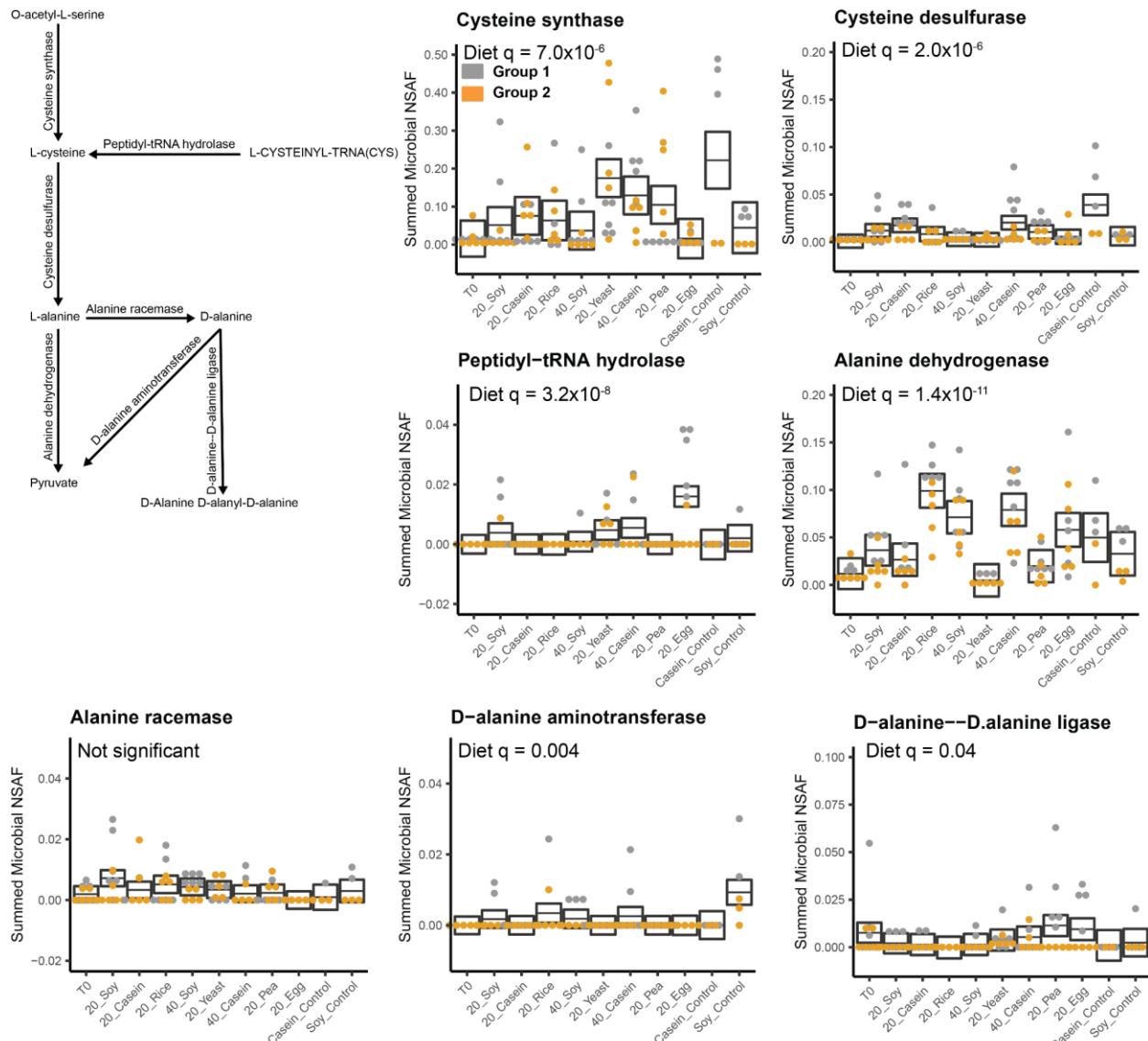
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221

**Supplementary Figure 7: Changes in glutamate and glutamine metabolism due to source of dietary protein:**  
 A reconstruction of the pathways involved in glutamate and glutamine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).



1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231

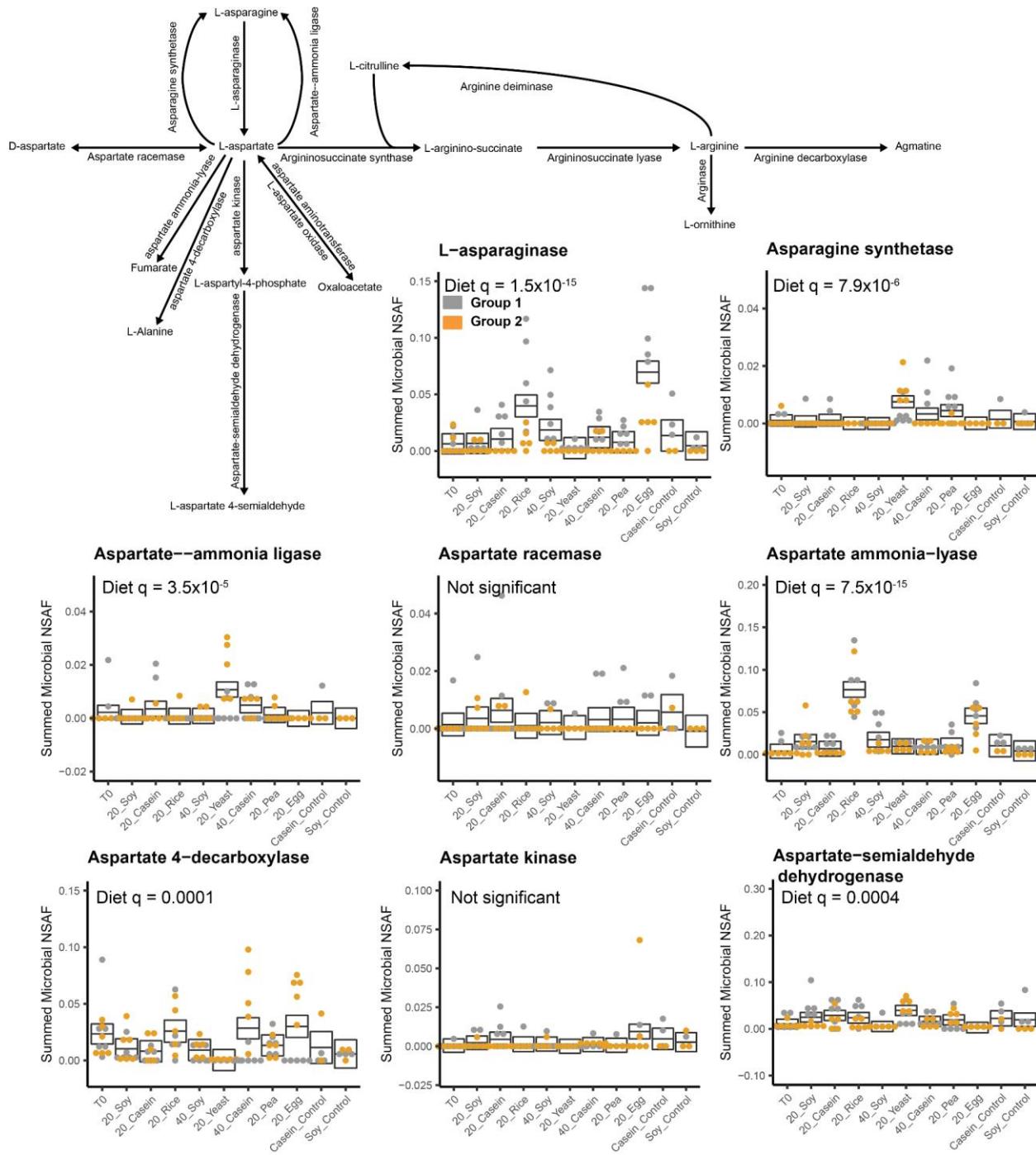
**Supplementary Figure 8: Changes in branched-chain amino acid metabolism due to source of dietary protein.** A reconstruction of the pathways involved in valine, leucine, and isoleucine metabolism based on enzymes detected in our metaproteomes. With the exception of leucine synthesis enzymes, the same enzymes act on all three of these amino acids so we did not try to distinguish them. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).



1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239

**Supplementary Figure 9: Changes in cysteine and alanine metabolism due to source of dietary protein:**

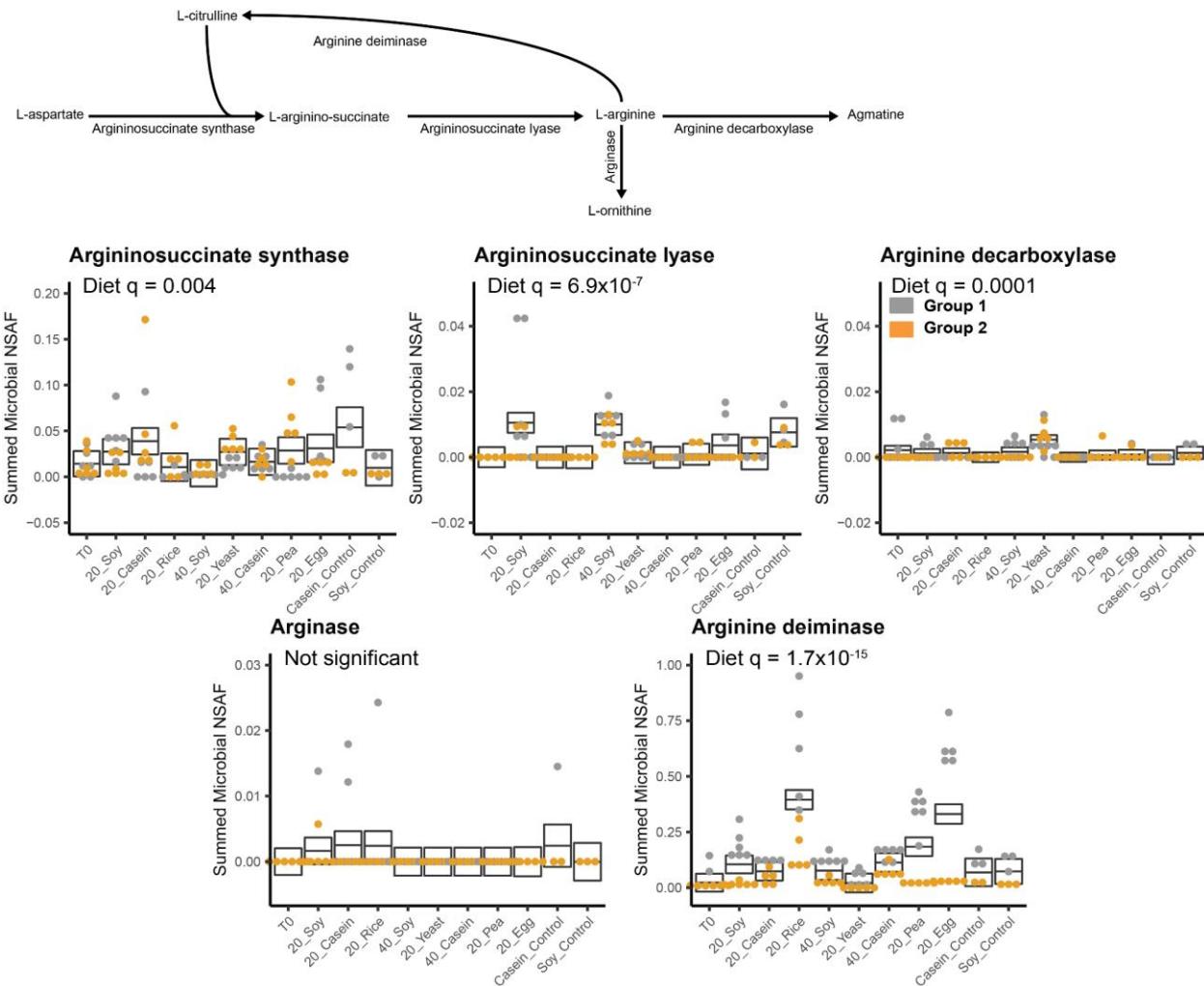
A reconstruction of the pathways involved in cysteine and alanine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. The boxes represent the 95% confidence interval of the mean (line) for each diet from 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).



1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249

**Supplementary Figure 10: Changes in asparagine, aspartate, and arginine metabolism due to source of dietary protein part 1:**

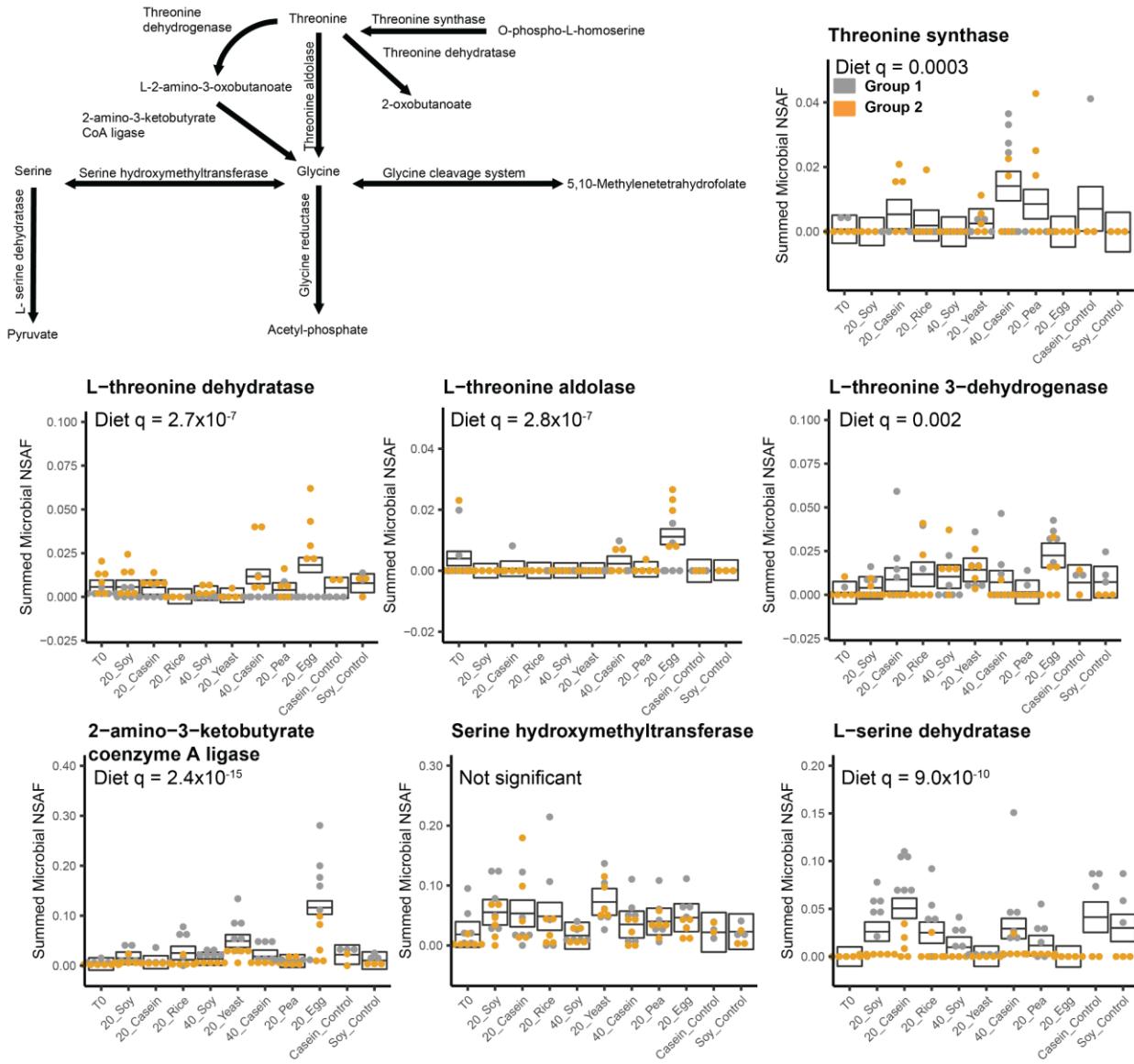
A reconstruction of the pathways involved in asparagine, aspartate, and arginine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. The boxes represent the 95% confidence interval of the mean (line) for 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).



1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258

**Supplementary Figure 11: Changes in asparagine, aspartate, and arginine metabolism due to source of dietary protein part 2:**

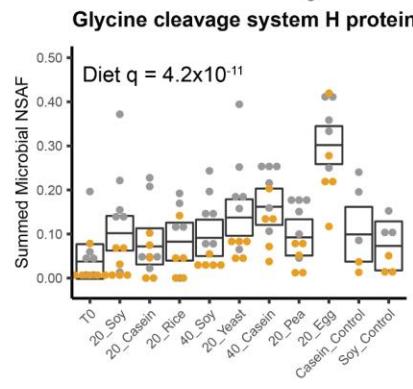
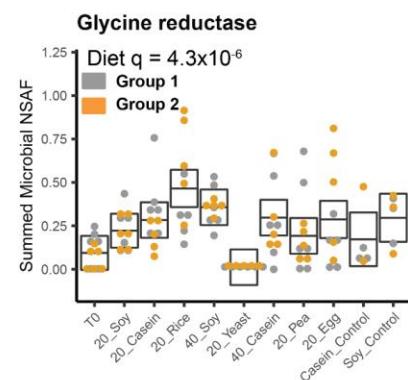
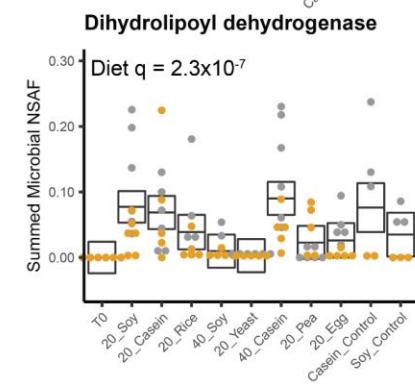
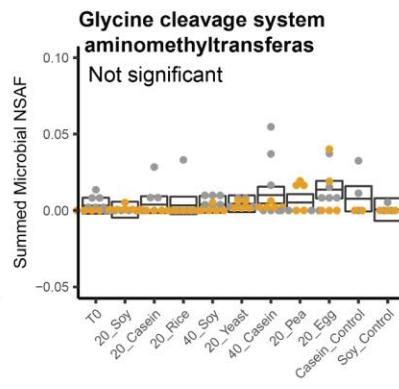
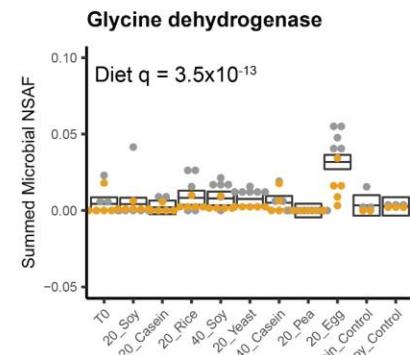
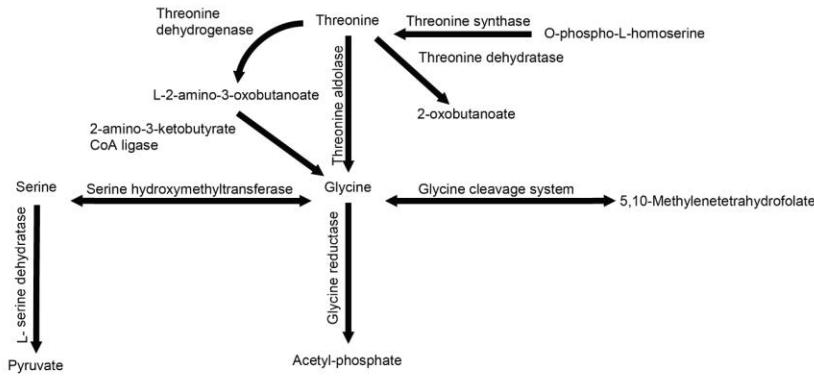
A reconstruction of the pathways involved in asparagine, aspartate, and arginine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. The boxes represent the 95% confidence interval of the mean (line) for 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).



1259  
1260

1261 **Supplementary Figure 12: Changes in threonine, glycine, and serine metabolism due to source of dietary**  
1262 **protein part 1:**

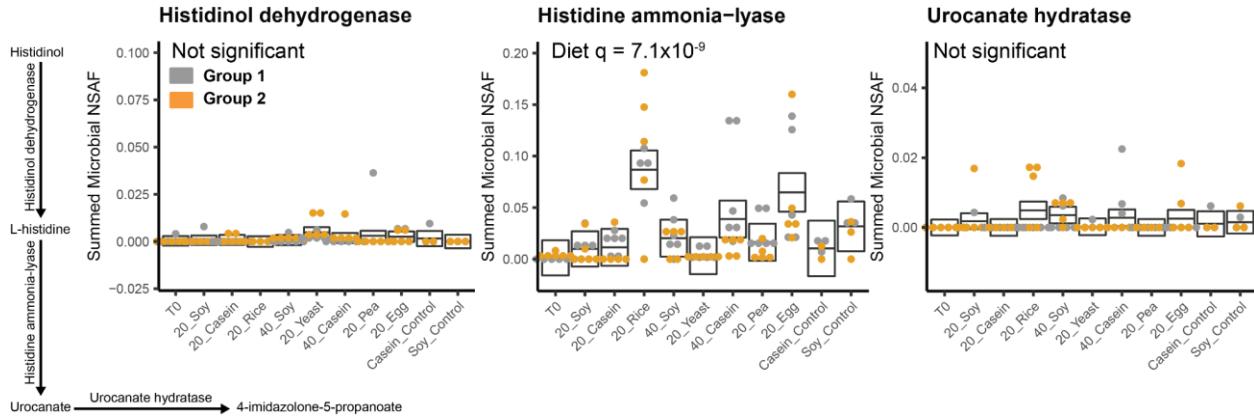
1263 A reconstruction of the pathways involved in threonine, glycine, and serine metabolism based on enzymes detected  
1264 in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the  
1265 pathway. The boxes represent the 95% confidence interval of the mean (line) for each diet from a complete mixed  
1266 effects model, and the q-values represent the FDR controlled p-values for the diet factor from an ANOVA on these  
1267 models (q<0.05 indicates significance)(Extended Data Tables 11-12).



1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275

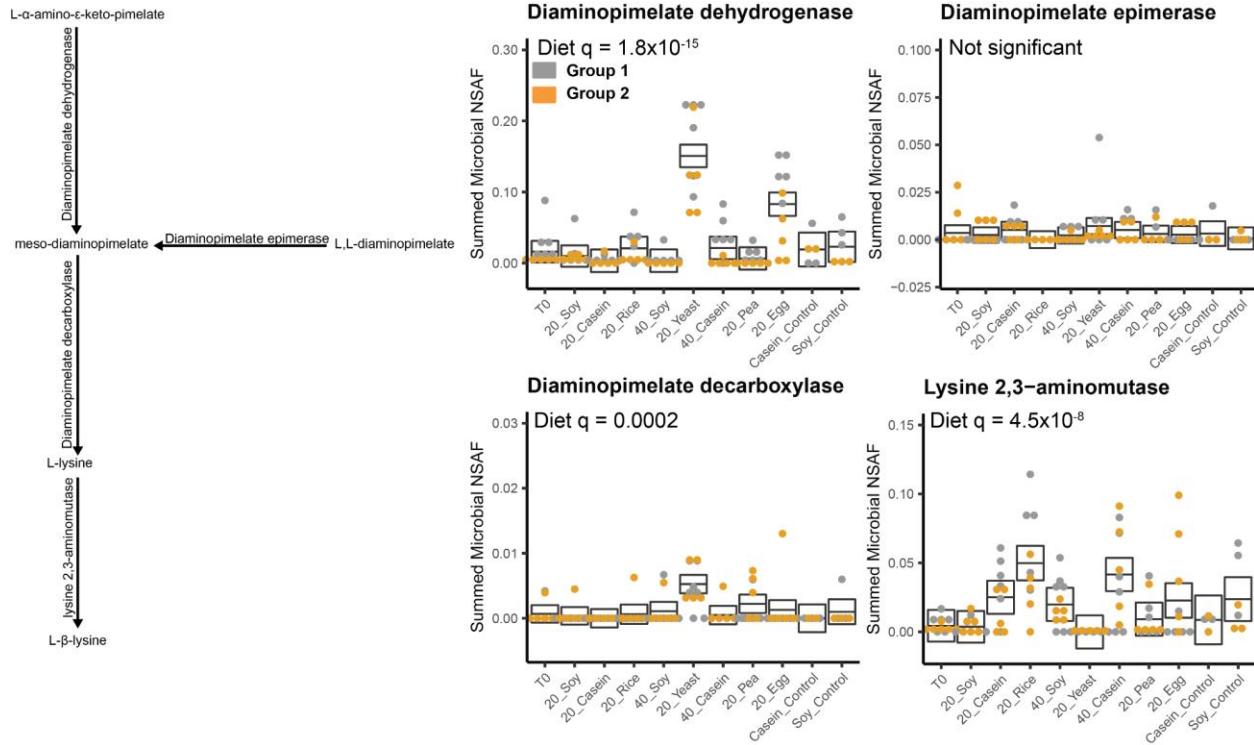
**Supplementary Figure 13: Changes in threonine, glycine, and serine metabolism due to source of dietary protein part 2:**

A reconstruction of the pathways involved in threonine, glycine, and serine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).



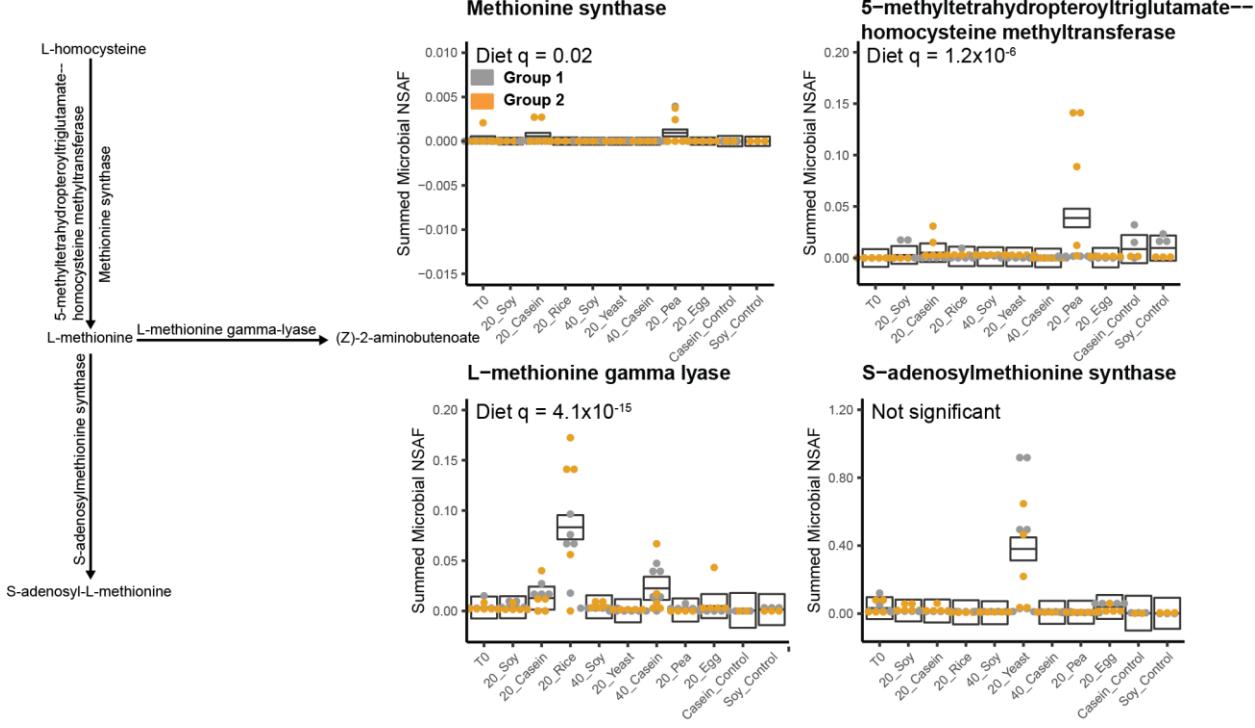
**Supplementary Figure 14: Changes in histidine metabolism due to source of dietary protein:**

A reconstruction of the pathways involved in histidine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).

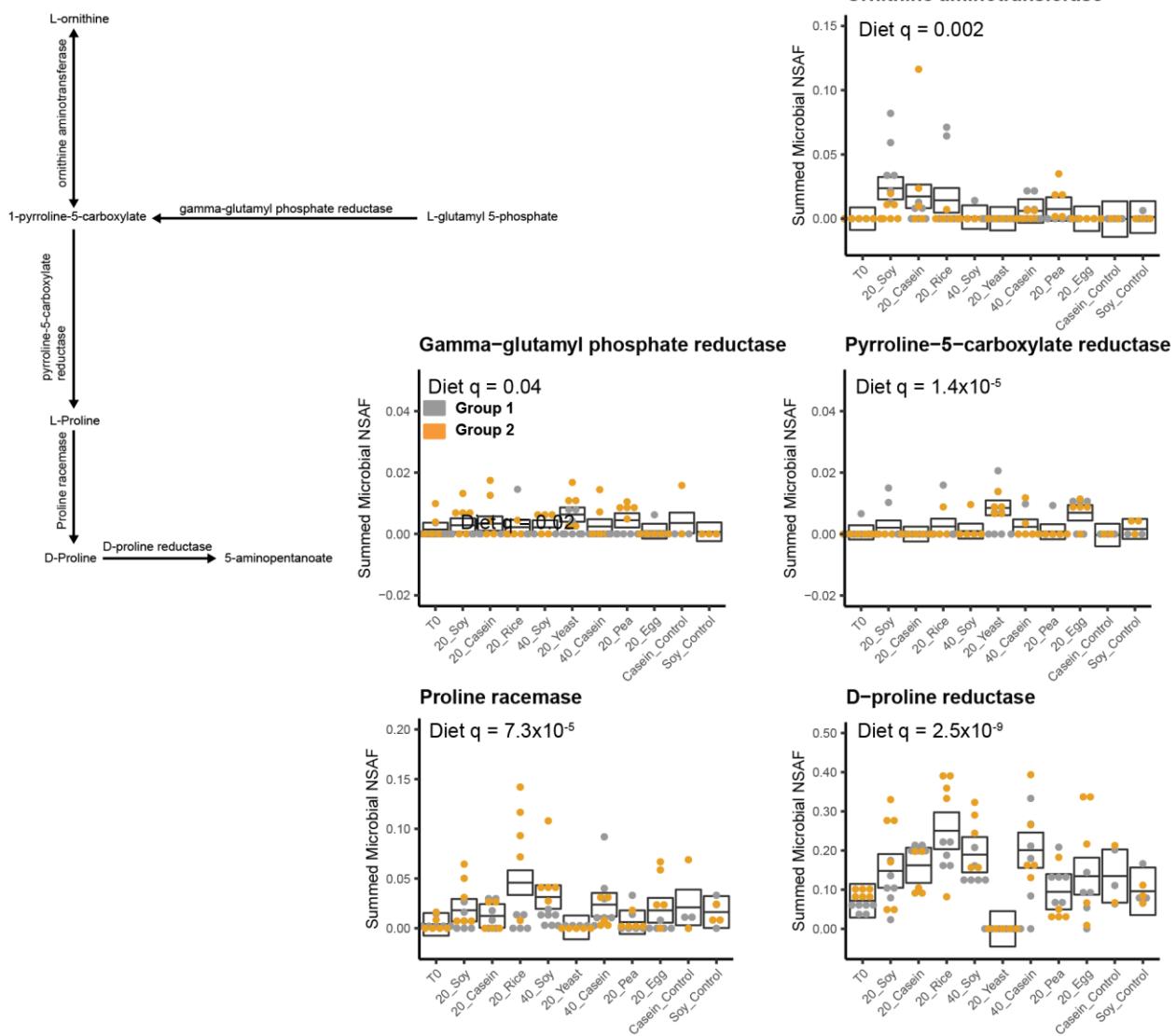


**Supplementary Figure 15: Changes in lysine metabolism due to source of dietary protein:**

A reconstruction of the pathways involved in lysine metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).



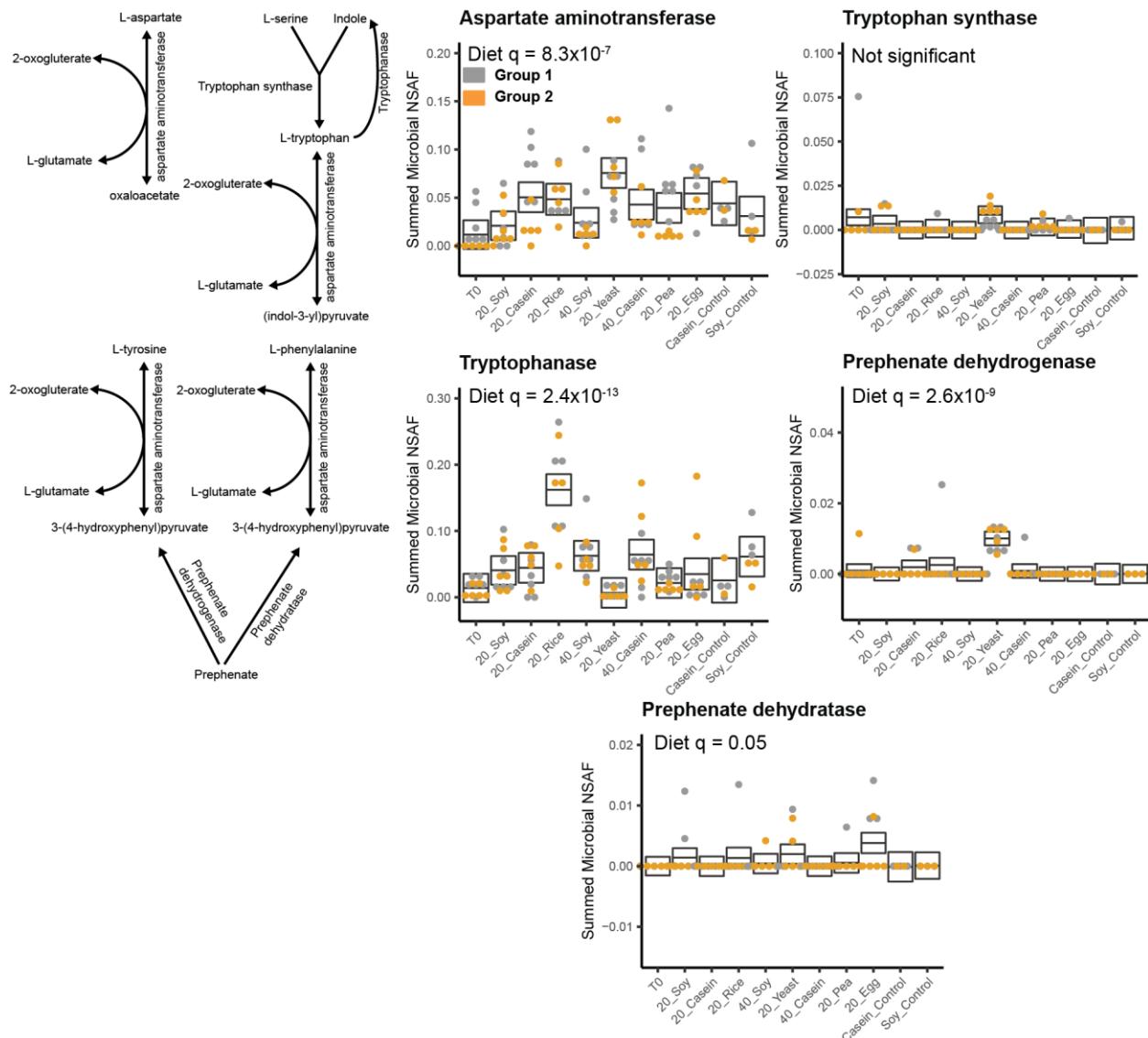
1292  
1293 **Supplementary Figure 16: Changes in methionine metabolism due to source of dietary protein:**  
1294 A reconstruction of the pathways involved in methionine metabolism based on enzymes detected in our  
1295 metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. The  
1296 boxes represent the 95% confidence interval of the mean (line) for each diet from a complete mixed effects model,  
1297 and the q-values represent the FDR controlled p-values for the diet factor from an ANOVA on these models ( $q < 0.05$   
1298 indicates significance)(Extended Data Tables 11-12).  
1299  
1300  
1301



1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311

**Supplementary Figure 17: Changes in proline metabolism due to sources of dietary protein:**

A reconstruction of the pathways involved in proline metabolism based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific enzymes involved in the pathway. 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).

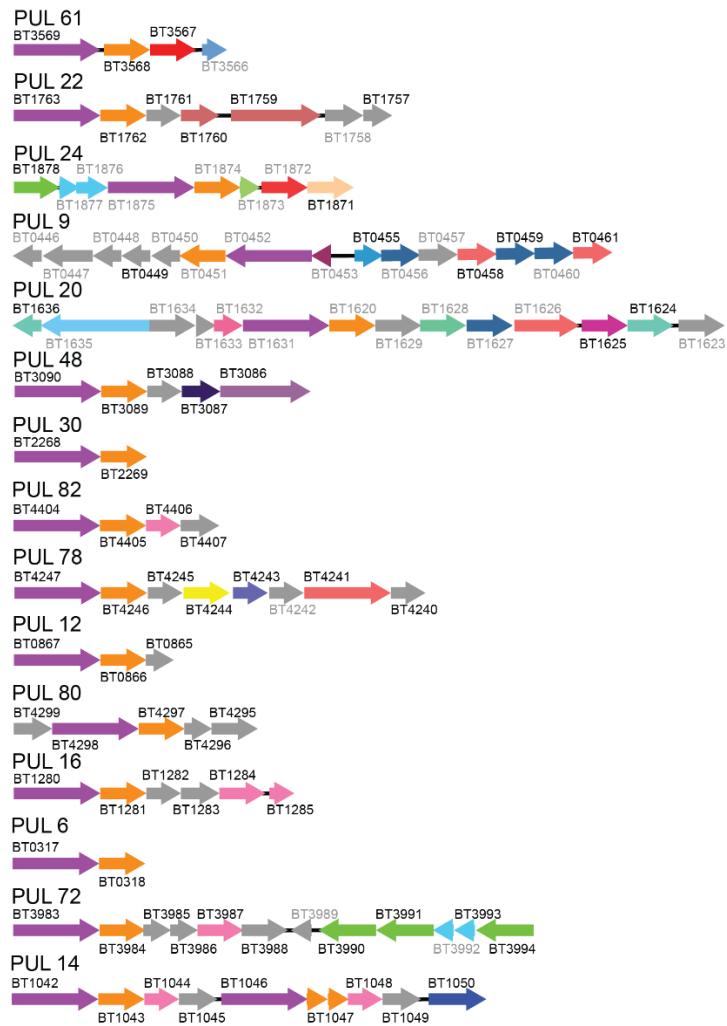


1312  
 1313 **Supplementary Figure 18: Changes in aspartate, serine, tryptophan, tyrosine, and phenylalanine metabolism**  
 1314 **due to source of dietary protein:**  
 1315 A reconstruction of the pathways involved in aspartate, serine, tryptophan, tyrosine, and phenylalanine metabolism  
 1316 based on enzymes detected in our metaproteomes. Box plots represent the aggregate abundance of the specific  
 1317 enzymes involved in the pathway. The boxes represent the 95% confidence interval of the mean (line) for each diet  
 1318 from a complete mixed effects model, and the q-values represent the FDR controlled p-values for the diet factor from  
 1319 an ANOVA on these models (q<0.05 indicates significance) (Extended Data Tables 11-12).  
 1320  
 1321  
 1322  
 1323  
 1324  
 1325

### a. Yeast associated PULs

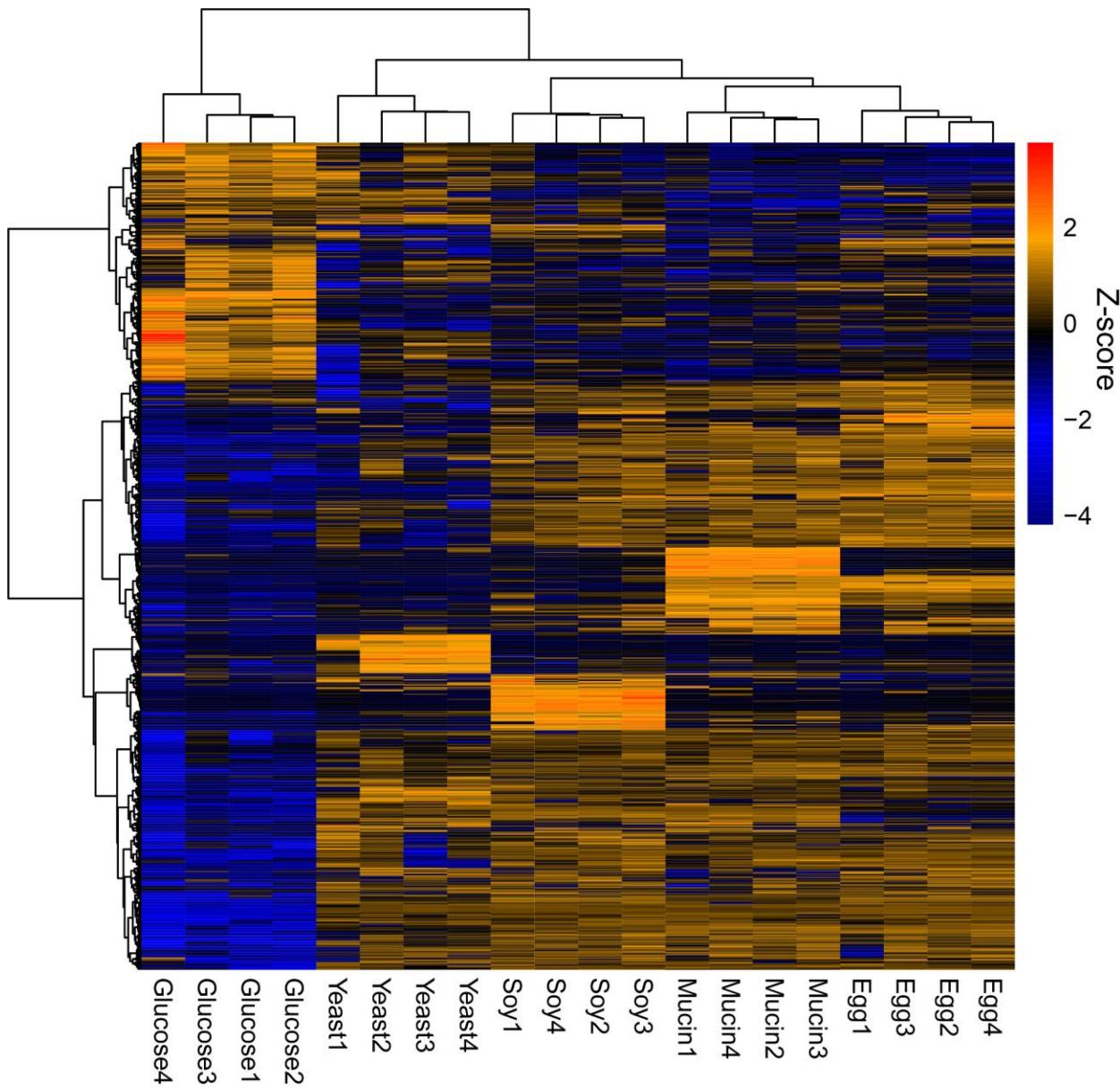


### b. Egg associated PULs



1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338

**Supplementary Figure 19: Distinct polysaccharide utilization loci (PULs) are expressed by *B. theta* between the egg white and yeast diet.** Graphical representation of PUL gene neighborhoods detected in the metaproteome. PUL identifiers are the literature derived PUL identifiers from PULDB<sup>8</sup>, but the PUL structure was verified in RAST using the *B. theta* genome assembled from our metagenome. Metagenome identifiers were cross referenced to BT numbers from previous PUL papers. If the BT number is black, it is detected in the metaproteome, if gray it is not detected in our metaproteome but detected in our genome, and if blue it means that we did not have those genes in our genome but instead detected homologs with the exact same gene neighborhood structure and similar gene percent identity. (a) PUL operons detected in our metaproteome that were increased in the yeast diet relative to the egg white diet. (c) PUL operons detected in our metaproteome that were increased in the egg white diet relative to the yeast diet.



1339  
1340  
1341  
1342  
1343

**Supplementary Figure 20: Significantly different proteins in *in vitro* proteomes of *B. theta* clustered by growth medium.** Clustered heatmap of the *in vitro* proteomes of *B. theta* after z-score standardization and removal of non-significant proteins after testing by ANOVA ( $q < 0.05$ ). We generated dendograms using the ward clustering algorithm on euclidean distances.

- 1344
- 1345 **References**
- 1346
- 1347 1. Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria
- 1348 robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J.*
- 1349 **282**, 2029–2044 (2015).
- 1350 2. Fasnacht, M. & Polacek, N. Oxidative Stress in Bacteria and the Central Dogma of Molecular
- 1351 Biology. *Front. Mol. Biosci.* **8**, (2021).
- 1352 3. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human
- 1353 gut microbiome. *Front. Genet.* **6**, (2015).
- 1354 4. Hahnke, R. L. *et al.* Genome-Based Taxonomic Classification of Bacteroidetes. *Front.*
- 1355 *Microbiol.* **7**, (2016).
- 1356 5. Cavaglieri, C. R. *et al.* Differential effects of short-chain fatty acids on proliferation and
- 1357 production of pro- and anti-inflammatory cytokines by cultured lymphocytes. *Life Sci.* **73**,
- 1358 1683–1690 (2003).
- 1359 6. Tedelind, S., Westberg, F., Kjerrulf, M. & Vidal, A. Anti-inflammatory properties of the short-
- 1360 chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel
- 1361 disease. *World J. Gastroenterol.* **13**, 2826–2832 (2007).
- 1362 7. Kleiner, M. *et al.* Assessing species biomass contributions in microbial communities via
- 1363 metaproteomics. *Nat. Commun.* **8**, 1558 (2017).
- 1364 8. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci.
- 1365 *Nucleic Acids Res.* **46**, D677–D683 (2018).
- 1366