Source of dietary protein alters the abundance of proteases, intestinal epithelial and immune proteins both directly and via interactions with the gut microbiota

Ayesha Awan[1]*, Alexandria Bartlett[1,2], J. Alfredo Blakeley-Ruiz[1], Tanner Richie[1], Amanda Ziegler[3], Manuel Kleiner[1]*

1: Department of Plant and Microbial Biology, College of Agriculture and Life Sciences, North Carolina State University, Raleigh, NC, USA
2: Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, NC, USA
3: Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA

*Correspondence should be addressed to:
Ayesha Awan - aawan2@ncsu.edu
Manuel Kleiner - manuel_kleiner@ncsu.edu

Abstract

Dietary protein has been shown to impact long-term health outcomes differentially depending on its amount and source. It has been suggested that interactions of the gut microbiota with dietary proteins mediate some of the effects of dietary protein on health outcomes. However, it remains unclear what specific host responses drive the health effects of dietary proteins from different plant and animal sources. Additionally, which specific host responses are mediated by interactions of dietary protein source with the gut microbiota and which host responses are caused by dietary proteins directly is not well understood. We used metaproteomics to quantify dietary, host, and microbial proteins in fecal samples of conventional and germ-free mice fed purified dietary protein from six different plant and animal sources, including casein, egg-white, soy, brown rice, pea, and yeast. We characterized differences in the host fecal proteome across the six dietary protein sources as well as between the conventional and germ-free mice for each source to determine how the host responds to the different dietary protein sources and the role of the gut microbiota in mediating these responses. We found that both the source of dietary protein and the presence or absence of the gut microbiota drive the host response to dietary protein source in the fecal host proteome. Host proteins pertaining to immune response, digestion, and barrier function were differentially abundant in different protein sources with and without the gut microbiota. These changes in the host response correlated with changes in microbial composition and differences in protein digestibility. Our results show how dietary protein sources, through their interactions with the gut microbiota, impact several aspects of host physiology.

Introduction

Dietary protein from different sources has been shown to affect long-term health outcomes differentially. For example, increased mortality associated with dietary proteins from animal sources, including red meat, eggs, and dairy, is significantly improved by substituting animal protein with plant protein, and animal-sourced high-protein diets worsen disease activity in colitis[1–5]. It has been suggested that these impacts on health are mediated by differences in the digestibility of dietary protein from different sources and the subsequent interaction between undigested protein residues and the gut microbiota[6]. Previous studies comparing casein and soy protein found that casein led to reduced gut microbial diversity and worsened inflammation in a mouse model of colitis[7]. Compared to casein, soy protein was associated with lower levels of secretory immunoglobulin A, IL-4, IL-13, mucin 1, and mucin 2[8]. Additionally, a study showed that mannans, branched polysaccharides associated with yeast cell wall proteins, can be metabolized by specific *Bacteroides* species in the gut, but the impact of these interactions on the host remains unknown[9]. We have previously shown that purified dietary proteins from different sources, including egg white, brown rice and yeast, are differentially degraded by the host and the gut microbiota, and strongly impact gut microbial composition and metabolism[10,11]. However, how these dynamic interactions between protein sources, host and gut microbiota impact specific aspects of host physiology and lead to cumulative effects on health outcomes is not well characterized.

Current approaches used to measure host response to diet-microbiota interactions in the gut range from molecular assays, which usually target a single protein or a preset panel, to -omics approaches such as transcriptomics and metabolomics[12]. The targeted approaches that measure a single feature or a pre-selected panel do not capture the entire complexity of host responses happening in the gastrointestinal tract. Targeted molecular assays such as ELISAs are also limited to commercially available antibodies and the standardization and sensitivity of relevant kits. Transcriptomics approaches measure mRNA abundance, which provides insight into the first step of gene expression, but is limited to analysis of gut tissue samples as mRNA is not actively secreted into the lumen and is generally unstable. Additionally, post-transcriptional regulation can lead to differences between transcript abundance levels and the actual levels of end products of gene expression (proteins and functional RNAs). Metabolomics can provide functional insights into host-microbiota interactions in the gut. However, due to the overlap in the metabolites produced by the host and the gut microbiota, differentiating host and microbial signals using this approach is difficult. Additionally, targeted molecular assays and transcriptomics can involve invasive sample collection procedures such as serum collection or tissue harvest, which make longitudinal analyses difficult or do not capture samples at the interface of the host-microbiota interactions.

The host releases proteins into the mucus layer and the intestinal lumen, which reflect its interactions with the gut microbiota and diet[13–16]. Quantifying these proteins offers a direct method to characterize the host's response to diet and microbial perturbation in the gut. Metaproteomics is a powerful approach that can simultaneously quantify the abundance of host, microbial, and dietary proteins in gut and fecal samples in an untargeted manner[17]. Thus allowing the direct characterization of the host's response to the diet itself and to the diet-mediated changes in gut microbial composition and function. Stool samples contain

hundreds of host proteins with various functions, including digestive enzymes, structural proteins, immunoglobulins, antimicrobial proteins, signaling proteins, and epithelial proteins, including cell adhesion and tight junction proteins. Therefore, fecal host proteomics provides an opportunity to determine changes in several aspects of host physiology in a single measurement[18]. Additionally, since the diet-host-microbiota interactions take place in the gut, characterizing host responses in the same matrix provides the most direct understanding of how these interactions impact host biology.

In this study, we used metaproteomics to measure the host proteome in fecal samples to characterize the impact of dietary protein from different plant, animal and microbial sources on the host. To decipher which effects on the host were directly linked to the dietary protein source and which effects were mediated by the impact of the dietary protein source on the gut microbiota, we fed defined diets with purified protein sources to male and female germ-free mice and mice with a conventional microbiota. The purified protein sources included casein, egg white, pea, brown rice, soy and yeast. Our results show that dietary protein source impacts the host directly as well as through its effects on the gut microbiota. We found source-specific effects on host proteins pertaining to digestion, intestinal epithelial integrity and immune function.

Results

More than 250 host proteins changed in abundance in response to at least one protein source, with the response of more than 150 proteins depending on the presence of a gut microbiota

To determine the overall impact of dietary protein source and the presence versus absence of the gut microbiota on the host fecal proteome, we compared the composition of host proteins detected in the fecal samples of germ-free and conventional mice fed purified dietary protein from casein, egg white, pea, brown rice, soy, and yeast. These dietary protein sources were selected based on their relevance to animal and human diets. Casein and soy are the main sources of dietary protein in the standard chow and defined diets fed to mice in lab settings[19]. Casein is also the main component of milk protein and soy is an important component of plant-based diets[20,21]. Egg white is considered to be a high-quality protein based on its amino acid composition and pea and brown rice are popular protein sources in plant-based diets and supplements[22–24]. Yeast protein is found in various fermented foods and dietary supplements. We found that the presence or absence of the gut microbiota significantly impacted the composition of the host fecal proteome across all eight diets according to a PERMANOVA analysis ($p<0.05$). PCA analysis showed that the fecal host proteomes of conventional and germ-free mice separated along the first component (Fig. 1A). Within the germ-free and conventional groups, the host proteome was significantly different across all six dietary protein sources (Fig. 1B and C). The male and female groups clustered separately only in the conventional mice fed the yeast, egg white and brown rice dietary protein (Suppl. Fig. 1B). No separation between the male and female groups was observed in any of the germ-free samples (Suppl. Fig. 1A).

We found that 277 host proteins were significantly different in at least one pairwise comparison of different dietary protein sources with pairwise comparisons done separately for germ-free and conventional mice. We found that the majority of host proteins (43%) that were significantly different between different protein sources were different in both the germ-free and conventional groups, indicating host responses to dietary protein sources that are independent of the presence or absence of a gut microbiota (Fig. 1D). 33.2% and 23.8% of all significant pairwise fecal host protein abundance comparisons were significant in only the germ-free or conventional groups, respectively. We determined the main functional categories of the differentially abundant host proteins using Gene Ontology and PANTHER protein classification. We found that the majority of proteins that were significant in both germ-free and conventional mice could be classified as proteases, defense/immunity-related proteins, structural proteins, protease inhibitors and cell adhesion molecules (Fig. 1E). Host proteins differentially abundant in only the germ-free group mostly fell in the structural, scaffold/adaptor, transporter, calcium-binding and protease categories. Host proteins differentially abundant across different diets in only the conventional group fell in the immunoglobulin, protease, defense/immunity, structural and protease inhibitor categories. Overall, we observed differences in protein categories in germ-free and conventional mice across the dietary protein sources. To explore this further, we focused on specific proteins within the categories that had the greatest number of differentially abundant proteins. These included proteases, protease inhibitors, immune proteins, and intestinal barrier proteins. The intestinal barrier proteins included proteins from several categories, including structural proteins, scaffold proteins and extracellular matrix proteins. More in-depth comparisons of proteins across dietary protein sources in these categories are presented below.

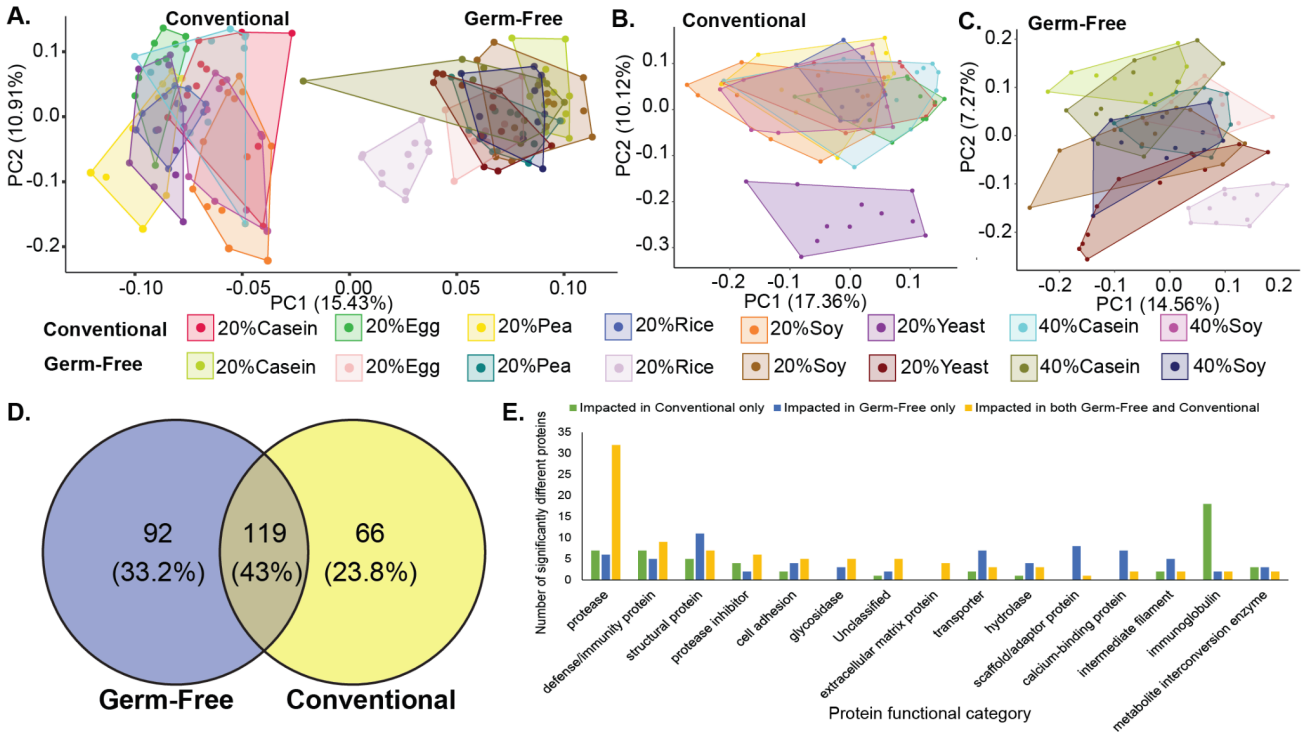

Figure 1. Dietary protein source impacts host fecal proteome in the presence and absence of the gut microbiota. A-C) PCA plots showing compositional differences between the host fecal proteomes of germ-free and conventional mice fed defined diets with a single protein source, the protein sources being 20% casein (n=11), 20% egg white (n=10), 20% pea (n=10), 20% brown rice (n=11), 20% soy (n=12), 20% yeast (n=12), 40% casein (n=10) and 40% soy (n=10). D) Venn diagram showing the number and percentage of host proteins significantly different in at least one pairwise comparison across the eight diets in the Germ-Free group only, the Conventional group only and in both the Germ-Free and Conventional groups. E) Barplot showing the number of significantly different proteins in various functional categories in the Germ-Free group only (blue), the Conventional group only (green) and in both the Germ-Free and Conventional groups (yellow). Only the top 10 categories containing the most differentially abundant proteins across the 3 groups are shown. Protein abundance data is available in Supplementary Table S1 and the underlying data with proteins in each category is available in Supplementary Table S2.

Dietary protein source impacts the abundance of host proteases and protease inhibitors differentially in the presence and absence of the gut microbiota

Since host proteases and protease inhibitors were particularly numerous among proteins differentially abundant between dietary protein sources in both the germ-free and conventional groups, we analyzed these proteins in detail. We focused on proteins that responded to multiple protein sources by identifying host proteases and protease inhibitors that were significantly enriched or depleted in a dietary protein source compared to at least 4 other dietary protein sources. We identified 12 different proteases that were significantly higher in abundance in egg white or brown rice protein groups in either germ-free or conventional mice (Fig. 2). These included serine protease 3B/trypsinogen 7 (Q9CPN9), three chymotrypsin-like elastase family members (1, Q91X79; 2A, P05208; and 3B, Q9CQ52), aminopeptidase N (P97449), carboxypeptidase A2 (Q504N0), chymotrypsinogen B (Q9CR35), trypsin 4(Q9R0T7), chymotrypsin A (Q9ER05), serine protease 3/trypsinogen 11 (Q792Z0), carboxypeptidase A1 (Q7TPZ8) and cytosolic non-specific dipeptidase (Q9D1A2). These proteases, except for the aminopeptidase and the cytosolic dipeptidase, are released into the small intestine as part of pancreatic secretions for digestion[16]. Aminopeptidase N is produced in the brush border of the small intestine and plays a role in protein digestion, while cytosolic dipeptidase is found within cells and plays a role in intracellular protein turnover[25,26].

We also found that the abundance of protease inhibitor serpin B6 was lowest in germ-free mice fed egg white protein and highest in mice fed yeast protein (Fig. 3A). This effect was not observed in the conventional mice (Fig. 3B). The abundance of serpin A3K was significantly lower in the germ-free mice fed the egg-white and brown rice protein, while in the conventional mice, it was present in the highest abundance in the yeast protein group. Serpin I2 was higher in the yeast protein group compared to all other dietary protein sources in the germ-free mice. However, in the conventional mice, serpin I2 was present in high abundance in

the pea protein group compared to the other dietary protein sources.

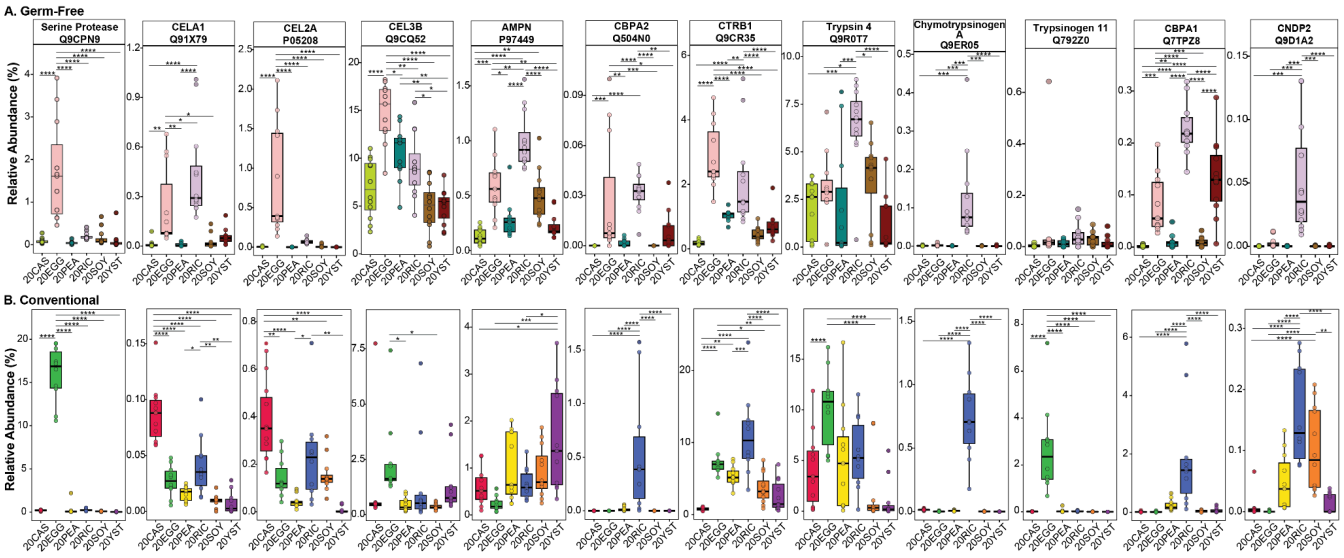

Figure 2. Dietary protein source impacts the abundance of host proteases in both conventional and germ-free mouse feces. Boxplots showing the relative abundance (%) of differentially abundant host proteases in fecal samples of A) Germ-Free and B) Conventional mice fed defined diets containing 20% protein from casein, egg white, pea, brown rice, soy and yeast. The header strip includes the protein name or gene code (where the name was too long) and its UniProt accession. Statistical significance as determined by ANOVA followed by Tukey's HSD post hoc test is indicated by horizontal lines and asterisks (✱= p<0.05, ✱✱ = p<0.01, ✱✱✱ = p<0.001, ✱✱✱✱ = p<0.0001). The complete underlying data set can be found in Supplemental Table S1.

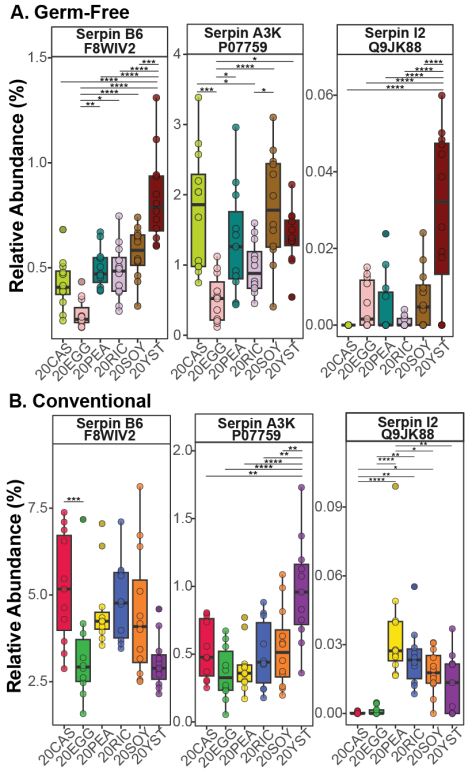

Figure 3. Dietary protein source impacts the abundance of protease inhibitors in both germ-free and conventional mouse feces. Boxplots showing the relative abundance (%) of differentially abundant host protease inhibitors in fecal samples of A) Germ-Free and B) Conventional mice fed defined diets containing 20% protein from casein, egg white, pea, brown rice, soy and yeast. The header strip includes the protein name or gene code (where the name was too long) and its UniProt accession. Statistical significance as determined by ANOVA followed by Tukey's HSD post hoc test is indicated by horizontal lines and asterisks (✱= p<0.05, ✱✱ = p<0.01, ✱✱✱ = p<0.001, ✱✱✱✱ = p<0.0001). The complete underlying data set can be found in Supplemental Table 1.

Abundances of mucus and epithelial proteins associated with the maintenance of the intestinal barrier are impacted by the dietary protein source and the presence of the gut microbiota

To determine the impact of dietary proteins from different sources on intestinal function, we identified proteins associated with the mucus and epithelial layers in the host fecal proteome and quantified changes in their abundance across the six dietary protein sources in the presence and absence of the gut microbiota. We focused on proteins that were significantly enriched or depleted in a dietary protein source compared to at least 4 other dietary protein sources in either the germ-free or conventional mice. We found that mucin 2, which is the main component of the intestinal mucus layer, was significantly higher in germ-free mice fed yeast protein compared to all the other protein sources and was lower in conventional mice fed egg white protein[27] (Fig. 4). Mucin 17, which is a transmembrane protein produced by mature epithelial cells, was significantly higher in germ-free mice fed brown rice protein and was lowest

in the group fed egg white protein[28]. This trend was absent in the conventional mice, where it was lowest in the groups fed the yeast protein. Trefoil factor 3, a goblet cell marker protein, was the most abundant in the conventional mice fed yeast protein, while in the germ-free group, yeast did not have this effect[29]. Fc gamma binding protein, secreted by goblet cells and part of the mucus layer, was significantly lower in the germ-free and conventional mice fed egg-white protein compared to other dietary protein sources[30].

The abundance of proteins associated with the development and maintenance of the intestinal epithelial barrier was also impacted by dietary protein source and the gut microbiota. Cadherin-related family proteins 2 and 5, which play a role in cell adhesion, were significantly lower in egg white and yeast protein fed groups and higher in the casein protein fed group in both germ-free and conventional mice[31] (Fig. 5). Both villin and plastin proteins, which are actin-bundling proteins that play a role in the maintenance and function of intestinal epithelial cells, were significantly higher in germ-free mice fed egg white and brown rice proteins[32,33]. However, this difference was not observed in the conventional mice. Annexin A4, A7 and A11 proteins, which are involved in cell migration, proliferation and apoptosis, were significantly higher in germ-free mice fed yeast protein[34]. This trend was not observed in the conventional mice where annexin A4 and A11 were lower in the egg-white protein fed group. Eps8-like protein 3 (epidermal growth factor receptor pathway substrate 8) protein, an actin remodeler that plays a role in the formation and function of microvilli, was significantly higher in germ-free mice fed brown rice protein[35]. However, it did not significantly change in abundance across dietary protein sources in the conventional mice.

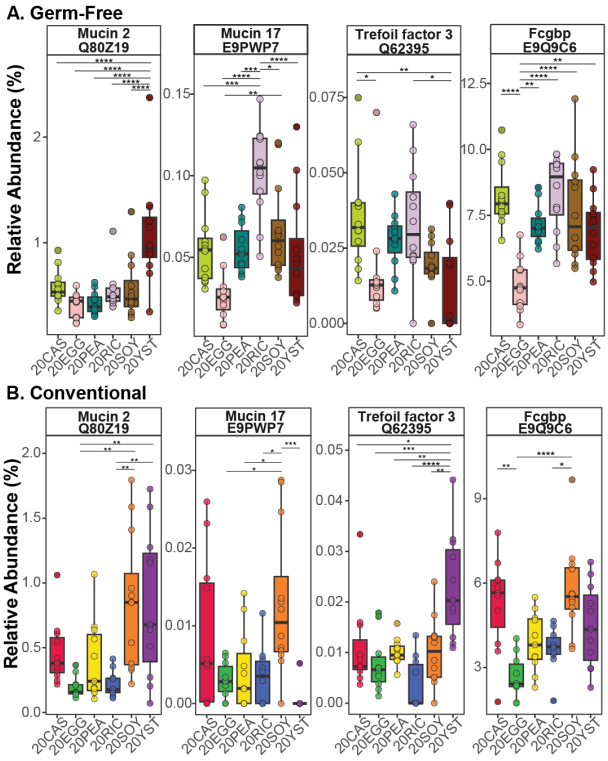

Figure 4. Mucins, goblet cell marker trefoil factor 3 and mucin-bound Fc gamma binding protein respond differently to dietary protein sources in the presence and absence of the gut microbiota. Boxplots showing the relative abundance (%) of differentially abundant host

proteins in fecal samples of A) Germ-Free and B) Conventional mice fed defined diets containing 20% protein from casein, egg white, pea, brown rice, soy and yeast. The header strip includes the protein name or gene code (where the name was too long) and its UniProt accession. Statistical significance as determined by ANOVA followed by Tukey's HSD post hoc test is indicated by horizontal lines and asterisks (✱= p<0.05, ✱✱ = p<0.01, ✱✱✱ = p<0.001, ✱✱✱✱ = p<0.0001).

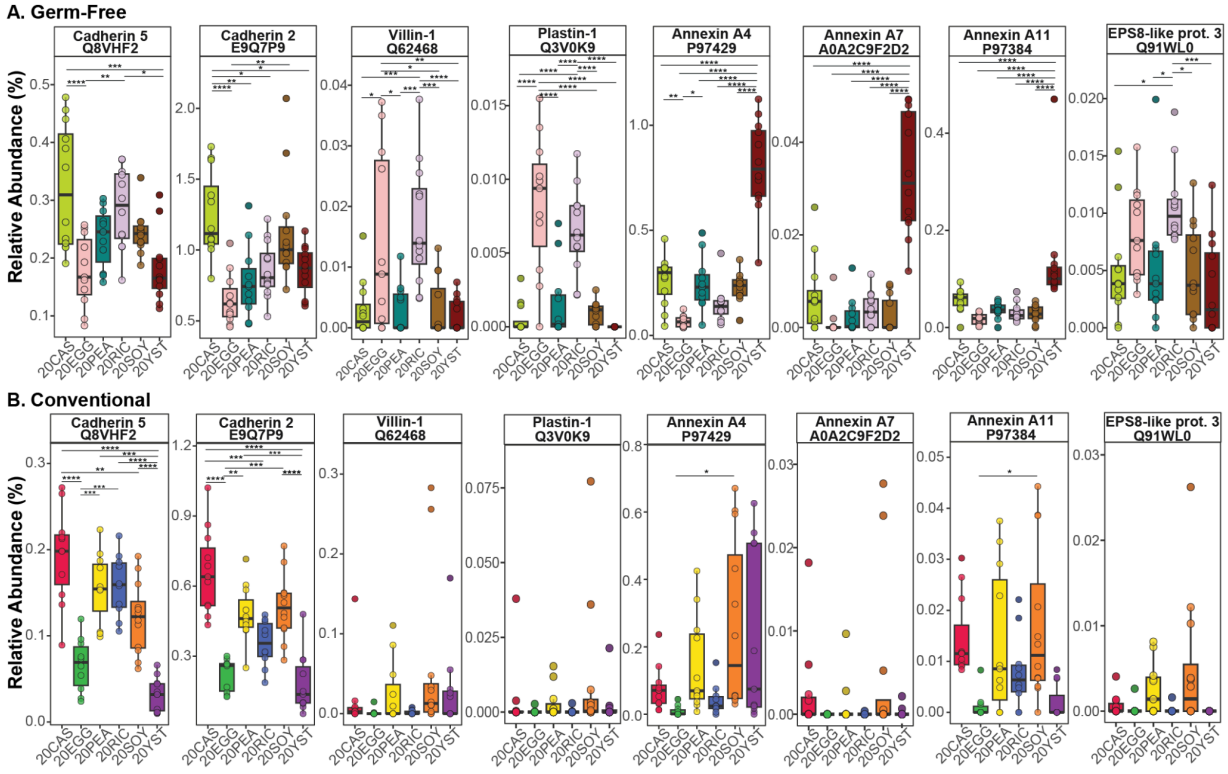

Figure 5. Proteins related to the maintenance of the intestinal epithelium change in abundance in response to different dietary protein sources and the presence or absence of the gut microbiota. Boxplots showing the relative abundance (%) of differentially abundant host proteins in fecal samples of A) Germ-Free and B) Conventional mice fed defined diets containing 20% protein from casein, egg white, pea, brown rice, soy and yeast. The header strip includes the protein name or gene code (where the name was too long) and its UniProt accession. Statistical significance as determined by ANOVA followed by Tukey's HSD post hoc test is indicated by horizontal lines and asterisks (✱= p<0.05, ✱✱ = p<0.01, ✱✱✱ = p<0.001, ✱✱✱✱ = p<0.0001).

Dietary protein source alters the abundance of antimicrobial and immune response proteins

To determine the impact of dietary protein source on the host immune response in the gut, we identified immune proteins in the fecal samples of germ-free and conventional mice that were significantly different in abundance across at least four dietary protein sources. We found that several antimicrobial or host defense proteins were impacted by changes in dietary protein source in the presence and absence of the gut microbiota. Immunoglobulin alpha chain C

protein was significantly enriched in the feces of conventional mice fed the yeast protein diet but did not change in abundance in the germ-free feces[36]. Antimicrobial proteins Reg3γ and β were significantly higher in mice fed the yeast protein compared to all other dietary protein sources in both the germ-free and conventional groups[15]. Pentraxin family member mptx2, which plays an important role in regulating gut microbiota composition and is a Paneth cell marker, was significantly higher in germ-free mice fed yeast and egg white protein compared to other protein sources[37,38]. In the conventional mice it was significantly higher in the yeast protein group compared to all other protein sources.

The calprotectin subunits S100-A8 and S100-A9, which have antimicrobial activity and are clinical markers of inflammation, were both significantly higher in the conventional mice fed yeast protein[39,40]. However, their abundance did not change significantly in the germ-free group. The abundances of antimicrobial proteins Ly6/PLAUR domain-containing protein 8 and glycoprotein 2 (zymogen granule membrane) were significantly higher in the germ-free mice fed yeast protein but did not change in the conventional group. Zymogen granule protein 16, an antimicrobial lectin-like protein, that plays a role in maintaining the intestinal barrier by limiting bacterial translocation, was significantly lower in conventional mice fed egg white protein. However, it did not change in response to different dietary protein sources in the germ-free mice. Peptidoglycan recognition protein 1 and intestinal-type alkaline phosphatase, changed in response to dietary protein source in both germ-free and conventional mice, but in different ways. Peptidoglycan recognition protein 1, which is an antimicrobial pattern recognition receptor, was significantly lower in conventional mice fed egg white protein, but in germ-free mice it was significantly lower in the yeast protein group. Intestinal-type alkaline phosphatase, which dephosphorylates lipopolysaccharide and plays a role in maintaining gut barrier function, was significantly higher in conventional mice fed casein and soy protein but was higher in germ-free mice fed yeast protein[41,42].

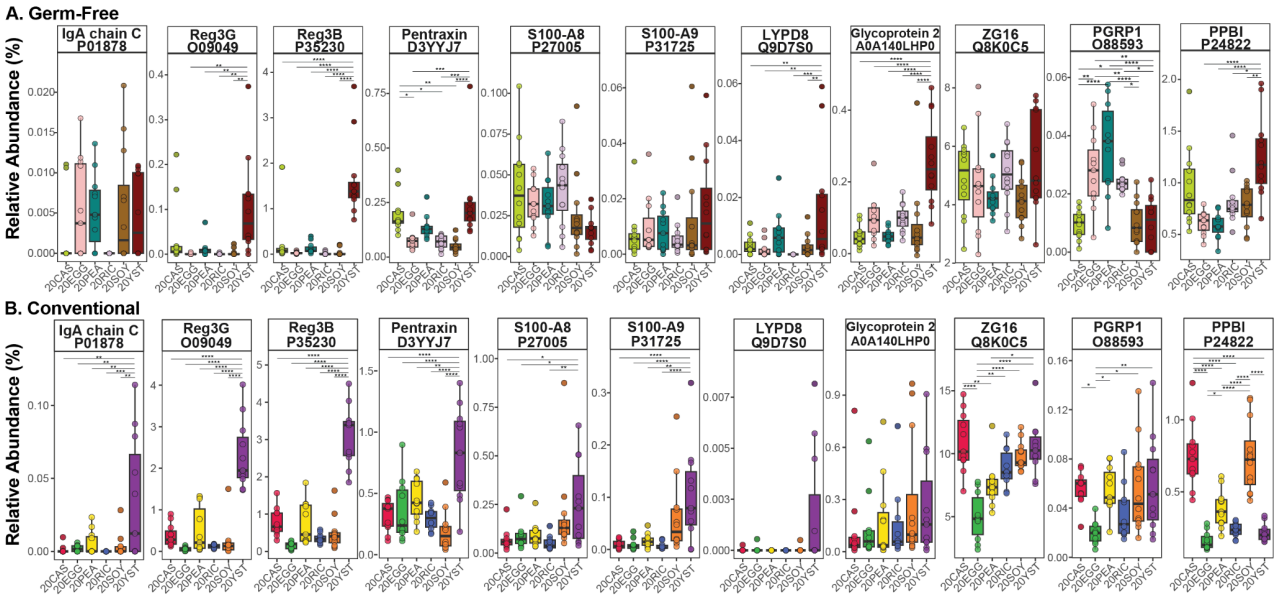

Figure 6. Immune proteins with antimicrobial activity change in abundance in response to different dietary protein sources and the presence or absence of the gut microbiota. Boxplots showing the relative abundance (%) of differentially abundant host proteins in fecal samples of A) Germ-Free and B) Conventional mice fed defined diets containing 20% protein

from casein, egg white, pea, brown rice, soy and yeast. The header strip includes the protein name or gene code (where the name was too long) and its UniProt accession. Statistical significance as determined by ANOVA followed by Tukey's HSD post hoc test is indicated by horizontal lines and asterisks (✱= p<0.05, ✱✱ = p<0.01, ✱✱✱ = p<0.001, ✱✱✱✱ = p<0.0001).

Discussion

Our goal in this study was to characterize the impact of dietary proteins from different sources on the host responses in the gut. We also aimed to decipher which effects were directly linked to diet-host interactions and which effects were mediated by the impact of dietary protein sources on the gut microbiota. Our results showed that dietary protein from different sources impacts the abundance of hundreds of host proteins, reflecting various aspects of host physiology both in the presence and absence of the gut microbiota, indicating that dietary protein source impacts the host differentially with and without interactions with the gut microbiota. We also found that male and female mice respond differently to specific dietary protein sources, including yeast, brown rice, and egg white, but this effect is only observed in the presence of the gut microbiota indicating that the different composition of the gut microbiota in the male and female mice, which we observed in a previous study, could be playing a role in mediating these differential effects.

Additionally, our results showed that egg white, brown rice, and yeast protein had the most impact on the fecal host proteome compared to other dietary protein sources in terms of significant pairwise comparisons across all six dietary protein sources. This is interesting because we previously found that egg white and brown rice proteins are less efficiently digested by the host in comparison to the other sources and are differentially degraded in the presence and absence of the gut microbiota[10]. We further established in a previous study that egg white, brown rice and yeast protein significantly impact gut microbial composition and function[11]. These previous findings, combined with our results in this study suggest that dietary proteins that are not efficiently degraded by the host and strongly influence the gut microbiota, impact the host response more than other sources of protein.

In terms of specific changes in the host fecal proteome, we found that the abundance of host proteases was significantly increased in the egg white and brown rice protein diets in both the presence and absence of the gut microbiota indicating that the presence of undigested protein residues in the gut impacts host protease production. We further found that the abundance of protease inhibitors was specifically lower in the egg white protein group. Proteases have been shown to have functions beyond digestion in the gut through their interactions with protease receptors and inhibitors[43]. These functions include extracellular matrix remodeling and maintaining gut immune homeostasis by performing proteolytic activity against chemokines and receptors. Increased host protease activity in the gut can lead to intestinal barrier disruption by upsetting the immune balance and leading to mucosal damage and has been previously linked to inflammatory bowel diseases.

Our results also showed that the abundance of host proteins associated with the maintenance and function of the intestinal epithelium changed significantly across the dietary protein sources. Egg white, brown rice and yeast protein impacted the abundance of these proteins differentially in the presence and absence of the gut microbiota, indicating that these

dietary protein sources directly, and their interactions with the gut microbiota, impact intestinal function. Additionally, we found differences in the abundance of immune and antimicrobial proteins, which play a role in maintaining intestinal barrier function and regulating host-microbial interactions in the gut in both germ-free and conventional mice. Yeast protein led to a higher abundance of the reg3 antimicrobial proteins in both the germ-free and conventional mice, while immunoglobulin A chain C protein and calprotectin only changed in response to yeast in the presence of the gut microbiota. These differences in response indicate that yeast protein impacts antimicrobial and immune proteins via direct effects on the host as well as through its effects on the gut microbiota.

Our study has several limitations that need to be addressed in future studies. The functional insights derived from changes in the abundance of host fecal proteins require further validation through complementary analyses, such as histological and tissue-specific studies, to determine whether the observed changes in protein abundance in fecal samples are reflective of differences in tissue morphology. For example, future studies should investigate whether changes in the abundance of mucus proteins in the feces correlate to changes in the thickness of the mucus layer. Moreover, our study does not distinguish host responses specific to different regions of the gastrointestinal tract.

Methods

Animals and housing

Animal handling and housing protocol as described in Bartlett et al., 2023. 12 germ-free (6 male, 6 female, NCSU gnotobiotic core) and 12 conventionally raised (6 male, 6 female, Jackson Labs, Bar Harbor) C57BL/J6 mice, aged 3-6 months were used for this study. The germ-free mice were housed in gnotobiotic isolators and were monitored for sterility. The male and female mice were housed separately in groups of three. The animals were kept at a 12-hour dark/light cycle at an average temperature of 70^{o}C and 35% humidity. The bedding was autoclaved and sterile water (Gibco) was used for both germ-free and conventional mice. The diets for both germ-free and conventional mice were from the same exact batch. The diets were gamma-irradiated and vacuum packaged. The outside of the vacuum sealed diet packages was further sterilized using peracetic acid for 30 minutes before introduction into germ-free mouse isolators to avoid autoclaving, which would have introduced differences between diets fed to germ-free and conventional mice. The animals were fed, weighed, and assessed daily by trained animal handlers and all cage changes were performed in a laminar flow hood for conventional mice. All animal experiments were conducted in the Association for the Assessment and Accreditation of Laboratory Animal Care International (AAALAC) accredited Laboratory Animal Facilities at the NCSU College of Veterinary Medicine. The animal care protocol was approved by NC State's Institutional Animal Care and Use Committee (Protocol # 18-034-B for conventionally raised mice and # 18-165-B for germ-free mice).

Animal diets and sample collection

Defined diets used in the study were obtained from Inotiv (previously Envigo Teklad). Components other than the source of the dietary protein did not change and were consistent and defined across the different diets. Diets were sterilized using gamma-irradiation and were vacuum packaged and sterilized on the outside before being introduced into the gnotobiotic isolators. The diets contained purified protein from a single source and were not supplemented with amino acids. Diets containing 20% soy, 20% casein, 20% brown rice, 40% soy, 20% yeast, 40% casein, 20% pea, and 20% egg white were fed to each mouse *ad libitum* for 7 days per diet in this sequence. At the end of the dietary sequence mice were divided into two groups and one group was fed the 20% casein and the other was fed the 20% soy diet again as a control to see if the effects observed in the beginning of the feeding sequence were reproducible. After feeding the defined diet for a week, fecal samples were collected from each mouse before switching to the next diet. All mice were handled following the approved IACUC protocol. NAP preservation solution (935 ml autoclaved MilliQ water, 700 g Ammonium Sulfate, 25 ml of 1 M Sodium Citrate and 40 ml of 0.5 M EDTA adjusted to pH 5.2 using 1 M H_2SO_4) [44] was used to preserve and store the fecal samples at a 1:10, sample-to-solution ratio. This method prevented sample modification, for example by proteases. Samples from germ-free mice and conventionally raised mice were treated identically to maintain consistency. Samples were roughly homogenized using a sterilized disposable pestle and then stored frozen at -80°C.

Protein extraction from fecal samples and peptide preparation

To extract proteins from the fecal samples, we first removed the NAP solution by centrifugation (21,000 x g, 5 min) and then placed the fecal pellets in Lysing Matrix E tubes (MP Biomedicals). The samples were then bead beat for 5 cycles of 45 s at 6.45 m/s with 1 min between cycles in SDT lysis buffer (4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT). Samples were heated to 95°C for 10 minutes and the resulting lysates were centrifuged (21,000 x g, 5 min) to remove the lysing beads and any other debris. To prepare peptides from the lysate, we followed the filter aided sample preparation (FASP) protocol. 60 µl of the lysate was mixed with 400 µL of UA solution (8 M urea in 0.1 M Tris/HCl pH 8.5). The mixture was loaded onto a 10 kDa 500 µL filter unit (VWR International) and centrifuged at 14,000 x g for 30 minutes. This step was repeated up to a total of 3 times to load the filter to capacity. The filters were washed using 200 µl of UA solution and centrifuged at 14,000 g for 40 min. The filters were then incubated with 100 µl IAA (0.05 M iodoacetamide in UA solution) for 20 min and centrifuged at 14,000 *g* for 20 min. Filters were washed with 100 uL of UA buffer 3 times, followed by 3 washes with 100 µl of ABC (50 mM Ammonium Bicarbonate) buffer to remove the UA buffer. 0.95 µg of MS grade trypsin (Thermo Scientific Pierce, Rockford, IL, USA) solubilized in 40 µl of ABC buffer was then added to the filters and incubated in a wet chamber at 37 °C to digest the proteins into peptides. Peptides were then eluted off the filter after 16 hours by centrifugation at 14,000 x g for 20 minutes. 50 uL of 0.5 M NaCl was added to the filters, followed by centrifugation at 14,000 x g for 20 minutes. The concentration of the resulting peptides was determined using the Pierce Micro BCA assay (Thermo Scientific Pierce) following the manufacturer's instructions.

LC-MS/MS analysis of peptides fecal samples

Peptides from diet and fecal samples were analyzed using 1D-LC-MS/MS, as previously described in Bartlett et al., 2023. 600 ng of peptides from each of the mouse fecal samples was loaded onto a 5 mm, 300 µm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific) with loading solvent A (2% acetonitrile, 0.05% TFA) in an UltiMate™ 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific). Peptides were separated on an EASY-Spray analytical column heated to 60°C (PepMap RSLC C18, 2 µm material, 75 cm× 75 µm, Thermo Fisher Scientific) using a 140 min gradient at a flow rate of 300 nl/min. The first 102 minutes of the gradient went from 95% eluent A (0.1% formic acid) to 31% eluent B (0.1% formic acid, 80% acetonitrile), then 18 min from 31 to 50% B, and the last 20 min at 99% B. A 100% acetonitrile wash run was performed between each sample to minimize carryover. The eluted peptides were ionized using an Easy-Spray source and analyzed in a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) with the following parameters: m/z 445.12003 lock mass, normalized collision energy equal to 24, 25 s dynamic exclusion, and exclusion of ions of +1 charge state. Full MS scans were acquired for 380 to 1600 m/z at a resolution of 60,000 and a max IT time of 200 ms and data-dependent MS^2 was performed for the 15 most abundant ions at a resolution of 15,000 and max IT of 100 ms.

Database construction and protein identification

The protein sequence database used to identify proteins consisted of dietary, host, and microbial proteins, as previously described in Bartlett et al., 2023. Custom databases specific to each diet/sample type were used in line with the literature guidelines. For the host component of the database, the reference proteome for the C57BL/6J mouse (*Mus musculus*, UP000000589) was downloaded from UniProt and added to the dietary and microbial components. The MS^2 spectra were searched using the sample-specific database for each diet group using run calibration and the Sequest HT node in Proteome Discoverer software version 2.3 (Thermo Fisher Scientific) with the following settings: trypsin (Full), maximum 2 missed cleavages, 10 ppm precursor mass tolerance, 0.1 Da fragment mass tolerance and maximum 3 equal dynamic modifications per peptide. Dynamic modifications, including oxidation on M (+15.995 Da), deamidation on N, Q, R (0.984 Da), and acetyl on the protein N terminus (+42.011 Da), and the static modification carbamidomethyl on C (+57.021 Da) were used. Proteins were quantified based on the area under the curve method in the Minora Feature Detector node with the following parameters: 5 minimum number of non-zero points in a chromatographic trace, 0.2 min maximum retention time of isotope pattern multiplets, and high PSM confidence. The percolator node in Proteome Discoverer was used to calculate peptide false discovery rate (FDR) with the following parameters: maximum Delta Cn 0.05, a strict target FDR of 0.01, a relaxed target FDR of 0.05, and validation based on q-value. The precursor ion quantifier node was used for quantification using the following settings: unique and razor peptides used for quantification and precursor abundance based on area. Only master proteins using the 5% FDR cutoff were included in the downstream data analysis.

Data processing, statistical analysis, and data visualization

Host proteins were filtered from the whole fecal metaproteome, and the relative abundance of host proteins present in at least 70% of all the replicates was normalized using total sum scaling. For the principal coordinate analysis and the PERMANOVA in Fig1 A, B, and C, we transformed the host protein data using a centered log-ratio transform using the compositions package (version 2.0.5) in R. ANOVA followed by the Tukey's HSD post hoc test (rstatix package in R) was used to determine differentially abundant proteins across the different groups. Protein functions were annotated and classified using UniProt, Panther, GeneOntology and an extensive literature review. R (R version 4.2.2) and Excel were used for all data processing and analysis. R (ggplot2, version 3.5.1) and Adobe Illustrator were used for data visualization.

Ethics statement

NC State's Institutional Animal Care and Use Committee approved the animal care protocol (Protocol # 18-034-B for conventionally raised mice and # 18-165-B for germ-free mice).

Data availability

Proteomics data, including raw mass-spectrometry files and databases used, were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD041586 (https://www.ebi.ac.uk/pride/archive/projects/PXD041586).

Supplemental materials

Supplemental Table S1: Table showing total sum scale normalized abundance of host proteins in all fecal samples identified in at least 70% of the replicates and filtered for master proteins at 5% FDR.
Supplemental Table S2: Table showing the protein functional categories in Fig. 1E, determined using panther protein class and gene ontology.

Declaration of interests

The authors declare no competing interests.

Author contributions

AA: Data processing, data analysis, writing the manuscript, editing
AB: Conceptualization of the study, experimental design, data collection, editing
JABR: Data processing, editing
TR: Data collection
AZ: Data analysis, editing
MK: Funding, conceptualization of the study, experimental design, data collection, data processing, data analysis, writing, editing

Acknowledgments

All LC-MS/MS measurements were made in the Molecular Education, Technology, and Research Innovation Center (METRIC) at North Carolina State University. The Gnotobiotic Core at the College of Veterinary Medicine, North Carolina State University is supported by the National Institutes of Health funded Center for Gastrointestinal Biology and Disease, NIH-NIDDK P30 DK034987. This work was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R35GM138362.

References

1. Chen, Z. *et al.* Dietary protein intake and all-cause and cause-specific mortality: results from the Rotterdam Study and a meta-analysis of prospective cohort studies. *Eur. J. Epidemiol.* **35**, 411–429 (2020).
2. Song, M. *et al.* Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality. *JAMA Intern. Med.* **176**, 1453 (2016).
3. Sun, Y. *et al.* Association of Major Dietary Protein Sources With All‐Cause and Cause‐Specific Mortality: Prospective Cohort Study. *J. Am. Heart Assoc.* **10**, e015553 (2021).
4. Budhathoki, S. *et al.* Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality in a Japanese Cohort. *JAMA Intern. Med.* **179**, 1509 (2019).
5. Jantchou, P., Morois, S., Clavel-Chapelon, F., Boutron-Ruault, M.-C. & Carbonnel, F. Animal Protein Intake and Risk of Inflammatory Bowel Disease: The E3N Prospective Study. *Am. J. Gastroenterol.* **105**, 2195–2201 (2010).
6. O'Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 691–706 (2016).
7. Kostovcikova, K. *et al.* Diet Rich in Animal Protein Promotes Pro-inflammatory Macrophage Response and Exacerbates Colitis in Mice. *Front. Immunol.* **10**, 919 (2019).
8. Zeng, B. *et al.* Dietary Soy Protein Isolate Attenuates Intestinal Immunoglobulin and Mucin Expression in Young Mice Compared with Casein. *Nutrients* **12**, 2739 (2020).

9. Abbott, D. W., Martens, E. C., Gilbert, H. J., Cuskin, F. & Lowe, E. C. Coevolution of yeast mannan digestion: Convergence of the civilized human diet, distal gut microbiome, and host immunity. *Gut Microbes* **6**, 334–339 (2015).
10. Awan, A. *et al.* Dietary protein from different sources escapes host digestion and is differentially modified by the microbiota. Preprint at https://doi.org/10.1101/2024.06.26.600830 (2024).
11. Blakeley-Ruiz, J. A. *et al.* Dietary protein source strongly alters gut microbiota composition and function. Preprint at https://doi.org/10.1101/2024.04.04.588169 (2024).
12. Lichtman, J. S., Sonnenburg, J. L. & Elias, J. E. Monitoring host responses to the gut microbiota. *ISME J.* **9**, 1908–1915 (2015).
13. Bevins, C. L., Martin-Porter, E. & Ganz, T. Defensins and innate host defence of the gastrointestinal tract. *Gut* **45**, 911–915 (1999).
14. Salvo Romero, E., Alonso Cotoner, C., Pardo Camacho, C., Casado Bedmar, M. & Vicario, M. The intestinal barrier function and its involvement in digestive disease. *Rev. Esp. Enfermedades Dig.* **108**, (2015).
15. Mukherjee, S. & Hooper, L. V. Antimicrobial Defense of the Intestine. *Immunity* **42**, 28–39 (2015).
16. Whitcomb, D. C. & Lowe, M. E. Human Pancreatic Digestive Enzymes. *Dig. Dis. Sci.* **52**, 1–17 (2007).
17. Salvato, F., Hettich, R. L. & Kleiner, M. Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLOS Pathog.* **17**, e1009245 (2021).
18. Lichtman, J. S., Marcobal, A., Sonnenburg, J. L. & Elias, J. E. Host-centric Proteomics of Stool: A Novel Strategy Focused on intestinal Responses to the Gut Microbiota. *Mol. Cell. Proteomics* **12**, 3310–3318 (2013).
19. Pellizzon, M. Choice of laboratory animal diet influences intestinal health. *Lab Anim.* **45**,

238–239 (2016).

20. Day, L., Cakebread, J. A. & Loveday, S. M. Food proteins from animals and plants: Differences in the nutritional and functional properties. *Trends Food Sci. Technol.* **119**, 428–442 (2022).
21. Hoffman, J. R. & Falvo, M. J. Protein - Which is Best? *J. Sports Sci. Med.* **3**, 118–130 (2004).
22. Hoogenkamp, H., Kumagai, H. & Wanasundara, J. P. D. Rice Protein and Rice Protein Products. in *Sustainable Protein Sources* 47–65 (Elsevier, 2017). doi:10.1016/B978-0-12-802778-3.00003-2.
23. Liu, N. *et al.* Pea albumin extracted from pea (Pisum sativum L.) seed protects mice from high fat diet-induced obesity by modulating lipid metabolism and gut microbiota. *J. Funct. Foods* **97**, 105234 (2022).
24. Puglisi, M. J. & Fernandez, M. L. The Health Benefits of Egg Protein. *Nutrients* **14**, 2904 (2022).
25. Olsen, J., Kokholm, K., Norén, O. & Sjöström, H. Structure and Expression of Aminopeptidase N. in *Cellular Peptidases in Immune Functions and Diseases* (eds. Ansorge, S. & Langner, J.) vol. 421 47–57 (Springer US, Boston, MA, 1997).
26. Bauer, K. Cytosol nonspecific dipeptidase. in *Handbook of Proteolytic Enzymes* 1020–1022 (Elsevier, 2004). doi:10.1016/B978-0-12-079611-3.50324-4.
27. Bansil, R. & Turner, B. S. The biology of mucus: Composition, synthesis and organization. *Adv. Drug Deliv. Rev.* **124**, 3–15 (2018).
28. Gum, J. R., Crawley, S. C., Hicks, J. W., Szymkowski, D. E. & Kim, Y. S. MUC17, a Novel Membrane-Tethered Mucin. *Biochem. Biophys. Res. Commun.* **291**, 466–475 (2002).
29. Kjellev, S. The trefoil factor family – small peptides with multiple functionalities. *Cell. Mol. Life Sci.* **66**, 1350–1369 (2009).
30. Gustafsson, J. K. & Johansson, M. E. V. The role of goblet cells and mucus in intestinal

homeostasis. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 785–803 (2022).

31. Crawley, S. W. *et al.* Intestinal Brush Border Assembly Driven by Protocadherin-Based Intermicrovillar Adhesion. *Cell* **157**, 433–446 (2014).
32. Khurana, S. Structure and Function of Villin. in *Advances in Molecular and Cell Biology* vol. 37 89–117 (Elsevier, 2006).
33. Shinomiya, H. Plastin Family of Actin-Bundling Proteins: Its Functions in Leukocytes, Neurons, Intestines, and Cancer. *Int. J. Cell Biol.* **2012**, 1–8 (2012).
34. Xi, Y., Ju, R. & Wang, Y. Roles of Annexin A protein family in autophagy regulation and therapy. *Biomed. Pharmacother.* **130**, 110591 (2020).
35. Zwaenepoel, I. *et al.* Ezrin regulates microvillus morphogenesis by promoting distinct activities of Eps8 proteins. *Mol. Biol. Cell* **23**, 1080–1095 (2012).
36. Raya Tonetti, F., Eguileor, A. & Llorente, C. Goblet cells: guardians of gut immunity and their role in gastrointestinal diseases. *eGastroenterology* **2**, e100098 (2024).
37. Yan, W. *et al.* Loss of Mptx2 alters bacteria composition and intestinal homeostasis potentially by impairing autophagy. *Commun. Biol.* **7**, 94 (2024).
38. Capdevila, C. *et al.* Cellular origins and lineage relationships of the intestinal epithelium. *Am. J. Physiol.-Gastrointest. Liver Physiol.* **321**, G413–G425 (2021).
39. Jukic, A., Bakiri, L., Wagner, E. F., Tilg, H. & Adolph, T. E. Calprotectin: from biomarker to biological function. *Gut* **70**, 1978–1988 (2021).
40. Aranda, C. J. *et al.* Calprotectin protects against experimental colonic inflammation in mice. *Br. J. Pharmacol.* **175**, 3797–3812 (2018).
41. Santos, G. M. *et al.* Intestinal Alkaline Phosphatase: A Review of This Enzyme Role in the Intestinal Barrier Function. *Microorganisms* **10**, 746 (2022).
42. Singh, S. B. & Lin, H. C. Role of Intestinal Alkaline Phosphatase in Innate Immunity. *Biomolecules* **11**, 1784 (2021).
43. Antalis, T. M., Shea-Donohue, T., Vogel, S. N., Sears, C. & Fasano, A. Mechanisms of

Disease: protease functions in intestinal mucosal pathobiology. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **4**, 393–402 (2007).

44. Mordant, A. & Kleiner, M. Evaluation of Sample Preservation and Storage Methods for Metaproteomics Analysis of Intestinal Microbiomes. *Microbiol. Spectr.* **9**, e01877-21 (2021).