

ABSTRACT

MORDANT, ANGIE LISA. Linking Intestinal Microorganisms to Their Diet-Derived Growth Substrates. (Under the direction of Dr. Manuel Kleiner).

The intestinal microbiota is a diverse and metabolically active community that has profound effects on its host. This complex community influences the health of its host by altering the availability of nutrients and the host's susceptibility to infection and disease. Diet is a key factor affecting the microbiota, and diet-microbiota interactions have been implicated in many diseases. Therefore, there is a growing interest in deliberately modulating the microbiota through diet to improve the host's health and mitigate the risk of diseases. Dietary components are thought to be the primary growth substrates of intestinal microorganisms. However, our current understanding of links between specific dietary components and their microbial metabolizers is primarily based on indirect evidence. Our study used metaproteomics and protein stable isotope fingerprinting (Protein-SIF) to acquire direct evidence of utilization of dietary components by intestinal bacteria. We conducted studies in gnotobiotic mice colonized with a community of human-derived bacterial strains. The mice were fed defined diets varying in fiber, protein, and fat source. We found that bacteria likely used a combination of growth substrates, and substrate choice appeared to be driven by competition. We demonstrated the potential of the protein-SIF approach in associating intestinal bacteria with the carbohydrates, proteins, and host-derived compounds they metabolize; and how the approach provides a unique angle to study the interactions between diet and intestinal microbiota.

© Copyright 2021 by Angie Mordant

All Rights Reserved

Linking Intestinal Microorganisms to Their Diet-Derived Growth Substrates

by
Angie Lisa Mordant

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Microbiology

Raleigh, North Carolina
2021

APPROVED BY:

Manuel Kleiner
Committee Chair

Casey Theriot

Amy Grunden

Benjamin Callahan

DEDICATION

To my parents and brother for all their support.

BIOGRAPHY

Angie Mordant obtained a B.S. in Marine Biology in 2017 from the University of West Florida. While at UWF, she was a member of Dr. Jim Spain's lab where she researched bacterial degradation of soil contaminants. After graduation, she continued her research until she moved to Raleigh, NC, in 2018 to pursue a graduate degree in Microbiology at North Carolina State University and joined the lab of Dr. Manuel Kleiner.

ACKNOWLEDGMENTS

I am sincerely grateful to all the following individuals.

Dr. Manuel Kleiner for providing excellent training and mentorship as my advisor, for being supportive, and for always challenging me to accomplish more.

Dr. Casey Theriot for serving on my committee, and providing me with scientific and career advice.

Dr. Amy Grunden for serving on my committee, inspiring me in her class, and providing insight on anaerobic cultivation of microbes.

Dr. Benjamin Callahan for serving on my committee and providing insight on data analyses.

Dr. Fernanda Salvato for providing me with extensive metaproteomics and LC-MS/MS training, and for being a wonderful teammate and friend.

Dr. Susan Tonkonogy for providing insight on gnotobiotics.

Karen Flores for all the work with the gnotobiotic mice and the very numerous sample collections.

Abigail Korenek for all her assistance in the lab, and for being a fantastic trainee.

Deniz Durmusoglu and Dr. Nathan Crook for helping me with the anaerobic cultivation of bacterial strains.

Ibrahim Al'Abri for the lengthy sample collections.

Alexandria Bartlett, Clara Tang, and Marlene Jensen for being wonderful teammates and friends, for always being willing to help me in the lab, and for the many fruitful discussions.

Mariska Thayagan and Tanner Russ for providing assistance in the lab.

All Kleiner Lab members for discussing experiments and providing helpful feedback during group meetings, and for making the lab an enjoyable work environment.

Hannah Wapshott for showing me on how to sparge serum bottles with anaerobic gas.

Alissa Rivera for providing me training on mouse necropsy.

Dr. Alfredo Blakeley-Ruiz and Dr. Michael McLaren for providing insight on data analyses.

The many professors of NC State for inspiring me in their classes.

Dr. Michael Hyman for providing career advice.

Dr. Jim Spain, for continuing to provide mentorship beyond my time at UWF.

My family, my boyfriend Mitchell, my friends (particularly Haley and Amelia), and my cat Batman for providing ample moral support during my graduate training at NC State.

TABLE OF CONTENTS

LIST OF TABLES	viii	
LIST OF FIGURES	ix	
CHAPTER 1: IN VIVO GROWTH SUBSTRATES OF INTESTINAL		
MICROORGANISMS		1
1.1 Introduction.....	1	
1.2 What are the nutrients available to the intestinal microbiota?	2	
1.3 Methods to investigate the in vivo growth substrates of intestinal microorganisms	5	
1.4 Established and hypothesized associations between intestinal microorganisms and growth-substrates	7	
REFERENCES	10	
CHAPTER 2: METAPROTEOMICS AND PROTEIN STABLE ISOTOPE		
FINGERPRINTING		17
2.1 Metaproteomics as a powerful tool to study the intestinal microbiome	17	
2.2 The metaproteomics workflow	17	
2.3 Protein Stable Isotope Fingerprinting	21	
REFERENCES	23	
CHAPTER 3: EVALUATION OF SAMPLE PRESERVATION AND STORAGE METHODS		
FOR METAPROTEOMICS ANALYSIS OF INTESTINAL MICROBIOMES.....		26
3.1 Abstract.....	27	
3.2 Introduction.....	28	
3.3 Materials and Methods.....	33	
3.3.1 Preparation of fecal master mix and preservation treatments	33	
3.3.2 Protein extraction and peptide preparation	36	
3.3.3 LC-MS/MS	37	
3.3.4 Protein identification database	38	
3.3.5 Protein identification and quantification.....	39	
3.3.6 Quality assessment and outlier analysis.....	40	
3.3.7 Data analyses	41	
3.3.8 Data availability	43	

3.4 Results.....	43
3.4.1 Minimal differences in total numbers of identified features for co-extracted samples.....	43
3.4.2 Treatments shared over 76% of protein identifications	45
3.4.3 Relative protein abundances were highly similar between all treatments except the ethanol treatment	47
3.4.4 Within-treatment variability of relative protein abundances was low	49
3.4.5 Small but significant differences in the taxonomic composition of the metaproteomes based on preservation method	49
3.4.6 The preservation method did not bias towards specific biochemical properties of proteins.....	51
3.5 Discussion.....	52
3.6 Conclusions.....	54
3.7 Declarations and Acknowledgments.....	54
REFERENCES	56
CHAPTER 4: LINKING INTESTINAL MICROORGANISMS TO THEIR DIET-DERIVED GROWTH SUBSTRATES.....	62
4.1 Introduction.....	63
4.2 Materials and Methods.....	65
4.2.1 General experimental design.....	65
4.2.2 Gnotobiotic models	66
4.2.3 Design of diets with known isotopic signatures	67
4.2.4 Diet oscillations and sample collection.....	70
4.2.5 Protein extraction and peptide preparation	71
4.2.6 LC-MS/MS	72
4.2.7 Protein identification database	73
4.2.8 Protein identification and quantification.....	74
4.2.9 Quality assessment and outlier analysis.....	75
4.2.10 Protein-SIF.....	75
4.2.11 Data analyses	77

4.3 Results.....	80
4.3.1 Changes in community composition.....	80
4.3.2 Defined diets caused a decrease in estimated total bacterial load	86
4.3.3 Stable isotope fingerprints	86
4.3.4 Overall changes in the metaproteomes	93
4.3.5 Differentially abundant proteins	97
4.4 Discussion.....	105
4.5 Acknowledgments.....	108
REFERENCES	110
APPENDICES	114
Appendix A: Chapter 3 Supplementary Information.....	115
Appendix B: Chapter 4 Supplementary Information	125
Appendix C: Other Contributions to Science	166

LIST OF TABLES

Table 3-1 Linear correlation of replicates..... 49

Table 4-1 Natural carbon isotopic composition of the diets 60

LIST OF FIGURES

Figure 2-1	Simplified Metaproteomics and Protein-SIF Workflow	20
Figure 3-1	There were no significant differences in total numbers of PSMs, peptides, proteins, and protein groups between samples co-extracted after 1 week of preservation and only minimal differences existed in samples co-extracted after 4 weeks	44
Figure 3-2	Over 76% of microbial, host and dietary protein identifications overlapped between treatments.....	46
Figure 3-3	Ethanol preserved samples were distinct from all other samples in their protein abundance profiles.....	48
Figure 3-4	Small but significant differences in the representation of microbial taxa in the metaproteomes based on the preservation method.....	50
Figure 3-5	Distribution of biochemical properties of identified proteins.	51
Figure 4-1	Experimental design: timeline and diet oscillations.....	70
Figure 4.2	Relative proteinaceous biomass contribution of <i>M. formatexigens</i> , <i>B. uniformis</i> , <i>A. muciniphila</i> , <i>B. thetaiotaomicron</i> , <i>B. ovatus</i> , <i>E. coli</i> , <i>R. intestinalis</i> , <i>B. caccae</i> , <i>E. rectale</i> , <i>B. intestinihominis</i> , <i>C. symbiosum</i> , and <i>C. aerofaciens</i> in mice fed the protein diets (Group 1).	83
Figure 4-3	Relative proteinaceous biomass contribution of <i>M. formatexigens</i> , <i>B. uniformis</i> , <i>A. muciniphila</i> , <i>B. thetaiotaomicron</i> , <i>B. ovatus</i> , <i>E. coli</i> , <i>R. intestinalis</i> , <i>B. caccae</i> , <i>E. rectale</i> , <i>B. intestinihominis</i> , <i>C. symbiosum</i> , and <i>C. aerofaciens</i> in mice fed the fiber and fat diets (Group 2).	85
Figure 4-4	Comparison of the bacteria to host proteinaceous biomass ratios between the baseline standard chow diet and the defined diets.	86
Figure 4-5	Stable isotope fingerprints (SIFs) of the mouse host.	88
Figure 4-6	Stable isotope fingerprints (SIFs) of <i>M. formatexigens</i> (A), <i>A. muciniphila</i> (B), <i>B. thetaiotaomicron</i> (C), and <i>B. uniformis</i> (D) in group 1 mice fed the protein diets.	90
Figure 4-7	Stable isotope fingerprints (SIFs) of <i>M. formatexigens</i> (A), <i>A. muciniphila</i> (B), <i>B. thetaiotaomicron</i> (C), and <i>B. uniformis</i> (D), <i>B. caccae</i> (E), <i>E. coli</i> (F) in group 2 mice fed the fiber and fat diets.....	92
Figure 4-8	Overall differences in the metaproteomes of group 1 mice (protein diets).....	94

Figure 4-9 Overall differences in the metaproteomes of group 2 mice (fiber and fat diets)..... 96

Figure 4-10 Top 20 differentially abundant *M. formatexigens* proteins between the protein diets. 99

Figure 4-11 Top 20 differentially abundant *M. formatexigens* proteins between the inulin and corn fiber diet 101

Figure 4-12 Top 20 differentially abundant *B. thetaiotaomicron* proteins between the fiber diets 104

CHAPTER 1: IN VIVO GROWTH SUBSTRATES OF INTESTINAL MICROORGANISMS

1.1 Introduction

The human intestinal tract is home to trillions of microorganisms that form a highly diverse and metabolically active community. This complex community influences the health of its host by altering the availability of nutrients (Rowland et al. 2018; Gentile and Weir 2018; Wikoff et al. 2009; Turnbaugh et al. 2006) and the host's susceptibility to infection and disease (Boursier et al. 2016; Theriot et al. 2014). The intestinal microbiota is integral to proper host immune function (Montgomery et al. 2020; Belkaid and Hand 2014; Hooper, Littman, and Macpherson 2012) and host well-being (Chevalier et al. 2020; Rogers et al. 2016).

The intestinal microbiota survives and thrives by consuming various diet-derived and host-derived nutrients. Dietary components that the host did not absorb are thought to be the primary growth substrates of gut microbes (Rowland et al. 2018). Therefore, diet is known to be a key factor influencing the intestinal microbiota, and as a consequence, the host's health. The microbiota process dietary nutrients into metabolites that have considerable implications for host health and metabolism. Diet-microbiota interactions can be positive for the host, in the case of fermentable fiber, for example, to mitigate IBD risks (Zmora, Suez, and Elinav 2019).

Polyphenols are another example. Evidence suggests that polyphenols may play a role in preventing cardiovascular and neurodegenerative diseases (Scalbert et al. 2005), but polyphenols are not easily accessible without the microbiota. The intestinal microbiota transforms them into products that the host can efficiently uptake (Ozdal et al. 2016). However, interactions can also be detrimental to the host. For example, studies of protein degradation by the gut microbiota reveal that fermentation of proteins can result in the formation of harmful byproducts associated

with colon cancer (Diether and Willing 2019; Portune et al. 2016). Therefore, there is a growing interest in deliberately modulating the microbiota through diet to improve the host's health (Zmora, Suez, and Elinav 2019; De Filippis et al. 2018). However, to modulate the microbiota through diet, we first require knowledge of the interactions of intestinal microorganisms with dietary components.

In this chapter, I discuss (1) the nutrients available to the microbiota, (2) some indirect and direct measurement methods to study diet-microbiota associations, and (3) the established and hypothesized associations between microbial species and dietary components, with a focus of dietary fats.

1.2 What are the nutrients available to the microbiota?

The quantity of diet-derived nutrients available to the microbiota is variable and depends on multiple factors. First, it depends in part on the source of the dietary component. For example, the fiber from plant-based foods can bind and sequester dietary proteins. This process decreases the host digestibility of the dietary proteins and thus more protein transits to the colon (Acton et al. 1982; David-Birman et al. 2013). The degree of processing of the foods significantly affects, for example, the amount of starch that transits to the colon (Muir et al. 1995). Dietary components reach the distal intestinal tract also when the amount consumed simply exceeds the rate of digestion (Oliphant and Allen-Vercoe 2019). Lastly, the nutrients available to the microbiota vary with the physiology of the host and whether the host may have absorption deficiencies.

Carbohydrates

Complex polysaccharides from plants are the dietary component most readily available to the microbiota, provided the host consumes a balanced diet including fruits, vegetables, and grains. Fiber resists host digestion and absorption due to the lack of necessary digestive enzymes and the insolubility of fiber (Flint et al. 2008; Gibson et al. 2004; Bäckhed et al. 2005). Starch, a polymer of glucose, is highly abundant in Western diets. Starch can be broken down to glucose and readily absorbed by the host. Some starch, however, resists host digestion and transits to the colon as a form of fiber called resistant starch. The amount of resistant starch depends largely on the degree of processing of the foods (Vonk et al. 2000). Unlike fiber, the host readily absorbs monosaccharides, but it is estimated that a portion of the simple sugars ingested by the host is available to the microbiota (Di Rienzi and Britton 2020). The disaccharide lactose resists digestion in individuals with primary lactase deficiency and thus transits to the colon (Boron and Boulpaep 2012; Zhong et al. 2004).

Fats

Fats are thought to influence the intestinal microbiota, but the fate of dietary fats in the colon is not well understood. Assuming the host is healthy and consumes a balanced diet, most dietary lipids are absorbed in the small intestine. Only a tiny fraction (~ 5 %, or 4-5 g daily) passes through the intestinal tract and reaches the microbiota (Iqbal and Hussain 2009; Morales et al. 2016). However, overconsumption of fats, which is common in Western diets, can lead to excess lipids passing through to the distal intestinal tract. Diseases such as celiac sprue can also cause more fat to be unabsorbed and thus available to the colonic microbiota (Boron and Boulpaep 2012). About 90% of dietary lipids are triglycerides, and the remaining are phospholipids and sterols. Triglycerides are molecules composed of three long-chain fatty acids

esterified to glycerol. Dietary lipids are broken down into free fatty acids by host lipases. Therefore, lipids in the distal intestinal tract are likely a mixture of undigested glycerides, free fatty acids, and glycerol. Glycerol is the backbone of triglycerides, and it can serve as a potential growth substrate for intestinal microbes (De Weirdt et al. 2010). More than 20 types of fatty acids are found in foods, which are characteristic of their food source. For example, animal fats are rich in stearic acid, a non-essential saturated fatty acid. In contrast, plant fats such as soybean oil are rich in linoleic acid, a polyunsaturated essential fatty acid (White 2009).

Proteins

A portion of dietary proteins ingested transits to the colon in the form of proteins, peptides, and free amino acids. Protein source and degree of processing impact the type and amount of protein available to the intestinal microbiota. Minimally processed plant-based proteins are thought to be most readily available to the microbiota because of polysaccharides sequestration and protease inhibitors that allow them to resist host digestion (Joye 2019; Barbé et al. 2014; Acton, Breyer, and Satterlee 1982).

Non-diet derived substrates

Non-diet derived compounds such as secreted host metabolites or the host-derived glycans are also available as potential growth substrates for intestinal microorganisms. The mucus layer lining up the intestinal wall is rich in glycoproteins and oligosaccharides (Schroeder 2019). Growth substrates may also be available through microbial cross-feeding (Henriques et al. 2020).

1.3 Methods to investigate diet-microbiota associations

The ability of intestinal microorganisms to degrade and use dietary components as growth substrates has been studied extensively *in vitro*. The use of defined growth media and fermentation assessments in pure cultures and consortia can be key to elucidating metabolic profiles of gut bacteria. *in vitro* methods can be instrumental in determining candidate *in vivo* substrates. Furthermore, the relative growth capacities of intestinal bacteria grown in consortia tend to reflect their relative abundance observed in the microbiomes of healthy humans (Tramontano et al. 2018). *In vivo*, diet-microbiota interactions have been studied a lot in mouse models and excellent work has been conducted in gnotobiotic mice especially (Patnode et al. 2019; Desai et al. 2016; E. D. Sonnenburg et al. 2010).

A majority of the microbe-growth substrate associations *in vivo* have been inferred from correlations data. Many studies have looked at abundance effects resulting from dietary perturbations. In the past two decades, abundance effects were largely assessed by sequencing-based approaches. Due to its relative simplicity and low cost, amplicon sequencing of phylogenetic marker genes, such as the 16S rRNA gene, is a popular method to determine community composition changes.

Multi-omics approaches are increasingly used to investigate associations of intestinal microbes and diet-derived substrates. Metagenomics has been used for genetic screening of metabolic potential. Metatranscriptomics and metaproteomics reveal which of the relevant genes are expressed in particular conditions. Metaproteomics is discussed in more detail in **chapter 2**. Metabolomics can also reveal substrate use by monitoring product release. However, strains cannot be linked to the product molecules, and thus it is challenging to use on complex communities such as the human gut microbiota. Combinations of -omics approaches have the

potential to provide a more holistic picture of the diet-microbiota interactions. For example, using a combination of metaproteomics and metabolomics, researchers established new linkages between resistant starch and the specific bacterial species that metabolize it (Maier et al. 2017).

Establishing direct linkages between intestinal microorganisms and their *in vivo* growth substrates can be complex. Indirect measurements are often crucial to narrow down the target. A recent study provided direct evidence for the use of fiber-derived glycans by three *Bacteroides* species (Patnode et al. 2019). They passed magnetic beads coated with glycans through the GI tract of gnotobiotic mice that were colonized with different combinations of the *Bacteroides* species. They then retrieved the beads and measured the remaining amount of glycans, showing how much of the glycans were consumed by specific bacterial species.

Direct measurements often rely on the use of isotopes. In this context, stable isotope probing (SIP) consists of administering a heavy labeled substrate to the host and monitoring the microbial populations that assimilate the isotopes (Berry and Loy 2018). SIP can be applied to DNA, RNA, or proteins. Labeled nucleic acids have a different density than unlabeled nucleic acids and thus can be separated and later analyzed with sequencing-based approaches. Protein-SIP evaluates assimilation of heavy labels with mass-spectrometry-based approaches. In our experiments (**chapter 4**), we used a protein stable isotope fingerprinting (Protein-SIF)(Kleiner et al. 2018) approach to evaluate the assimilation of natural isotopes instead of heavy labels. Protein-SIF is discussed in **chapter 2**.

1.4 Established and hypothesized associations between microbial species and dietary components

Carbohydrates

There is ample evidence for the use of carbohydrates by intestinal bacteria, particularly fiber-derived glycans. The genomes of *Bacteroides* species encode many carbohydrate-degrading enzymes that allow for the use of a variety of polysaccharides and resistant starch (Flint et al. 2008). The degradation of carbohydrates by intestinal bacteria has been extensively studied, and knowledge has been compiled into a database (CAZy database <https://www.cazy.org>) (Lombard et al. 2014).

Fats

The interaction between dietary fats and intestinal bacteria is thought to be a more indirect one. Dietary fats promote changes in the bile acid profile of the host, and the bile acids interact with the microbiota causing compositional and functional shifts. For example, In a mouse study, diets high in saturated fats (milk derived) caused a substantial increase of taurocholic acid which subsequently caused a bloom of *B. wadsworthia* - a microbe associated with colitis (Devkota et al. 2012).

Some intestinal microorganisms can metabolize polyunsaturated free fatty acids to convert them into less toxic saturated fatty acids (Polan et al. 1964). For example, Lactic acid bacteria such as *Lactobacillus* species can metabolize linoleic acid, a polyunsaturated fatty acid (Schoeler and Caesar 2019, Ogawa et al. 2005). However, it is unclear whether intestinal bacteria can use fatty acids as a growth substrate.

In vitro, several species belonging to the Proteobacteria phylum can sustain growth in media containing almost exclusively purified fatty acids (Agans et al. 2018). *Escherichia coli*

and *Enterobacter cloacae* encode genes for enzymes of the fatty acid B-oxidation pathway. Species of genera *Bilophila* that belong to the Bacteroidetes phylum can also sustain growth on media devoid of carbohydrates and proteins (Agans et al. 2018).

In vivo studies have employed indirect measurements to relate fats and intestinal microorganisms. High-fat diets have been associated with large compositional changes in humans and mouse models, with typically an increase in the proportion of Firmicutes and Proteobacteria (Abulizi et al. 2019; Hildebrandt et al. 2009; Zmora, Suez, and Elinav 2019). The type of fat introduced has the potential of aggravating the compositional changes. For example, the proportion of Firmicutes increases significantly more when diets are rich in saturated fatty acids than diets rich in polyunsaturated fatty acids (T. Liu et al. 2012; Wit et al. 2012).

At the species level, dietary fats influence the growth of *Akkermansia muciniphila* which thrives in the presence of unsaturated fatty acids (Li et al., 2017; Patrone et al., 2018). Celiac sprue causes more fat to be unabsorbed and thus available to the colonic microbiota. *Akkermansia* was found higher in abundance in the microbiota of patients with celiac sprue as compared to pre-diseased controls (Bodkhe et al. 2019). *Bilophila* bacteria were observed to increase in abundance in mice were fed a high-fat diet (Schneeberger et al. 2015).

A study explored the changes in the microbiota composition when artificially increasing the number of lipids passing through the GI tract by administration of a drug that induces malabsorption by inhibiting lipases (Morales et al. 2016). They demonstrated that there was a significantly higher amount of unabsorbed fat in the fecal samples of subjects that were given the medication but it did not cause any significant change in the composition of the microbiota. Intestinal bacteria may not be able to degrade lipids in the form of unabsorbed glycerides

(Cândido et al. 2018). Rather, it is likely to be the free fatty acids interacting with the microbes that cause shifts in community composition.

Protein

Faith et al. found that dietary protein may be a primary limiting factor for bacterial biomass (Faith et al. 2011). They found that in gnotobiotic mice colonized with a 10-species microbiota, the bacterial community biomass (measured as DNA/ μ g fecal material) increased with increases in dietary protein content, suggesting protein may be an important growth substrate for intestinal bacteria. Protein degradation has been described mostly in Clostridia species (Dai et al. 2015).

Host-derived compounds

There have been observations of intestinal bacteria feeding on the host mucin and its associated glycans (Desai et al. 2016; Berry et al. 2013) The mucus layer serves as a substrate to bacteria that encode the glycoside enzymes to degrade complex mucus glycans, but it is not always a preferred substrate. While some organisms, such as *Akkermansia muciniphila*, are considered mucus specialists (Desai et al. 2016), others, such as *Bacteroides thetaiotaomicron*, forage host-derived glycans only when plant polysaccharides are absent from the diet (Li et al. 2015; Bjursell, Martens, and Gordon 2006; J. L. Sonnenburg et al. 2005). The ability of using host-derived glycans is thought to confer a competitive advantage in organisms such as *Marvinbryantia formatexigens* (Rey et al. 2010).

REFERENCES

- Abulizi, Nijiati, Candice Quin, Kirsty Brown, Yee Kwan Chan, Sandeep K. Gill, and Deanna L. Gibson. 2019. "Gut Mucosal Proteins and Bacteriome Are Shaped by the Saturation Index of Dietary Lipids." *Nutrients* 11 (2): 418. <https://doi.org/10.3390/nu11020418>.
- Acton, J. C., L. Breyer, and L. D. Satterlee. 1982. "Effect of Dietary Fiber Constituents on the In Vitro Digestibility of Casein." *Journal of Food Science* 47 (2): 556–60. <https://doi.org/10.1111/j.1365-2621.1982.tb10122.x>.
- Agans, Richard, Alex Gordon, Denise Lynette Kramer, Sergio Perez-Burillo, José A. Rufián-Henares, and Oleg Paliy. 2018. "Dietary Fatty Acids Sustain the Growth of the Human Gut Microbiota." *Applied and Environmental Microbiology* 84 (21). <https://doi.org/10.1128/AEM.01525-18>.
- Bäckhed, Fredrik, Ruth E. Ley, Justin L. Sonnenburg, Daniel A. Peterson, and Jeffrey I. Gordon. 2005. "Host-Bacterial Mutualism in the Human Intestine." *Science* 307 (5717): 1915–20. <https://doi.org/10.1126/science.1104816>.
- Barbé, Florence, Steven Le Feunteun, Didier Rémond, Olivia Ménard, Julien Jardin, Gwénaële Henry, Béatrice Laroche, and Didier Dupont. 2014. "Tracking the in Vivo Release of Bioactive Peptides in the Gut during Digestion: Mass Spectrometry Peptidomic Characterization of Effluents Collected in the Gut of Dairy Matrix Fed Mini-Pigs." <https://pubag.nal.usda.gov/catalog/5432125>.
- Belkaid, Yasmine, and Timothy W. Hand. 2014. "Role of the Microbiota in Immunity and Inflammation." *Cell* 157 (1): 121–41. <https://doi.org/10.1016/j.cell.2014.03.011>.
- Berry, David, and Alexander Loy. 2018. "Stable-Isotope Probing of Human and Animal Microbiome Function." *Trends in Microbiology* 26 (12): 999–1007. <https://doi.org/10.1016/j.tim.2018.06.004>.
- Berry, David, Bärbel Stecher, Arno Schintlmeister, Jochen Reichert, Sandrine Brugiroux, Birgit Wild, Wolfgang Wanek, et al. 2013. "Host-Compound Foraging by Intestinal Microbiota Revealed by Single-Cell Stable Isotope Probing." *Proceedings of the National Academy of Sciences of the United States of America* 110 (12): 4720–25. <https://doi.org/10.1073/pnas.1219247110>.
- Bjursell, Magnus K., Eric C. Martens, and Jeffrey I. Gordon. 2006. "Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, *Bacteroides Thetaiotaomicron*, to the Suckling Period *." *Journal of Biological Chemistry* 281 (47): 36269–79. <https://doi.org/10.1074/jbc.M606509200>.
- Boron, Walter F., and Emile L. Boulpaep. 2012. *Medical Physiology, 2e Updated Edition : With STUDENT CONSULT Online Access*. Philadelphia, UNITED STATES: Elsevier. <http://ebookcentral.proquest.com/lib/ncsu/detail.action?docID=1430142>.
- Boursier, Jérôme, Olaf Mueller, Matthieu Barret, Mariana Machado, Lionel Fizanne, Felix

- Araujo-Perez, Cynthia D. Guy, et al. 2016. “The Severity of Nonalcoholic Fatty Liver Disease Is Associated with Gut Dysbiosis and Shift in the Metabolic Function of the Gut Microbiota.” *Hepatology* 63 (3): 764–75. <https://doi.org/10.1002/hep.28356>.
- Cândido, Flávia Galvão, Flávia Xavier Valente, Łukasz Marcin Grześkowiak, Ana Paula Boroni Moreira, Daniela Mayumi Usuda Prado Rocha, and Rita de Cássia Gonçalves Alfenas. 2018. “Impact of Dietary Fat on Gut Microbiota and Low-Grade Systemic Inflammation: Mechanisms and Clinical Implications on Obesity.” *International Journal of Food Sciences and Nutrition* 69 (2): 125–43. <https://doi.org/10.1080/09637486.2017.1343286>.
- Chevalier, Grégoire, Eleni Siopi, Laure Guenin-Macé, Maud Pascal, Thomas Laval, Aline Rifflet, Ivo Gomperts Boneca, et al. 2020. “Effect of Gut Microbiota on Depressive-like Behaviors in Mice Is Mediated by the Endocannabinoid System.” *Nature Communications* 11 (1): 1–15. <https://doi.org/10.1038/s41467-020-19931-2>.
- David-Birman, Tatyana, Alan Mackie, and Uri Lesmes. 2013. “Impact of Dietary Fibers on the Properties and Proteolytic Digestibility of Lactoferrin Nano-Particles.” *Food Hydrocolloids* 31 (1): 33–41. <https://doi.org/10.1016/j.foodhyd.2012.09.013>.
- De Filippis, Francesca, Paola Vitaglione, Rosario Cuomo, Roberto Berni Canani, and Danilo Ercolini. 2018. “Dietary Interventions to Modulate the Gut Microbiome—How Far Away Are We From Precision Medicine.” *Inflammatory Bowel Diseases* 24 (10): 2142–54. <https://doi.org/10.1093/ibd/izy080>.
- De Weirdt, Rosemarie, Sam Possemiers, Griet Vermeulen, Tanja C. W. Moerdijk-Poortvliet, Henricus T. S. Boschker, Willy Verstraete, and Tom Van de Wiele. 2010. “Human Faecal Microbiota Display Variable Patterns of Glycerol Metabolism.” *FEMS Microbiology Ecology* 74 (3): 601–11. <https://doi.org/10.1111/j.1574-6941.2010.00974.x>.
- Desai, M. S., A. M. Seekatz, N. M. Koropatkin, N. Kamada, C. A. Hickey, M. Wolter, N. A. Pudlo, et al. 2016. “A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility.” *Cell* 167 (5): 1339–1353.e21. <https://doi.org/10.1016/j.cell.2016.10.043>.
- Devkota, Suzanne, Yunwei Wang, Mark W. Musch, Vanessa Leone, Hannah Fehlner-Peach, Anuradha Nadimpalli, Dionysios A. Antonopoulos, Bana Jabri, and Eugene B. Chang. 2012. “Dietary-Fat-Induced Taurocholic Acid Promotes Pathobiont Expansion and Colitis in Il10^{-/-} Mice.” *Nature* 487 (7405): 104–8. <https://doi.org/10.1038/nature11225>.
- Di Rienzi, Sara C., and Robert A. Britton. 2020. “Adaptation of the Gut Microbiota to Modern Dietary Sugars and Sweeteners.” *Advances in Nutrition* 11 (3): 616–29. <https://doi.org/10.1093/advances/nmz118>.
- Diether, Natalie E., and Benjamin P. Willing. 2019. “Microbial Fermentation of Dietary Protein: An Important Factor in Diet–Microbe–Host Interaction.” *Microorganisms* 7 (1): 19. <https://doi.org/10.3390/microorganisms7010019>.
- Flint, Harry J., Edward A. Bayer, Marco T. Rincon, Raphael Lamed, and Bryan A. White. 2008.

- “Polysaccharide Utilization by Gut Bacteria: Potential for New Insights from Genomic Analysis.” *Nature Reviews Microbiology* 6 (2): 121–31.
<https://doi.org/10.1038/nrmicro1817>.
- Gentile, Christopher L., and Tiffany L. Weir. 2018. “The Gut Microbiota at the Intersection of Diet and Human Health.” *Science* 362 (6416): 776–80.
<https://doi.org/10.1126/science.aau5812>.
- Gibson, Glenn R., Hollie M. Probert, Jan Van Loo, Robert A. Rastall, and Marcel B. Roberfroid. 2004. “Dietary Modulation of the Human Colonic Microbiota: Updating the Concept of Prebiotics.” *Nutrition Research Reviews* 17 (2): 259–75.
- Henriques, Sílvia F., Darshan B. Dhakan, Lúcia Serra, Ana Patrícia Francisco, Zita Carvalho-Santos, Célia Baltazar, Ana Paula Elias, et al. 2020. “Metabolic Cross-Feeding in Imbalanced Diets Allows Gut Microbes to Improve Reproduction and Alter Host Behaviour.” *Nature Communications* 11 (1): 4236. <https://doi.org/10.1038/s41467-020-18049-9>.
- Hildebrandt, Marie A., Christian Hoffmann, Scott A. Sherrill–Mix, Sue A. Keilbaugh, Micah Hamady, Ying-Yu Chen, Rob Knight, Rexford S. Ahima, Frederic Bushman, and Gary D. Wu. 2009. “High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity.” *Gastroenterology* 137 (5): 1716-1724.e2.
<https://doi.org/10.1053/j.gastro.2009.08.042>.
- Holscher, Hannah D., Andrew M. Taylor, Kelly S. Swanson, Janet A. Novotny, and David J. Baer. 2018. “Almond Consumption and Processing Affects the Composition of the Gastrointestinal Microbiota of Healthy Adult Men and Women: A Randomized Controlled Trial.” *Nutrients* 10 (2): 126. <https://doi.org/10.3390/nu10020126>.
- Hooper, L. V., D. R. Littman, and A. J. Macpherson. 2012. “Interactions between the Microbiota and the Immune System.” *Science* 336 (6086): 1268–73.
<https://doi.org/10.1126/science.1223490>.
- Iqbal, Jahangir, and M. Mahmood Hussain. 2009. “Intestinal Lipid Absorption.” *American Journal of Physiology. Endocrinology and Metabolism* 296 (6): E1183-1194.
<https://doi.org/10.1152/ajpendo.90899.2008>.
- Joye, Iris. 2019. “Protein Digestibility of Cereal Products.” *Foods* 8 (6): 199.
<https://doi.org/10.3390/foods8060199>.
- Kleiner, Manuel, Xiaoli Dong, Tjorven Hinzke, Juliane Wippler, Erin Thorson, Bernhard Mayer, and Marc Strous. 2018. “Metaproteomics Method to Determine Carbon Sources and Assimilation Pathways of Species in Microbial Communities.” *Proceedings of the National Academy of Sciences* 115 (24): E5576–84.
<https://doi.org/10.1073/pnas.1722325115>.
- Li, Hai, Julien P. Limenitakis, Tobias Fuhrer, Markus B. Geuking, Melissa A. Lawson, Madeleine Wyss, Sandrine Brugiroux, et al. 2015. “The Outer Mucus Layer Hosts a

- Distinct Intestinal Microbial Niche.” *Nature Communications* 6 (1): 1–13.
<https://doi.org/10.1038/ncomms9292>.
- Liu, Tianyu, Helen Hougen, Amy C. Vollmer, and Sara M. Hiebert. 2012. “Gut Bacteria Profiles of *Mus Musculus* at the Phylum and Family Levels Are Influenced by Saturation of Dietary Fatty Acids.” *Anaerobe* 18 (3): 331–37.
<https://doi.org/10.1016/j.anaerobe.2012.02.004>.
- Liu, Zhibin, Wei Wang, Guangwei Huang, Wen Zhang, and Li Ni. 2016. “In Vitro and in Vivo Evaluation of the Prebiotic Effect of Raw and Roasted Almonds (*Prunus Amygdalus*).” *Journal of the Science of Food and Agriculture* 96 (5): 1836–43.
<https://doi.org/10.1002/jsfa.7604>.
- Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. “The Carbohydrate-Active Enzymes Database (CAZy) in 2013.” *Nucleic Acids Research* 42 (Database issue): D490.
<https://doi.org/10.1093/nar/gkt1178>.
- Maier, Tanja V., Marianna Lucio, Lang Ho Lee, Nathan C. VerBerkmoes, Colin J. Brislawn, Jörg Bernhardt, Regina Lamendella, et al. 2017. “Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome.” *MBio* 8 (5).
<https://doi.org/10.1128/mBio.01343-17>.
- Mandalari, G., C. Nueno-Palop, G. Bisignano, M. S. J. Wickham, and A. Narbad. 2008. “Potential Prebiotic Properties of Almond (*Amygdalus Communis* L.) Seeds.” *Applied and Environmental Microbiology* 74 (14): 4264–70. <https://doi.org/10.1128/AEM.00739-08>.
- Montgomery, Theresa L., Axel Künstner, Josephine J. Kennedy, Qian Fang, Lori Asarian, Rachel Culp-Hill, Angelo D’Alessandro, Cory Teuscher, Hauke Busch, and Dimitry N. Kremontsov. 2020. “Interactions between Host Genetics and Gut Microbiota Determine Susceptibility to CNS Autoimmunity.” *Proceedings of the National Academy of Sciences* 117 (44): 27516–27. <https://doi.org/10.1073/pnas.2002817117>.
- Morales, Pamela, Sayaka Fujio, Paola Navarrete, Juan A. Ugalde, Fabien Magne, Catalina Carrasco-Pozo, Karina Tralma, et al. 2016. “Impact of Dietary Lipids on Colonic Function and Microbiota: An Experimental Approach Involving Orlistat-Induced Fat Malabsorption in Human Volunteers.” *Clinical and Translational Gastroenterology* 7 (4): e161. <https://doi.org/10.1038/ctg.2016.20>.
- Muir, J. G., A. Birkett, I. Brown, G. Jones, and K. O’Dea. 1995. “Food Processing and Maize Variety Affects Amounts of Starch Escaping Digestion in the Small Intestine.” *The American Journal of Clinical Nutrition* 61 (1): 82–89.
<https://doi.org/10.1093/ajcn/61.1.82>.
- Oliphant, Kaitlyn, and Emma Allen-Vercoe. 2019. “Macronutrient Metabolism by the Human Gut Microbiome: Major Fermentation by-Products and Their Impact on Host Health.” *Microbiome* 7 (1): 91. <https://doi.org/10.1186/s40168-019-0704-8>.

- Ozidal, Tugba, David A. Sela, Jianbo Xiao, Dilek Boyacioglu, Fang Chen, and Esra Capanoglu. 2016. "The Reciprocal Interactions between Polyphenols and Gut Microbiota and Effects on Bioaccessibility." *Nutrients* 8 (2). <https://doi.org/10.3390/nu8020078>.
- Patnode, Michael L., Zachary W. Beller, Nathan D. Han, Jiye Cheng, Samantha L. Peters, Nicolas Terrapon, Bernard Henrissat, et al. 2019. "Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans." *Cell* 179 (1): 59–73. <https://doi.org/10.1016/j.cell.2019.08.011>.
- Polan, C. E., J. J. McNeill, and S. B. Tove. 1964. "Biohydrogenation of Unsaturated Fatty Acids by Rumen Bacteria." *Journal of Bacteriology* 88 (4): 1056–64.
- Portune, Kevin J., Martin Beaumont, Anne-Marie Davila, Daniel Tomé, François Blachier, and Yolanda Sanz. 2016. "Gut Microbiota Role in Dietary Protein Metabolism and Health-Related Outcomes: The Two Sides of the Coin." *Trends in Food Science & Technology*, Unravelling the role of the gut microbiome in energy balance and brain development and function: the European project MyNewGut, 57 (November): 213–32. <https://doi.org/10.1016/j.tifs.2016.08.011>.
- Rey, Federico E., Jeremiah J. Faith, James Bain, Michael J. Muehlbauer, Robert D. Stevens, Christopher B. Newgard, and Jeffrey I. Gordon. 2010. "Dissecting the in Vivo Metabolic Potential of Two Human Gut Acetogens." *The Journal of Biological Chemistry* 285 (29): 22082. <https://doi.org/10.1074/jbc.M110.117713>.
- Rogers, G. B., D. J. Keating, R. L. Young, M.-L. Wong, J. Licinio, and S. Wesselingh. 2016. "From Gut Dysbiosis to Altered Brain Function and Mental Illness: Mechanisms and Pathways." *Molecular Psychiatry* 21 (6): 738. <https://doi.org/10.1038/mp.2016.50>.
- Rowland, Ian, Glenn Gibson, Almut Heinken, Karen Scott, Jonathan Swann, Ines Thiele, and Kieran Tuohy. 2018. "Gut Microbiota Functions: Metabolism of Nutrients and Other Food Components." *European Journal of Nutrition* 57 (1): 1–24. <https://doi.org/10.1007/s00394-017-1445-8>.
- Scalbert, Augustin, Claudine Manach, Christine Morand, Christian Rémésy, and Liliana Jiménez. 2005. "Dietary Polyphenols and the Prevention of Diseases." *Critical Reviews in Food Science and Nutrition* 45 (4): 287–306. <https://doi.org/10.1080/1040869059096>.
- Schneeberger, Marc, Amandine Everard, Alicia G. Gómez-Valadés, Sébastien Matamoros, Sara Ramírez, Nathalie M. Delzenne, Ramon Gomis, Marc Claret, and Patrice D. Cani. 2015. "Akkermansia Muciniphila Inversely Correlates with the Onset of Inflammation, Altered Adipose Tissue Metabolism and Metabolic Disorders during Obesity in Mice." *Scientific Reports* 5. <https://doi.org/10.1038/srep16643>.
- Schroeder, Bjoern O. 2019. "Fight Them or Feed Them: How the Intestinal Mucus Layer Manages the Gut Microbiota." *Gastroenterology Report* 7 (1): 3. <https://doi.org/10.1093/gastro/goy052>.
- Sonnenburg, Erica D., Hongjun Zheng, Payal Joglekar, Steven K. Higginbottom, Susan J.

- Firbank, David N. Bolam, and Justin L. Sonnenburg. 2010. "Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations." *Cell* 141 (7): 1241–52. <https://doi.org/10.1016/j.cell.2010.05.005>.
- Sonnenburg, Justin L., Jian Xu, Douglas D. Leip, Chien-Huan Chen, Benjamin P. Westover, Jeremy Weatherford, Jeremy D. Buhler, and Jeffrey I. Gordon. 2005. "Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont." *Science* 307 (5717): 1955–59. <https://doi.org/10.1126/science.1109051>.
- Theriot, Casey M., Mark J. Koenigsknecht, Paul E. Carlson, Gabrielle E. Hatton, Adam M. Nelson, Bo Li, Gary B. Huffnagle, Jun Z. Li, and Vincent B. Young. 2014. "Antibiotic-Induced Shifts in the Mouse Gut Microbiome and Metabolome Increase Susceptibility to Clostridium Difficile Infection." *Nature Communications* 5 (1): 3114. <https://doi.org/10.1038/ncomms4114>.
- Turnbaugh, Peter J., Ruth E. Ley, Michael A. Mahowald, Vincent Magrini, Elaine R. Mardis, and Jeffrey I. Gordon. 2006. "An Obesity-Associated Gut Microbiome with Increased Capacity for Energy Harvest." *Nature* 444 (7122): 1027–31. <https://doi.org/10.1038/nature05414>.
- Ukhanova, Maria, Xiaoyu Wang, David J. Baer, Janet A. Novotny, Marlene Fredborg, and Volker Mai. 2014. "Effects of Almond and Pistachio Consumption on Gut Microbiota Composition in a Randomised Cross-over Human Feeding Study." *British Journal of Nutrition* 111 (12): 2146–52. <https://doi.org/10.1017/S0007114514000385>.
- Vonk, Roel J., Renate E. Hagedoorn, Rynate de Graaff, Henk Elzinga, Saskia Tabak, Yue-Xin Yang, and Frans Stellaard. 2000. "Digestion of So-Called Resistant Starch Sources in the Human Small Intestine." *The American Journal of Clinical Nutrition* 72 (2): 432–38. <https://doi.org/10.1093/ajcn/72.2.432>.
- White, Brett. 2009. "Dietary Fatty Acids." *American Family Physician* 80 (4): 345–50.
- Wikoff, William R., Andrew T. Anfora, Jun Liu, Peter G. Schultz, Scott A. Lesley, Eric C. Peters, and Gary Siuzdak. 2009. "Metabolomics Analysis Reveals Large Effects of Gut Microflora on Mammalian Blood Metabolites." *Proceedings of the National Academy of Sciences* 106 (10): 3698–3703. <https://doi.org/10.1073/pnas.0812874106>.
- Wit, Nicole de, Muriel Derrien, Hanneke Bosch-Vermeulen, Els Oosterink, Shohreh Keshtkar, Caroline Duval, Johan de Vogel-van den Bosch, Michiel Kleerebezem, Michael Müller, and Roelof van der Meer. 2012. "Saturated Fat Stimulates Obesity and Hepatic Steatosis and Affects Gut Microbiota Composition by an Enhanced Overflow of Dietary Fat to the Distal Intestine." *American Journal of Physiology-Gastrointestinal and Liver Physiology*, September. <https://doi.org/10.1152/ajpgi.00488.2011>.
- Zhong, Yan, Marion G. Priebe, Roel J. Vonk, Cheng-Yu Huang, Jean-Michel Antoine, Tao He, Hermie J. M. Harmsen, and Gjal W. Welling. 2004. "The Role of Colonic Microbiota in Lactose Intolerance." *Digestive Diseases and Sciences* 49 (1): 78–83.

<https://doi.org/10.1023/B:DDAS.0000011606.96795.40>.

Zmora, Niv, Jotham Suez, and Eran Elinav. 2019. "You Are What You Eat: Diet, Health and the Gut Microbiota." *Nature Reviews Gastroenterology & Hepatology* 16 (1): 35–56.
<https://doi.org/10.1038/s41575-018-0061-2>.

CHAPTER 2: METAPROTEOMICS AND PROTEIN STABLE ISOTOPE FINGERPRINTING

2.1 Metaproteomics as a powerful tool to study the intestinal microbiome

Metaproteomics is a powerful tool to study interactions in the intestinal microbiome and the microbiome's influence on host health (Salvato, Hettich, and Kleiner 2021; Kleiner 2019, 20; Starr et al. 2017). Metaproteomics allows for the identification and quantification of large numbers of microbial, dietary, and host proteins in microbiome samples in a high-throughput fashion (Xiong et al. 2015; Verberkmoes et al. 2009; Zhang et al. 2017). Because proteins are central to all biological processes, metaproteomics provides direct evidence of the activities and functions of microbial community members and their contributions to disease (Lee et al. 2017). For example, metaproteomics revealed protein biomarkers of disease in Inflammatory Bowel Disease (IBD) and colorectal cancer and provided insights into the role of the microbiome in such diseases (Long et al. 2020; Zhang, Deeke, et al. 2018). In addition to quantifying differences in protein abundances between samples, metaproteomics can also be used to assess microbial community structure based on proteinaceous biomass (Kleiner et al. 2017; Pible et al. 2020; Gouveia et al. 2020; Van Den Bossche et al. 2021) and track incorporation of specific substrates using stable isotope content of peptides (Kleiner et al. 2021; 2018; Justice et al. 2014; Jehmlich et al. 2008).

2.2 The Metaproteomics Workflow

Metaproteomics is still in its early stages of development. Consequently, there is no standardized workflow for the metaproteomics analyses of intestinal microbiome samples. Nonetheless, a majority of the individual steps of the workflow have been optimized in the past.

A common workflow includes the following steps: sample collection and preservation, cell lysis, protein extraction, enzymatic digestion of proteins to peptides, separation of peptides by liquid chromatography, data acquisition by tandem mass spectrometry, and database searching. Here I discuss the specific workflow that we used in the experiments described in this thesis, but other approaches exist.

First, we typically collect microbiome samples as fecal pellets that we immediately immerse in the preservative solution RNeasy Lysis Buffer™ to preserve sample integrity and prevent protein profiles from changing. The choice of sample preservation and storage method was decided based on experimental optimization (Chapter 3 of this thesis). Next, we remove the preservation solution and extract the proteins in the fecal samples, which belong to the dietary components, the host, and the microorganisms. To do so, we first lyse the cells using heat, sodium dodecyl sulfate (SDS), and bead-beating. Some sample types may only require chemical lysis with SDS; however, microbiome samples that contain gram-positive bacteria require physical lysis as well (Zhang, Li, et al. 2018)(and unpublished data from the Kleiner lab supporting bead-beating). Then, we apply the extracted protein mixtures to the Filter-Aided Sample Preparation (FASP) protocol. The FASP protocol is a commonly used high-throughput protocol that allows for the clean-up and modification of the extracted protein mixture (Wiśniewski et al. 2009).

The FASP protocol is also used in the final step of sample preparation: digestion of proteins to peptides with trypsin. Trypsin is a highly specific enzyme that cleaves proteins at known residues and therefore produces predictable peptide fragments. This digestion is performed in parallel *in silico* from the protein sequences of the database. The *in silico* digestion aims to predict all potential peptides present in the samples. Later, the experimental peptide data

is matched to the *in silico* generated peptide data to identify and quantify peptides. Proteins are then inferred from the identified peptides. Protein inference is most successful when the data contain many “protein unique peptides (PUPs).” While some peptides are shared between several proteins, PUPs have distinguishable proteins of origin. Therefore, PUPs are often used to filter datasets and increase confidence in protein identifications. Identified proteins are also typically filtered with a 5 % False Discovery Rate (FDR) threshold. FDR is computed using a decoy database of randomized protein sequences.

Confident protein identification is also dependent on the protein database used. The ideal database includes all proteins potentially present in a sample. Best practice is to create a matched metagenomic database (Tanca et al. 2013). As an alternative, we conducted experiments in gnotobiotic mice colonized with a community of fully-sequenced strains (Chapter 4). In this case, a database can be assembled from the publicly available proteomes.

Lastly, identified and quantified proteins are typically normalized to %Normalized Spectral Abundance Factors (%NSAFs) or Centered-log-ratio (CLR) transformed counts.

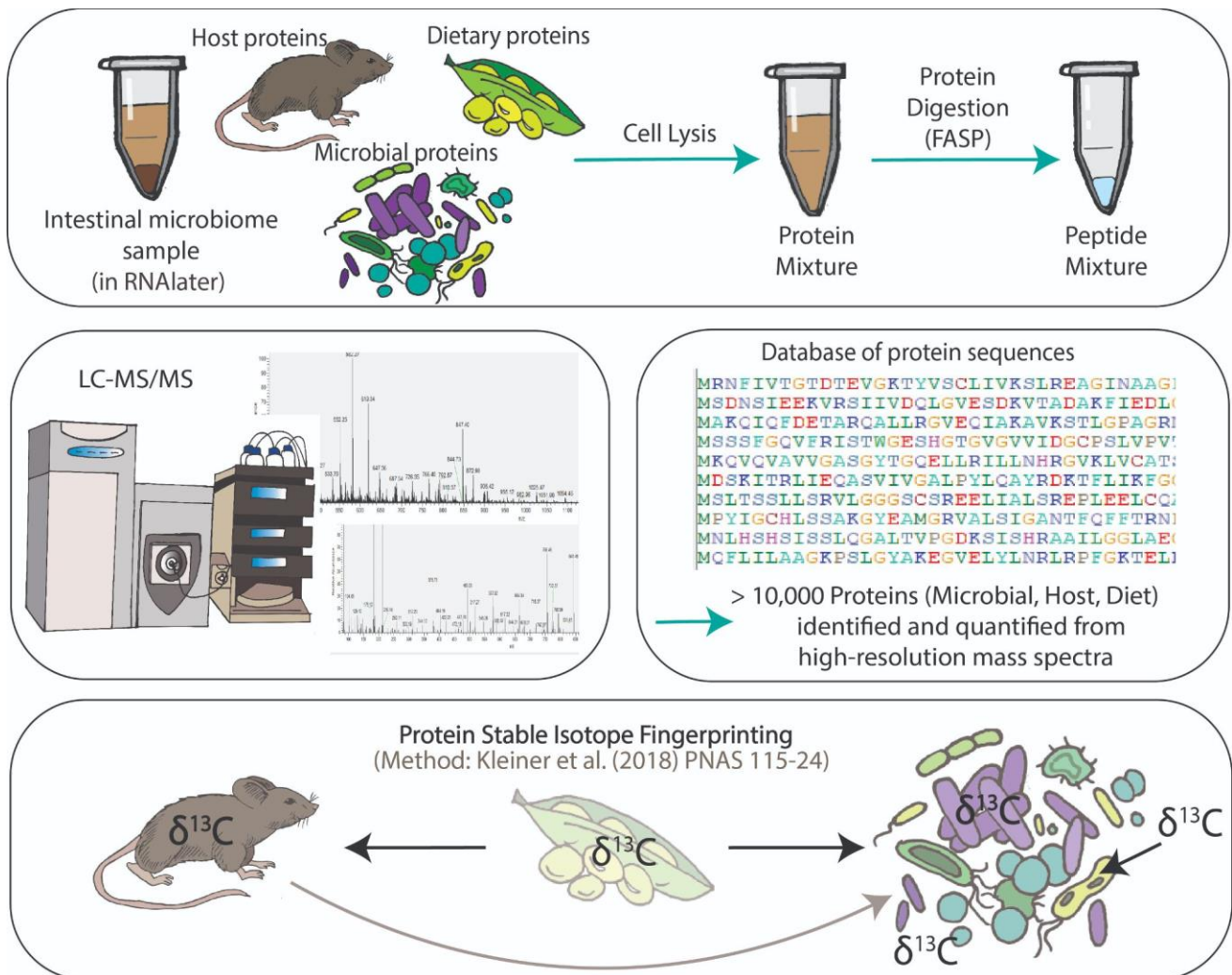


Figure 2-1. Simplified Metaproteomics and Protein-SIF Workflow. First, cells are lysed using physical and chemical lysis methods (bead-beating, heat, and SDS) to extract the proteins contained in the fecal samples. The extracted protein mixture contains dietary proteins, host proteins, and microbial proteins. We then apply the Filter Aided Sample Preparation (FASP) to clean up the protein mixture and to digest the proteins enzymatically into predictable peptide fragments. We then separate the peptides by liquid chromatography and analyze them by tandem mass spectrometry to obtain tens of thousands of high-resolution mass spectra. Next, the mass spectra are searched against a protein sequence database containing proteomes of organisms that may be present in the samples (the host, intestinal microorganisms, and organisms from which the dietary components originate). After we have identified and quantified proteins contained in the samples, we apply the protein stable isotope fingerprinting (Protein-SIF) to estimate the natural isotopic signatures of the organisms present in the samples.

2.3 The protein-SIF approach: a metaproteomics-based approach to investigate growth substrates of intestinal microorganisms

Heterotrophic organisms, such as the majority of intestinal bacteria, have a similar natural carbon isotopic signature ($\delta^{13}\text{C}$) as their food source (Pelz et al. 1998). Thus, natural carbon isotopic signatures can be used to associate intestinal bacteria with their growth substrates.

The protein stable isotope fingerprinting (Protein-SIF) method estimates carbon isotope ratios ($^{13}\text{C}/^{12}\text{C}$, or $\delta^{13}\text{C}$) of organisms using proteomic data (Kleiner et al. 2017). Isotopic signatures are typically measured by Isotope Ratio Mass Spectrometry (IRMS). Unlike IRMS, Protein-SIF has the advantage that it can estimate the isotopic signatures of multiple organisms within the same sample and therefore is a valuable tool to study complex systems such as the microbiota.

SIFs are computed from metaproteomic data using the calis-p 2.0 software (Kleiner et al. 2021). The Calis-p software requires two input files: raw spectral files produced by the LC-MS/MS, and the corresponding peptide-spectrum match (PSM) files containing the protein identifications and quantifications. The calis-p software then performs fully automatically. Calis-p performs two main steps: isotopic pattern extraction and SIF computation. The software first filters out “ambiguous” peptides of low identification confidence from the input PSM file. Then, for each remaining peptide identification, the software finds the corresponding mass spectrum and extracts the isotopic pattern. After some clean-up and filtering steps, the software compares the experimentally derived isotopic pattern to *in silico* derived isotopic patterns to infer $\delta^{13}\text{C}$ values for the peptide. This step is repeated for many peptides of each organism. Finally, the median $\delta^{13}\text{C}$ of the peptides from an organism is used to estimate that organism’s signature. At least 30 peptides are required to estimate an organism’s SIF accurately.

The resulting values need to be corrected for fractionation caused by the mass spectrometer using internal standards. $\delta^{13}\text{C}$ values for the standards (typically human hair) are obtained both with protein-SIF and IRMS to calculate the offset between the two measurement methods.

REFERENCES

- Gouveia, Duarte, Olivier Pible, Karen Culotta, Virginie Jouffret, Olivier Geffard, Arnaud Chaumot, Davide Degli-Esposti, and Jean Armengaud. 2020. "Combining Proteogenomics and Metaproteomics for Deep Taxonomic and Functional Characterization of Microbiomes from a Non-Sequenced Host." *Npj Biofilms and Microbiomes* 6 (1): 1–6. <https://doi.org/10.1038/s41522-020-0133-2>.
- Jehmlich, Nico, Frank Schmidt, Mathias Hartwich, Martin von Bergen, Hans-Hermann Richnow, and Carsten Vogt. 2008. "Incorporation of Carbon and Nitrogen Atoms into Proteins Measured by Protein-based Stable Isotope Probing (Protein-SIP)." *Rapid Communications in Mass Spectrometry* 22 (18): 2889–97. <https://doi.org/10.1002/rcm.3684>.
- Justice, Nicholas, Z Li, Y Wang, S. E. Spaulding, A. C. Mosier, R. L. Hettich, C. Pan, and J. F. Banfield. 2014. "(15)N- and (2)H Proteomic Stable Isotope Probing Links Nitrogen Flow to Archaeal Heterotrophic Activity." *Environmental Microbiology* 16 (10): 3224–37. <https://doi.org/10.1111/1462-2920.12488>.
- Kleiner, Manuel. 2019. "Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities." *MSystems* 4 (3): e00115-19. <https://doi.org/10.1128/mSystems.00115-19>.
- Kleiner, Manuel, Xiaoli Dong, Tjorven Hinzke, Juliane Wippler, Erin Thorson, Bernhard Mayer, and Marc Strous. 2018. "Metaproteomics Method to Determine Carbon Sources and Assimilation Pathways of Species in Microbial Communities." *Proceedings of the National Academy of Sciences* 115 (24): E5576–84. <https://doi.org/10.1073/pnas.1722325115>.
- Kleiner, Manuel, Angela Kouris, Marlene Jensen, Yihua Liu, Janine McCaLder, and Marc Strous. 2021. "Ultra-Sensitive Protein-SIP to Quantify Activity and Substrate Uptake in Microbiomes with Stable Isotopes." *BioRxiv*, March, 2021.03.29.437612. <https://doi.org/10.1101/2021.03.29.437612>.
- Kleiner, Manuel, Erin Thorson, Christine E. Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. 2017. "Assessing Species Biomass Contributions in Microbial Communities via Metaproteomics." *Nature Communications* 8 (1): 1558. <https://doi.org/10.1038/s41467-017-01544-x>.
- Lee, Pey Yee, Siok-Fong Chin, Hui-min Neoh, and Rahman Jamal. 2017. "Metaproteomic Analysis of Human Gut Microbiota: Where Are We Heading?" *Journal of Biomedical Science* 24 (1): 1–8. <https://doi.org/10.1186/s12929-017-0342-z>.
- Long, Shuping, Yi Yang, Chengpin Shen, Yiwen Wang, Anmei Deng, Qin Qin, and Liang Qiao. 2020. "Metaproteomics Characterizes Human Gut Microbiome Function in Colorectal Cancer." *Npj Biofilms and Microbiomes* 6 (1): 1–10. <https://doi.org/10.1038/s41522-020-0123-4>.

- Pelz, Oliver, Luis A. Cifuentes, Beth T. Hammer, Cheryl A. Kelley, and Richard B. Coffin. 1998. "Tracing the Assimilation of Organic Compounds Using $\Delta^{13}\text{C}$ Analysis of Unique Amino Acids in the Bacterial Peptidoglycan Cell Wall." *FEMS Microbiology Ecology* 25 (3): 229–40. <https://doi.org/10.1111/j.1574-6941.1998.tb00475.x>.
- Pible, Olivier, François Allain, Virginie Jouffret, Karen Culotta, Guylaine Miotello, and Jean Armengaud. 2020. "Estimating Relative Biomasses of Organisms in Microbiota Using 'Phylopeptidomics.'" *Microbiome* 8 (1): 30. <https://doi.org/10.1186/s40168-020-00797-x>.
- Salvato, Fernanda, Robert L. Hettich, and Manuel Kleiner. 2021. "Five Key Aspects of Metaproteomics as a Tool to Understand Functional Interactions in Host-Associated Microbiomes." *PLOS Pathogens* 17 (2): e1009245. <https://doi.org/10.1371/journal.ppat.1009245>.
- Starr, Amanda E., Shelley A. Deeke, Leyuan Li, Xu Zhang, Rachid Daoud, James Ryan, Zhibin Ning, et al. 2017. "Proteomic and Metaproteomic Approaches to Understand Host–Microbe Interactions" 90 (1): 86–109. <https://doi.org/10.1021/acs.analchem.7b04340>.
- Tanca, Alessandro, Antonio Palomba, Massimo Deligios, Tiziana Cubeddu, Cristina Fraumene, Grazia Biosa, Daniela Pagnozzi, Maria Filippa Addis, and Sergio Uzzau. 2013. "Evaluating the Impact of Different Sequence Databases on Metaproteome Analysis: Insights from a Lab-Assembled Microbial Mixture." *PLOS ONE* 8 (12): e82981. <https://doi.org/10.1371/journal.pone.0082981>.
- Van Den Bossche, Tim, Benoit J. Kunath, Kay Schallert, Stephanie S. Schäpe, Paul E. Abraham, Jean Armengaud, Magnus Ø Arntzen, et al. 2021. "Critical Assessment of Metaproteome Investigation (CAMPI): A Multi-Lab Comparison of Established Workflows." *BioRxiv*, March, 2021.03.05.433915. <https://doi.org/10.1101/2021.03.05.433915>.
- Verberkmoes, Nathan C., Alison L. Russell, Manesh Shah, Adam Godzik, Magnus Rosenquist, Jonas Halfvarson, Mark G. Lefsrud, et al. 2009. "Shotgun Metaproteomics of the Human Distal Gut Microbiota." *The ISME Journal* 3 (2): 179–89. <https://doi.org/10.1038/ismej.2008.108>.
- Wiśniewski, Jacek R., Alexandre Zougman, Nagarjuna Nagaraj, and Matthias Mann. 2009. "Universal Sample Preparation Method for Proteome Analysis." *Nature Methods* 6: 359. <https://doi.org/10.1038/nmeth.1322>.
- Xiong, Weili, Paul E. Abraham, Zhou Li, Chongle Pan, and Robert L. Hettich. 2015. "Microbial Metaproteomics for Characterizing the Range of Metabolic Functions and Activities of Human Gut Microbiota." *Proteomics* 15 (20): 3424–38. <https://doi.org/10.1002/pmic.201400571>.
- Zhang, Xu, Wendong Chen, Zhibin Ning, Janice Mayne, David Mack, Alain Stintzi, Ruijun Tian, and Daniel Figeys. 2017. "Deep Metaproteomics Approach for the Study of Human Microbiomes" 89 (17): 9407–15. <https://doi.org/10.1021/acs.analchem.7b02224>.
- Zhang, Xu, Shelley A. Deeke, Zhibin Ning, Amanda E. Starr, James Butcher, Jennifer Li, Janice

Mayne, et al. 2018. “Metaproteomics Reveals Associations between Microbiome and Intestinal Extracellular Vesicle Proteins in Pediatric Inflammatory Bowel Disease.” *Nature Communications* 9 (1): 1–14. <https://doi.org/10.1038/s41467-018-05357-4>.

Zhang, Xu, Leyuan Li, Janice Mayne, Zhibin Ning, Alain Stintzi, and Daniel Figeys. 2018. “Assessing the Impact of Protein Extraction Methods for Human Gut Metaproteomics.” *Journal of Proteomics, Proteomics in Infectious Diseases*, 180 (May): 120–27. <https://doi.org/10.1016/j.jprot.2017.07.001>.

**CHAPTER 3 - EVALUATION OF SAMPLE PRESERVATION AND STORAGE
METHODS FOR METAPROTEOMICS ANALYSIS OF INTESTINAL MICROBIOMES**

Authors:

Angie Mordant¹, Manuel Kleiner¹

Authors' Affiliations:

1: Department of Plant and Microbial Biology, North Carolina State University, Raleigh NC

***Corresponding Author Information**

Angie Mordant - almordan@ncsu.edu

Manuel Kleiner - manuel_kleiner@ncsu.edu

Publication Status: accepted by Microbiology Spectrum

3.1 Abstract

A critical step in studies of the intestinal microbiome using meta-omics approaches is the preservation of samples before analysis. Preservation is particularly important for approaches that measure gene expression, such as metaproteomics, which is used to identify and quantify proteins in microbiomes. The "gold standard" for preserving fecal samples is flash-freezing followed by storage at -80°C , but some experimental set-ups do not allow for immediate freezing of samples. For example, flash-freezing is not possible in gnotobiotic isolators or with individuals sampling at home or in remote locations around the globe. This study aimed to develop and evaluate methods to preserve fecal microbiome samples for metaproteomics analyses when flash-freezing is not an option.

We collected fecal samples from conventional adult C57BL/6 mice and stored them for 1 and 4 weeks using the following preservation methods: flash-freezing in liquid nitrogen, immersion in *RNAlater*TM, immersion in 95% ethanol, immersion in a homemade *RNAlater*-like buffer, and combinations of these methods. After storage, we extracted protein and prepared peptides for LC-MS/MS analysis to identify and quantify peptides and proteins. All samples produced highly similar metaproteomes, regardless of the preservation method, except for ethanol-preserved samples that were distinct from all other samples in terms of protein identifications and protein abundance profiles. Flash-freezing and *RNAlater*TM (or *RNAlater*-like treatments) produced metaproteomes that differed only slightly, with less than 0.7% of identified proteins differing in abundance. On the other hand, ethanol preservation resulted in an average of 9.5% of the identified proteins differing in abundance between ethanol and the other treatments.

Our results suggest that preservation at room temperature in *RNAlater*TM, or an *RNAlater*-like solution, performs as well as flash-freezing to preserve intestinal microbiome samples before metaproteomics analyses.

3.2 Introduction

The intestinal microbiome is a highly diverse and metabolically active community that profoundly affects its host (Sommer and Bäckhed 2013). This complex community influences the health of its host by altering the availability of nutrients (Rowland et al. 2018; Gentile and Weir 2018; Wikoff et al. 2009; Turnbaugh et al. 2006) and the host's susceptibility to infection and disease (Boursier et al. 2016; Theriot et al. 2014). The intestinal microbiome is integral to proper host immune function (Montgomery et al. 2020; Belkaid and Hand 2014; Hooper, Littman, and Macpherson 2012) and host well-being (Chevalier et al. 2020; Rogers et al. 2016). So far, most studies have used DNA sequencing and taxonomy-based approaches to study the intestinal microbiome providing critical insights into taxonomic shifts in the community-related host genotype, diet, and disease state (Ley et al. 2005; David et al. 2014; Turnbaugh et al. 2009; Bajaj et al. 2014). Taxonomic shifts, however, have been found to not always measure important functional shifts in the microbiome, as different taxa can perform the same function (Blakeley-Ruiz et al. 2019) and highly similar strains can perform different functions by encoding a few unique gene clusters (Patnode et al. 2019). Therefore the use of function-focused multi-omics approaches is essential for understanding the role of the intestinal microbiome in health and disease (Wang et al. 2020; Heintz-Buschart and Wilmes 2018; Jansson and Baker 2016; Hettich et al. 2013).

Metaproteomics is a valuable tool to study interactions in the intestinal microbiome and the microbiome's influence on host health (Salvato, Hettich, and Kleiner 2021; Kleiner 2019; Starr et al. 2017). Metaproteomics allows for the identification and quantification of large numbers of microbial, dietary, and host proteins in microbiome samples in a high-throughput fashion (Xiong et al. 2015; Verberkmoes et al. 2009; Zhang et al. 2017). Because proteins are central to all biological processes, metaproteomics provides direct evidence of the activities and functions of microbial community members and their contributions to disease (Lee et al. 2017). For example, metaproteomics revealed protein biomarkers of disease in Inflammatory Bowel Disease (IBD) and colorectal cancer and provided insights into the role of the microbiome in such diseases (Long et al. 2020; Zhang, Deeke, et al. 2018). In addition to quantifying differences in protein abundances between samples, metaproteomics can also be used to assess microbial community structure based on proteinaceous biomass (Kleiner et al. 2017; Pible et al. 2020; Gouveia et al. 2020; Van Den Bossche et al. 2021) and track incorporation of specific substrates using stable isotope content of peptides (Kleiner et al. 2021; 2018; Justice et al. 2014; Jehmlich et al. 2008).

Metaproteomics workflows are complex and variability can be introduced at every step, from sample preparation to data acquisition by LC-MS/MS and data processing (Van Den Bossche et al. 2021). There is no standardized workflow for metaproteomics of intestinal microbiome samples but some of the individual steps have been optimized in the past, such as protein extraction (Zhang, Li, et al. 2018), database creation (Tanca et al. 2016), and database searching (Tyanova et al. 2016). A critical step that has not yet been optimized is the storage and preservation of samples before analysis. Adequate storage of samples is critical because exposure to environmental changes could induce changes in protein profiles of species in the

samples and thus provide misleading study results. For example, exposing samples to air can strongly bias colorectal cancer studies because oxidative stress and enrichment of bacterial superoxide dismutase enzymes that will occur from air exposure are also characteristics of colorectal cancer in the intestinal tract (Long et al. 2020). Therefore, appropriately storing samples immediately upon collection helps to avoid post-collection changes in protein abundances. A suitable storage method should preserve the information contained in the microbiome at the time of sampling without introducing substantial bias.

Typically, microbiome samples are frozen immediately upon collection, with flash-freezing in liquid nitrogen followed by storage at -80°C . However, flash-freezing is not always possible, and very little is known about suitable alternatives to flash-freezing to preserve microbiome samples before metaproteomics analyses. For example, clinical or diet studies involving human subjects usually require the subjects to sample themselves at their homes where they do not have access to liquid nitrogen (Jordan et al. 2017; Kim et al. 2017). It can be challenging to maintain sample integrity also in resource-limited fieldwork conditions (Carruthers et al. 2019) or out in a wild animal's environment with no access to liquid nitrogen and cold storage (Hale et al. 2015). Even in the laboratory, immediate freezing in liquid nitrogen can be difficult. One specific case is work with gnotobiotic animals, which are invaluable models to study and manipulate the microbiota in a controlled system. Gnotobiotic animals reside in isolators where everything (food, bedding, etc.) entering the isolators is sterilized through autoclaving, irradiation, or strong chemicals before being introduced through a two-ended port (Williams 2014). Removing samples from the isolators causes long delays between sampling and sample storage, thus exposing samples to environmental changes (air exposure, temperature change, nutrient depletion, etc.) before they can be adequately stored.

Several studies have been conducted to test the effects of preservation methods on nucleic acids, but the effects of such methods on proteins are poorly understood. Freezing at -80°C is thought to maintain sample integrity similar to fresh samples (Fouhy et al. 2015; Carroll et al. 2012). The effectiveness of storage at room temperature is typically evaluated based on comparisons to frozen treatments. *RNAlater*TM is a popular storage solution that has been shown to be effective at preserving DNA in gut microbiome studies, with negligible differences compared to freezing (Tap et al. 2019; Blekhman et al. 2016). Ethanol (95% or absolute) may also be suitable for the preservation of nucleic acids in microbiome samples before taxonomic profiling as long as it is used consistently (Carruthers et al. 2019; Hale et al. 2015; Song et al. 2016). However, there are studies in which these storage solutions significantly biased the downstream results (Choo, Leong, and Rogers 2015), particularly in RNA sequencing studies (Passow et al. 2018) and thus great care should be taken when selecting a preservation method. While the effects of sample preservation on nucleic acids have been extensively studied, to the best of our knowledge, only two studies have investigated the effects of sample preservation on protein profiles. First, Saito and colleagues demonstrated that *RNAlater*TM has the potential to preserve proteomes as effectively as immediate freezing (Saito et al. 2011). Their results are promising, however, their study was performed on a single marine microorganism (cyanobacterium *Synechococcus* WH8102) and thus does not indicate whether *RNAlater*TM would preserve samples as complex as those from the intestinal microbiome. Furthermore, their study was performed in earlier stages of proteomics when replication was expensive and effort-intensive. For that reason, the researchers only included technical replication and therefore they could not assess the robustness of *RNAlater*TM in terms of within-treatment consistency. Second, Hickl and colleagues observed vast differences in the identifications and relative abundances of

proteins from human fecal samples depending on the preservation and storage procedure applied to the samples (Hickl et al. 2019). They tested two preservation and storage methods: a flash-freezing-based approach “FF” and RNAlater, “RL”. The first method “FF”, consisted of flash-freezing in liquid nitrogen followed by storage at -80°C, cryomilling, and storage at -20°C for 16 h while immersed in RNAlaterICE®. The “RL” method simply consisted of immersion in refrigerated RNAlater™ for 6 h. They found less than 50% overlap in protein identification between the two treatments. Of the overlapping proteins, they found roughly 2,000 proteins to significantly differ in abundance by more than 1.5 fold between the two treatments. The majority of the differences they observed were attributed to taxonomy. For example, class Actinobacteria represented about 20% of the composition of the FF samples while Actinobacteria only made up less than 2% of the composition of the RL samples. However, one cannot attribute the differences Hickl et al. observed to a specific aspect of the preservation and storage process because of the many variables in the sample processing. For example, the flash-frozen samples were cryomilled which could favor lysis of gram-positive bacteria, such as Actinobacteria, and may explain the large difference in the relative abundance of this taxon (de Boer et al. 2010).

This study aimed to (1) compare the effects of different sample preservation methods on intestinal microbiome metaproteomes, (2) evaluate comparable aspects of sample processing by limiting the number of variables, (3) assess within-treatment variability, and (4) evaluate the methods over a longer period of preservation/storage time.

3.3 Materials and methods

3.3.1 Preparation of fecal master mix and preservation treatments

Fresh fecal pellets were collected from 12 conventional 5-month old C57BL/6 mice obtained from The Jackson Laboratory. The pellets were pooled and homogenized using a spatula to make a fecal master mix and thus remove inter-individual variation as a variable. The master mix was split into aliquots of 8 mg each. The aliquots were either resuspended in 200 μ L of a preservation solution and stored at room temperature ($\sim 22^{\circ}$ C) in the dark or were flash-frozen using liquid nitrogen and stored at -80° C. The preservation treatments are discussed in detail below. In brief, we tested six preservation treatments: flash-freezing in liquid nitrogen followed by storage at -80° C (“FF” treatment); preservation at room temperature in RNAlater™ (RNAlater™ Stabilization Solution, Invitrogen; “R” treatment); a combination of RNAlater™ and flash freezing (“RF” treatment); an RNAlater™-like preservation buffer (“NAP buffer”) as described in (Menke et al. 2017) (“N” treatment); this same “NAP buffer” autoclaved (“AN” treatment); and 95% ethanol (“E” treatment). We tested the effectiveness of the treatments at preserving microbiome samples over two storage durations: 1 week and 4 weeks. We prepared four replicate samples per treatment and time point, with each replicate being an 8 mg aliquot of the fecal master mix described above.

Flash-freezing and storage at -80° C (“FF”)

Immediate freezing followed by storage at -80° C is the method most frequently used to preserve biological specimens and is regarded as the “gold standard” approach. Fouhy and colleagues observed that immediate freezing retains information similar to fresh samples in a 16S rRNA gene amplicon sequencing experiment of healthy human fecal samples (Fouhy et al. 2015). The only significant differences they observed between frozen and fresh samples were in

the relative abundances of the genera *Faecalibacterium* and *Leuconostoc*; however, the differences were subtle and may be attributable to a batch effect in DNA extraction rather than sample preservation. The effectiveness of storage solutions used at room temperature is typically evaluated based on comparisons to frozen treatments (Saito et al. 2011).

Immersion in RNAlater™ and storage at room temperature (“R”)

RNAlater™ Stabilization Solution is a popular storage reagent. Its effectiveness can be attributed to its ability to permeate tissue to stabilize and protect RNA quickly. RNAlater™ is effective at preserving nucleic acids for intestinal microbiome studies, with negligible bias compared to freezing (Tap et al. 2019). In addition, RNAlater™ has the potential to preserve proteins because its main component is ammonium sulfate, and ammonium sulfate precipitates proteins which can later be re-solubilized without degradation. Saito et al. demonstrated that RNAlater™ is effective at preserving the proteome of the marine cyanobacterium *Synechococcus* WH8102 (Saito et al. 2011). In our study, we immersed samples in RNAlater™ (Invitrogen) in a 1:10 sample: solution ratio and then stored them at room temperature (~22° C) in the dark.

Immersion in RNAlater™ and flash-freezing followed by storage at room temperature (“RF”)

To determine whether the use of a storage solution makes a difference as compared to storing samples dry at -80° C, we immersed RF samples in RNAlater™ (Invitrogen) and flash-froze the tubes in liquid nitrogen before storing them at -80° C. Observed differences between R and RF samples would provide evidence regarding the effects of freezing on sample integrity.

Immersion in NAP buffer and storage at room temperature (“N”)

The major limitation of RNAlater™ is its high cost. It has been demonstrated that RNAlater™-like buffers work as effectively as the commercially available solution (Camacho-

Sanchez et al. 2013). Menke and colleagues even argue that their RNAlater-like solution called “Nucleic Acid Preservation (NAP) buffer” outperformed commercial RNAlater™ in preserving DNA for 16S rRNA gene amplicon sequencing experiments based on comparisons with immediately-frozen controls (Menke et al. 2017). “NAP buffer” was included as a treatment in this study and was prepared as previously described by Camacho-Sanchez et al. (Camacho-Sanchez et al. 2013). Briefly, 1.5 L of NAP buffer (pH 5.2) contained 935 ml of ultrapure water, 700 g of ammonium sulfate, 25 ml of 1 M sodium citrate, and 40 ml of 0.5 M ethylenediaminetetraacetic acid (EDTA). We prepared the solution fresh two days before the experiment. We immersed the samples in the NAP buffer solution in a 1:10 sample: solution ratio before storing them at room temperature (~22° C) in the dark.

Immersion in Autoclaved NAP buffer and storage at room temperature (“AN”)

RNAlater™ and RNAlater-like buffers do not need to be autoclaved because their chemical composition inhibits the growth of contaminants. Manufacturers of RNAlater™ recommend against autoclaving the reagent. However, in some cases, such as when working with gnotobiotic isolators, the solution needs to be autoclaved to prevent the introduction of microorganisms into the isolators. We tested an autoclaved version of the NAP buffer as an additional treatment to simulate real experimental conditions with gnotobiotic isolators. The same solution described above as the N treatment, from the same batch, was autoclaved (60 min at 121.5° C) two days before the start of the experiment. We immersed the samples in the autoclaved NAP buffer solution in a 1:10 sample: solution ratio before storing them at room temperature (~22° C) in the dark.

Immersion in 95% Ethanol and storage at room temperature (“E”)

Alcohol preservation is a common method in which biological specimens are preserved by dehydration. Hale and colleagues (Hale et al. 2015) found that absolute ethanol worked as well as immediate freezing of DNA for preserving samples prior to metagenomics analysis. The effectiveness of ethanol as a preservation treatment depends on its concentration. Sinha et al. (Sinha et al. 2016) observed low stability of microbial DNA when preserved in 70% ethanol. Saito et al. (Saito et al. 2011) observed that 90% was not ideal for the preservation of the proteome of the marine cyanobacterium *Synechococcus* WH8102 as only ~75% of the proteins were recovered as compared to flash-freezing. Because the organism studied by Saito et al. (Saito et al. 2011) is very different from the intestinal microbiome, we included ethanol (95%) as a treatment. We prepared 95% ethanol by mixing pure anhydrous (200 proof/100%) ethyl alcohol (Koptec) with ultrapure water (Optima™ LC/MS Grade, Fisher Chemical™).

3.3.2 Protein extraction and peptide preparation

We prepared samples for metaproteomics analysis at two time-points: after storing the samples for 1 week and after 4 weeks. We removed the storage solutions from the samples by centrifugation at 21,000 x g for 5 min and then resuspended the samples in 400 µl of SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT]. Cells were lysed by bead-beating in lysing matrix E tubes (MP Biomedicals) with a Bead Ruptor Elite (Omni International) for 5 cycles of 45 sec at 6.45 m/s with 1 min dwell time between cycles, followed by heating at 95° C for 10 min. The lysates were centrifuged for 5 min at 21,000 x g to remove cell debris. We prepared peptides according to the filter-aided sample preparation (FASP) protocol described by (Wiśniewski et al. 2009). All centrifugations mentioned below were performed at 14,000 x g.

Samples were loaded onto 10 kDa MWCO 500 µl centrifugal filters (VWR International) by combining 60 µl of lysate with 400 µl of Urea solution (8 M urea in 0.1 M Tris/HCl pH 8.5) and centrifuging for 30 min. This step was repeated twice until the filter capacity was reached. Filters were washed twice by applying 200 µl of urea solution followed by 40 min of centrifugation. 100 µl IAA solution (0.05 M iodoacetamide in Urea solution) was then added to filters for a 20 min incubation followed by centrifugation for 20 min. The filters were washed three times with 100 µl of urea solution and 20 min centrifugations, followed by buffer exchange to ABC (50 mM Ammonium Bicarbonate). Buffer exchange was accomplished by adding 100 µl of ABC and centrifuging three times followed by centrifugation for 20 min. Tryptic digestion was performed by adding 0.85 µg of MS grade trypsin (Thermo Scientific Pierce, Rockford, IL, USA) in 40 µl of ABC to the filters and incubating for 16 hours in a wet chamber at 37° C. The tryptic peptides were eluted by adding 50 µl of 0.5 M NaCl and centrifuging for 20 min. Peptide concentrations were determined with the Pierce Micro BCA assay (Thermo Fisher Scientific) following the manufacturer's instructions.

3.3.3 LC-MS/MS

Samples were analyzed by 1D-LC-MS/MS using a published method (Speare et al. 2020) with several modifications. The samples were blocked and randomized according to Oberg and Vitek's method (Oberg and Vitek 2009) to control for batch effects. For each sample, 600 ng of tryptic peptides were loaded with an UltiMate™ 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) in loading solvent A (2 % acetonitrile, 0.05 % trifluoroacetic acid) onto a 5 mm, 300 µm ID C18 Acclaim® PepMap100 pre-column and desalted (Thermo Fisher Scientific). Peptides were then separated on a 75 cm x 75 µm analytical EASY-Spray column

packed with PepMap RSLC C18, 2 μm material (Thermo Fisher Scientific) heated to 60 °C via the integrated column heater at a flow rate of 300 nl min^{-1} using a 140 min gradient going from 95 % buffer A (0.1 % formic acid) to 31 % buffer B (0.1 % formic acid, 80 % acetonitrile) in 102 min, then to 50 % B in 18 min, to 99 % B in 1 min and ending with 99 % B. Carryover was reduced by wash runs (injection of 20 μl acetonitrile with 99% eluent buffer B) between samples.

The analytical column was connected to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) via an Easy-Spray source. Eluting peptides were ionized via electrospray ionization (ESI). MS^1 spectra were acquired by performing a full MS scan at a resolution of 60,000 on a 380 to 1600 m/z window. MS^2 spectra were acquired using a data-dependent approach by selecting for fragmentation the 15 most abundant ions from the precursor MS^1 spectra. A normalized collision energy of 25 was applied in the HCD cell to generate the peptide fragments for MS^2 spectra. Other settings of the data-dependent acquisition included: a maximum injection time of 100 ms, a dynamic exclusion of 25 sec, and exclusion of ions of +1 charge state from fragmentation. About 60,000 MS/MS spectra were acquired per sample.

3.3.4 Protein Identification Database

We constructed a protein sequence database for identifying proteins from the four main components of the sample: the host, wheat (the main component of mouse chow), the microbiota, and potential contaminants. Protein sequences of the mouse host, *Mus musculus*, were downloaded from Uniprot (<https://www.uniprot.org/proteomes/UP000000589>). Protein sequences of wheat, *Triticum aestivum*, were downloaded from Uniprot

(<https://www.uniprot.org/proteomes/UP000019116>). A public database constructed by Xiao et al. (Xiao et al. 2015) was used for the microbiota sequences. While the use of such a reference database is not recommended for studies that address specific biological questions, as it has been shown that such reference databases can lead to lower identification rates and species and protein miss assignments (Timmins-Schiffman et al. 2017; Tanca et al. 2016), it is sufficient for determining the overall effects of sample preservation and preparation methods. The database from Xiao et al. contains ~2.6 million “non-redundant” genes from metagenomic sequencing of fecal material from 184 mice. The corresponding annotated protein sequences were downloaded from <http://gigadb.org/dataset/view/id/100114/token/mZIMYJIF04LshpgP>. The taxonomy (available as a separate file) was integrated into the string of the sequence descriptions using the join command in Linux. Most (67.8%) of the sequences were assigned a taxonomy at the phylum level, and 9.8% of the sequences were assigned at the genus level (Xiao et al. 2015). Initial analyses suggested the presence of sequences that were too similar for adequate discrimination in the downstream workflow so the protein sequences were clustered with an identity threshold of 95% using the CD-HIT tool (Li and Godzik 2006). About 8% of the sequences were combined into clusters, while the remaining ~92% remained as individual sequences. Also included in the database were sequences of common laboratory contaminants (<http://www.thegpm.org/crap/>). The database contained a total of 2,396,591 protein sequences and is included with the PRIDE submission for this study (PXD024115).

3.3.5 Protein identification and quantification

For peptide and protein identification, MS data were searched against the above-described database using the Sequest HT node in Proteome Discoverer version 2.3.0.523

(Thermo Fisher Scientific) with the following parameters: digestion with trypsin (Full), maximum of 2 missed cleavages, 10 ppm precursor mass tolerance, 0.1 Da fragment mass tolerance and maximum 3 equal dynamic modifications per peptide. We considered the following dynamic modifications: oxidation on M (+15.995 Da), carbamidomethyl on C (+57.021 Da), and acetyl on the protein N terminus (+42.011 Da). Peptide false discovery rate (FDR) was calculated using the Percolator node in Proteome Discoverer and only peptides identified at a 5% FDR were retained for protein identification. Proteins were inferred from peptide identifications using the Protein-FDR Validator node in Proteome Discoverer with a target FDR of 5%. From the generated multiconsensus dataset, we removed contaminant (cRAP sequences) and low confidence proteins (> 5 % FDR) and kept proteins that were identified by at least 1 protein unique peptide. To decrease the number of “one-hit wonder” proteins, we removed proteins that were not detected in a total of at least 3 samples (n = 47 samples in total), which is the minimum number of replicates in one treatment and time-point. The dataset contained 6,086 proteins after applying these filtering steps. The data were then normalized by calculating normalized spectral abundance factors (NSAFs, (Zybailov et al. 2006)) and multiplied by 100, to give the relative protein abundance in %.

3.3.6 Quality assessment and outlier analysis

We assessed data quality by first inspecting raw data in the Xcalibur™ Software (Thermo Fisher Scientific), then comparing the number of peptide spectrum matches (PSMs), peptides, proteins, and protein groups identified in each sample individually. We tested for statistical significance using a Student’s t-test (2-tailed, equal variability, FDR of 0.05). Samples in the dataset had on average 25,220 +/- 4.954 proteins identified at 5% FDR and 3,499 +/- 662

identified protein groups. Assuming the numbers of proteins identified per sample were normally distributed data, about 99.7% of the samples in the dataset were expected to have at least 10,358 detected proteins and 1,513 protein groups. These numbers correspond to the means stated above minus three standard deviations. One ethanol-preserved sample (“E6”) was deemed an outlier and was removed from the dataset because it had only 767 proteins and 84 protein groups. We suspect the protein extraction for that particular sample failed due to leaks in the filter unit during sample preparation.

3.3.7 Data analyses

To investigate the degree of overlap in protein identifications between treatments, we used the filtered dataset of 6,086 proteins. If a protein was detected in at least one sample of a treatment within this dataset, it was counted as identified in that treatment. We imported the accession codes of the identified proteins into Venny 2.1 (Oliveros, J.C. 2015) to create Venn diagrams representing the overlap between treatments in terms of identified proteins.

To identify differentially abundant proteins between treatments that are statistically significant, we performed a centered-log-ratio (CLR) transformation in R (version 4.0.2, `compositions_2.0-1` package)(Aitchison 1982) on peptide spectrum matches (PSMs) before performing statistical tests. We added 1 to every PSM value before performing the CLR transformation to protect against issues with missing values. Although CLR-normalized counts lose interpretability, CLR is a method better suited for statistical analyses of compositional data such as metaproteomic data (Fernandes et al. 2014; Gloor et al. 2017). Pairwise comparisons of all treatments were performed in the Perseus software platform (version 1.6.12.0) (Tyanova et al.

2016) using a Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5% ($S_0=0.1$, both sides, not paired).

We used a principal component analysis (PCA) to visualize how samples separate or cluster based on relative protein abundances. We performed the analysis in the Perseus software platform (version 1.6.12.0) (Tyanova et al. 2016) on the CLR-transformed dataset described above.

We investigated whether the preservation treatment affected the measured abundances of specific taxa by comparing the relative biomass contributions of the taxa. Biomass contributions of specific taxa were assessed at the phylum and genus levels using the method described in Kleiner et al. (Kleiner et al. 2017). Briefly, proteins were filtered for at least 2 protein unique peptides to increase the confidence in taxonomic identifications, and PSMs summed by taxon were used to estimate the biomass contribution of each taxon in the metaproteomes.

We investigated whether preservation treatments were biased towards proteins with specific biochemical characteristics such as isoelectric point (pI), molecular weight, or presence of transmembrane domains. First, we retrieved the pI and molecular weight associated with each identified protein from Proteome Discoverer and detected transmembrane domains by searching sequences of the identified proteins on the TMHMM 2.0 Server (Krogh et al. 2001). Then, we compared the distributions of these properties in each treatment as histograms with defined ranges.

Lastly, we assessed the amount of within-treatment variation using linear regression scatterplots in R (version 4.0.2; `psych_2.1.3` package). We fit the scatterplots onto the percent normalized spectral abundance factors (%NSAFs) for each pair of replicates that received the

same treatment (n=4, except for the ethanol treatment time point 4 weeks: n=3 because sample E6 was removed). We then compared the Pearson correlation coefficients.

3.3.8 Data availability

The mass spectrometry metaproteomics data and protein sequence database were deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository with the dataset identifier PXD024115.

3.4 Results

A fecal master mix (homogenate) was prepared from fecal samples of healthy adult conventional C57BL/6 mice. Aliquots of the master mix were randomized and then preserved using different methods. After 1 week and 4 weeks of storage, proteins were extracted and analyzed by LC-MS/MS. Preservation methods were assessed based on the variability between replicates and the degree of bias compared to the other methods, particularly flash freezing.

3.4.1 Minimal differences in total numbers of identified features for co-extracted samples

We compared the number of peptide spectrum matches (PSMs), peptides, proteins, and protein groups identified at a false discovery rate (FDR) of 5% between the different treatments and time-points (**Figure 1**) to determine whether the preservation treatment impacted the number of features identified. Samples preserved for 1 week and co-extracted as part of the first extraction batch did not significantly differ in their total counts, regardless of the preservation method. The numbers of features for the 4-week time-point (2nd extraction batch) and 1-week time-point differed slightly, but differences did not test significant except for the samples

preserved at -80°C in RNAlater™ (“RF” samples). This difference is likely due to batch effects in sample preparation and peptide quantification via microBCA assay because the 1-week and 4-week samples were prepared separately. At the 4-week time-point, flash-frozen (FF) samples preserved for 4 weeks at -80°C were significantly lower in their total counts compared to RF samples or samples preserved in the NAP buffer (N) or autoclaved NAP buffer (AN) at room temperature.

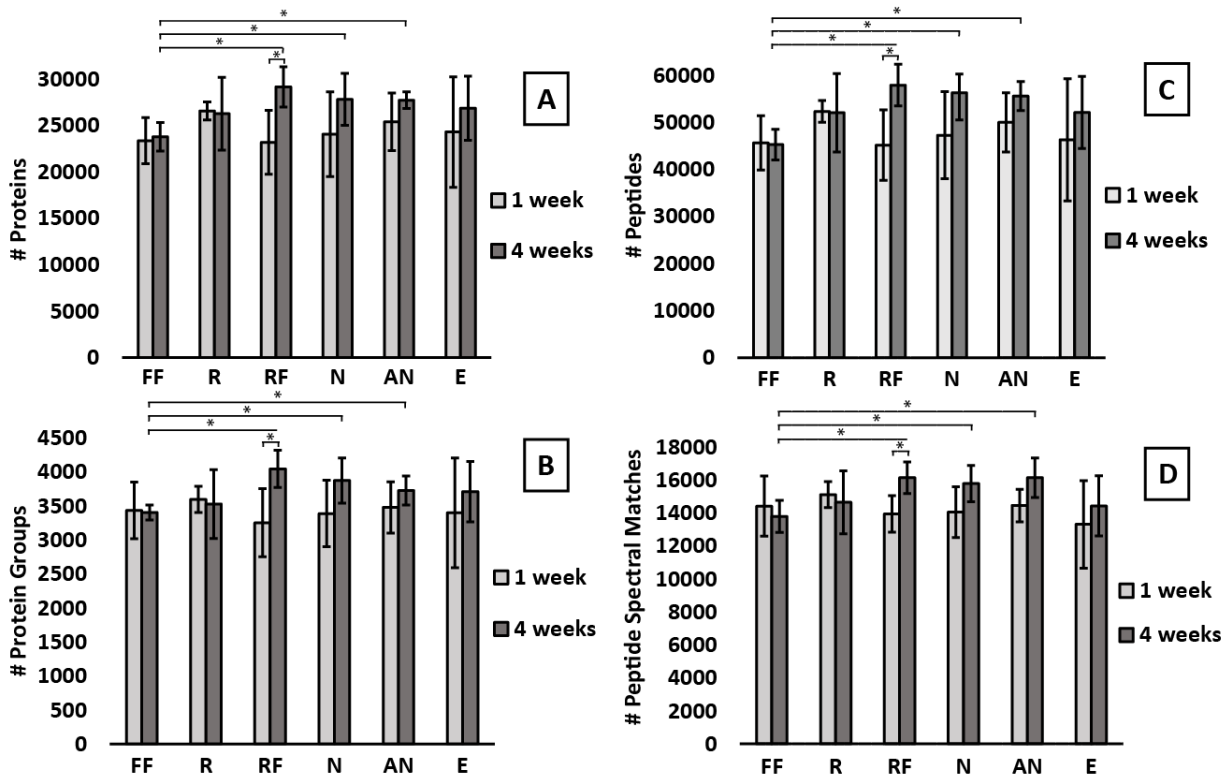


Figure 3-1. There were no significant differences in total numbers of PSMs, peptides, proteins, and protein groups between samples co-extracted after 1 week of preservation and only minimal differences existed in samples co-extracted after 4 weeks. FF = Flash-freezing, R = RNAlater, RF = RNAlater + flash freezing, N = NAP buffer, AN = Autoclaved NAP buffer, E = 95% Ethanol. 1 week = preserved for one week and first extraction batch. 4 weeks = preserved for 4 weeks and second extraction batch. Bars represent the arithmetic mean ($n = 4$ for all except E - 4 weeks where $n = 3$). Error bars represent standard deviation. Asterisks indicate statistical significance (Student's t-test, two-sided, $p < .05$). **A)** Total proteins identified at 5% FDR include the microbial, host, and dietary proteins. **B)** Total protein groups identified at 5% FDR. **C)** Total peptides identified at 5% FDR. **D)** Total peptide spectrum matches (PSMs) identified at 5% FDR.

3.4.2 Treatments shared over 76% of protein identifications

Comparing and quantifying the proteins identified by multiple treatments showed that most proteins were detected in every treatment. **Figure 2A** shows a Venn diagram of the four most different treatments in terms of physical and chemical properties: R, E, FF, and N. Each of these four treatments produced metaproteomes that identified the same 4,641 proteins (76.3% of the dataset) and uniquely identified ~1% of proteins. In the E treatment, 281 proteins or 4.6% of the protein identifications were not detected; this was the largest proportion of undetected proteins, followed by the FF treatment that did not detect 198 proteins or 3.3% of the protein identifications. **Figure 2B** represents the overlap of proteins between chemically similar treatments: R, RF, N, and AN. These four treatments produced metaproteomes that identified the same 4,878 proteins (80.4% of the dataset) and uniquely identified ~1% of proteins. The RF treatment was the most distinct of the comparison shown in **Figure 2B** (R, RF, N, AN) with 74 unique protein identifications (~1.2%) that were not detected in the other treatments.

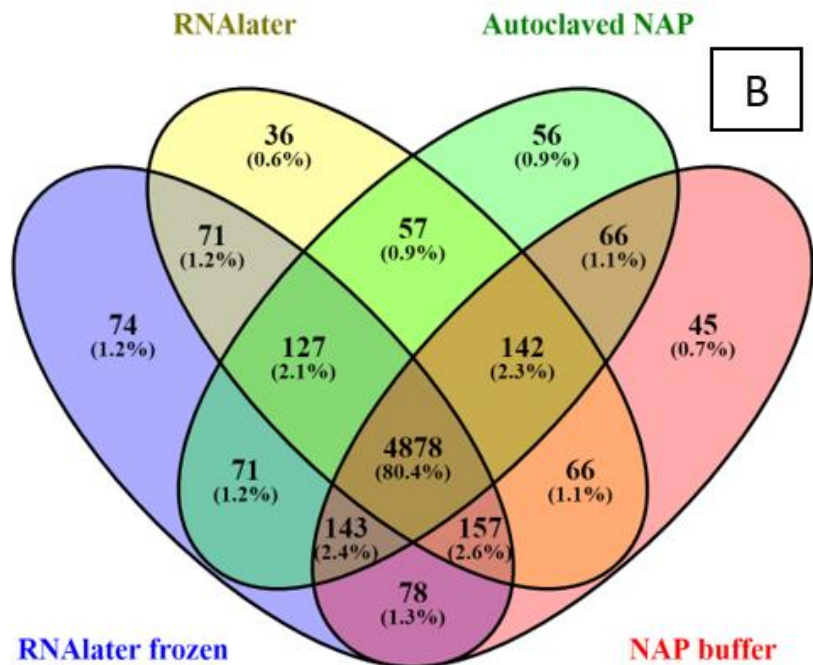
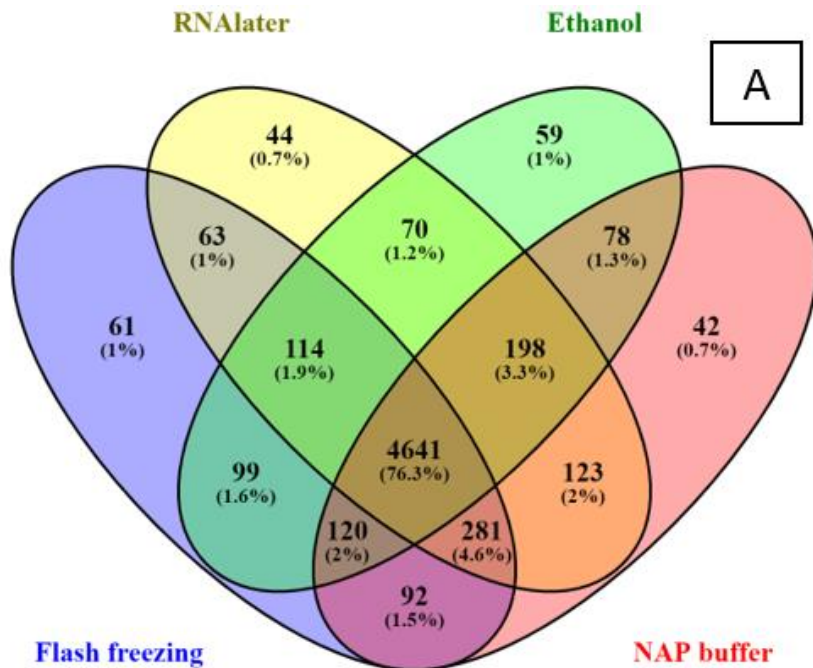
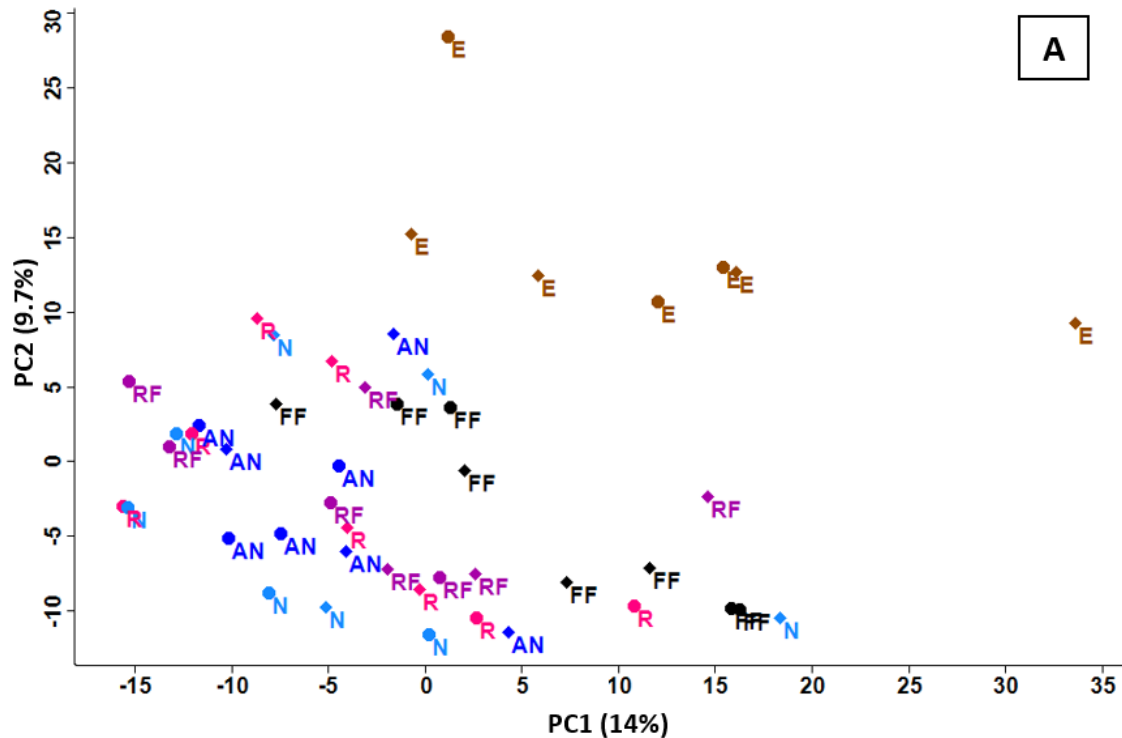


Figure 3-2. Over 76% of microbial, host and dietary protein identifications overlapped between treatments. Both time points were combined (n = 8 samples per treatment, except for Ethanol where n = 7). Proteins were included if they were identified with FDR <5%, at least one protein unique peptide, and present in at least three samples in the whole dataset. **A)** Comparison of treatments that differed most in terms of physical/chemical properties: Flash freezing (FF), RNAlater (R), 95% Ethanol (E), and NAP buffer (N). **B)** Comparison of the chemically similar treatments (RNAlater™ and RNAlater-like treatments): RNAlater (R), RNAlater frozen (RF), NAP buffer (N), Autoclaved NAP (AN).

3.4.3 Relative protein abundances were highly similar between all treatments except for the ethanol treatment

Principal component analysis (PCA), performed on the Centered-Log-ratio (CLR) transformed dataset of relative protein abundances, showed that the ethanol-preserved samples clustered together and clearly separated from samples of all other treatments (**Figure 3A**). If the preservation treatment did not affect protein profiles, we would expect to see no clustering, but rather the samples would be randomly distributed over the PCA plot.

We then compared CLR-transformed relative protein abundances between treatments using t-tests corrected for multiple hypothesis testing (two-sided, FDR of 0.05 and SO of 0.1) to identify proteins that significantly differed in abundance based on the preservation treatment (**Figure 3B**). We found no significant differences between RNAlater™(R) and the NAP buffer (N) or between the NAP buffer (N) and the autoclaved NAP buffer (AN). The proteins not shared between these treatments (**Fig 2A**) were sparse and lowly abundant proteins that were not significantly different from an undetected protein (**Additional file 2**). The ethanol treatment was the most distinct treatment with, on average, about 9.5% of the proteins significantly differing in abundance, with 247 proteins being less abundant, as compared to the other treatments (**Additional file 2**). Flash-frozen samples and the samples preserved in RNAlater or RNAlater-like solutions produced metaproteomes that differed only minimally (< 1% of proteins with different abundances).



	RF	R	N	AN	E
FF	17 (~0.28%)	41 (~0.67%)	16 (~0.26%)	30 (~0.49%)	474 (~7.79%)
RF		0	2 (~0.03%)	0	499 (~8.20%)
R			0	0	556 (~9.14%)
N				0	683 (~11.22%)
AN					680 (~11.17%)

Figure 3-3. Ethanol preserved samples were distinct from all other samples in their protein abundance profiles. **A)** Principal component analysis (PCA) of the relative protein abundances from each sample (CLR-transformed). E = ethanol; FF = flash-freezing; R = RNAlater™; RF = RNAlater™ and flash-freezing; N = NAP buffer; AN = autoclaved NAP buffer. Diamonds = 1-week; Circles = 4-weeks. **B)** Number of significant differences between each treatment (two-sided t-tests, permutation-based FDR of 0.05 and S0 of 0.1). A significant difference represents one protein that is more abundant in one treatment over the other for each paired comparison (refer to Additional file 2 for directionality). Percentages in parentheses indicate the percentage of significant proteins out of the total proteins considered (n = 6,086).

3.4.4 Within-treatment variability of relative protein abundances was low

We assessed the amount of within-treatment variability in terms of quantified protein abundances by fitting linear scatter plots for all replicates against all replicates and evaluating the Pearson correlation coefficients (**Additional file 6: Figures S1 - 6**). The means of the Pearson correlation coefficients (**Table 1**) showed a high correlation between replicates, indicating within treatment variability was low for all treatments.

Table 3-1. Linear correlation of replicates.

Mean Pearson coefficient of the linear correlation between replicates of the same treatment and preservation duration (n = 4, except for E 4 weeks: n = 3). We fit a linear model for each pair of samples within a treatment and time point in R (version 4. 0. 2; psych_2.1.3 package) using the dataset of percent normalized spectral abundance factors (%NSAFs).

	1 week	4 weeks
NAP buffer	0.937 +/- 0.024	0.963 +/- 0.006
Autoclaved NAP buffer	0.965 +/- 0.008	0.958 +/- 0.02
RNAlater™	0.957 +/- 0.01	0.957 +/- 0.017
RNAlater™ + Flash-freezing	0.938 +/- 0.034	0.967 +/- 0.008
Flash Freezing	0.961 +/- 0.009	0.943 +/- 0.019
95% Ethanol	0.957 +/- 0.01	0.95 +/- 0.015

3.4.5 Small but significant differences in the taxonomic composition of the metaproteomes based on preservation method

The relative taxonomic composition of the samples in terms of proteinaceous biomass contribution was consistent across replicates and preservation treatments (**Additional text and**

additional. fig. S7). The biomass contribution is shown per phylum in **Fig. 4A** and per genus in **Fig. 4B** for the most abundant genera: *Clostridium*, *Eubacterium*, *Butyrivibrio*, *Lactobacillus*, *Turcibacter*, *Blautia*, *Roseburia*, and *Coprococcus*. The abundances of specific taxa significantly differed at the phylum and genus levels. At the phylum level, Firmicutes was overrepresented in the ethanol-preserved samples as compared to the flash-frozen and NAP-buffer preserved samples. At the genus level, *Clostridium* and *Blautia* were subtly but significantly overrepresented in the ethanol-preserved samples as compared to all other treatments (t-test, paired, 2-tailed, $p < 0.05$). Furthermore, NAP buffer and Ethanol-preserved samples differed in their representation of the genus *Butyrivibrio*, and RNAlater and Ethanol-preserved samples differed in their representation of the genus *Roseburia*.

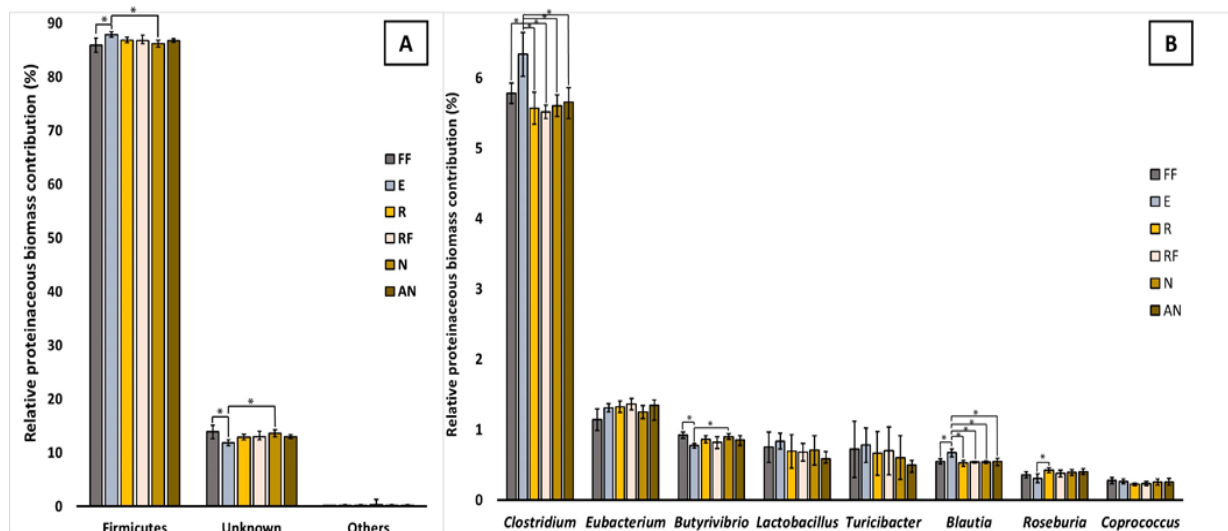


Figure 3-4. Small but significant differences in the representation of microbial taxa in the metaproteomes based on preservation methods. Bars represent the mean percent proteinaceous biomass for each taxon at the phylum level (A) or the genus level (B). Error bars represent the standard deviation ($n=8$, except $n=7$ for the Ethanol treatment). Asterisks represent statistical significance (t-test, paired, 2-tailed, $p < 0.05$). E = ethanol; FF = flash-freezing; R = RNAlater; RF = RNAlater and flash-freezing; N = NAP buffer; AN = autoclaved NAP buffer. The eight most abundant genera are displayed in the figure. Percentages are low because genus-level taxonomy could be assigned for $11.1 \pm 0.53\%$ ($n = 47$) of the total proteinaceous biomass in our samples, distributed over 28 different microbial genera.

3.4.6 The preservation methods did not bias towards specific biochemical properties of proteins.

We investigated whether the preservation treatment biased towards or against proteins with a specific isoelectric point (pI), molecular weight (**Additional Data 3**), or transmembrane domains (**Additional Data 4**) by comparing the distributions of these properties in each treatment. Distributions did not differ between treatments, indicating no bias (**Fig. 5**).

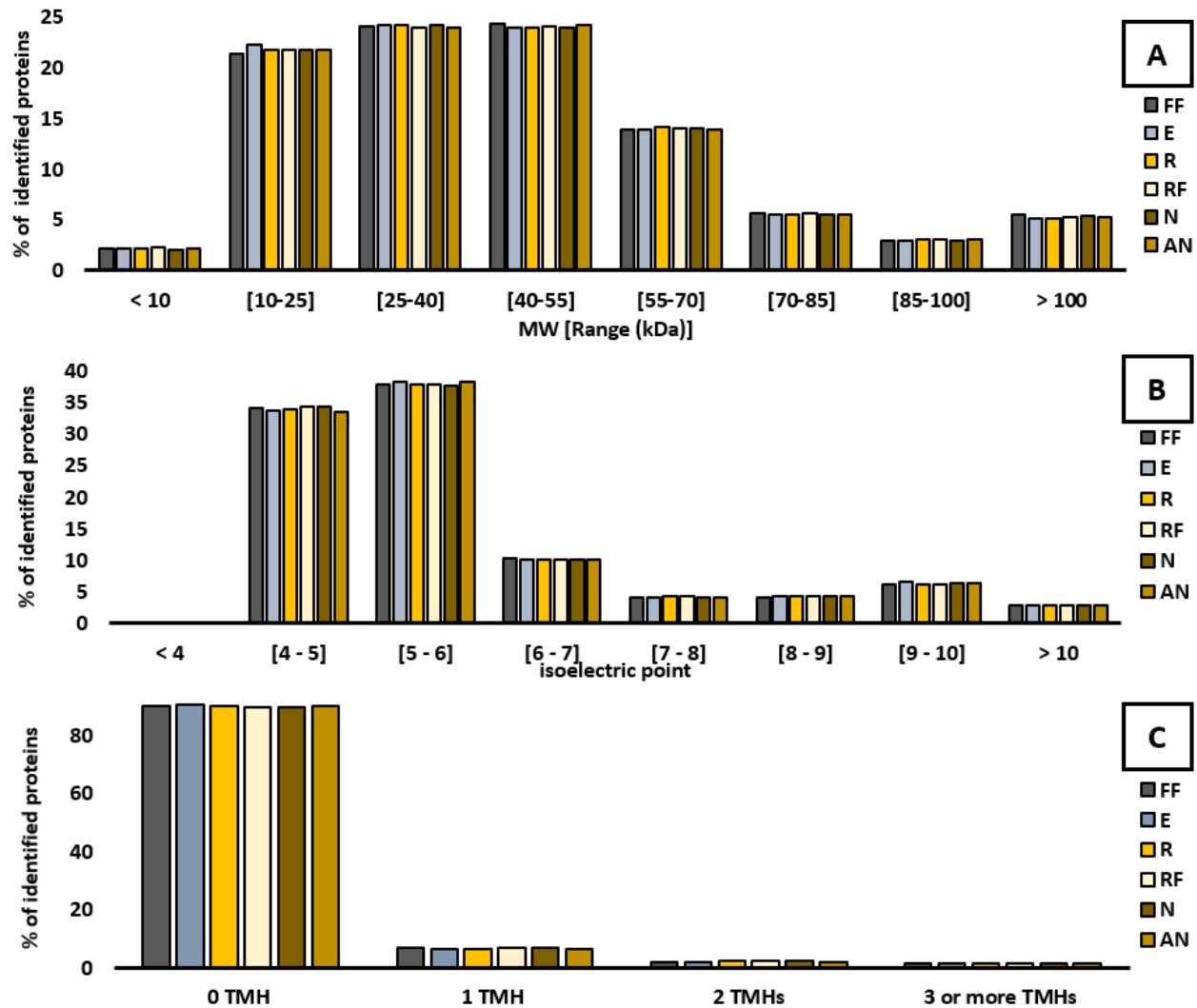


Figure 3-5. Distribution of biochemical properties of identified proteins. A) molecular weight (kDa) B) isoelectric point C) number of predicted transmembrane helices. Bars represent the proportion (%) of identified proteins belonging in each range. E = ethanol; FF = flash-freezing; R = RNAlater; RF = RNAlater and flash-freezing; N = NAP buffer; AN = autoclaved NAP buffer.

3.5 Discussion

This study evaluated the effects of different preservation treatments on the metaproteomes of intestinal microbiome samples to identify a preservation method suitable to use when flash-freezing is not an option. The data show that the metaproteomes of samples preserved at room temperature while immersed in *RNAlater*[™] or *RNAlater*-like solutions (NAP buffer and autoclaved NAP buffer) were highly similar to the metaproteomes of samples preserved by flash-freezing and storage at -80°C, with only negligible differences. On the other hand, samples preserved by immersion in 95% ethanol differed substantially from the flash-frozen samples and all the other samples. Because methods sharing the largest number of discoveries/values with the majority of the other methods tested may be more likely to produce valid results (Calgaro et al. 2020), our results suggest that the 95% ethanol treatment creates the largest bias in the metaproteomes, while the flash freezing, *RNAlater*[™], and *RNAlater*-like treatments are most likely to represent the protein profiles at the time of collection accurately. The majority of the differences of the ethanol treatment were found at the protein abundance level. Roughly 9.5% of all the proteins in the ethanol preserved samples differed in abundance from all other samples, suggesting that ethanol preservation could strongly influence study results. Taxonomic abundances also differed under the ethanol treatment, with the genera *Clostridium* and *Blautia* being overrepresented compared to all other treatment samples. However, although ethanol preservation introduces biases, within-treatment variability was low in ethanol-preserved samples. These results suggest that ethanol may be appropriate in some studies if it is used consistently.

While we tested a diversity of popular preservation methods, there are potentially other storage solutions and methods that could be used in addition to or instead of those tested and

described herein. There are, for example, a range of commercially available “microbiome” storage solutions designed for the preservation of fecal material for amplicon sequencing. These solutions could also be used for metaproteomics; however, their compatibility with the proteomic workflow and quality of preservation would have to be carefully tested, particularly as compatibility issues could arise if a preservation reagent is incompatible with standard proteomic workflows. For example, some preservatives contain guanidinium chloride, which forms a solid if it contacts SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT] that is used in many metaproteomic workflows.

Our results concur with the study by Saito et al. (Saito et al. 2011), which investigated sample preservation effects on the proteome of the cyanobacterium *Synechococcus* WH8102. They found *RNAlater*TM to be effective at preserving the proteome of a pure culture when compared to frozen storage. Here, we demonstrated that *RNAlater*TM can also be effective at preserving the metaproteomes of complex microbiome samples while minimizing storage effects. Furthermore, it appears that using the cost-effective *RNAlater*-like solution “NAP buffer” is a suitable alternative to the commercial *RNAlater*TM solution. Menke et al. (Menke et al. 2017) previously demonstrated that the NAP buffer effectively preserves the DNA of intestinal microbiome samples. Here we demonstrated that the NAP buffer effectively preserved proteins of intestinal microbiome samples and did not affect protein abundances when compared to samples preserved in *RNAlater*TM. Furthermore, autoclaving the NAP buffer did not significantly affect the metaproteome, suggesting that an autoclaved NAP buffer could be used in studies that require sterile material (e.g., gnotobiotic isolators).

3.6 Conclusions

This study evaluated the effects of different preservation treatments on the metaproteomes of intestinal microbiome samples. Based on our results, we recommend preserving intestinal microbiome samples by flash-freezing, or in *RNAlater*TM or in an *RNAlater*-like solution before metaproteomics analyses. The consistent use of these methods appears to minimize storage effects and thus improve the reliability of metaproteomics studies of the intestinal microbiome.

3.7 Acknowledgments

We thank Deniz Durmusoglu for providing the fresh mouse fecal samples, Dr. Fernanda Salvato for providing A.M. with extensive metaproteomics training, Ibrahim Al'Abri and Alexandria Bartlett for providing mouse fecal samples for prior optimization rounds of the study, Marlene Jensen, Clara Tang, and Amelia Ellis who assisted in the lab, everyone in the Kleiner Lab for discussing the experimental design and results, and Dr. Heather Maughan for feedback on the manuscript. We made all LC-MS/MS measurements in the Molecular Education, Technology, and Research Innovation Center (METRIC) at NC State University.

Declarations

Ethics Approval

The protocols for husbandry and experimentation of all mice used in this study were approved by the Institutional Animal Care and Use Committee at North Carolina State University (Institution reference: D16-00214).

Availability of Data

The mass spectrometry metaproteomics data and protein sequence database were deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository with the dataset identifier PXD024115.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R35GM138362 and the Foundation for Food and Agriculture Research (FFAR) Grant ID: 593607

Author contributions

A.M. and M.K. designed the study. A.M. performed the experiments. A.M. and M.K. analyzed the data and wrote the manuscript.

REFERENCES

1. Sommer F, Bäckhed F. The gut microbiota — masters of host development and physiology. *Nat Rev Microbiol.* 2013;11:227–38.
2. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr.* 2018;57:1–24.
3. Gentile CL, Weir TL. The gut microbiota at the intersection of diet and human health. *Science.* 2018;362:776–80.
4. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci.* 2009;106:3698–703.
5. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature.* 2006;444:1027–31.
6. Boursier J, Mueller O, Barret M, Machado M, Fizanne L, Araujo-Perez F, et al. The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. *Hepatology.* 2016;63:764–75.
7. Theriot CM, Koenigsnecht MJ, Carlson PE, Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat Commun.* 2014;5:3114.
8. Montgomery TL, Künstner A, Kennedy JJ, Fang Q, Asarian L, Culp-Hill R, et al. Interactions between host genetics and gut microbiota determine susceptibility to CNS autoimmunity. *Proc Natl Acad Sci.* 2020;117:27516–27.
9. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell.* 2014;157:121–41.
10. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science.* 2012;336:1268–73.
11. Chevalier G, Siopi E, Guenin-Macé L, Pascal M, Laval T, Rifflet A, et al. Effect of gut microbiota on depressive-like behaviors in mice is mediated by the endocannabinoid system. *Nat Commun.* 2020;11:1–15.
12. Rogers GB, Keating DJ, Young RL, Wong M-L, Licinio J, Wesselingh S. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. *Mol Psychiatry.* 2016;21:738.
13. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci.* 2005;102:11070–5.

14. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505:559–63.
15. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med*. 2009;1:6ra14.
16. Bajaj JS, Heuman DM, Hylemon PB, Sanyal AJ, White MB, Monteith P, et al. Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J Hepatol*. 2014;60:940–7.
17. Blakeley-Ruiz JA, Erickson AR, Cantarel BL, Xiong W, Adams R, Jansson JK, et al. Metaproteomics reveals persistent and phylum-redundant metabolic functional stability in adult human gut microbiomes of Crohn’s remission patients despite temporal variations in microbial taxa, genomes, and proteomes. *Microbiome*. 2019;7:1–15.
18. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell*. 2019;179:59–73.
19. Wang Y, Zhou Y, Xiao X, Zheng J, Zhou H. Metaproteomics: A strategy to study the taxonomy and functionality of the gut microbiota. *J Proteomics*. 2020;219:103737.
20. Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. *Trends Microbiol*. 2018;26:563–74.
21. Jansson JK, Baker ES. A multi-omic future for microbiome studies. *Nat Microbiol*. 2016;1:1–3.
22. Hettich RL, Pan C, Chourey K, Giannone RJ. Metaproteomics: Harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal Chem*. 2013;85:4203–14.
23. Salvato F, Hettich RL, Kleiner M. Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLOS Pathog*. 2021;17:e1009245.
24. Kleiner M. Metaproteomics: Much more than measuring gene expression in microbial communities. *mSystems*. 2019;4:e00115-19.
25. Starr AE, Deeke SA, Li L, Zhang X, Daoud R, Ryan J, et al. Proteomic and metaproteomic approaches to understand host–microbe interactions. *Anal Chem*. 2017;90:86–109.
26. Xiong W, Abraham PE, Li Z, Pan C, Hettich RL. Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics*. 2015;15:3424–38.
27. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, et al.

- Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 2009;3:179–89.
28. Zhang X, Chen W, Ning Z, Mayne J, Mack D, Stintzi A, et al. Deep metaproteomics approach for the study of human microbiomes. *Anal Chem.* 2017;89:9407–15.
 29. Lee PY, Chin S-F, Neoh H, Jamal R. Metaproteomic analysis of human gut microbiota: where are we heading? *J Biomed Sci.* 2017;24:1–8.
 30. Long S, Yang Y, Shen C, Wang Y, Deng A, Qin Q, et al. Metaproteomics characterizes human gut microbiome function in colorectal cancer. *Npj Biofilms Microbiomes.* 2020;6:1–10.
 31. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, Li J, et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun.* 2018;9:1–14.
 32. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun.* 2017;8:1558.
 33. Pible O, Allain F, Jouffret V, Culotta K, Miotello G, Armengaud J. Estimating relative biomasses of organisms in microbiota using “phylopeptidomics.” *Microbiome.* 2020;8:30.
 34. Gouveia D, Pible O, Culotta K, Jouffret V, Geffard O, Chaumot A, et al. Combining proteogenomics and metaproteomics for deep taxonomic and functional characterization of microbiomes from a non-sequenced host. *Npj Biofilms Microbiomes.* 2020;6:1–6.
 35. Van Den Bossche T, Kunath BJ, Schallert K, Schäpe SS, Abraham PE, Armengaud J, et al. Critical assessment of metaproteome investigation (CAMPI): a multi-lab comparison of established workflows. *bioRxiv.* 2021;:2021.03.05.433915.
 36. Kleiner M, Kouris A, Jensen M, Liu Y, McCaldler J, Strous M. Ultra-sensitive Protein-SIP to quantify activity and substrate uptake in microbiomes with stable isotopes. *bioRxiv.* 2021;:2021.03.29.437612.
 37. Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, et al. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc Natl Acad Sci.* 2018;115:E5576–84.
 38. Justice N, Li Z, Wang Y, Spaulding SE, Mosier AC, Hettich RL, et al. (15)N- and (2)H proteomic stable isotope probing links nitrogen flow to archaeal heterotrophic activity. *Environ Microbiol.* 2014;16:3224–37.
 39. Jehmlich N, Schmidt F, Hartwich M, Bergen M von, Richnow H-H, Vogt C. Incorporation of carbon and nitrogen atoms into proteins measured by protein-based stable isotope probing (Protein-SIP). *Rapid Commun Mass Spectrom.* 2008;22:2889–97.
 40. Zhang X, Li L, Mayne J, Ning Z, Stintzi A, Figeys D. Assessing the impact of protein

- extraction methods for human gut metaproteomics. *J Proteomics*. 2018;180:120–7.
41. Tanca A, Palomba A, Fraumene C, Pagnozzi D, Manghina V, Deligios M, et al. The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome*. 2016;4:1–13.
 42. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13:731–40.
 43. Jordan S, Baker B, Dunn A, Edwards S, Ferranti E, Mutic AD, et al. Maternal–child microbiome: specimen collection, storage and implications for research and practice. *Nurs Res*. 2017;66:175.
 44. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017;5:1–14.
 45. Carruthers LV, Moses A, Adriko M, Faust CL, Tukahebwa EM, Hall LJ, et al. The impact of storage conditions on human stool 16S rRNA microbiome composition and diversity. *PeerJ*. 2019;7:e8133.
 46. Hale VL, Tan CL, Knight R, Amato KR. Effect of preservation method on spider monkey (*Ateles geoffroyi*) fecal microbiota over 8 weeks. *J Microbiol Methods*. 2015;113:16–26.
 47. Williams SCP. Gnotobiotics. *Proc Natl Acad Sci*. 2014;111:1661.
 48. Fouhy F, Deane J, Rea MC, O’Sullivan Ó, Ross RP, O’Callaghan G, et al. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLOS ONE*. 2015;10:e0119355.
 49. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. *PLOS ONE*. 2012;7:e46953.
 50. Tap J, Cools-Portier S, Pavan S, Druesne A, Öhman L, Törnblom H, et al. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. *Sci Rep*. 2019;9:1–9.
 51. Blekhman R, Tang K, Archie EA, Barreiro LB, Johnson ZP, Wilson ME, et al. Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Sci Rep*. 2016;6:1–5.
 52. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, et al. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems*. 2016;1:e00021-16.
 53. Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep*. 2015;5:1–10.

54. Passow CN, Kono TJY, Stahl BA, Jaggard JB, Keene AC, McGaugh SE. RNA later and flash freezing storage methods nonrandomly influence observed gene expression in RNAseq experiments. *bioRxiv*. 2018;:379834.
55. Saito MA, Bulygin VV, Moran DM, Taylor C, Scholin C. Examination of microbial proteome preservation techniques applicable to autonomous environmental sample collection. *Front Microbiol*. 2011;2:215–215.
56. Hickl O, Heintz-Buschart A, Trautwein-Schult A, Hercog R, Bork P, Wilmes P, et al. Sample preservation and storage significantly impact taxonomic and functional profiles in metaproteomics studies of the human gut microbiome. *Microorganisms*. 2019;7:367.
57. de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods*. 2010;80:209–11.
58. Menke S, Gillingham MAF, Wilhelm K, Sommer S. Home-Made cost effective preservation buffer is a better alternative to commercial preservation methods for microbiome research. *Front Microbiol*. 2017;8:102.
59. Camacho-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA. Preservation of RNA and DNA from mammal samples under field conditions. *Mol Ecol Resour*. 2013;13:663–73.
60. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, et al. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol Biomark Prev*. 2016;25:407–16.
61. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6:359.
62. Speare L, Smith S, Salvato F, Kleiner M, Septer AN. Environmental viscosity modulates interbacterial killing during habitat transition. *mBio*. 2020;11:e03060-19.
63. Oberg AL, Vitek O. Statistical design of quantitative mass spectrometry-based proteomic experiments. *J Proteome Res*. 2009;8:2144–56.
64. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, et al. A catalog of the mouse gut metagenome. *Nat Biotechnol*. 2015;33:1103–8.
65. Timmins-Schiffman E, May DH, Mikan M, Riffle M, Frazar C, Harvey HR, et al. Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *ISME J*. 2017;11:309–14.
66. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma Oxf Engl*. 2006;22:1658–9.
67. Zybailov B, Mosley AL, Sardu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *saccharomyces cerevisiae*. *J*

- Proteome Res. 2006;5:2339–47.
68. Oliveros, J.C. Venny. An interactive tool for comparing lists with Venn’s diagrams. 2015. <https://bioinfogp.cnb.csic.es/tools/venny/index.html>.
 69. Aitchison J. The statistical analysis of compositional data. *J R Stat Soc.* 1982;44:139–77.
 70. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* 2014;2:1–13.
 71. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 2017;8:2224.
 72. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
 73. Vizcaíno JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 2016;44:D447-56.
 74. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of single cell RNA-seq statistical methods on microbiome data. *bioRxiv.* 2020;:2020.01.15.907964.

CHAPTER 4 – LINKING INTESTINAL MICROORGANISMS TO THEIR GROWTH SUBSTRATES

Authors:

Angie Mordant¹, Manuel Kleiner¹

Authors' Affiliations:

1: Department of Plant and Microbial Biology, North Carolina State University, Raleigh NC

***Corresponding Author Information**

Angie Mordant - almordan@ncsu.edu

Manuel Kleiner - manuel_kleiner@ncsu.edu

Publication Status: manuscript in preparation

4.1 Introduction

The intestinal microbiota is a diverse and metabolically active community that has profound effects on its host. This complex community influences the health of its host by altering the availability of nutrients (Rowland et al. 2018; Gentile and Weir 2018; Wikoff et al. 2009; Turnbaugh et al. 2006) and the host's susceptibility to infection and disease (Boursier et al. 2016; Theriot et al. 2014).

Diet is highly influential to the microbiota because unabsorbed dietary components are thought to be the primary growth substrates of intestinal microbes (Rowland et al. 2018). The microbiota process dietary nutrients into metabolites that have considerable implications for host health and metabolism. Diet-microbiota interactions can be positive for the host, such as fermentable fiber, to mitigate IBD risks (Zmora, Suez, and Elinav 2019). However, the metabolism of dietary components by the microbiota can also be detrimental to the host. For example, studies of protein degradation by the gut microbiota reveal that fermentation of proteins can result in the formation of harmful byproducts associated with colon cancer (Diether and Willing 2019; Portune et al. 2016). Therefore, there is a growing interest in deliberately modulating the microbiota through diet to improve the host's health and mitigate the risk of diseases (Zmora, Suez, and Elinav 2019; De Filippis et al. 2018).

To modulate the microbiota through diet, we first require knowledge of the interactions of intestinal microorganisms with dietary components. However, current understanding of assimilation of diet-derived substrates by intestinal microorganisms is primarily based on indirect evidence. In this study, we investigated the use of dietary components as growth substrates by intestinal bacteria using protein stable isotope fingerprinting (Protein-SIF)(Kleiner et al. 2018). The protein-SIF approach relies on the principle that heterotrophic organisms, such

as most of the intestinal bacteria, have a similar natural isotopic signature as their food source (Pelz et al. 1998). Thus, if we know the natural isotopic signatures of dietary components and the signatures of the bacterial cells, we could potentially infer diet-bacteria associations. We combined this approach with metaproteomics analyses of differentially abundant proteins to provide a unique angle in studying diet-microbiota interactions.

We searched for direct evidence of the use of dietary protein and dietary fat as growth substrates by intestinal bacteria in an *in vivo* system. Our system consisted of gnotobiotic mice colonized with a defined community of bacteria representative of a healthy human microbiota, as done in (Desai et al. 2016). We fed the gnotobiotic mice with defined diets containing known natural isotopic signatures, and we monitored the assimilation of these signatures in the bacteria.

The use of dietary fats as growth substrates has been hypothesized (Schoeler and Caesar 2019), but there is no evidence of it happening *in vivo*. *In vitro*, several species belonging to the Proteobacteria phylum can sustain growth in media containing almost exclusively purified fatty acids (Agans et al. 2018). In addition, species of genera *Bilophila* which belong to the Bacteroidetes phylum can also sustain growth on media devoid of carbohydrates and proteins (Agans et al. 2018). In this study, we searched for direct evidence for the use of dietary fats as growth substrates by intestinal bacteria. Previous studies have linked fats rich in polyunsaturated fatty acids with increased growth of specific bacterial species, such as *Akkermansia muciniphila* (Li et al., 2017; Patrone et al., 2018). Therefore, in our study, we included dietary fat sources rich in polyunsaturated fatty acids: corn oil, soybean oil, and sunflower oil.

The degradation of complex polysaccharides by intestinal bacteria has been studied extensively and has been demonstrated with direct evidence (Patnode et al. 2019). However, the use of dietary proteins as growth substrates has been inferred mostly indirectly. Therefore we

searched for direct evidence of bacteria using proteins of different sources (casein, egg whites, and soy protein) as growth substrates. We also included dietary fiber in our study as potential proof of principle of the methods. Furthermore, many carbohydrate-targeting enzymes have been characterized and compiled into a database (CAZy database <https://www.cazy.org>) (Lombard et al. 2014) and, therefore, identifying CAZy enzymes in metaproteomes could complement Protein-SIF results. We administered three types of fibers: cellulose, inulin, and corn fiber. Cellulose is a glucose polymer, and inulin is a fructose polymer. Corn fiber is more complex as it contains a broader variety of monomers: arabinose, mannose, galactose, glucose, and xylose.

4.2 Materials and Methods

4.2.1 Overall experimental design

We conducted two experiments in parallel to investigate the assimilation of dietary macronutrients by intestinal bacteria. The first experiment had 5 gnotobiotic mice ("group 1 mice") fed diets differing only in the source of dietary protein. The dietary protein sources we tested were egg whites, casein, and soy protein. The protein source had a distinct natural isotopic signature from the other components of the diets.

The second experiment included 6 gnotobiotic mice ("group 2 mice") fed diets that differed either only in their source of dietary fiber or in their source of dietary fat. For the "fiber" diets, we tested corn fiber, cellulose, and inulin. These diets had corn oil as their source of fat. For the "fat" diets, we tested corn oil, soybean oil, and sunflower oil. These diets had corn fiber as their source of fiber. For each diet, the fiber or the fat source respective of the diet group ("fiber" or "fat" group) had a distinct isotopic signature from the other components of the diets.

Every mouse in this study was colonized with a defined community of 13 human commensal bacterial strains (Community used in (Desai et al. 2016) and acquired from Dr. Martens's lab; **suppl. table 1**). The colonized mice were fed one diet per week, and samples were collected on the 7th day after introducing a diet. We collected baseline samples at the beginning when the mice were still fed their standard chow (Lab diet 5010), 21 days post gavage. We collected samples at the end of the experiments when the mice were returned to their standard chow.

We applied metaproteomics and the protein stable isotope fingerprinting (Protein-SIF) method described in **chapter 2** to link the bacterial strains to the dietary components they assimilated.

4.2.2 Gnotobiotic models

We used 11 germ-free C57BL/J6 mice (5 in group 1 and 6 in group 2, all females). The NCSU Gnotobiotic core supplied and housed the mice. The mice were housed in groups of 3 throughout the experiment, except for one cage, which had 2 mice. All animal experiments followed protocols approved by the Institutional Animal Care and Use Committee of North Carolina State University.

We used a defined microbiota consisting of 13 fully sequenced human intestinal bacteria (**Suppl. table 4-1**). The 13 strains belonged to the community assembled by Desai et al. (Desai et al. 2016) from common human intestinal bacteria representing the 5 major phyla: Bacteroidetes, Firmicutes, Actinobacteria, Verrucomicrobia, and Proteobacteria.

The germ-free mice were colonized with this defined community at 9 - 11 weeks of age with freshly prepared bacterial inocula. Bacteria were grown in individual cultures in their

respective media (same media used by (Desai et al. 2016)). The cultures were grown anaerobically in Hungate tubes pre-reduced with balanced nitrogen. The cultures were incubated at 37 C for 1 -2 days depending on the strain until they reached optical densities (OD, absorbance at 600 nm) ranging from about 0.2 to > 2 (**Suppl. table 4-2**). Bacterial cultures were mixed in equal volumes. These strains niche-partition once in the intestinal tract of their mouse host, and therefore mixing in equal volumes regardless of OD is adequate (Desai et al. 2016). We gavaged each mouse with 200 uL of the bacterial mixture for three consecutive days at about the same time each morning. Identities of the bacterial strains were confirmed by sequencing of the 16S rRNA gene, followed by comparison to sequences in the NCBI database. Colonization status was monitored throughout the experiment using microscopy and cultivation approaches. The mice were maintained on the standard chow diet for 21 days before introducing the defined diets.

4.2.3 Design of diets with known isotopic signatures

We designed diets with known isotopic signatures. Every diet had one ingredient with a distinct signature from the other components in the diet, except for the control diet. To design the diets, we first measured the natural isotopic signatures of purified ingredients. We obtained purified ingredients from Envigo Teklad® and Amazon. We weighed the ingredients into tin capsules (~ 1 mg per sample) and analyzed them by Elemental Analysis Isotope Ratio Mass Spectrometry (EA-IRMS). The ingredients were compared to a Vienna Pee Dee Belemnite (VPDB) standard using area under the curve calculations to determine the carbon isotope ratios. We analyzed each purified ingredient in duplicates because IRMS measurements are highly robust (list of ingredients used in the study and corresponding natural isotopic signatures in **suppl. table 4-3**).

The defined diets we designed had the same nutritional composition as the standard AIN-93 rodent diet (Reeves, Nielsen, and Fahey 1993), with minor modifications. We increased the fat content to make up 12 % of the diets by weight, or about 27% kcal. As a result, the corn starch content was slightly reduced to make room for the fat (~392 g/Kg in the AIN-93 and ~352 g/Kg in our diets). We designed 7 diets for this study (**formulas in Suppl. S1-7**). All 7 diets contained the same source of starches and simple sugars: corn starch, maltodextrin, and sucrose. We included 3 groups of diets: “protein diets”, “fiber diets”, and “fat diets”. In each group, we compared 3 sources of protein, fiber, or fat. Depending on the diet group, either the protein, or the fiber, or the fat had a distinct natural isotopic signature from the other components of the diets (**Table 4-1**). One of the three sources in each comparison belonged to the control diet in which none of the signatures were distinct from the rest. This control diet contained egg white solids as the source of protein, corn fiber as the fiber source, and corn oil as the fat source. This diet was introduced to both groups of mice and used as a comparison in each diet group. The protein diets included the following sources of dietary protein: casein, soy protein, and egg whites (egg whites from the control diet). The fiber diets included the following sources of dietary fiber: cellulose, inulin, and corn fiber (corn fiber from the control diet). The fat diets included the following sources of dietary fat: soybean oil, sunflower oil, and corn oil (corn oil from the control diet). All defined diets were purchased from Envigo Teklad®. All defined diets were irradiated.

We verified the composition of the diets by metaproteomics analysis as described below (**Sections 4.3.4- 8; suppl. table S4-4**).

Table 4-1. Natural carbon isotopic composition of the diets.

Signatures of dietary components are averaged between two replicates measured by EA-IRMS (measurement uncertainties of ± 0.42 ‰ or less; ± 0.13 ‰ averaged uncertainty).

Diet	Protein source	Protein signature $\delta^{13}\text{C}$ (‰)	Fat source	Fat signature $\delta^{13}\text{C}$ (‰)	Fiber source	Fiber signature $\delta^{13}\text{C}$ (‰)	Starches & sugar	Signatures $\delta^{13}\text{C}$ (‰)
Control (no distinct signature)	Egg whites	-17.19	Corn oil	-16.92	Corn fiber	-11.17	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Fiber diet 1: Cellulose diet	Egg whites	-17.19	Corn oil	-16.92	Cellulose	-26.55	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Fiber diet 2: Inulin diet	Egg whites	-17.19	Corn oil	-16.92	Inulin	-26.88	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Protein diet 1: Casein diet	Casein	-27.81	Corn oil	-16.92	Corn fiber	-11.17	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Protein diet 2: Soy protein diet	Soy protein	-26.33	Corn oil	-16.92	Corn fiber	-11.17	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Fat diet 1: Soybean oil diet	Egg whites	-17.19	Soybean oil	-32.05	Corn fiber	-11.17	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3
Fat diet 2: Sunflower oil diet	Egg whites	-17.19	Sunflower oil	-28.7	Corn fiber	-11.17	Corn Starch, Maltodextrin, Sucrose	-10.75, -10.58, -12.3

All mice were fed the LabDiet® autoclavable rodent diet 5010 before and after the experiment (**Formula in Suppl. Fig. S4-8**). We refer to this diet as the “standard chow.” We used the standard chow as a baseline diet for before and after comparisons and assessments of reproducibility.

4.2.4 Diet oscillations and sample collection

We fed the mice diets in one-week increments and collected samples on the 7th day after introducing a diet. David et al. observed that the microbiota goes back to its baseline structure in as little as two days following dietary interventions (David et al. 2014). Diets were fed to the mice in the order shown in **fig. 4-1**.

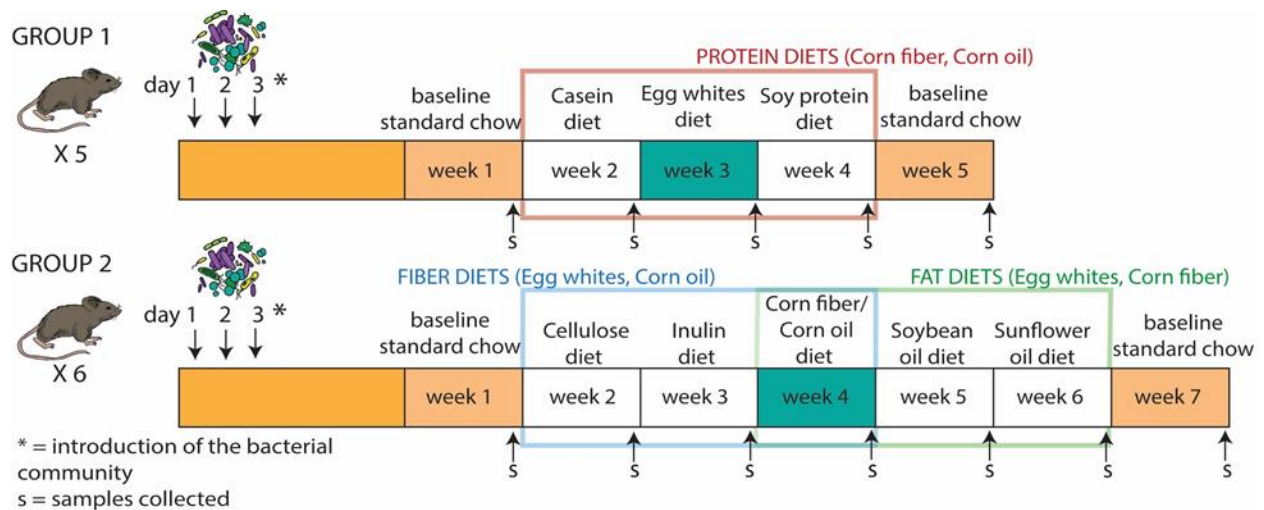


Figure 4-1. Experimental design: timeline and diet oscillations. Both groups of mice were gavaged with the defined community for three consecutive days while fed the “standard chow diet.” We started the experiment two weeks after introducing the community and confirming successful colonization. Group 1 mice were fed the protein diets, and group 2 mice were fed the fiber and fat diets. Both groups were fed the control diet, colored in green in the figure. We collected fecal samples from every mouse on the 7th day after the introduction of a diet. Samples were collected in RNA_{later} and later frozen at -80 °C. We collected samples before introducing the defined diets when the mice were fed the standard chow, and at the end of the experiment when the mice were returned to the standard chow.

4.2.5 Protein extraction and peptide preparation

We performed sample preparations for metaproteomic analyses on the diets, purified casein and purified egg white solids, and fecal microbiome samples. We extracted the diets and purified ingredient samples in triplicates. Each replicate consisted of 100 mg of diet that we powdered down from diet pellets or 100 mg of a purified ingredient. For the fecal microbiome samples, we extracted one sample per mouse per diet. We removed the storage solutions from the samples (if applicable, i.e., fecal pellets but not diets) by centrifugation at 21,000 x g for 5 min. We then resuspended the samples in 400 µl of SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT]. Cells were lysed by bead-beating in lysing matrix E tubes (MP Biomedicals) with a Bead Ruptor Elite (Omni International) for 5 cycles of 45 sec at 6.45 m/s with 1 min dwell time between cycles, followed by heating at 95° C for 10 min. Bead-beating was applied to the diet and the fecal samples but not the purified protein powders, which were simply heated at 95° C for 10 min. The lysates were centrifuged for 5 min at 21,000 x g to remove cell debris. We prepared peptides according to the filter-aided sample preparation (FASP) protocol described by [61]. All centrifugations mentioned below were performed at 14,000 x g. Samples were loaded onto 10 kDa MWCO 500 µl centrifugal filters (VWR International) by combining 60 µl of lysate with 400 µl of Urea solution (8 M urea in 0.1 M Tris/HCl pH 8.5) and centrifuging for 30 min. This step was repeated twice until the filter capacity was reached. Filters were washed twice by applying 200 µl of urea solution followed by 40 min of centrifugation. 100 µl IAA solution (0.05 M iodoacetamide in Urea solution) was then added to filters for a 20 min incubation followed by centrifugation for 20 min. The filters were washed three times with 100 µl of urea solution and 20 min centrifugations, followed by buffer exchange to ABC (50 mM Ammonium Bicarbonate). Buffer exchange was accomplished by

adding 100 µl of ABC and centrifuging three times, followed by centrifugation for 20 min. Tryptic digestion was performed by adding 0.85 µg of MS grade trypsin (Thermo Scientific Pierce, Rockford, IL, USA) in 40 µl of ABC to the filters and incubating for 16 hours in a wet chamber at 37° C. The tryptic peptides were eluted by adding 50 µl of 0.5 M NaCl and centrifuging for 20 min. Peptide concentrations were determined with the Pierce Micro BCA assay (Thermo Fisher Scientific) following the manufacturer's instructions.

We mixed the peptides from the purified casein and the purified egg white solids in equal concentrations to create a casein and egg sample that was later used as an internal standard for the Protein-SIF approach (more details in **section 4.3.9**).

4.2.6 LC-MS/MS

Samples were analyzed by 1D-LC-MS/MS using a published method (Speare et al. 2020) with several modifications. The samples were blocked and randomized according to Oberg and Vitek's method (Oberg and Vitek 2009) to control batch effects. We included two types of Protein-SIF standards at the beginning and end of the run sequence, alongside the fecal microbiome samples. Every sample was run as four consecutive technical replicates to increase the number of peptides available for Protein-SIF. Only the human hair Protein-SIF standard was run as a single replicate as it contained enough peptides. For each sample replicate, 600 ng of tryptic peptides were loaded with an UltiMate™ 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) in loading solvent A (2 % acetonitrile, 0.05 % trifluoroacetic acid) onto a 5 mm, 300 µm ID C18 Acclaim® PepMap100 pre-column and desalted (Thermo Fisher Scientific). Peptides were then separated on a 75 cm x 75 µm analytical EASY-Spray column packed with PepMap RSLC C18, 2 µm material (Thermo Fisher Scientific) heated to 60 °C via

the integrated column heater at a flow rate of 300 nl min⁻¹ using a 140 min gradient going from 95 % buffer A (0.1 % formic acid) to 31 % buffer B (0.1 % formic acid, 80 % acetonitrile) in 102 min, then to 50 % B in 18 min, to 99 % B in 1 min and ending with 99 % B. Carryover was reduced by wash runs (injection of 20 µl acetonitrile with 99% eluent buffer B) between samples.

The analytical column was connected to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) via an Easy-Spray source. Eluting peptides were ionized via electrospray ionization (ESI). MS1 spectra were acquired by performing a full MS scan at a resolution of 60,000 on a 380 to 1600 m/z window. MS2 spectra were acquired using a data-dependent approach by selecting for fragmentation the 15 most abundant ions from the precursor MS1 spectra. A normalized collision energy of 25 was applied in the HCD cell to generate the peptide fragments for MS2 spectra. Other settings of the data-dependent acquisition included: a maximum injection time of 100 ms, a dynamic exclusion of 25 sec, and exclusion of ions of +1 charge state from fragmentation. About 60,000 MS/MS spectra were acquired per sample.

4.2.7 Protein identification database

We constructed a protein sequence database for identifying proteins from the components of the microbiome samples: the microbiota, the host, the dietary components, and potential contaminants. Protein sequences of the mouse host, *Mus musculus*, were downloaded from Uniprot (<https://www.uniprot.org/proteomes/UP000000589>). The strains used in this study are fully sequenced, and their complete proteomes are on Uniprot (list of strains and corresponding Uniprot IDs in **Suppl table S4-1**). To identify dietary proteins, we included the proteomes of the

following organisms: *Gallus gallus* (Chicken UP000000539 - egg whites), *Glycine max* (Soybean UP000008827 - soy protein and soybean oil), *Bos taurus* (Cow UP000009136 - casein), *Zea mays* (Corn - corn fiber, corn oil, cornstarch), *Helianthus annuus* (Sunflower - sunflower oil), and *Beta vulgaris* (Sugar beet - sucrose). Proteomes of the mouse host and the organisms representing dietary components contain too similar sequences for adequate discrimination (isoforms). Therefore, for each proteome, the protein sequences were clustered with an identity threshold of 95% using the CD-HIT tool (Li and Godzik 2006). The proteomes of the bacterial strains were not modified. Also included in the database were sequences of common laboratory contaminants (<http://www.thegpm.org/crap/>). The database contained a total of 324,982 protein sequences and *will be* included with the PRIDE submission.

4.2.8 Protein identification and quantification

For peptide and protein identification, MS data were searched against the above-described database using the Sequest HT node in Proteome Discoverer version 2.3.0.523 (Thermo Fisher Scientific) with the following parameters: digestion with trypsin (Full), maximum of 2 missed cleavages, 10 ppm precursor mass tolerance, 0.1 Da fragment mass tolerance and maximum 3 equal dynamic modifications per peptide. We considered the following dynamic modifications: oxidation on M (+15.995 Da), carbamidomethyl on C (+57.021 Da), and acetyl on the protein N terminus (+42.011 Da). Peptide false discovery rate (FDR) was calculated using the Percolator node in Proteome Discoverer, and only peptides identified at a 5% FDR were retained for protein identification. Proteins were inferred from peptide identifications using the Protein-FDR Validator node in Proteome Discoverer with a target FDR of 5%. We generated files of individual samples by combining the four replicate LC-

MS/MS-produced files in the search. We used the resulting file with no modifications for the Protein-SIF method described in **4.3.10** below.

From the generated multiconsensus dataset, we removed contaminant (cRAP sequences) and low confidence proteins (> 5 % FDR) and kept proteins that were identified by at least 1 protein unique peptide. In addition, to decrease the number of "one-hit wonder" proteins, we removed proteins that were not detected in a total of at least 5 and 6 samples for group 1 and for group 2, respectively, which is the minimum number of replicates per diet. The data were then normalized by calculating normalized spectral abundance factors (NSAFs, (Zybailov et al. 2006)) and multiplied by 100 to give the relative protein abundance in %.

4.2.9 Quality assessment and filtering

We assessed data quality by first inspecting raw data in the Xcalibur™ Software (Thermo Fisher Scientific), then comparing the number of peptide spectrum matches (PSMs), peptides, proteins, and protein groups identified in each sample individually.

4.2.10 Protein stable isotope fingerprinting (Protein-SIF)

The protein-SIF approach requires many peptides from the organism of interest to infer an accurate organism SIF. For this reason, it is recommended to collect a large number of high-resolution mass spectra by either increasing the gradient length or including technical replication. We experimentally determined that running each sample as four consecutive technical replicates with each a 140 min elution gradient allows for the accurate estimation of signatures of the most abundant organisms in the samples (**Suppl. Figure S4-9**).

We used the calis-p 2.0 software (Kleiner et al. 2021) to estimate stable isotopic fingerprints of the organisms in the fecal microbiome samples. The Calis-p software requires two input files: raw spectral files produced by the LC-MS/MS, and the peptide-spectrum match (PSM) files containing the protein identifications and quantifications. Raw files were converted into mzML format using the MSConvertGUI tool via ProteoWizard (Chambers et al. 2012) with the following options: Binary encoding precision: 64-bit, Write index: checked, TPP compatibility: checked, Filter: Peak Picking, Algorithm: Vendor, MS Levels: 1. The PSM files were generated as described above (4.2.8) and input into the Calis-p software as tab-delimited text files.

We then ran the calis-p software, which performs fully automatically. Calis-p performs two main steps: isotopic pattern extraction and SIF computation. Details can be found in (Kleiner et al. 2018; 2021). Briefly, the software first filtered out “ambiguous” peptides of low identification confidence from the input PSM file. Then, for each remaining peptide identification, the software found the corresponding mass spectrum in the mzML and extracted the isotopic pattern. After some clean-up and filtering steps, the software compared the experimentally derived isotopic pattern to *in silico* derived isotopic patterns to infer $\delta^{13}\text{C}$ values for the peptide. This step was repeated for many peptides. Finally, the average $\delta^{13}\text{C}$ of the peptides from an organism was used to estimate that organism’s signature. We filtered the results to retain only SIF values computed from at least 30 peptides because that is the threshold required by the calis-p 2.0 software to accurately estimate an organism’s SIF.

We corrected for the offset introduced by the mass spectrometer using both the human hair standard and the casein and egg standard. $\delta^{13}\text{C}$ values for the standards were obtained both with protein-SIF and EA-IRMS, so we calculated the offset between the two measurement

methods. We then used the averaged offset value (**Suppl. Table S4-7**) to correct the protein-SIF values of the organisms in the microbiome samples.

We compared the organism SIFs to the signatures of the dietary components fed to the mice. We looked for correspondence between signatures of organisms and signatures of dietary components to hypothesize on the assimilation of dietary constituents by the organisms. We also looked at how the organisms' SIFs changed over time due to the different diet inputs to inform further data analyses. We identified significant differences ($p < .05$) using pairwise t-tests corrected for multiple hypotheses (Benjamini-Hochberg correction), computed in R version 4.0.2 (script in Suppl. text). We included the SIF values of the standard chow samples in the results to assess reproducibility. However, we did not know the signatures of the dietary constituents of the standard chow diet, and thus we did not test for significant differences between the standard chow and the defined diets. We prepared plots for organisms that had at least 3 data points in a minimum of two defined diets so that we could perform statistical analyses. All offset-corrected SIF values collected that passed the 30 peptide threshold were reported in the **supplementary tables S4-8 and S4-9**.

4.2.11 Data analyses

First, we looked at how the different diets affected the measured abundances of each bacterial species by comparing the relative biomass contributions of each organism. Each organism's relative biomass contribution was assessed using the method described in Kleiner et al. (Kleiner et al. 2017). Briefly, proteins were filtered for at least 2 protein unique peptides to increase the confidence in taxonomic identifications, and PSMs summed by organism were used to estimate the biomass contribution of each organism in the metaproteomes. We tested for

significant changes in relative abundances using a one-way ANOVA followed by a Tukey's honestly significant difference (HSD) post hoc test, computed in R (version 4.0.2). We assigned letters to group means that are similar using the “rcompanion” package (version 2.4.1; script in supplementary text).

Next, we investigated whether the total bacterial load was affected by the change from the standard chow to the defined diets. We estimated the bacterial load based on the bacterial to host ratio of proteinaceous biomass contribution, assuming that the host biomass contribution remained approximately constant. We obtained ratios by dividing the number of total bacterial PSMs by the number of mouse PSMs (filtered for at least 2 protein unique peptides, as described above). We tested for significant differences between the standard chow and defined diets using a Welch's t-test (two-sided, $p < 0.05$).

To identify differentially abundant proteins between diets that are statistically significant, we performed a centered-log-ratio (CLR) transformation in R (version 4.0.2, `compositions_2.0-1` package)(Aitchison 1982) on peptide spectrum matches (PSMs) before performing statistical tests. Before performing the CLR transformation, we removed the baseline standard chow samples and all proteins belonging to the diets (proteins from the cow proteome, proteins from the chicken proteome, etc.) from the dataset. We added 1 to every PSM value before performing the CLR transformation to protect against issues with missing values. Although CLR-normalized counts lose interpretability, CLR is a method better suited for statistical analyses of compositional data such as metaproteomic data (Fernandes et al. 2014; Gloor et al. 2017). Pairwise comparisons of all defined diets were performed in the Perseus software platform (version 1.6.12.0) (Tyanova et al. 2016) using a Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5% ($S_0=0.1$, both sides, not paired).

We used hierarchical clustering and heat map analysis to visualize how samples separated or clustered based on relative protein abundances. We performed the analysis in the Perseus software platform (version 1.6.12.0) ([Tyanova et al. 2016](#)) on the CLR-transformed dataset described above.

Informed by the Protein-SIF results (method described above, 4.3.9), we investigated differences in the protein profiles of *M. formatexigens* and *B. thetaiotaomicron*, and *A. muciniphila*. For *M. formatexigens*, we looked at proteins that differed in abundance in the following comparisons: soy protein diet vs. egg whites diet, soy protein diet vs. casein diet, and inulin vs. corn fiber. For *B. thetaiotaomicron*, we looked at proteins that differed in abundance in these comparisons: cellulose vs. inulin, and inulin vs. corn fiber. For *A. muciniphila*, we looked at proteins in these comparisons: casein vs. egg whites, and corn oil vs. sunflower oil. If no significant difference existed in a comparison, we reported the most abundant proteins. We chose comparisons based on significant changes in the SIFs. To account for differences in relative abundance, we normalized protein abundances within the species of interest to %orgNSAFs. %orgNSAFs represent peptide spectral matches (PSMs) normalized by protein length and to the total number of PSMs detected from that organism (Mueller et al. 2010). We averaged the normalized spectral abundance factors per diet for each protein and calculated ratios of abundance between diets. Then, to identify differentially abundant proteins between diets that are statistically significant, we performed a centered-log-ratio (CLR) transformation in R (version 4.0.2, `compositions_2.0-1` package) ([Aitchison 1982](#)) on peptide spectrum matches (PSMs) before performing statistical tests. We added 1 to every PSM value before computing the CLR transformation to protect against issues with missing values. Pairwise comparisons were performed in the Perseus software platform (version 1.6.12.0) ([Tyanova et al. 2016](#)) using a

Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5% ($S_0 = 0.1$, both sides, not paired). Within the dataset of proteins that tested significant, we sorted the proteins based on the difference ratios described above and based on the average abundance of the protein. We sorted by two levels: most different and then most abundant to obtain the top 10 upregulated proteins in the diet in question. Each comparison was between two diets, so we obtained a set of the top 20 differentially abundant proteins in the comparison of interest. We hypothesized the interactions of upregulated proteins and dietary components based on searches in the Uniprot, KEGG, and CAZy database (<http://www.cazy.org/>) We leveraged results with the literature. Interpretation of CAZy entries was largely based on findings of Desai et al. who performed in vitro growth assays and transcriptional profiling experiments on the species of the defined community (Desai et al. 2016).

Plots were prepared in Origin 2018b or in the Perseus software platform (version 1.6.12.0) (Tyanova et al. 2016), and enhanced in Adobe Illustrator 2021.

4.3 Results

4.3.1 Changes in community composition

The relative taxonomic composition of the samples in terms of proteinaceous biomass contribution was consistent across the mice, with minor exceptions such as *E. coli* in the group 2 mice (**Fig. 4-2 and 4-3**). Five species made up about 80 % of the relative proteinaceous biomass. These five species represented 3 phyla: Bacteroidetes (*Bacteroides uniformis* and *Bacteroides ovatus*), Firmicutes (*Marvinbryantia formatexigens*), and Verrucomicrobia (*Akkermansia muciniphila*). These five species dominated the biomass except in the samples of mice consuming the cellulose or inulin diets in which *Bacteroides caccae* made up ~ 14 % and 21 %

of the biomass. *B. caccae* was relatively low in abundance in the other samples. *B. caccae* along with Firmicutes species *Eubacterium rectale*, *Roseburia intestinalis*, and Proteobacterium *E. coli*, collectively made up about 17% of the biomass. The smallest fraction, collectively about 3 % of the total biomass included Actinobacteria species *Collinsella aerofaciens*, Firmicutes species *Clostridium symbiosum*, and Bacteroidetes species *Barnesiella intestinihominis*. The Firmicutes *Faecalibacterium prausnitzii* was mostly undetected (**Suppl. tables S4-5 and S4-6**).

We compared the baseline standard chow (SC) samples to assess reproducibility in the relative abundances of the bacteria. In both groups of mice, the baseline “SC start” samples were collected 21 days after introducing the bacterial community to the germ-free mice, before any diet change. In mouse group 1, the baseline “SC end” samples were collected 4 weeks later, after three diet oscillations and after the mice were returned to the standard chow diet for 7 days. In mouse group 2, the baseline “SC end” samples were collected 6 weeks later, after five diet oscillations and after the mice were returned to the standard chow for 7 days. The taxonomic composition of the baseline “SC start” samples were highly similar were to the “SC end” samples for both groups, with minor exceptions. *B. ovatus*’s relative abundance was slightly lower at the end in the group 1 mice ($p = .03$) *E. rectale*’s relative abundance was slightly higher at the end in group 1 and in group 2 ($p = .02$; $p = .001$). *R. intestinalis*’s relative abundance was slightly lower at the end in the group 1 mice ($p = .00007$). All other differences between baseline samples were not significant.

We observed large changes in relative abundances between the baseline standard chow samples and the defined diets. *B. thetaiotaomicron*, *B. uniformis*, *A. muciniphila*, *B. caccae*, *B. intestinihominis*, *E. coli*, and *C. aerofaciens* all increased in relative abundance on the defined diets while *B. ovatus*, *E. rectale*, *M. formatexigens*, *R. intestinalis*, *C. symbiosum* decreased on

the defined diets compared to the standard chow diet. This decrease was drastic in *R. intestinalis*. *R. intestinalis* made up about 7% of the biomass when the mice were fed the standard chow diet and dropped below 1% when the mice were fed the defined diets.

The changes in relative abundance within the defined diets were more subtle, but there were exceptions. In group 1, the relative abundances of *B. uniformis*, *B. thetaiotaomicron*, *A. muciniphila*, *R. intestinalis*, and *E. rectale* remained constant across the casein, egg whites, and soy protein diets. The relative abundance of *B. intestinihominis* was significantly higher on the egg whites diet than the other two protein diets ($p = .00007$, $p = .00001$) The relative abundance of *B. caccae* was higher on the egg whites diet than on the soy protein diet. The relative abundance of *M. formatexigens* was significantly higher on the casein diet compared to the soy protein diet ($p = .02$). The relative abundance of *E. coli* was higher on the soy protein diet than the other two protein diets ($p = .002$, $p = .0005$). The relative abundance of *B. ovatus* was higher on the soy protein diet compared to the casein diet ($p = .0004$).

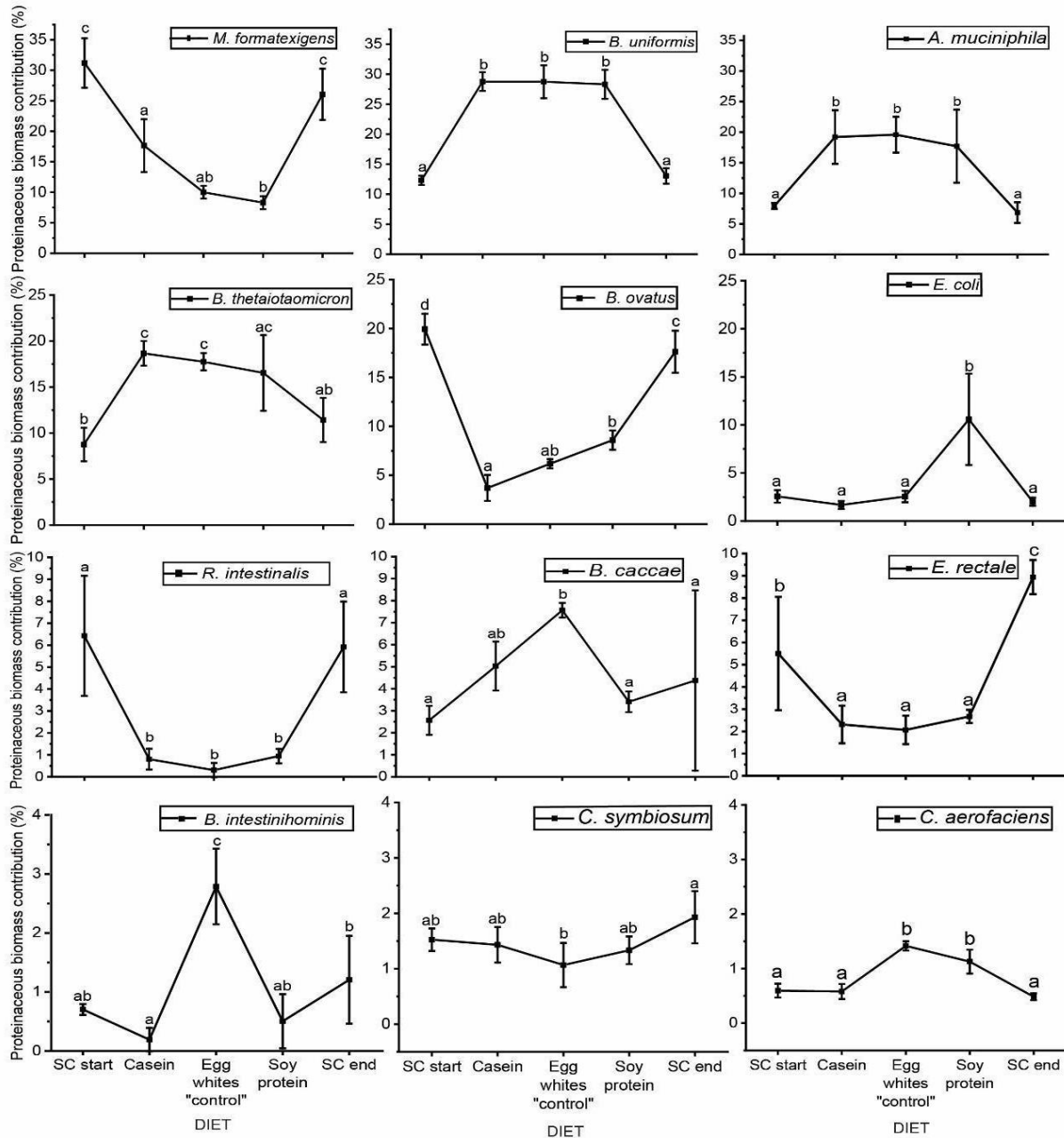


Figure 4-2. Relative proteinaceous biomass contribution of *M. formatexigens*, *B. uniformis*, *A. muciniphila*, *B. thetaiotaomicron*, *B. ovatus*, *E. coli*, *R. intestinalis*, *B. caccae*, *E. rectale*, *B. intestinihominis*, *C. symbiosum*, and *C. aerofaciens* in mice fed the protein diets (Group 1). Each point represents the relative biomass contribution of the organism after the mice were fed the diet indicated on the x-axis for seven days. Relative abundances were averaged and error bars indicate standard deviation (n = 5). Diets are ordered on the x-axis in chronological order fed to the mice. SC = standard chow diet. Casein = diet with Casein as the protein source; Egg whites (“control”) = diet with egg whites as the protein source; Soy protein = diet with soy protein as the protein source. Please note that the y-axis scale differs per row. Significant differences are displayed as different letters (a, b, c, d; based on one-way ANOVA and Tukey’s HSD post hoc test, p < .05).

In the group 2 mice, the different fiber diets caused larger shifts in relative abundances while the bacteria's relative abundances were mostly unaffected by changes in the fat source (**Fig. 4-3**). Only in *A. muciniphila*, we observed a subtle but statistically significant decrease (~ 5 %) when the mice were switched from the corn oil diet to the sunflower oil diet ($p = .002$).

B. thetaiotaomicron and *B. caccae* were the organisms most greatly affected by the changes in fiber source. *B. thetaiotaomicron* made up ~20 +/- 2 % of the biomass when mice were consuming corn fiber and ~17 +/- 2 % when the mice were consuming cellulose, but made up only ~6 +/- 1 % of the biomass when mice consumed inulin as the sole source of fiber. Levels of *B. caccae* were highest when mice consumed the inulin diet (~21 +/- 3%), followed by the cellulose diet (~14 +/- 1%), and lastly the corn fiber (~8 +/- 1%).

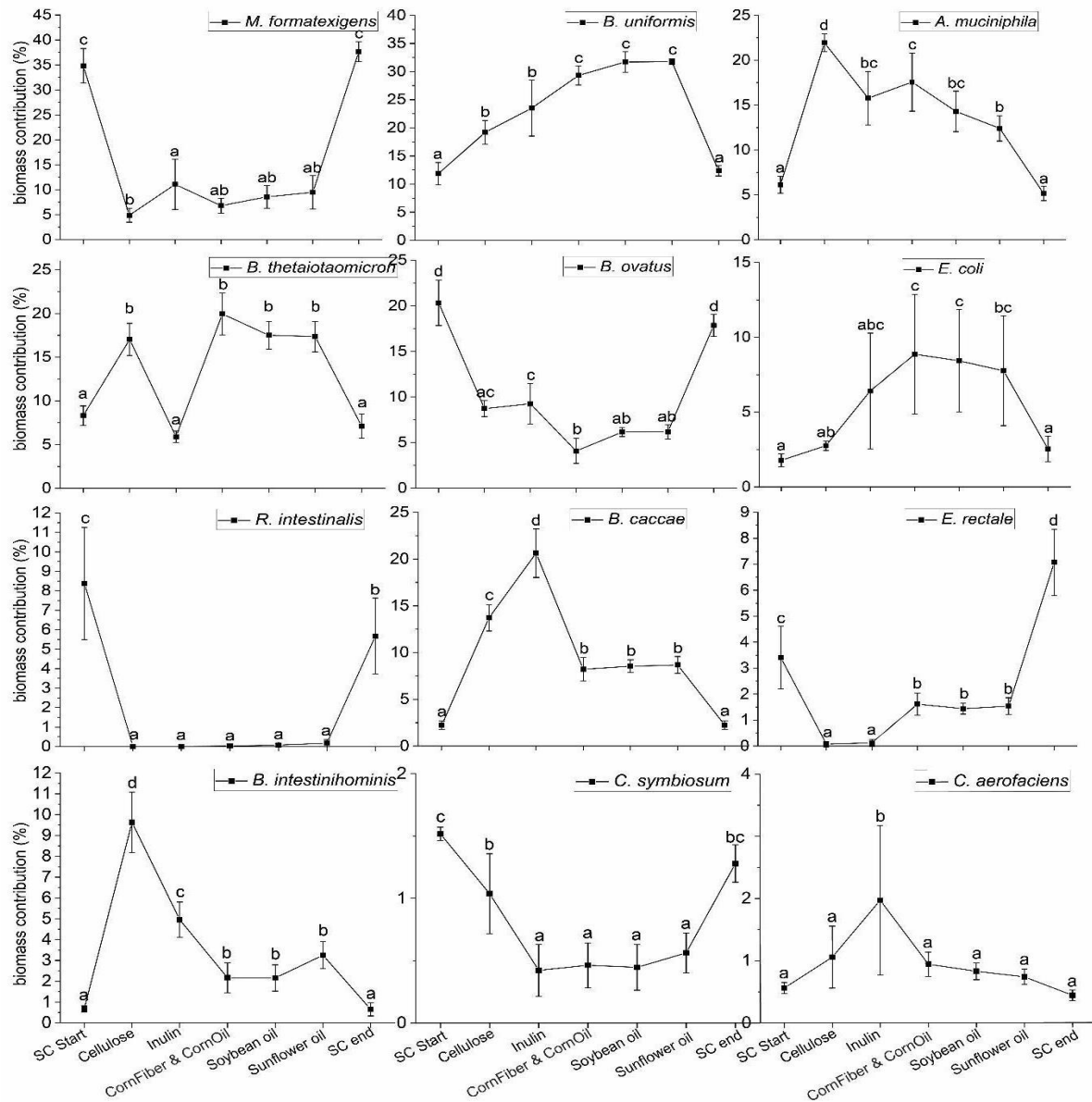


Figure 4-3. Relative proteinaceous biomass contribution of *M. formatexigens*, *B. uniformis*, *A. muciniphila*, *B. thetaiotaomicron*, *B. ovatus*, *E. coli*, *R. intestinalis*, *B. caccae*, *E. rectale*, *B. intestinihominis*, *C. symbiosum*, and *C. aerofaciens* in mice fed the fiber and fat diets (Group 2). Each point represents the relative biomass contribution of the organism after the mice were fed the diet indicated on the x-axis for seven days. Relative abundances were averaged and error bars indicate standard deviation (n = 6). Diets are ordered on the x-axis in chronological order fed to the mice. SC = standard chow diet. Cellulose = diet with cellulose as the fiber source; Corn fiber & Corn oil (“control”) = diet with corn fiber as the fiber source and corn oil as the fat source; Soybean oil = diet with soybean oil as the fat source. Sunflower oil = diet with sunflower oil as the fat source. Please note that the y-axis scale differs per row. Significant differences are displayed as different letters (a, b, c, d; based on one-way ANOVA and Tukey’s HSD post hoc test, p < .05).

4.3.2 Defined diets caused a decrease in estimated total bacterial load

We compared the bacterial to host biomass contribution in the fecal samples between mice fed the standard chow diet to mice fed the defined diets to estimate changes in total bacterial load. We found that the estimated bacterial load decreased substantially and significantly when mice were fed the defined diets (Welch's t-test, two-sided, $p < .05$). The average ratio of bacteria to host was 6.59 ± 1.62 and 6.05 ± 2.02 in the standard chow samples in group 1 ($n = 10$) and group 2 mice ($n = 12$), respectively. The ratio dropped to 2.94 ± 1.22 and 2.94 ± 0.86 in the defined diets in group 1 ($n = 15$) and group 2 mice ($n = 30$) (Fig. 4-4).

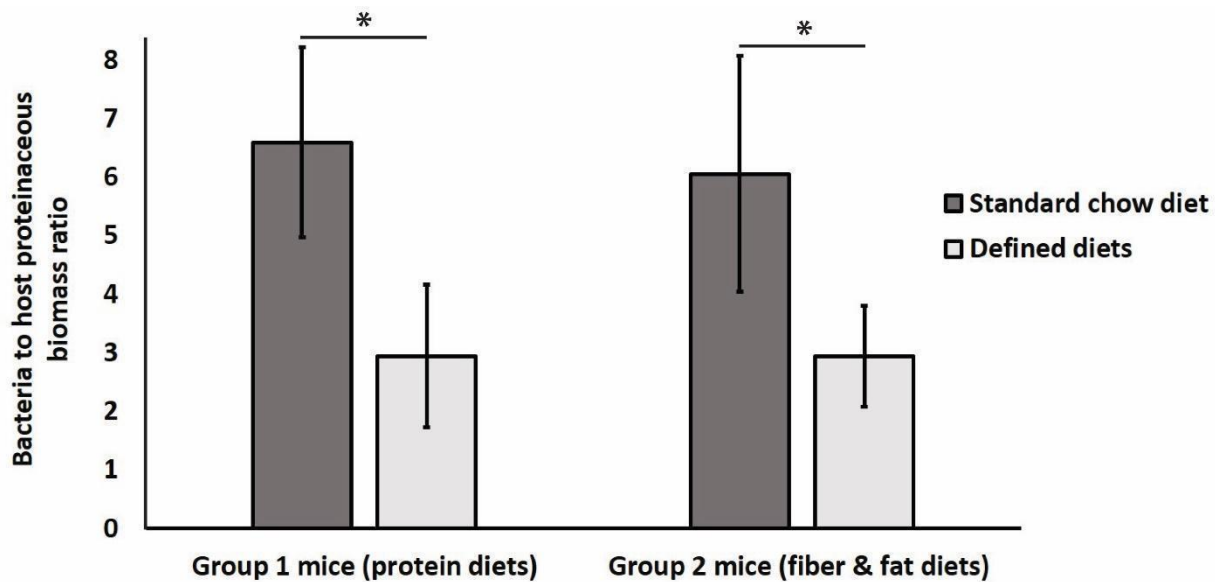


Figure 4-4. Comparison of the bacteria to host proteinaceous biomass ratios between the baseline standard chow diet and the defined diets. Bars represent the averaged biomass ratio of bacteria to host. Error bars = standard deviation ($n = 10$ in standard chow group 1, $n = 15$ in defined diets group 2; $n = 12$ in standard chow group 1, $n = 30$ in standard chow group 2).

4.3.3. Stable Isotope Fingerprints

We used the Protein Stable Isotope Fingerprinting (Protein-SIF) approach to estimate the natural isotopic signatures of the organisms in the microbiome samples as the mice were fed

different diets. Each diet contained one ingredient with a distinct signature. We related these changes in the dietary components' signatures to the changes in the signatures of the organisms to link dietary components to the organisms that may have assimilated them as growth substrates.

The accurate estimation of SIFs is dependent on the number of peptides available per organism, and therefore we obtained SIF values for the host mice (**Fig. 4-5**) and for the most abundant organisms in the community: *Marvinbryantia formatexigens*, *Bacteroides uniformis*, *Bacteroides thetaiotaomicron*, and *Akkermansia muciniphila* (**Fig. 4-6** for group 1 and **Fig. 4-7** for group 2). In the second group (fiber and fat diets), we also obtained SIFs for *Bacteroides caccae* and *E. coli*. We only plotted SIFs for organisms at least 3 data points in a minimum of two defined diets so that we could perform statistical analyses. Some sparse values were obtained for additional organisms and are reported in supplementary tables **S4-8** and **S4-9**.

We first computed SIF values for the organisms in the baseline standard chow samples to estimate reproducibility. We found no significant difference in the SIFs of the organisms in the baseline samples when the mice were fed the standard chow diet at the beginning of the experiment to when the mice were returned to their same standard chow diet at the end (Student's t-test, two-sided, BH correction, $p < .05$), with the exception of *A. muciniphila* ($p = .04$). Note that we did not know the signatures of the ingredients in the standard chow diet. Therefore we only used SIF values in the standard chow samples for reproducibility assessments and did not test for significant differences between standard chow and defined diets.

We observed a trend in the signature of the mouse host that was fed diets with different protein sources. Please note that we refer here to host-derived compounds found in the fecal samples and not host tissue. In the first group of mice, the host's SIF shifted in a pattern

resembling the pattern of the dietary protein sources (**Fig. 4-5 A**). The signature of the host was significantly different on the egg white diet than on the casein ($p = .0002$) or soy protein diet ($p = .008$). In the second group of mice, the host's SIF was neither affected by changes in dietary fiber source nor dietary fat source (**Fig. 4-5 B**). Its SIF remained constant around -11 ‰ , approximately lining up with the signature of the starch and sucrose.

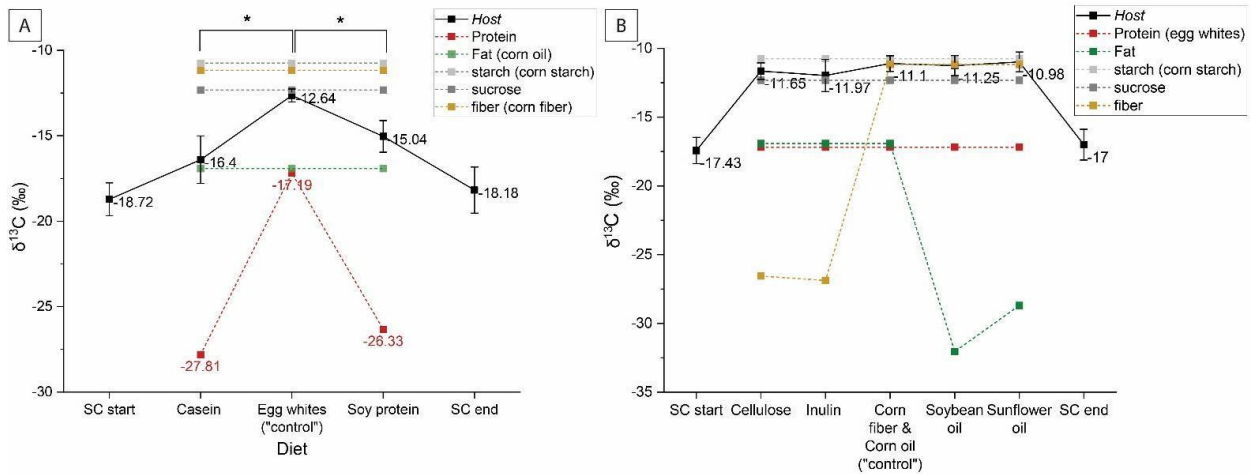


Figure 4-5. Stable isotope fingerprints (SIFs) of the mouse host. The signatures of the dietary components (protein, fat, starch, sucrose, and fiber) are the average between two replicates measured by EA-IRMS (measurement uncertainties of $\pm 0.42\text{ ‰}$ or less). **A) Group 1 mice fed the protein diets.** The signatures of the fat source (corn oil), starch (cornstarch), sucrose, and fiber (corn fiber) remained constant across all defined diets. The signature of the dietary protein varied based on the protein source (casein, egg whites, or soy protein; indicated on the x-axis). **B) Group 2 mice fed the fiber and fat diets.** The signatures of the starch (cornstarch) and sucrose remained constant across all defined diets. The signature of the fiber varied based on the fiber source (cellulose, inulin, or corn fiber; indicated on the x-axis). The signature of the fat varied based on the fat source (corn oil, soybean oil, or sunflower oil; indicated on the x-axis). Each point represents the mice's SIF after they were fed the diet indicated on the x-axis for seven days. SIFs were averaged across mice and error bars indicate standard deviation ($n = 5$ in group 1, $n = 6$ in group 2). Diets are ordered on the x-axis in chronological order fed to the mice. SC = standard chow diet (baseline). Signatures of the dietary components in the SC diet were unknown. Asterisks = significant difference ($p < .05$, t-tests corrected for multiple hypothesis testing, two-sided, Benjamini-Hochberg correction). We tested for statistical significance between the defined diets only because of the unknown signatures in the standard chow diet.

The majority of the bacterial SIF values in group 1 were lighter than the signature of the carbohydrates but heavier than the signatures of the dietary proteins. In group 1, changes in dietary protein source resulted in statistically significant changes in the SIFs of *M. formatexigens*, and *A. muciniphila* (t-tests, two-sided, BH correction, $p < .05$) (**Fig. 4-6**).

The SIF of *M. formatexigens* was similar in the casein diet and egg white diets samples, at around -18 ‰ and shifted up to about -8.5 ‰ on the soy protein diet (**Fig 4-6 A**). On the casein diet, the SIF of *M. formatexigens* was close to the signature of the fat source (~-17 ‰) and the signature of its mouse host (~16 ‰) and was in between the signatures of the casein (~ -28 ‰) and that of the carbohydrates (~ 11 ‰). On the egg whites diet, the SIF of *M. formatexigens* lined up with the signature of the protein source. However, the signature of the egg whites (~17 ‰) was not very distinct from the other dietary components. The introduction of the soy protein diet caused a significant shift ($p = .0004$ Soy vs. Casein; $p = 0.001$ Soy vs. Egg) in *M. formatexigens*'s SIF. Its SIF was around -8.5 ‰ which was closest to the starch, fiber, and sucrose signatures. The SIF of *B. uniformis* followed a similar pattern has the SIF of *M. formatexigens* (**Fig. 4-6 D**). The SIF of *A. muciniphila* followed the same pattern as the SIF of the host (**Fig. 4-6 B** compared to the host on **Fig. 4-5 A**). Its SIF differed significantly between the casein and egg whites samples ($p = .007$). Changes in the SIF of *B. thetaiotaomicron* did not test significant ($p > .05$)(**Fig. 4-6 C**).

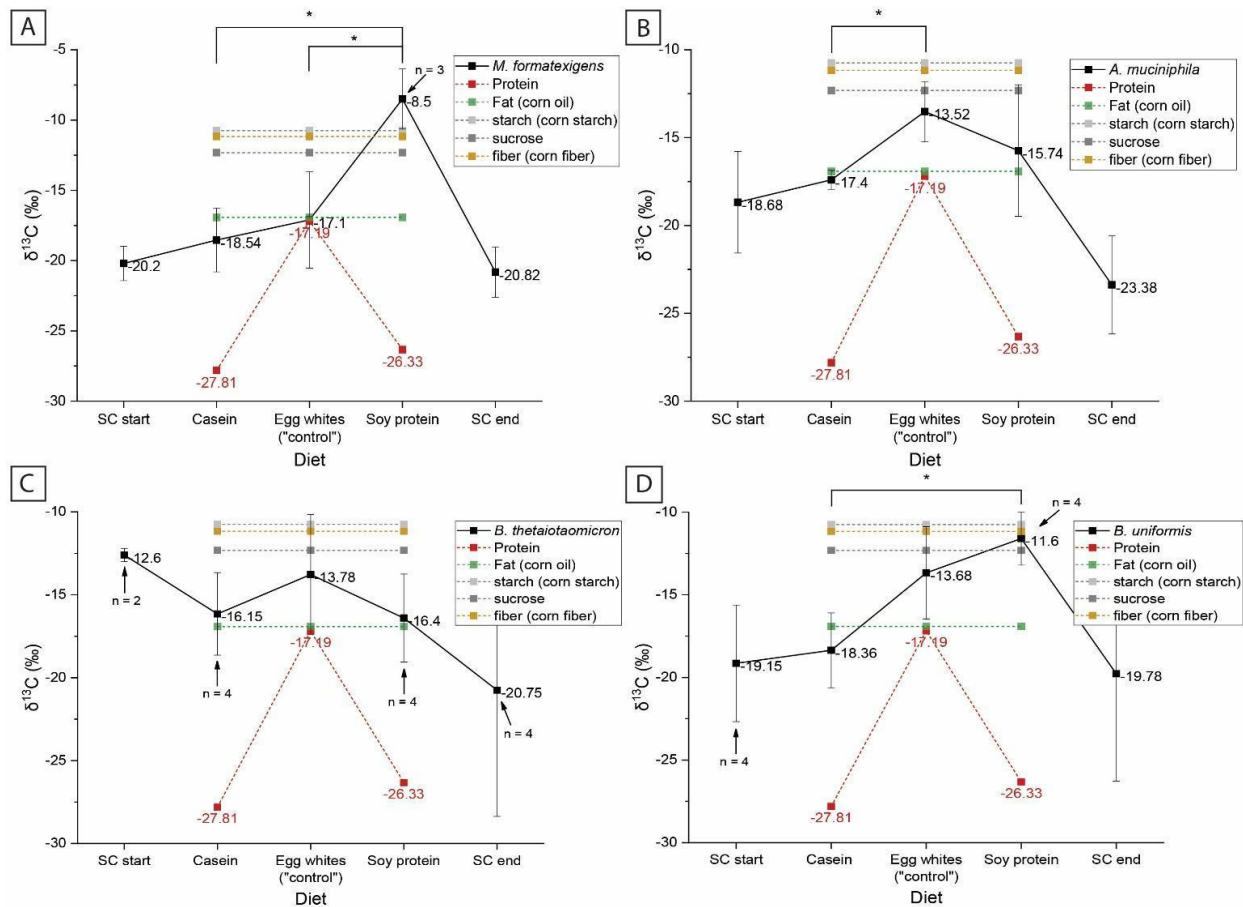


Figure 4-6. Stable isotope fingerprints (SIFs) of *M. formatexigens* (A), *A. muciniphila* (B), *B. thetaiotaomicron* (C), and *B. uniformis* (D) in group 1 mice fed the protein diets. The signatures of the dietary components (protein, fat, starch, sucrose, and fiber) are the average between two replicates measured by EA-IRMS (measurement uncertainties of ± 0.35 ‰ or less). The signatures of the starch (cornstarch) and sucrose remained constant across all defined diets. The signature of the fiber varied based on the fiber source (cellulose, inulin, or corn fiber; indicated on the x-axis). The signature of the fat varied based on the fat source (corn oil, soybean oil, or sunflower oil; indicated on the x-axis). Each point represents the SIF of the organism after the mice were fed the diet indicated on the x-axis for seven days. SIFs were averaged across mice and error bars indicate standard deviation ($n = 6$, unless indicated on the plots). Diets are ordered on the x-axis in chronological order fed to the mice. SC = standard chow diet (baseline diet). Signatures of the dietary components in the SC diet were unknown. Asterisks = significant difference ($p < .05$, t-tests corrected for multiple hypothesis testing, two-sided, Benjamini-Hochberg correction). We tested for statistical significance between the defined diets only because of the unknown signatures in the standard chow diet.

In group 2, changes in dietary fiber source resulted in statistically significant changes in the SIFs of *M. formatexigens* and *B. thetaiotaomicron* (t-tests, two-sided, BH correction, $p < .05$)

(**Fig. 4-7 A and C**). Changes in dietary fat source resulted in a significant change in the SIF of *A. muciniphila* (**Fig. 4-7 B**).

The SIF of *M. formatexigens* on the inulin diet was about – 18 ‰ which is lighter than the starch and sucrose but heavier than the inulin. It was closest to the signature of the protein and fat. The SIF of *M. formatexigens* significantly shifted up to around -11 ‰ with the introduction of the corn fiber diet ($p = .001$). The SIF of *M. formatexigens* appeared different but because of the low abundance of *M. formatexigens* on the cellulose diet, we only obtained two data points and thus could not properly test for significance. The SIF of *B. thetaiotaomicron* followed a similar pattern to the SIF of *M. formatexigens* in the fiber samples ($p = .03$ inulin vs. corn fiber; $p = .04$ inulin vs. cellulose). While the changes in the SIF of *A. muciniphila* between the different fiber sources did not test significant, its SIF differed significantly between the corn oil diet and the sunflower oil diet ($p = .007$). Changes in the SIFs of *B. uniformis*, *B. caccae*, and *E. coli* did not test significant ($p > .05$) (**Fig. 4-7 D, E, F**).

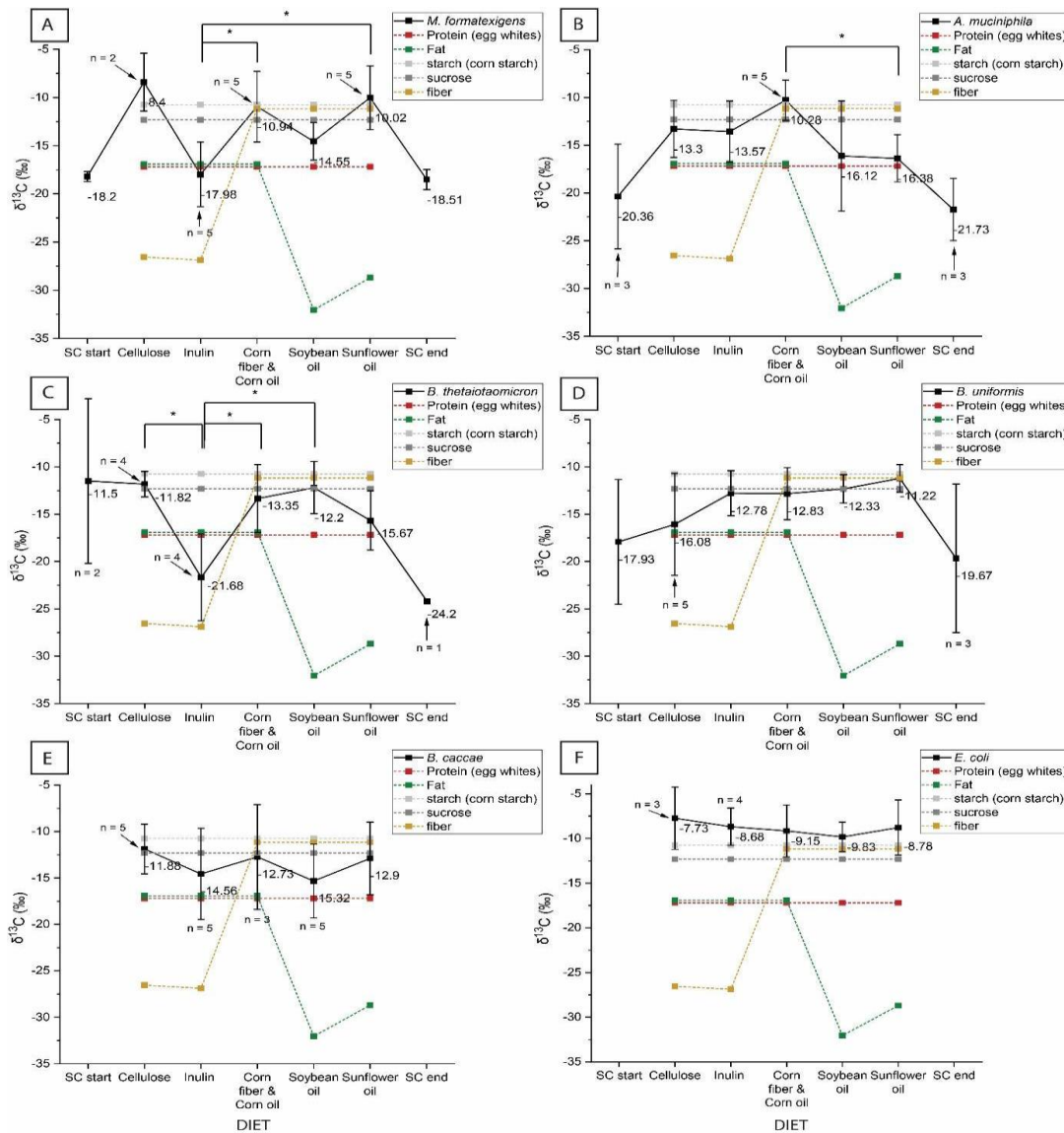


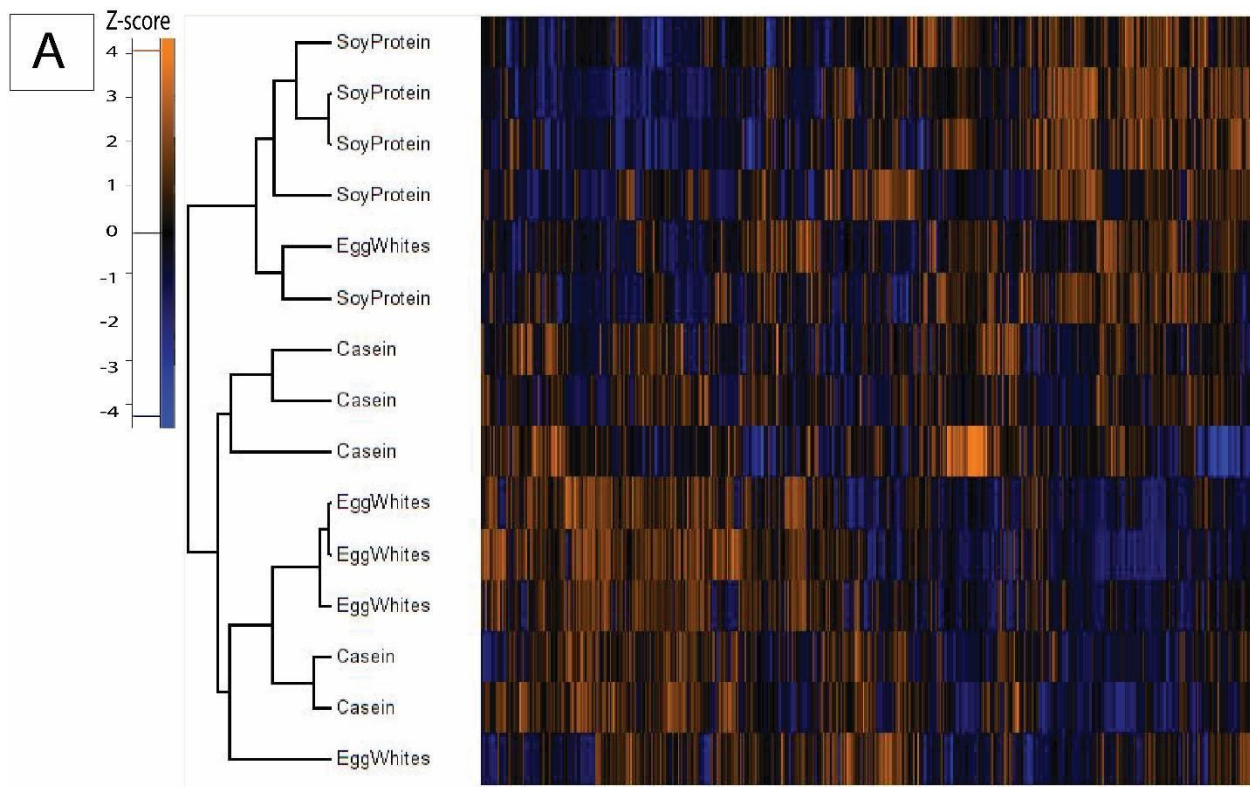
Figure 4-7. Stable isotope fingerprints (SIFs) of *M. formatexigens* (A), *A. muciniphila* (B), *B. thetaiotaomicron* (C), and *B. uniformis* (D), *B. caccae* (E), *E. coli* (F) in group 2 mice fed the fiber and fat diets. The signatures of the dietary components (protein, fat, starch, sucrose, and fiber) are the average between two replicates measured by EA-IRMS (measurement uncertainties of ± 0.42 ‰ or less). The signatures of the fat source (corn oil), starch (cornstarch), sucrose, and fiber (corn fiber) were constant in all defined diets. The signature of the dietary protein varied on the protein source (casein, egg whites, or soy protein; indicated on the x-axis). Each point represents the SIF of the organism after the mice were fed the diet indicated on the x-axis for seven days. SIFs were averaged across mice and error bars indicate standard deviation ($n = 5$, unless indicated on the plots). Diets are ordered on the x-axis in chronological order fed to the mice. SC = standard chow diet (baseline). Signatures of the dietary components in the SC diet were unknown. Asterisks = significant difference ($p < .05$, t-tests corrected for multiple hypothesis testing, two-sided, Benjamini-Hochberg correction). We tested for statistical significance between the defined diets only because of the unknown signatures in the standard chow diet.

4.3.4 Overall differences in Metaproteomes

Before diving into the specific changes that may explain the SIF results, we first looked at how the different diets shaped the metaproteomes of the microbial community.

We performed hierarchical clustering analysis (Pearson correlation, average linkage) on the metaproteomes of the group 1 mice fed the protein diets and found that the metaproteomes formed two main clusters (**Fig. 4-8 A**). One cluster contained all Soy protein samples and one Egg whites sample and the second cluster contained the remaining Egg whites samples and all Casein samples. Protein source alone did not seem to fully dictate how the metaproteomes clustered.

We next investigated how many proteins differed significantly in abundance between the metaproteomes of mice fed the different protein diets (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%; $S_0 = 0.1$, both sides, not paired). As reflected in the hierarchical clusters mentioned above, the Soy protein and Casein samples were the most different with 717 proteins that differed significantly in abundance (**Fig. 4-8 B**). 90 of these proteins were from the host and 627 were bacterial proteins. Despite the one Egg white sample clustered with the Soy protein samples, Egg whites and Soy protein samples were also very different, with 694 proteins that differed significantly in abundance. 116 proteins were from the host and 578 were bacterial proteins. The Casein and Egg whites diets samples differed slightly with 62 bacterial proteins that were significantly different in abundance.



	Egg Whites diet	Soy protein diet
Casein diet	62 proteins (all bacterial)	717 (90 host proteins, 627 bacterial)
Egg Whites diet		694 (116 host proteins, 578 bacterial)

Figure 4-8 Overall differences in the metaproteomes of group 1 mice (protein diets). **A)** Hierarchical clustering and heatmap based on Pearson correlation and average linkage, plotted on centered-log-ratio transformed data, and z-score normalized. **B)** Table displaying number of microbial and host proteins that differed significantly in relative abundance between the diets (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%; $S_0 = 0.1$, both sides, not paired)

We performed hierarchical clustering analysis (Pearson correlation, average linkage) on the metaproteomes of the group 2 mice fed the fiber and fat diets and found that the metaproteomes formed two main clusters (**Fig. 4-9 A**). One cluster contained all Corn fiber samples (Corn fiber & corn oil diet, soybean oil diet, and sunflower oil diet) while the second cluster contained the Cellulose and Inulin samples. Within this second cluster, the Inulin and

Cellulose samples formed two sub-clusters based on fiber. While some degree of clustering was based on the fat source within the first cluster (4 corn oil samples clustering together), fat source alone did not seem to dictate how the metaproteomes clustered.

We next investigated how many proteins differed significantly in abundance between the metaproteomes of mice fed the different fiber and fat diets (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%; $S_0 = 0.1$, both sides, not paired) (**Fig. 4-9 B**). While we found that changing the fiber source caused substantial changes in the metaproteomes, changing both the fat and fiber sources (at the same time) caused the most differences in protein abundances (group 2 mice). However, changing the fat source alone appeared to have only minimally affected the metaproteomes.

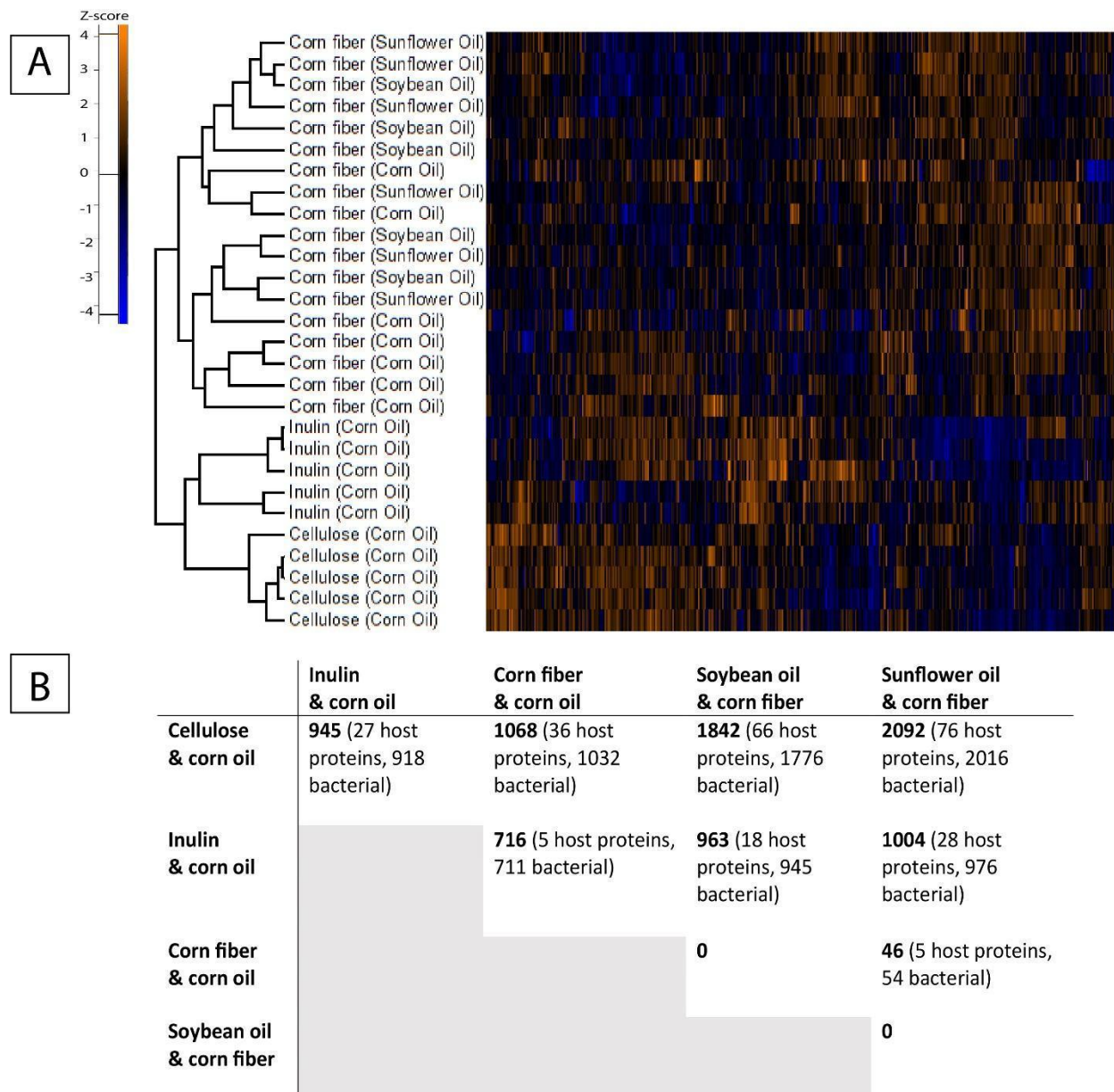


Figure 4-9. Overall differences in the metaproteomes of group 2 mice (fiber and fat diets)
A) Hierarchical clustering and heat map based on Pearson correlation and average linkage, plotted on centered-log-ratio transformed data, and z-score normalized. **B)** Table displaying the numbers of microbial and host proteins that differed significantly in relative abundance between the diets (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%; $S_0 = 0.1$, both sides, not paired).

Some of the numbers could be overinflated due to changes in the species' relative abundances, which caused some proteins to be more abundant without being a sign of functional

shifts. For this reason, we used the %orgNSAF (Mueller et al. 2010) approach in the next analyses to normalize counts per species.

4.3.5 Differentially abundant proteins

We investigated changes in the proteins profiles of *M. formatexigens*, *B. thetaiotaomicron*, and *A. muciniphila* between diets that gave significant Protein-SIF results.

M. formatexigens and dietary protein

The SIF of *M. formatexigens* differed significantly on the soy protein diet compared to the casein and egg whites diets (**Fig. 4-6 A**). Therefore, we identified and compared the most differentially abundant *M. formatexigens* proteins between the soy protein diet and the other two protein diets. All proteins showed on the plots of **Fig. 4-10** tested significant (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired).

We identified a highly abundant Sugar ABC transporter (Uniprot identifier C6LF52; undefined sugar) making up roughly 6% (**Suppl. table S4-10**) of the *M. formatexigens* proteins in the soy protein samples. This sugar ABC transporter was about 34 times more abundant in the soy protein diet than the casein diet and about 4 times more abundant than the egg white diet. Nine others *M. formatexigens* proteins were commonly more abundant in the soy protein diet than both the casein and egg white diets. For example, an uncharacterized protein (Uniprot ID C6LB56) was 96 times and 34 times more abundant in the soy protein diet than the casein and egg white diets, respectively.

Six *M. formatexigens* proteins were significantly upregulated in both the casein and egg white diets (**suppl. table S4-11**). The six proteins included three ABC transporters (Uniprot IDs C6LAI2 and C6L8Y8) and two enzymes that likely target host-derived N-glycans: Glucosamine-

6-phosphate deaminase (Uniprot ID C6LCU0), and N-acetylglucosamine-6-phosphate deacetylase (Uniprot ID C6LCT9) (**Fig. 4-10 A & B**).

M. formatexigens in the casein diet samples had a third abundant N-glycan-targeting enzyme: Putative N-acetylmannosamine-6-phosphate 2-epimerase (Uniprot ID C6L9A5) and several upregulated glycosyl hydrolases that were not upregulated in the egg white diets. Glycosyl hydrolases of families 4 and 31 are thought to be starch-targeting while glycosyl hydrolases of family 3 are thought to target hemicelluloses, B-glucans, or host-derived N-linked glycans.

M. formatexigens proteins that were upregulated in the egg diet samples but not the casein samples included two upregulated branch-chain amino-acid ABC transporters (C6LDZ4 and C6LL31) and an L-arabinose isomerase (Uniprot ID C6LHQ2). Egg whites are high in branch-chain amino acids, and arabinose is found in corn fiber, the fiber source of the protein diets.

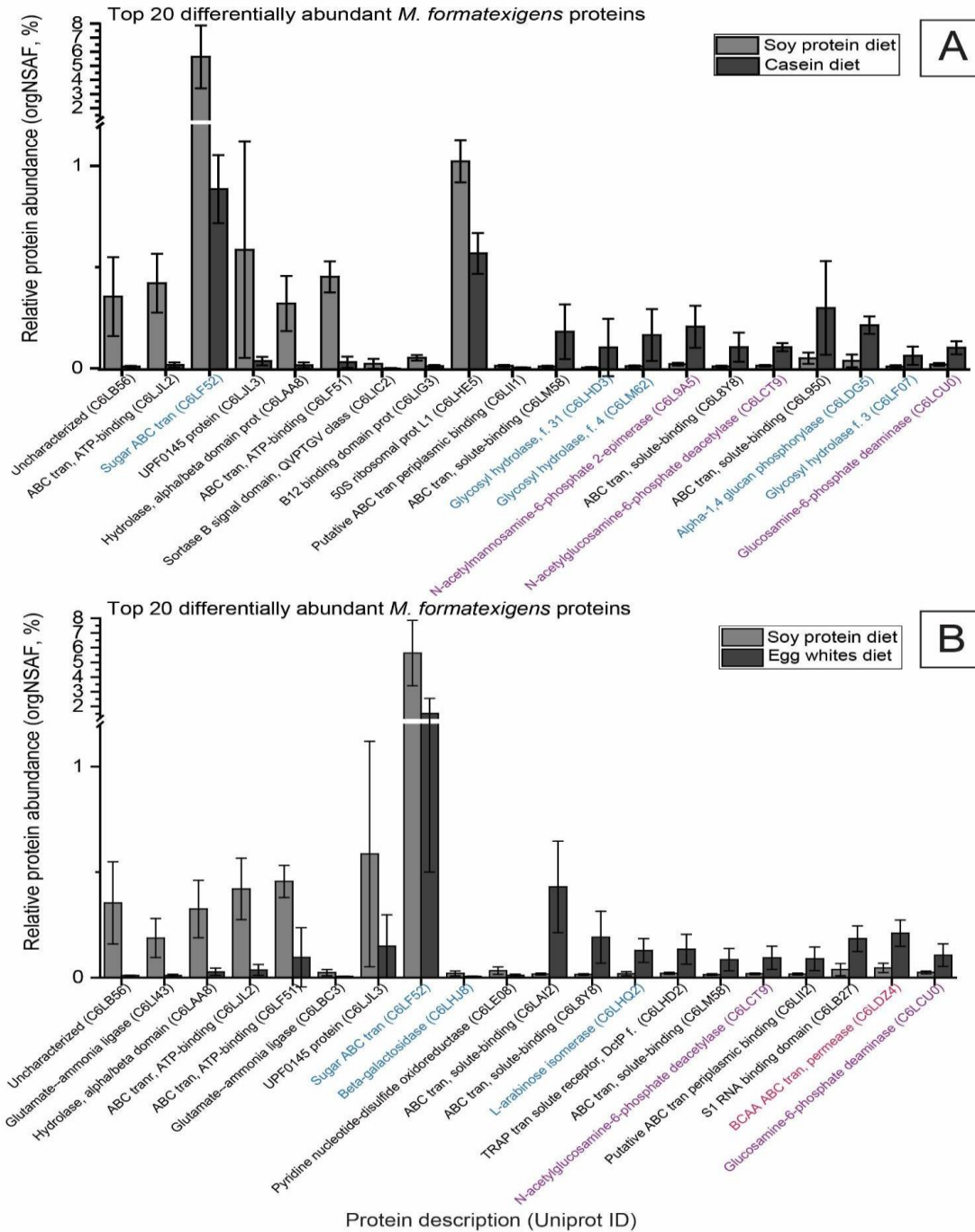


Figure 4-10. Top 20 differentially abundant *M. formatexigens* proteins between the protein diets. A and B) The 10 proteins on the left of the plot were significantly upregulated in the soy protein diet. **A)** The 10 proteins on the right were significantly upregulated in the casein diet. **B)** The 10 proteins on the right were significantly upregulated in the egg white diet. (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired). Error bars: standard deviation ($n = 5$). Blue indicates proteins involved in carbohydrate metabolism. Purple indicates hypothesized amino sugar metabolism. Red highlights a protein hypothesized to be involved in dietary protein uptake (based on Uniprot, KEGG and CAZy entries). Please note the break on the y-axis at 1.5%.

***M. formatexigens* and dietary fiber**

The SIF of *M. formatexigens* differed significantly between the inulin and corn fiber diets (**Fig. 4-7 A**). Therefore, we identified and compared the most differentially abundant *M. formatexigens* proteins between the inulin and corn fiber samples. All proteins showed on **Fig. 4-11** tested significant (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired).

There was no evidence of inulin degradation by *M. formatexigens* based on differentially abundant proteins. However, two starch degradation enzymes appeared upregulated in the inulin diet samples (alpha-amylases C6LAL3 and C6L913). Additionally, a branch-chain-amino-acid interacting protein was higher in abundance in these samples (C6LMB4). The most highly abundant protein was an undefined ABC transporter (C6LAL7).

As for the corn fiber samples, we identified the same Sugar ABC transporter (Uniprot identifier C6LF52; undefined sugar) highly abundant in the Soy protein diet (**Fig. 4-10**) be highly abundant in the Corn fiber diet as well. We also identified four abundant glycosyl hydrolases of family 3 (C6LJU1, C6L988, C6LJM4, C6LM74) and an O-glycosyl hydrolase of family 30 (C6LMK7). Family 3 glycosyl hydrolases may target hemicelluloses, B-glucans, or host-derived N-linked glycans.

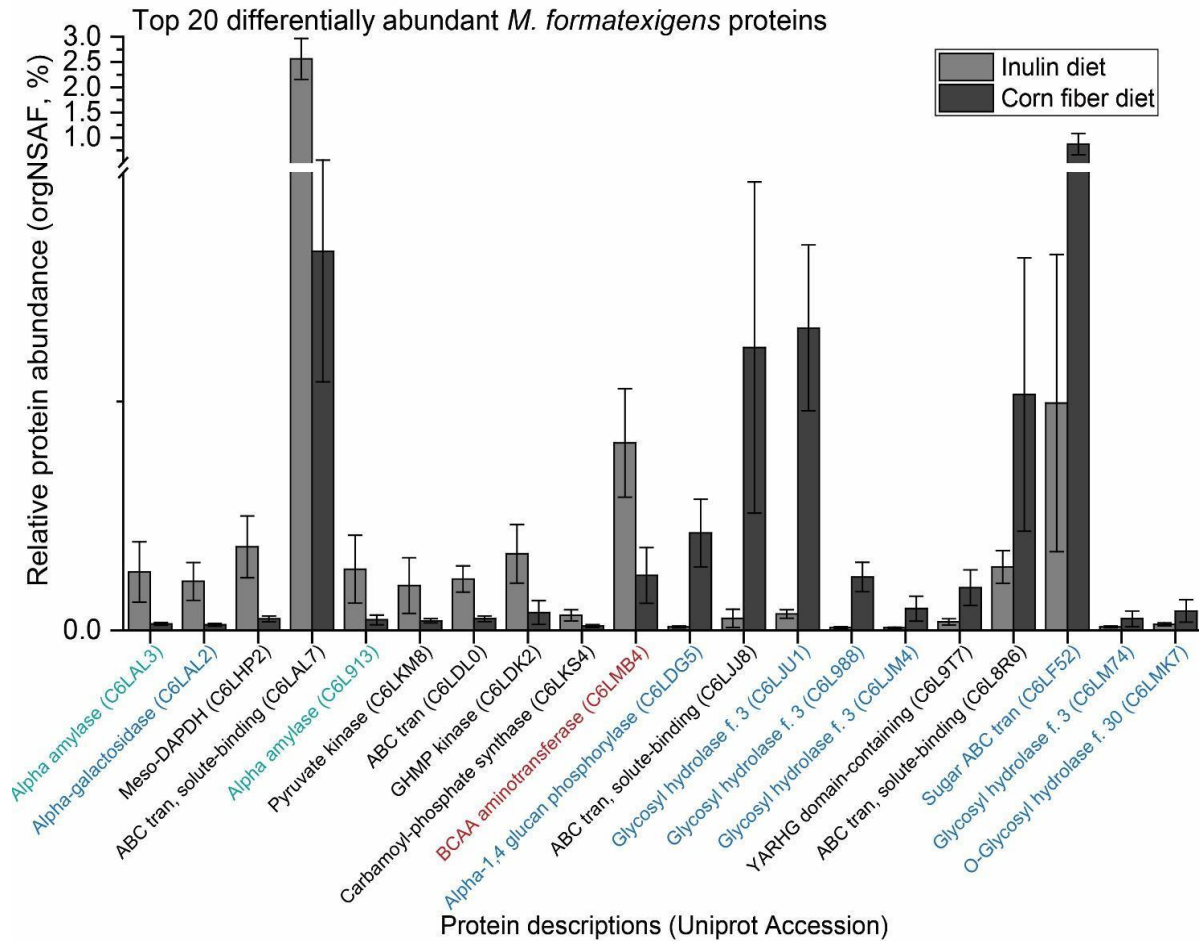


Figure 4-11. Top 20 differentially abundant *M. formatexigens* proteins between the inulin and corn fiber diet. The 10 proteins on the left of the plot were significantly upregulated in the inulin diet and the 10 proteins on the right were significantly upregulated in the corn fiber diet. (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired). Error bars: standard deviation ($n = 5$). Blue indicates proteins involved in carbohydrate metabolism and in green starch degradation in particular. Red highlights a protein hypothesized to be involved in dietary protein uptake. Please note the break on the y-axis at 0.75%.

A. *muciniphila* and dietary protein

The SIF of *A. muciniphila* differed significantly in the Soy protein samples compared to the Casein samples (**Fig. 4-6 B**). However, we identified many differentially abundant *A. muciniphila* proteins between the Egg white samples and Soy protein samples, but we did find any significant difference in the protein profiles of *A. muciniphila* in the casein samples compared to the soy protein samples (comparison of interest, based on SIF results). Both

proteomes contained several abundant mucus targeting enzymes, including the highly abundant N-acetylglucosamine-6-phosphate deacetylase (B2UQP5). Glyceraldehyde-3-phosphate dehydrogenase (B2UKW7) was the most abundant *A. muciniphila* protein across all protein diets.

***A. muciniphila* and dietary fat**

The SIF of *A. muciniphila* differed significantly on the sunflower oil diet compared to the corn oil diet (**Fig 4-7 B**). However, we identified many differentially abundant *A. muciniphila* proteins when changing both the source of fat and fiber (Cellulose/Corn oil diet vs. Corn fiber/Sunflower oil diet, for example) but we did not find any significant difference in the protein profiles of *A. muciniphila* between the sunflower oil diet and the corn oil diet (comparison of interest, based on SIF results). Again, we observed several abundant mucus targeting enzymes, including highly-abundant N-acetylglucosamine-6-phosphate deacetylase (B2UQP5), ~0.94% & glucosamine-6-phosphate deaminase (B2UN29) ~0.65%. Glyceraldehyde-3-phosphate dehydrogenase (B2UKW7) was the most abundant *A. muciniphila* protein across all fat diets, making up about 5.31% of the *A. muciniphila* proteins. Additionally, several peptidases were abundant in the fat diets.

***B. thetaiotaomicron* and fiber**

The SIF of *B. thetaiotaomicron* differed significantly between the inulin diet and the other two fiber diets (**Fig. 4-7 C**). Therefore, we identified and compared the most differentially abundant *B. thetaiotaomicron* proteins in these comparisons. All proteins showed on Fig. 4-12 tested significant (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired).

We identified several starch-targeting enzymes that were significantly more commonly more abundant in the cellulose and the corn fiber samples than in the inulin samples, such as a starch-binding protein (Q8A1G2) and the protein complex SusD/SusC. Additionally, the cellulose sample contained a highly abundant host-derived N-linked glycans–targeting enzyme (Q8ABW9) from *B. thetaiotaomicron*. Among the top differentially abundant proteins, we found one that may target cellulose (alpha-glucosidase Q8ABY2) but was not as abundant as the starch-targeting enzymes.

As for the inulin samples, we identified three proteins that likely indicate inulin degradation: glycosyl hydrolase family 32 (Q8A6W6), fructokinase (Q8A6W6), and levanase (Q8A6W7). GH family 32 encodes the inulinases. Levanase also has the potential to degrade inulin. Fructokinase may interact with fructose monomers of inulin. Furthermore, Sonnenburg et al. identified these proteins to form a putative fructan utilization locus in *B. thetaiotaomicron* (Sonnenburg et al. 2010).

As a side note, we found a differentially abundant bile acid-targeting enzyme (choloylglycine hydrolase C8A600) as additional evidence of interactions between the bacteria and the host-derived compounds.

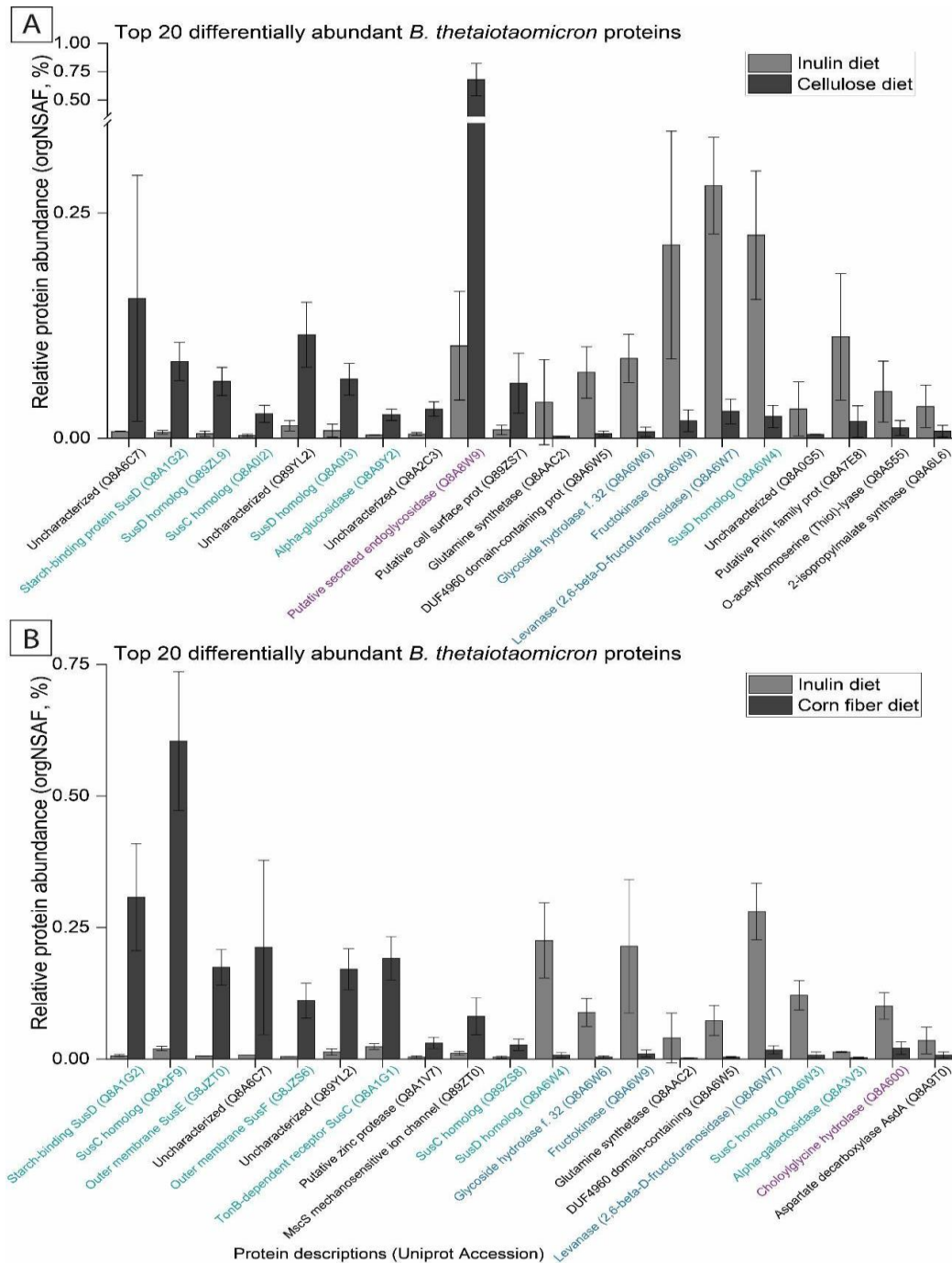


Figure 4-12. Top 20 differentially abundant *B. thetaiotaomicron* proteins between the fiber diets. A) The 10 proteins on the left of the plot were significantly upregulated in the cellulose diet and the 10 proteins on the right were significantly upregulated in the inulin diet (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$, both sides, not paired). Error bars: standard deviation ($n = 6$). Green indicates proteins likely involved in starch degradation. Purple indicates hypothesized metabolism of host compounds. Blue indicates proteins hypothesized to be involved in inulin degradation (based on Uniprot, KEGG and CAZy entries).

4.4 Discussion

Our current understanding of diet and intestinal microbiota interactions is largely based on studies correlating relative abundance to dietary inputs. However, measurements of relative abundance can yield misleading results. For example, in our study, *B. thetaiotaomicron* substantially decreased in abundance in gnotobiotic mice fed a diet containing inulin as the source of fiber. However, we found substantial evidence for the use of inulin by *B. thetaiotaomicron*, demonstrating the need for functional-based methods and more direct measurements.

We attempted to directly link intestinal bacteria to their diet-derived growth substrates using Metaproteomics and Protein-SIF. Our unique approach consisted of comparing natural isotopic signatures of intestinal bacteria to signatures of dietary components consumed by their host, followed by an analysis of upregulated proteins. We found that bacteria likely used a combination of growth substrates, and substrate choice appeared to be driven by competition. While we did not find enough evidence for the use of dietary fats by the intestinal bacteria, we demonstrated the potential of our methods to associate intestinal bacteria with the carbohydrates, proteins, and host-derived compounds they metabolize. Additionally, we found that defined diets with low ingredient diversity, such as the AIN-93 diet, are not favorable to the bacterial community in our study. Thus, some of the species may have to rely partially on host-derived compounds for substrates.

Apart from the dietary components, we detected an additional signature that we compared to the signatures of the bacteria: the signature of the mouse host. Our Protein-SIF results suggest that within a week, sufficient amounts of carbon are assimilated by the host to largely influence signatures of host proteins that are turned over quickly (mucin, digestive enzymes, etc.).

The Protein-SIF approach gave another layer of evidence for host foraging by *M. formatexigens* and *A. muciniphila*. In the metaproteomes of the gnotobiotic mice fed the defined diets, we detected highly abundant *M. formatexigens* and *A. muciniphila* enzymes that target N-linked glycans, likely host-derived glycans from the mucus layer. The signature of *A. muciniphila* in the group 1 mice (fed the protein diets) overlapped with the host's signature and followed the same trend. The signature of *M. formatexigens* also followed the same trend as the host's signature, but its signature deviated from the host's signature in the Soy protein samples. *M. formatexigens* enzymes that likely targeted host-derived glycans were present at much lower levels in these samples.

The organisms likely used host-derived compounds because the defined diets were not favorable to their growth. We observed a substantial decrease in the estimated bacterial load of the community in mice fed the defined diets. One may argue that the community adapted to the standard chow during the first three weeks of the experiment, and the switch to new diets caused the decrease in biomass. However, the community was exposed to the defined diets afterward for three weeks in group 1 and five weeks in group 2, and we observed the same biomass results when the mice were later returned to the standard chow. Differences between the standard chow diet composition and the composition of the defined diets likely explain this decrease in biomass. The overall nutritional composition differed mainly in the quantity of dietary protein. The standard chow contained about 25 % protein (formula suppl. S8), while the defined diets contained about 16 % protein (formulas suppl. S1-7). The fat and fiber contents were similar between the two diet groups, making up about 12 % and 4 - 5 % of both diet types. Faith et al. found that dietary protein may be a primary limiting factor for bacterial biomass (Faith et al. 2011). They found that in gnotobiotic mice colonized with a 10-species microbiota, the bacterial

community biomass (measured as DNA/ μ g fecal material) increased with increases in dietary protein content. Additionally, they found the species behaved differently to increased dietary protein; *Marvinbryantia formatexigens* and *Eubacterium rectale* decreased in abundance. In a different study, *E. rectale* also decreased in relative abundance in microbiomes of human subjects fed high-protein diets, as well as *R. intestinalis* (Russell et al. 2011). We, however, observed that *M. formatexigens*, *E. rectale*, and *R. intestinalis* had higher levels of relative abundance in the standard chow samples. This contradictory result suggests that the quantity of dietary protein was not the explanatory factor to the decreased biomass in the defined diets.

Instead, a more likely explanation is that the defined diets were less diverse in ingredient sources. Species may be competing for substrates, thus inhibiting bacteria from growing to their full potential. An example from our data illustrating this idea would be *B. thetaiotaomicron*'s substantial decrease in relative abundance in the microbiomes of mice fed the inulin diet. A previous study with gnotobiotic mice bi-colonized with *B. thetaiotaomicron* and *M. formatexigens* showed that the presence of *M. formatexigens* caused a decrease in the abundance of *B. thetaiotaomicron* compared to mono-colonized controls (Rey et al. 2010). In our study, *B. thetaiotaomicron* seemed to favor starch over fiber as a substrate in corn fiber samples, based on the analysis of differentially abundant proteins combined with the protein-SIF evidence. *M. formatexigens* appear to use corn fiber as a substrate. Without the corn fiber, *M. formatexigens* switched to consuming starch in the inulin diet, potentially forcing *B. thetaiotaomicron* to use inulin instead. However, inulin is not an ideal substrate for *B. thetaiotaomicron*. The organism does not grow well on inulin *in vitro* (Desai et al. 2016), which may explain its drop in relative abundance in our study.

Our interpretation is limited to the microorganisms that were high enough in abundance that we could compute a SIF. The other organisms in the community likely played a role in the competition for substrates. Furthermore, because competition appeared to play a major role in the substrate choice of bacteria, having a small defined community may not represent the substrate choices of these species in the human intestinal tract. Inter-individual variation is high in intestinal microbiomes (Blakeley-Ruiz et al. 2019; Carruthers et al. 2019), and therefore, growth-substrates may depend on community composition.

We demonstrated the potential of the Protein-SIF approach and how it provides a unique angle to study the interactions between diet and intestinal microbiota. Furthermore, the Protein-SIF approach relies on natural isotopic signatures of dietary components with no heavy labels, and therefore, the method could be applied to human subjects in the future.

4.5 Acknowledgments

We are sincerely grateful to Karen Flores and Dr. Susan Tonkonogy for their expertise with gnotobiotic mice and for performing all manipulations of the mice. We thank Nicholas Pudlo and Dr. Eric Martens for providing the bacterial strains. We are grateful to Deniz Durmusoglu and Dr. Nathan Crook for their assistance with growing the anaerobic strains. We thank Abigail Korenek for her assistance in preparing culture media. We thank Roxane Bowden and Dr. Christopher Osbourn for the IRMS measurements of the ingredients. We thank Dr. Fernanda Salvato for providing A.M. with extensive metaproteomics training and everyone in the Kleiner Lab for discussing the experimental design and results, particularly Ali Bartlett and Alfredo Blakeley-Ruiz. We made all LC-MS/MS measurements in the Molecular Education, Technology, and Research Innovation Center (METRIC) at NC State University.

Declarations

Ethics Approval

The protocols for husbandry and experimentation of all mice used in this study were approved by the Institutional Animal Care and Use Committee at North Carolina State University (Institution reference: D16-00214).

Availability of Data

The mass spectrometry metaproteomics data and protein sequence database *will be* deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R35GM138362 and the Foundation for Food and Agriculture Research (FFAR) Grant ID: 593607. The Gnotobiotic Core at the College of Veterinary Medicine, North Carolina State University is supported by the National Institutes of Health, funded by the Center for Gastrointestinal Biology and Disease, NIH-NIDDK P30 DK034987.

REFERENCES

- Agans, Richard, Alex Gordon, Denise Lynette Kramer, Sergio Perez-Burillo, José A. Rufián-Henares, and Oleg Paliy. 2018. “Dietary Fatty Acids Sustain the Growth of the Human Gut Microbiota.” *Applied and Environmental Microbiology* 84 (21). <https://doi.org/10.1128/AEM.01525-18>.
- Aitchison, J. 1982. “The Statistical Analysis of Compositional Data.” *Journal of the Royal Statistical Society, Methodological*, 44 (2): 139–77.
- Blakeley-Ruiz, J. Alfredo, Alison R. Erickson, Brandi L. Cantarel, Weili Xiong, Rachel Adams, Janet K. Jansson, Claire M. Fraser, and Robert L. Hettich. 2019. “Metaproteomics Reveals Persistent and Phylum-Redundant Metabolic Functional Stability in Adult Human Gut Microbiomes of Crohn’s Remission Patients despite Temporal Variations in Microbial Taxa, Genomes, and Proteomes.” *Microbiome* 7 (1): 1–15. <https://doi.org/10.1186/s40168-019-0631-8>.
- Boursier, Jérôme, Olaf Mueller, Matthieu Barret, Mariana Machado, Lionel Fizanne, Felix Araujo-Perez, Cynthia D. Guy, et al. 2016. “The Severity of Nonalcoholic Fatty Liver Disease Is Associated with Gut Dysbiosis and Shift in the Metabolic Function of the Gut Microbiota.” *Hepatology* 63 (3): 764–75. <https://doi.org/10.1002/hep.28356>.
- Carruthers, Lauren V., Arinaitwe Moses, Moses Adriko, Christina L. Faust, Edridah M. Tukahebwa, Lindsay J. Hall, Lisa C. Ranford-Cartwright, and Poppy H. L. Lamberton. 2019. “The Impact of Storage Conditions on Human Stool 16S rRNA Microbiome Composition and Diversity.” *PeerJ* 7 (December): e8133. <https://doi.org/10.7717/peerj.8133>.
- Chambers, Matthew C., Brendan Maclean, Robert Burke, Dario Amodei, Daniel L. Ruderman, Steffen Neumann, Laurent Gatto, et al. 2012. “A Cross-Platform Toolkit for Mass Spectrometry and Proteomics.” *Nature Biotechnology* 30 (10): 918. <https://doi.org/10.1038/nbt.2377>.
- David, Lawrence A., Corinne F. Maurice, Rachel N. Carmody, David B. Gootenberg, Julie E. Button, Benjamin E. Wolfe, Alisha V. Ling, et al. 2014. “Diet Rapidly and Reproducibly Alters the Human Gut Microbiome.” *Nature* 505 (7484): 559–63. <https://doi.org/10.1038/nature12820>.
- De Filippis, Francesca, Paola Vitaglione, Rosario Cuomo, Roberto Berni Canani, and Danilo Ercolani. 2018. “Dietary Interventions to Modulate the Gut Microbiome—How Far Away Are We From Precision Medicine.” *Inflammatory Bowel Diseases* 24 (10): 2142–54. <https://doi.org/10.1093/ibd/izy080>.
- Desai, M. S., A. M. Seekatz, N. M. Koropatkin, N. Kamada, C. A. Hickey, M. Wolter, N. A. Pudlo, et al. 2016. “A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility.” *Cell* 167 (5): 1339–1353.e21. <https://doi.org/10.1016/j.cell.2016.10.043>.

- Diether, Natalie E., and Benjamin P. Willing. 2019. "Microbial Fermentation of Dietary Protein: An Important Factor in Diet–Microbe–Host Interaction." *Microorganisms* 7 (1): 19. <https://doi.org/10.3390/microorganisms7010019>.
- Faith, Jeremiah J., Nathan P. McNulty, Federico E. Rey, and Jeffrey I. Gordon. 2011. "Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice." *Science (New York, N.Y.)* 333 (6038): 101–4. <https://doi.org/10.1126/science.1206025>.
- Fernandes, Andrew D., Jennifer NS Reid, Jean M. Macklaim, Thomas A. McMurrough, David R. Edgell, and Gregory B. Gloor. 2014. "Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-Seq, 16S RRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis." *Microbiome* 2 (1): 1–13. <https://doi.org/10.1186/2049-2618-2-15>.
- Gentile, Christopher L., and Tiffany L. Weir. 2018. "The Gut Microbiota at the Intersection of Diet and Human Health." *Science* 362 (6416): 776–80. <https://doi.org/10.1126/science.aau5812>.
- Gloor, Gregory B., Jean M. Macklaim, Vera Pawlowsky-Glahn, and Juan J. Egozcue. 2017. "Microbiome Datasets Are Compositional: And This Is Not Optional." *Frontiers in Microbiology* 8: 2224. <https://doi.org/10.3389/fmicb.2017.02224>.
- Kleiner, Manuel, Xiaoli Dong, Tjorven Hinzke, Juliane Wippler, Erin Thorson, Bernhard Mayer, and Marc Strous. 2018. "Metaproteomics Method to Determine Carbon Sources and Assimilation Pathways of Species in Microbial Communities." *Proceedings of the National Academy of Sciences* 115 (24): E5576–84. <https://doi.org/10.1073/pnas.1722325115>.
- Kleiner, Manuel, Angela Kouris, Marlene Jensen, Yihua Liu, Janine McCalder, and Marc Strous. 2021. "Ultra-Sensitive Protein-SIP to Quantify Activity and Substrate Uptake in Microbiomes with Stable Isotopes." *BioRxiv*, March, 2021.03.29.437612. <https://doi.org/10.1101/2021.03.29.437612>.
- Kleiner, Manuel, Erin Thorson, Christine E. Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. 2017. "Assessing Species Biomass Contributions in Microbial Communities via Metaproteomics." *Nature Communications* 8 (1): 1558. <https://doi.org/10.1038/s41467-017-01544-x>.
- Li, Weizhong, and Adam Godzik. 2006. "Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences." *Bioinformatics (Oxford, England)* 22 (13): 1658–59. <https://doi.org/10.1093/bioinformatics/btl158>.
- Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2014. "The Carbohydrate-Active Enzymes Database (CAZy) in 2013." *Nucleic Acids Research* 42 (Database issue): D490. <https://doi.org/10.1093/nar/gkt1178>.
- Mueller, Ryan S., Vincent J. Deneff, Linda H. Kalnejais, K. Blake Suttle, Brian C. Thomas, Paul

- Wilmes, Richard L. Smith, et al. 2010. "Ecological Distribution and Population Physiology Defined by Proteomics in a Natural Microbial Community." *Molecular Systems Biology* 6 (1): 374. <https://doi.org/10.1038/msb.2010.30>.
- Oberg, Ann L., and Olga Vitek. 2009. "Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments." *Journal of Proteome Research* 8 (5): 2144–56. <https://doi.org/10.1021/pr8010099>.
- Patnode, Michael L., Zachary W. Beller, Nathan D. Han, Jiye Cheng, Samantha L. Peters, Nicolas Terrapon, Bernard Henrissat, et al. 2019. "Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans." *Cell* 179 (1): 59–73. <https://doi.org/10.1016/j.cell.2019.08.011>.
- Pelz, Oliver, Luis A. Cifuentes, Beth T. Hammer, Cheryl A. Kelley, and Richard B. Coffin. 1998. "Tracing the Assimilation of Organic Compounds Using $\Delta^{13}\text{C}$ Analysis of Unique Amino Acids in the Bacterial Peptidoglycan Cell Wall." *FEMS Microbiology Ecology* 25 (3): 229–40. <https://doi.org/10.1111/j.1574-6941.1998.tb00475.x>.
- Portune, Kevin J., Martin Beaumont, Anne-Marie Davila, Daniel Tomé, François Blachier, and Yolanda Sanz. 2016. "Gut Microbiota Role in Dietary Protein Metabolism and Health-Related Outcomes: The Two Sides of the Coin." *Trends in Food Science & Technology*, Unravelling the role of the gut microbiome in energy balance and brain development and function: the European project MyNewGut, 57 (November): 213–32. <https://doi.org/10.1016/j.tifs.2016.08.011>.
- Reeves, Philip G., Forrest H. Nielsen, and George C. Fahey. 1993. "AIN-93 Purified Diets for Laboratory Rodents: Final Report of the American Institute of Nutrition Ad Hoc Writing Committee on the Reformulation of the AIN-76A Rodent Diet." *The Journal of Nutrition* 123 (11): 1939–51. <https://doi.org/10.1093/jn/123.11.1939>.
- Rey, Federico E., Jeremiah J. Faith, James Bain, Michael J. Muehlbauer, Robert D. Stevens, Christopher B. Newgard, and Jeffrey I. Gordon. 2010. "Dissecting the in Vivo Metabolic Potential of Two Human Gut Acetogens *." *Journal of Biological Chemistry* 285 (29): 22082–90. <https://doi.org/10.1074/jbc.M110.117713>.
- Rowland, Ian, Glenn Gibson, Almut Heinken, Karen Scott, Jonathan Swann, Ines Thiele, and Kieran Tuohy. 2018. "Gut Microbiota Functions: Metabolism of Nutrients and Other Food Components." *European Journal of Nutrition* 57 (1): 1–24. <https://doi.org/10.1007/s00394-017-1445-8>.
- Russell, Wendy R., Silvia W. Gratz, Sylvia H. Duncan, Grietje Holtrop, Jennifer Ince, Lorraine Scobbie, Garry Duncan, et al. 2011. "High-Protein, Reduced-Carbohydrate Weight-Loss Diets Promote Metabolite Profiles Likely to Be Detrimental to Colonic Health." *The American Journal of Clinical Nutrition* 93 (5): 1062–72. <https://doi.org/10.3945/ajcn.110.002188>.
- Schoeler, Marc, and Robert Caesar. 2019. "Dietary Lipids, Gut Microbiota and Lipid Metabolism." *Reviews in Endocrine and Metabolic Disorders* 20 (4): 461–72.

<https://doi.org/10.1007/s11154-019-09512-0>.

- Sonnenburg, Erica D., Hongjun Zheng, Payal Joglekar, Steven K. Higginbottom, Susan J. Firbank, David N. Bolam, and Justin L. Sonnenburg. 2010. "Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations." *Cell* 141 (7): 1241–52. <https://doi.org/10.1016/j.cell.2010.05.005>.
- Speare, Lauren, Stephanie Smith, Fernanda Salvato, Manuel Kleiner, and Alecia N. Septer. 2020. "Environmental Viscosity Modulates Interbacterial Killing during Habitat Transition." *MBio* 11 (1): e03060-19. <https://doi.org/10.1128/mBio.03060-19>.
- Theriot, Casey M., Mark J. Koenigskecht, Paul E. Carlson, Gabrielle E. Hatton, Adam M. Nelson, Bo Li, Gary B. Huffnagle, Jun Z. Li, and Vincent B. Young. 2014. "Antibiotic-Induced Shifts in the Mouse Gut Microbiome and Metabolome Increase Susceptibility to Clostridium Difficile Infection." *Nature Communications* 5 (1): 3114. <https://doi.org/10.1038/ncomms4114>.
- Turnbaugh, Peter J., Ruth E. Ley, Michael A. Mahowald, Vincent Magrini, Elaine R. Mardis, and Jeffrey I. Gordon. 2006. "An Obesity-Associated Gut Microbiome with Increased Capacity for Energy Harvest." *Nature* 444 (7122): 1027–31. <https://doi.org/10.1038/nature05414>.
- Tyanova, Stefka, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y. Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. 2016. "The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data." *Nature Methods* 13 (9): 731–40. <https://doi.org/10.1038/nmeth.3901>.
- Wikoff, William R., Andrew T. Anfora, Jun Liu, Peter G. Schultz, Scott A. Lesley, Eric C. Peters, and Gary Siuzdak. 2009. "Metabolomics Analysis Reveals Large Effects of Gut Microflora on Mammalian Blood Metabolites." *Proceedings of the National Academy of Sciences* 106 (10): 3698–3703. <https://doi.org/10.1073/pnas.0812874106>.
- Zmora, Niv, Jotham Suez, and Eran Elinav. 2019. "You Are What You Eat: Diet, Health and the Gut Microbiota." *Nature Reviews Gastroenterology & Hepatology* 16 (1): 35–56. <https://doi.org/10.1038/s41575-018-0061-2>.
- Zybailov, Boris, Amber L. Mosley, Mihaela E. Sardu, Michael K. Coleman, Laurence Florens, and Michael P. Washburn. 2006. "Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces Cerevisiae." *Journal of Proteome Research* 5 (9): 2339–47. <https://doi.org/10.1021/pr060161n>.

APPENDICES

APPENDIX A – CHAPTER 3 SUPPLEMENTARY INFORMATION

Supplementary information for “Evaluation of sample preservation and storage methods for metaproteomics analysis of intestinal microbiomes”

For each treatment, we plotted the relative abundance of each protein, expressed as percent Normalized Spectral Abundance Factors (%NSAFs, [2]) between every replicate pair of samples. This showed that within-treatment variability was low. Figures S1 - S6 are scatterplot matrices showing the linear correlations of replicates and the corresponding Pearson correlation coefficients. We prepared the plots in R (version 4.0.2; psych_2.1.3 package).

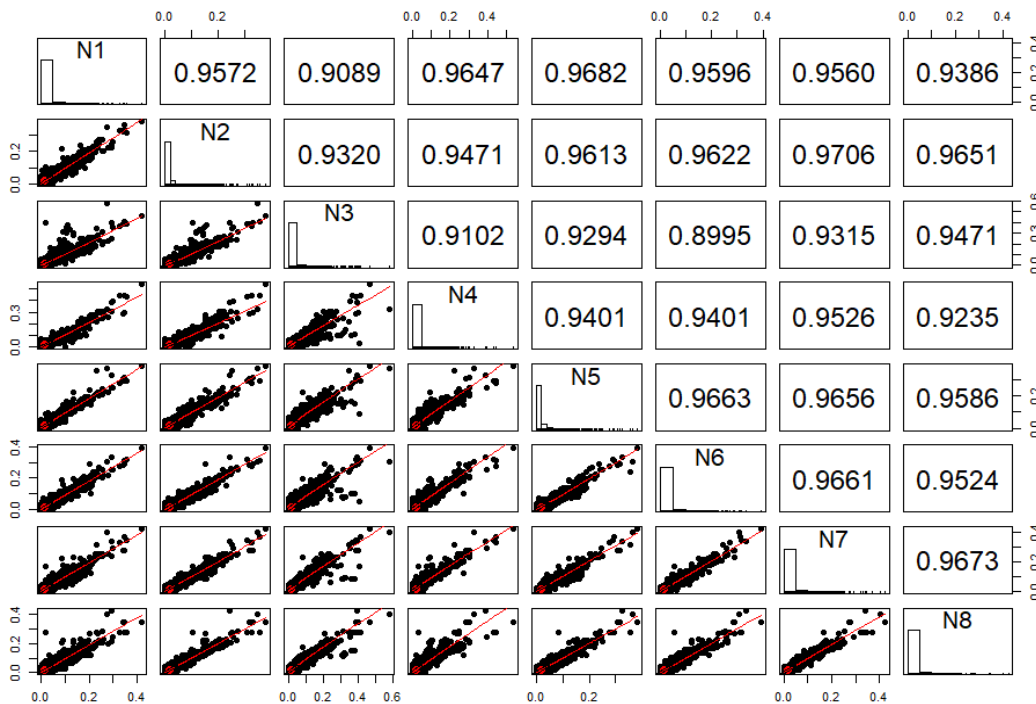


Figure S1. Linear correlation of NAP buffer (N) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

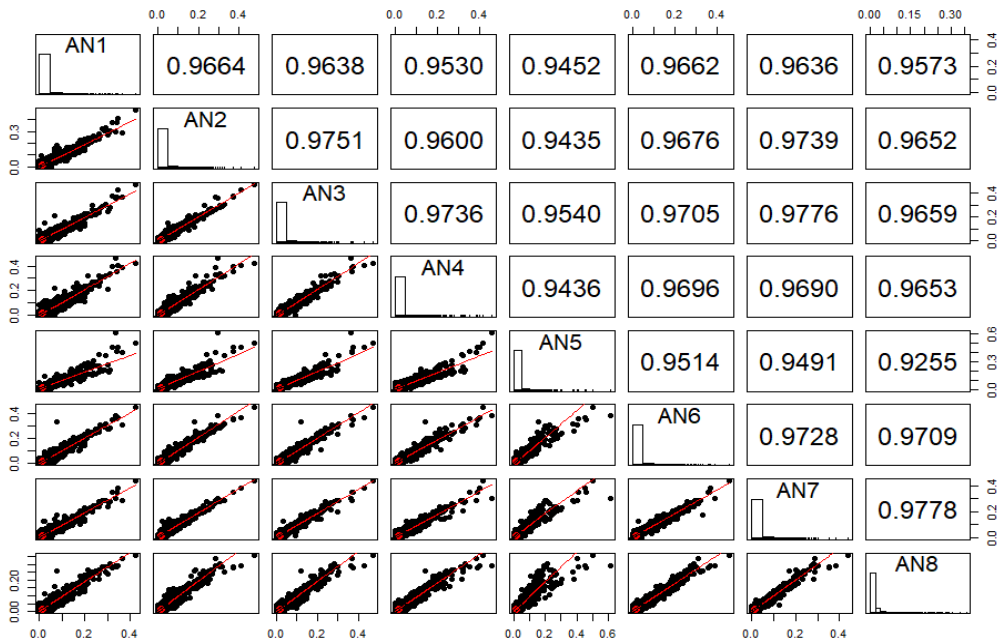


Figure S2. Linear correlation of Autoclaved NAP buffer (AN) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

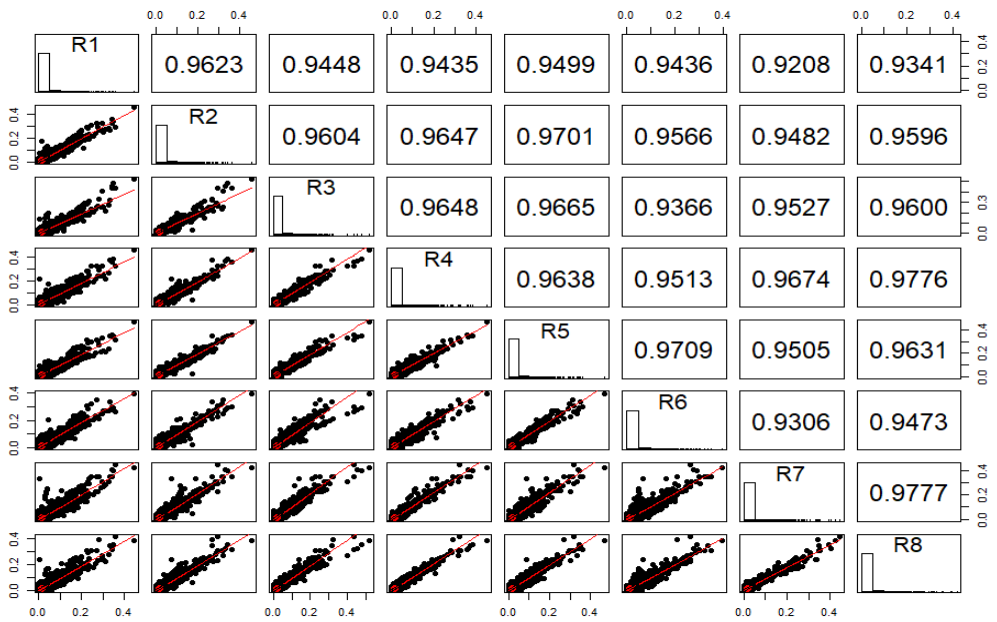


Figure S3. Linear correlation of RNAlater (R) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

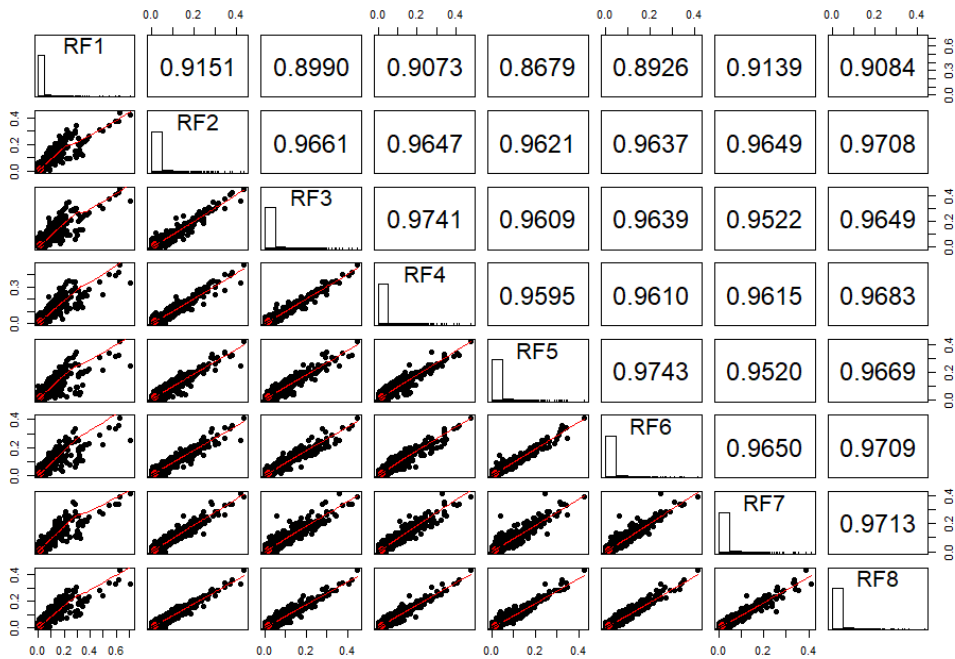


Figure S4. Linear correlation of RNAlater + flash-freezing (RF) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

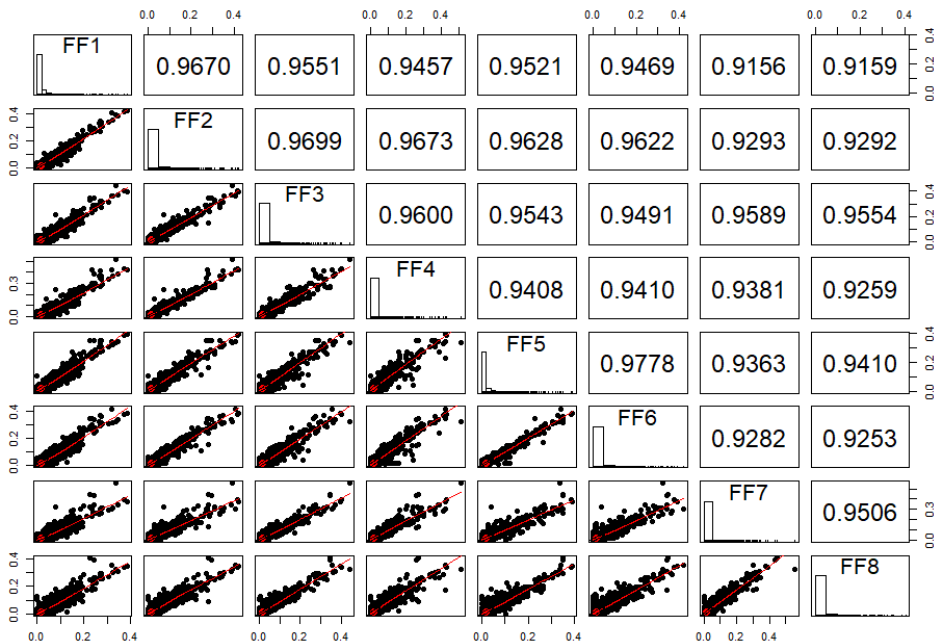


Figure S5. Linear correlation of Flash-frozen (FF) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

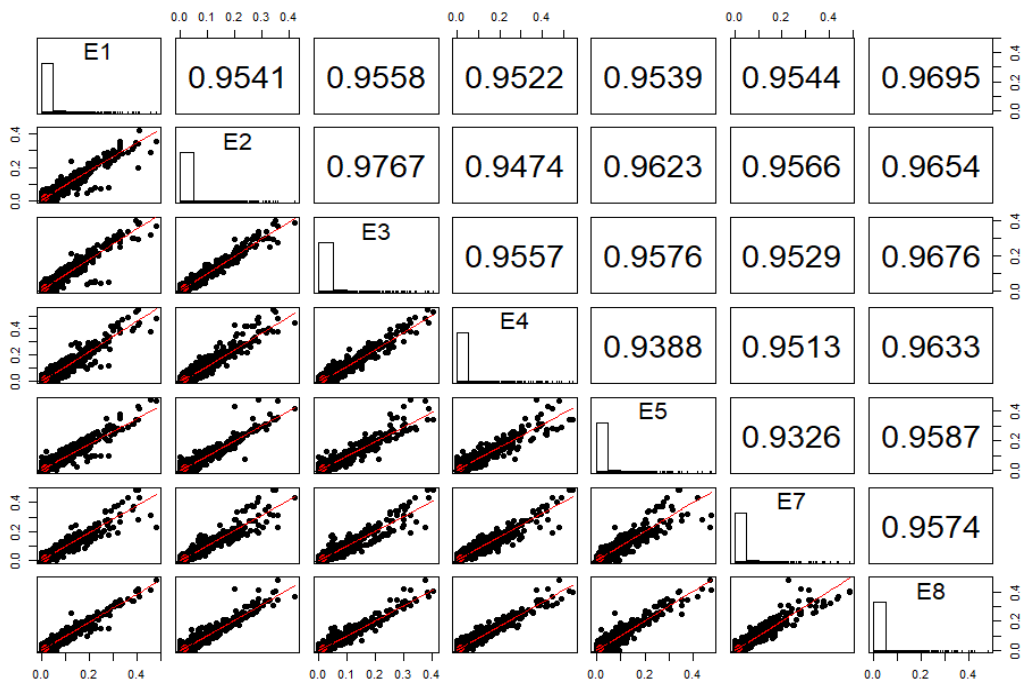


Figure S6. Linear correlation of ethanol (E) replicates. Plots are displayed in the lower panel and the corresponding Pearson correlation coefficients are displayed in the upper panel.

We analyzed the metaproteome-based taxonomic compositions of individual samples using the method described in [3]. Briefly, we filtered proteins to include only those with at least 2 protein unique peptides to increase the confidence in taxonomic identifications, and we summed Peptide Spectrum Matches (PSMs) by taxon; sums were then used to estimate the biomass contribution of each taxon in the metaproteomes. This analysis showed that host proteins contributed 5.7 +/- 1.6% (n = 47) of each metaproteome, while the microbial proteins contributed 94.2 +/- 1.6% (n = 47) and dietary proteins from wheat, the main component of the mice's diet, contributed roughly 0.1 +/- 0.2 % (n = 47). Taxonomic profiles in terms of biomass contribution were consistent across all replicates, with minor variability (as shown in Figure S7). Figures S7 A and S7 B show the representation of the detected phyla in each sample. The

phylum Firmicutes was dominant in these samples, making up 86.7 +/- 0.92 % (n = 47) of the total proteinaceous biomass. Microorganisms of unknown taxonomy made up 13.1 +/- 0.92 % (n = 47) of the total proteinaceous biomass of the samples. The number of detected proteins of known taxonomy was higher than expected given that only 67.8% of the microbial protein sequences in the database were assigned a taxonomy at the phylum level and 9.8% of the microbial sequences were annotated to the genus level [4]. We found that 11.1 +/- 0.53 % (n = 47) of the total proteinaceous biomass in our samples had a taxonomy assigned at the genus level, representing 28 microbial genera. Over half of these genera have very few Peptide Spectrum Matches (PSMs). The most abundant genera detected in the samples included *Clostridium*, *Eubacterium*, *Butyrivibrio*, *Lactobacillus*, *Turicibacter*, *Blautia*, *Roseburia*, and *Coprococcus* (Figure S2 C).

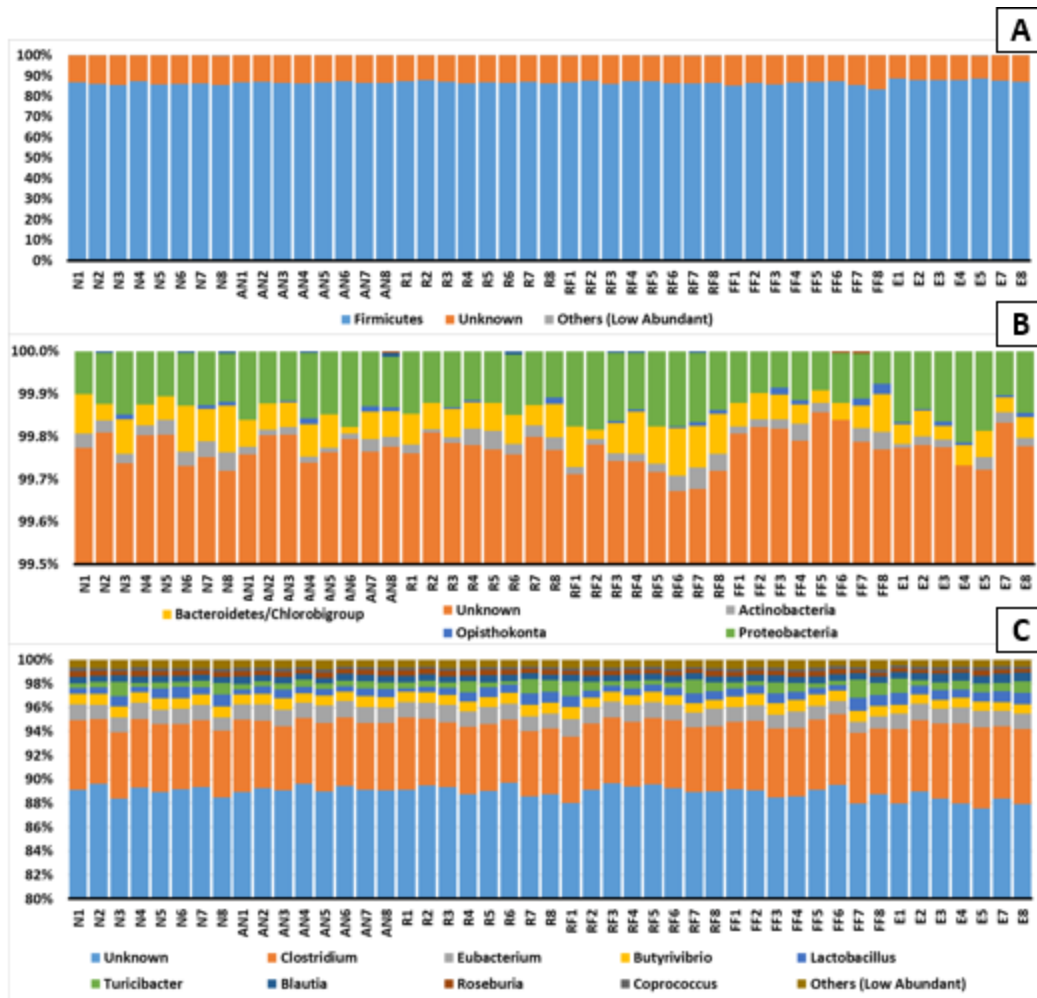


Figure S7: Taxonomic composition of individual samples as proteinaceous biomass contribution. (A) Relative abundances of the major phyla: Firmicutes, Unknown, and Others. “Others” corresponds to the sum of the phyla that were detected with very few Peptide Spectrum Matches (PSMs). The category “Others” made up less than 0.4% of the total biomass contribution of the samples. (B) Relative abundances of the lowly abundant phyla; i.e. composition of the category “others” shown in panel A. (C) Relative abundances of the most abundant genera. N = NAP buffer; AN = Autoclaved NAP buffer; R = RNAlater; RF = RNAlater + Flash-freezing; FF = Flash-freezing; E = Ethanol.

Table S1: Overview of the metaproteomic data provided with this study (dataset PXD024115 on the ProteomeXchange Consortium).

Sample name	Treatment	Storage time	Peptide concentration (ng/uL)	Peptide load (ng)	Gradient length (min)
AN1	Autoclaved NAP buffer	1 week	542.81	600	140
AN2	Autoclaved NAP buffer	1 week	551.851	600	140
AN3	Autoclaved NAP buffer	1 week	694.574	600	140
AN4	Autoclaved NAP buffer	1 week	366.24	600	140
AN5	Autoclaved NAP buffer	4 weeks	904.853	600	140
AN6	Autoclaved NAP buffer	4 weeks	473.039	600	140
AN7	Autoclaved NAP buffer	4 weeks	703.288	600	140
AN8	Autoclaved NAP buffer	4 weeks	533.189	600	140
E1	95% Ethanol	1 week	261.009	600	140
E2	95% Ethanol	1 week	439.121	600	140
E3	95% Ethanol	1 week	391.333	600	140
E4	95% Ethanol	1 week	143.315	600	140
E5	95% Ethanol	4 weeks	461.439	600	140
E6	95% Ethanol	4 weeks	191.788	600	140
E7	95% Ethanol	4 weeks	329.92	600	140
E8	95% Ethanol	4 weeks	325.898	600	140
FF1	Flash Freezing	1 week	609.099	600	140
FF2	Flash Freezing	1 week	671.558	600	140
FF3	Flash Freezing	1 week	461.606	600	140
FF4	Flash Freezing	1 week	628.12	600	140
FF5	Flash Freezing	4 weeks	283.113	600	140
FF6	Flash Freezing	4 weeks	228.295	600	140
FF7	Flash Freezing	4 weeks	253.664	600	140

Table S1 (continued).

FF8	Flash Freezing	4 weeks	237.902	600	140
N1	NAP buffer	1 week	607.684	600	140
N2	NAP buffer	1 week	587.895	600	140
N3	NAP buffer	1 week	599.197	600	140
N4	NAP buffer	1 week	841.462	600	140
N5	NAP buffer	4 weeks	608.211	600	140
N6	NAP buffer	4 weeks	543.518	600	140
N7	NAP buffer	4 weeks	634.832	600	140
N8	NAP buffer	4 weeks	490.3	600	140
R1	RNAlater	1 week	694.166	600	140
R2	RNAlater	1 week	633.792	600	140
R3	RNAlater	1 week	871.938	600	140
R4	RNAlater	1 week	555.872	600	140
R5	RNAlater	4 weeks	649.313	600	140
R6	RNAlater	4 weeks	508.285	600	140
R7	RNAlater	4 weeks	333.275	600	140
R8	RNAlater	4 weeks	563.871	600	140
RF1	RNAlater + Flash Freezing	1 week	665.051	600	140
RF2	RNAlater + Flash Freezing	1 week	413.117	600	140
RF3	RNAlater + Flash Freezing	1 week	375.914	600	140
RF4	RNAlater + Flash Freezing	1 week	521.545	600	140
RF5	RNAlater + Flash Freezing	4 weeks	479.529	600	140
RF6	RNAlater + Flash Freezing	4 weeks	525.108	600	140
RF7	RNAlater + Flash Freezing	4 weeks	390	600	140
RF8	RNAlater + Flash Freezing	4 weeks	568.535	600	140

ADDITIONAL FILES SUBMITTED TO JOURNAL (and referenced in the text)

Additional file 1: Quantification data for identified proteins.

Data sheets containing quantitative data for all identified proteins (i) expressed as Peptide Spectrum Matches (PSMs), (ii) expressed as percent Normalized Spectral Abundance Factors (%NSAFs), and (iii) expressed as Centered-Log-Ratio (CLR) transformed values. **(XLSX 7 MB)**

Additional file 2: Differentially abundant proteins.

Data sheet containing protein accessions, $-\log(p\text{-value})$, and fold difference for proteins that were significantly different between preservation methods (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, $S_0=0.1$). **(XLSX 142 KB)**

Additional file 3: List of proteins detected per treatment and their corresponding molecular weights and isoelectric points. (XLSX 3 MB)

Additional file 4: Number of predicted transmembrane domains for the proteins identified per treatment.

Data sheet displays the output of the TMHMM 2.0 Server [1] from searching sequences of the identified proteins. **(XLSX 1 MB)**

Additional file 5: Measured microbiome composition in each treatment based on metaproteomes.

Taxonomic composition at the (i) phylum and at the (ii) genus level. **(XLSX 15 KB)**

REFERENCES

1. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
2. Zybilov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J Proteome Res.* 2006;5:2339–47.
3. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun.* 2017;8:1558.
4. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, et al. A catalog of the mouse gut metagenome. *Nat Biotechnol.* 2015;33:1103–8.

APPENDIX B – CHAPTER 4 SUPPLEMENTARY INFORMATION

Supplemental table 4-1. Organisms/strains of the defined community

Species	Strain	Original Source (before sent to us from Dr. Eric Martens's laboratory)	Uniprot ID
<i>Bacteroides ovatus</i>	1896, Type strain	DSMZ	UP000005475
<i>Bacteroides uniformis</i>	8492	ATCC	UP000004110
<i>Bacteroides thetaiotaomicron</i>	2079, Type strain	DSMZ	UP000001414
<i>Bacteroides caccae</i>	19024, Type strain (ATCC43185)	DSMZ	UP000003325
<i>Barnesiella intestinihominis</i>	YIT11860 (JCM15079)	ATCC	UP000006044
<i>Roseburia intestinalis</i>	14610, Type strain L1-82	DSMZ	UP000004828
<i>Eubacterium rectale</i>	17629, A1-86	DSMZ	UP000007057
<i>Faecalibacterium prausnitzii</i>	17677, A2-165	DSMZ	UP000004619
<i>Marvinbryantia formatexigens</i>	14469, Type strain I-52	DSMZ	UP000005561
<i>Clostridium symbiosum</i>	934, Type strain, designation 2	DSMZ	UP000002970
<i>Collinsella aerofaciens</i>	3979, Type strain	DSMZ	UP000002979
<i>Escherichia coli</i>	HS	ATCC	NCBI CP000802.1
<i>Akkermansia muciniphila</i>	22959, Type strain, Muc	DSMZ	UP000001031

Supplemental table 4-2. Optical densities at 600 nm (OD600) of each organism before assembly into the inocula used to gavage the germ-free mice.

Organism	OD600 - Day 1	OD600 - Day 2	OD600 - Day 3
<i>Bacteroides ovatus</i> (BOV)	0.95	0.82	1.12
<i>Bacteroides uniformis</i> (BUN)	0.26	0.63	0.86
<i>Bacteroides thetaiomicon</i> (BTH)	0.56	0.13	0.16
<i>Bacteroides caccae</i> (BCC)	0.79	1.40	1.16
<i>Barnesiella intestinihominis</i> (BIN)	0.21	0.22	0.20
<i>Roseburia intestinalis</i> (RIN)	0.24	0.76	0.77
<i>Eubacterium rectale</i> (ERC)	1.75	1.66	1.70
<i>Faecalibacterium prausnitzii</i> (FPR)	1.05	0.77	1.06
<i>Marvinbryantia formatexigens</i> (MFR)	0.76	0.54	0.86
<i>Clostridium symbiosum</i> (CSM)	1.48	1.54	1.18
<i>Collinsella aerofaciens</i> (CAR)	1.03	> 2	1.74
<i>Escherichia coli</i> (ECL)	1.51	1.53	1.57
<i>Akkermansia muciniphila</i> (AMC)	0.86	0.94	0.89

Supplementary table 4-3. EA-IRMS results of dietary components used in this study.

Unless indicated, all ingredient samples were obtained from Envigo Teklad.

10-Jan-19	mg C	mg N	$\delta^{13}\text{C}$	$\delta^{15}\text{N}$	C:N	Avg. $\delta^{13}\text{C}$ (n=2)	Diff. reps (n = 2)
cornstarch	0.073		-10.84			-10.75	0.13
cornstarch	0.122		-10.66				
sucrose	0.061		-12.18			-12.32	0.19
sucrose	0.204		-12.46				
cellulose	0.193		-26.41			-26.55	0.20
cellulose	0.144		-26.69				
inulin	0.125		-26.92			-26.88	0.06

Supplementary table 4-3 (continued).


inulin	0.092		-26.83				
maltodextrin maltrin	0.238		-10.40			-10.40	0.01
maltodextrin maltrin	0.172		-10.39				
maltodextrin lodex	0.166		-10.58			-10.65	0.09
maltodextrin lodex	0.101		-10.71				
casein	0.091	0.02829	-26.46	6.02	3.7	-26.56	0.15
casein	0.095	0.02832	-26.67	5.19	3.9		
soy protein	0.112	0.03342	-26.36	-0.22	3.9	-26.33	0.05
soy protein	0.070	0.02152	-26.29	-0.06	3.8		
egg white solids	0.147	0.04316	-17.03	4.31	4.0	-17.19	0.22
egg white solids	0.113	0.03324	-17.35	4.05	4.0		
soybean oil	0.137		-32.06			-32.05	0.01
soybean oil	0.315		-32.05				
corn oil	0.142		-17.16			-16.92	0.35
corn oil	0.240		-16.67				
April 2019	mg C	mg N	d¹³C	C:N	d¹³C Avg	Stdev/ Diff reps	
Corn fiber Amazon B00NAD0IVU	0.091	0.0005	-11.11		-11.17	0.07	
Corn fiber Amazon B00NAD0IVU	0.180	0.0005	-11.15				
Corn fiber Amazon B00NAD0IVU	0.255	0.0005	-11.24				
Sunflower oil Amazon B0792GCNWW	0.148		-30.95		-31.24	0.42	
Sunflower oil Amazon B0792GCNWW	0.240		-31.54				

Teklad Custom Diet

TD.200545

EWS and 12% CO Diet (Corn Fiber, 93G)



Formula	g/Kg	Key Features
Egg White Solids, spray-dried	200.0	+ Purified Diet
Corn Starch	351.371	+ Egg White Solids
Maltodextrin	132.0	+ Corn Soluble Fiber
Sucrose	100.0	+ Corn Oil
Corn Oil	120.0	
Soluble Corn Fiber, customer supplied	50.0	
Mineral Mix, w/o Ca & P (98057)	13.39	
Calcium Phosphate, dibasic	12.02	
Calcium Carbonate	3.43	
Vitamin Mix, AIN-93-VX (94047)	15.0	
Choline Bitartrate	2.75	
Biotin	0.003	
Thiamin (81%)	0.01	
Vitamin K ₁ , phylloquinone	0.002	
TBHQ, antioxidant	0.024	
Footnote		Key Planning Information
Modification of AIN-93G to replace casein, soybean oil and cellulose with egg white solids, corn oil and soluble corn fiber. Total fat is increased to 12% by weight. Vitamins are increased to make the diet more suitable for irradiation.		+ Products are made fresh to order
		+ Store product at 4°C or lower
		+ Use within 6 months (applicable to most diets)
		+ Box labeled with product name, manufacturing date, and lot number
		+ Replace diet at minimum once per week <i>More frequent replacement may be advised</i>
		+ Lead time:
		- 2 weeks non-irradiated
		- 4 weeks irradiated
		Product Specific Information
		+ 1/2" Pellet or Powder (free flowing)
		+ Minimum order 3 Kg
		+ Irradiation available upon request
Selected Nutrient Information¹		Options (fees will apply)
	% by weight	% kcal from
Protein	16.1	16.2
Carbohydrate	56.3	56.7
Fat	12.0	27.2
Kcal/g	4.0	
¹ Values are calculated from ingredient analysis or manufacturer data		
Speak With A Nutritionist		Contact Us
+ (800) 483-5523		Obtain pricing · Check order status
+ askanutritionist@envigo.com		+ tekklad@envigo.com
		+ (800) 483-5523
		
		International Inquiry (outside USA or Canada)
		+ askanutritionist@envigo.com
		Place Your Order (USA & Canada)
		Please Choose One
		+ www.envigo.com/tekklad-orders
		+ tekkladorders@envigo.com
		+ (800) 483-5523
		+ (608) 277-2066 facsimile
<i>Teklad diets are designed & manufactured for research purposes only.</i>		

© 2015 Envigo

06/15/2015

Envigo Teklad Diets + Madison WI + envigo.com + tekkladinfo@envigo.com + (800) 483-5523

Supplemental figure S4-1. “Control diet” – Egg whites, Corn fiber, Corn oil diet

Teklad Custom Diet

TD.200546



EWS and 12% CO Diet (Cellulose, 93G)

Formula	g/Kg
Egg White Solids, spray-dried	200.0
Corn Starch	351.371
Maltodextrin	132.0
Sucrose	100.0
Corn Oil	120.0
Cellulose	50.0
Mineral Mix, w/o Ca & P (98057)	13.39
Calcium Phosphate, dibasic	12.02
Calcium Carbonate	3.43
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.75
Biotin	0.003
Thiamin (81%)	0.01
Vitamin K ₁ , phylloquinone	0.002
TBHQ, antioxidant	0.024

Footnote
 Modification of AIN-93G to replace casein and soybean oil with egg white solids and corn oil. Total fat is increased to 12% by weight. Vitamins are increased to make the diet more suitable for irradiation.

Selected Nutrient Information¹

	% by weight	% kcal from
Protein	16.1	16.3
Carbohydrate	55.6	56.3
Fat	12.0	27.3
Kcal/g	3.9	

¹ Values are calculated from ingredient analysis or manufacturer data

Speak With A Nutritionist
 + (800) 483-5523
 + askanutritionist@envigo.com

Teklad diets are designed & manufactured for research purposes only.

Key Features

- + Purified Diet
- + Egg White Solids
- + Corn Oil
- + Cellulose

Key Planning Information

- + Products are made fresh to order
- + Store product at 4°C or lower
- + Use within 6 months (applicable to most diets)
- + Box labeled with product name, manufacturing date, and lot number
- + Replace diet at minimum once per week
More frequent replacement may be advised
- + Lead time:
 - 2 weeks non-irradiated
 - 4 weeks irradiated


Product Specific Information

- + 1/2" Pellet or Powder (free flowing)
- + Minimum order 3 Kg
- + Irradiation available upon request

Options (fees will apply)

- + Rush order (pending availability)
- + Irradiation (see Product Specific Information)
- + Vacuum packaging (1 and 2 Kg)

Contact Us
 Obtain pricing - Check order status
 + teklad@envigo.com
 + (800) 483-5523



International Inquiry (outside USA or Canada)
 + askanutritionist@envigo.com

Place Your Order (USA & Canada)
Please Choose One
 + www.envigo.com/teklad-orders
 + tekladorders@envigo.com
 + (800) 483-5523
 + (608) 277-2066 facsimile

Supplemental figure S4-2. Cellulose diet

EWS and 12% CO Diet (Inulin, 93G)

Formula	g/Kg
Egg White Solids, spray-dried	200.0
Corn Starch	351.371
Maltodextrin	132.0
Sucrose	100.0
Corn Oil	120.0
Inulin	50.0
Mineral Mix, w/o Ca & P (98057)	13.39
Calcium Phosphate, dibasic	12.02
Calcium Carbonate	3.43
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.75
Biotin	0.003
Thiamin (81%)	0.01
Vitamin K ₁ , phylloquinone	0.002
TBHQ, antioxidant	0.024

Footnote
 Modification of AIN-93G to replace casein, soybean oil and cellulose with egg white solids, corn oil and inulin. Total fat is increased to 12% by weight. Vitamins are increased to make the diet more suitable for irradiation.

Selected Nutrient Information ¹		
	% by weight	% kcal from
Protein	16.1	16.1
Carbohydrate	57.1	57.0
Fat	12.0	26.9
Kcal/g	4.0	

¹ Values are calculated from ingredient analysis or manufacturer data

Speak With A Nutritionist
 + (800) 483-5523
 + askanutritionist@envigo.com

Teklad diets are designed & manufactured for research purposes only.

Key Features

- + Purified Diet
- + Egg White Solids
- + Corn Oil
- + Inulin

Key Planning Information

- + Products are made fresh to order
- + Store product at 4°C or lower
- + Use within 6 months (applicable to most diets)
- + Box labeled with product name, manufacturing date, and lot number
- + Replace diet at minimum once per week
More frequent replacement may be advised
- + Lead time:
 - 2 weeks non-irradiated
 - 4 weeks irradiated


Product Specific Information

- + 1/2" Pellet or Powder (free flowing)
- + Minimum order 3 Kg
- + Irradiation available upon request

Options (fees will apply)

- + Rush order (pending availability)
- + Irradiation (see Product Specific Information)
- + Vacuum packaging (1 and 2 Kg)

Contact Us
 Obtain pricing · Check order status
 + tekklad@envigo.com
 + (800) 483-5523



International Inquiry (outside USA or Canada)
 + askanutritionist@envigo.com

Place Your Order (USA & Canada)
 Please Choose One
 + www.envigo.com/tekklad-orders
 + tekkladorders@envigo.com
 + (800) 483-5523
 + (608) 277-2066 facsimile

Supplemental figure S4-3. Inulin diet.

Teklad Custom Diet

TD.200548

EWS and 12% SunO Diet (Corn Fiber, 93G)



Formula	g/Kg
Egg White Solids, spray-dried	200.0
Corn Starch	351.371
Maltodextrin	132.0
Sucrose	100.0
Sunflower Oil, blended, customer supplied	120.0
Soluble Corn Fiber, customer supplied	50.0
Mineral Mix, w/o Ca & P (98057)	13.39
Calcium Phosphate, dibasic	12.02
Calcium Carbonate	3.43
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.75
Biotin	0.003
Thiamin (81%)	0.01
Vitamin K ₁ , phyloquinone	0.002
TBHQ, antioxidant	0.024

Footnote
 Modification of AIN-93G to replace casein, soybean oil and cellulose with egg white solids, sunflower oil and soluble corn fiber. Total fat is increased to 12% by weight. Vitamins are increased to make the diet more suitable for irradiation.

Selected Nutrient Information¹

	% by weight	% kcal from
Protein	16.1	16.2
Carbohydrate	56.3	56.7
Fat	12.0	27.2
Kcal/g	4.0	

¹ Values are calculated from ingredient analysis or manufacturer data

Speak With A Nutritionist
 + (800) 483-5523
 + askanutritionist@envigo.com

Teklad diets are designed & manufactured for research purposes only.

Key Features

- + Purified Diet
- + Egg White Solids
- + Corn Soluble Fiber
- + Sunflower Oil

Key Planning Information

- + Products are made fresh to order
- + Store product at 4°C or lower
- + Use within 6 months (applicable to most diets)
- + Box labeled with product name, manufacturing date, and lot number
- + Replace diet at minimum once per week
More frequent replacement may be advised
- + Lead time:
 - 2 weeks non-irradiated
 - 4 weeks irradiated


Product Specific Information

- + 1/2" Pellet or Powder (free flowing)
- + Minimum order 3 Kg
- + Irradiation available upon request

Options (fees will apply)

- + Rush order (pending availability)
- + Irradiation (see Product Specific Information)
- + Vacuum packaging (1 and 2 Kg)

Contact Us
 Obtain pricing · Check order status
 + tekklad@envigo.com
 + (800) 483-5523



International Inquiry (outside USA or Canada)
 + askanutritionist@envigo.com

Place Your Order (USA & Canada)
 Please Choose One
 + www.envigo.com/tekklad-orders
 + tekkladorders@envigo.com
 + (800) 483-5523
 + (608) 277-2066 facsimile

© 2015 Envigo

08/05/20

Supplemental figure S4-4. Sunflower diet.

Formula	g/Kg
Isolated Soy Protein	200.0
Corn Starch	353.749
Maltodextrin	132.0
Sucrose	100.0
Corn Oil	120.0
Soluble Corn Fiber, customer supplied	50.0
Mineral Mix, w/o Ca & P (98057)	13.39
Calcium Phosphate, dibasic	5.6
Calcium Carbonate	7.475
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.75
Thiamin (81%)	0.01
Vitamin K ₁ , phylloquinone	0.002
TBHQ, antioxidant	0.024

Footnote		
Modification of AIN-93G to replace casein, soybean oil and cellulose with soy protein isolate, corn oil and soluble corn fiber. Total fat is increased to 12% by weight. Vitamins are increased to make the diet more suitable for irradiation.		

Selected Nutrient Information ¹		
	% by weight	% kcal from
Protein	17.4	17.1
Carbohydrate	56.7	55.8
Fat	12.2	27.0
Kcal/g	4.1	

¹ Values are calculated from ingredient analysis or manufacturer data

Speak With A Nutritionist
+ (800) 483-5523
+ askanutritionist@envigo.com

Teklad diets are designed & manufactured for research purposes only.

Key Features

- + Purified Diet
- + Whey Protein
- + Corn Oil
- + Soluble Corn Fiber

Key Planning Information

- + Products are made fresh to order
- + Store product at 4°C or lower
- + Use within 6 months (applicable to most diets)
- + Box labeled with product name, manufacturing date, and lot number
- + Replace diet at minimum once per week
 - More frequent replacement may be advised*
- + Lead time:
 - 2 weeks non-irradiated
 - 4 weeks irradiated

Product Specific Information

- + 1/2" Pellet or Powder (free flowing)
- + Minimum order 3 Kg
- + Irradiation available upon request

Options (fees will apply)

- + Rush order (pending availability)
- + Irradiation (see Product Specific Information)
- + Vacuum packaging (1 and 2 Kg)

Contact Us

Obtain pricing - Check order status

- + teklad@envigo.com
- + (800) 483-5523

**International Inquiry (outside USA or Canada)**

- + askanutritionist@envigo.com

Place Your Order (USA & Canada)

Please Choose One

- + www.envigo.com/teklad-orders
- + tekladorders@envigo.com
- + (800) 483-5523
- + (608) 277-2066 facsimile

Laboratory Autoclavable Rodent Diet 5010*

DESCRIPTION

Laboratory Autoclavable Rodent Diet 5010 is a complete life cycle diet formulated using managed formulation, delivering Constant Nutrition®. This is paired with the selection of highest quality ingredients to assure minimal inherent biological variation in long-term studies. It is formulated for life-cycle nutrition; however, it is not designed for maximizing production in mouse breeding colonies. Please consult us for mouse breeding diet options. It has been formulated with extra nutrients to compensate for the nutrient losses that occur during autoclaving and ensure nutritional adequacy. The product is coated with a small amount of silicon dioxide to reduce clumping during the autoclaving process.

Features and Benefits

- **Managed Formulation delivers Constant Nutrition®**
- Processed with silicon dioxide to reduce sticking and clumping
- Fortified with extra nutrients to compensate for losses during autoclaving
- High quality animal protein added to create a superior balance of amino acids for optimum performance
- Similar to Laboratory Rodent Diet 5001 in nutrient composition and animal performance – same guaranteed analysis
- Designed for rats, mice, hamsters and gerbils

Product Forms Available

- Oval pellet, 3/8" x 5/8" x 1" length
- Meal (ground pellets), special order

Catalog #
0001326

GUARANTEED ANALYSIS

Crude protein not less than	23.0%
Crude fat not less than	4.5%
Crude fiber not more than	6.0%
Ash not more than	8.0%
Moisture not more than	12.0%

AUTOCLAVING SUGGESTIONS

To autoclave the pellets, place on trays, in small bags, or in larger bags, to a depth of no more than 3 inches. When steam autoclaved, the pellets swell and exert force on adjacent pellets. Confinement by a bag or container creates additional pressure, which may result in sticking.

Assay before and after autoclaving: Conditions of sterilization must be determined for each autoclaving unit. Microbiological evaluation should be done to insure sterilization is achieved. It is best to assay the diet before and after sterilization to determine nutrient losses.

INGREDIENTS

Ground corn, dehulled soybean meal, wheat middlings, fish meal, whole wheat, wheat germ, brewers dried yeast, ground oats, dehydrated alfalfa meal, porcine animal fat preserved with BHA and citric acid, ground soybean hulls, calcium carbonate, dried beet pulp, salt, soybean oil, DL-methionine, pyridoxine hydrochloride, choline chloride, menadione dimethylpyrimidinol bisulfite (source of vitamin K), thiamine mononitrate, cholecalciferol, dicalcium phosphate, silicon dioxide, vitamin A acetate, folic acid, biotin, dl-alpha tocopheryl acetate, (form of vitamin E) calcium pantothenate, riboflavin, nicotinic acid, vitamin B₁₂ supplement, manganese oxide, zinc oxide, ferrous carbonate, copper sulfate, zinc sulfate, calcium iodate, cobalt carbonate.

FEEDING DIRECTIONS

Feed ad libitum to rodents. Plenty of fresh, clean water should be available to the animals at all times.

Rats- All rats will eat varying amounts of feed depending on their genetic origin. Larger strains will eat up to 30 grams per day. Smaller strains will eat up to 15 grams per day. Feeders in rat cages should be designed to hold two to three days supply of feed at one time.

Mice- Adult mice will eat up to 5 grams of pelleted ration daily. Some of the larger strains may eat as much as 8 grams per day per animal. Feed should be available on a free choice basis in wire feeders above the floor of the cage.

Hamsters- Adults will eat up to 14 grams per day.

NOTE: Do not feed this or any other autoclavable diet prior to autoclaving.

For information regarding shelf life please visit www.labdiet.com.

05/21/15

CHEMICAL COMPOSITION¹

Nutrients²

Protein, %	24.6
Arginine, %	1.55
Cystine, %	0.42
Glycine, %	1.21
Histidine, %	0.63
Isoleucine, %	1.02
Leucine, %	1.86
Lysine, %	1.45
Methionine, %	0.56
Phenylalanine, %	1.08
Tyrosine, %	0.73
Threonine, %	0.93

Tryptophan, %	0.28
Valine, %	1.14
Serine, %	1.17
Aspartic Acid, %	2.64
Glutamic Acid, %	4.80
Alanine, %	1.48
Proline, %	1.52
Taurine, %	0.05

Fat (ether extract), %	5.0
Fat (acid hydrolysis), %	6.4
Cholesterol, ppm	283
Linoleic Acid, %	1.42
Linolenic Acid, %	0.14
Arachidonic Acid, %	0.02
Omega-3 Fatty Acids, %	0.42
Total Saturated Fatty Acids, %	1.17
Total Monounsaturated Fatty Acids, %	1.29

Fiber (Crude), %	4.2
Neutral Detergent Fiber ³ , %	14.6
Acid Detergent Fiber ⁴ , %	5.3

Nitrogen-Free Extract

(by difference), %	50.0
Starch, %	29.4
Glucose, %	0.23
Fructose, %	0.27
Sucrose, %	1.16
Lactose, %	0.00

Total Digestible Nutrients, % 74.8

Gross Energy, kcal/gm 4.17

Physiological Fuel Value⁵, kcal/gm 3.43

Metabolizable Energy, kcal/gm 3.02

Minerals

Ash, %	6.1
Calcium, %	1.00
Phosphorus, %	0.79
Phosphorus (non-phytate), %	0.46
Potassium, %	0.99
Magnesium, %	0.22

Sulfur, %	0.31
Sodium, %	0.29
Chloride, %	0.48
Fluorine, ppm	15
Iron, ppm	250
Zinc, ppm	130
Manganese, ppm	120
Copper, ppm	19
Cobalt, ppm	0.60
Iodine, ppm	1.6
Chromium (added), ppm	0.01
Selenium, ppm	0.47

Vitamins

Carotene, ppm	1.3
Vitamin K, ppm	3.4
Thiamin Hydrochloride, ppm	81
Riboflavin, ppm	16
Niacin, ppm	120
Pantothenic Acid, ppm	26
Choline Chloride, ppm	2200
Folic Acid, ppm	6.1
Pyridoxine, ppm	17
Biotin, ppm	0.30
B ₁₂ , mcg/kg	50
Vitamin A, IU/gm	24
Vitamin D ₃ (added), IU/gm	4.6
Vitamin E, IU/kg	61
Ascorbic Acid, mg/gm	—

Calories provided by:

Protein, %	28.668
Fat (ether extract), %	13.111
Carbohydrates, %	58.221

*Product Code

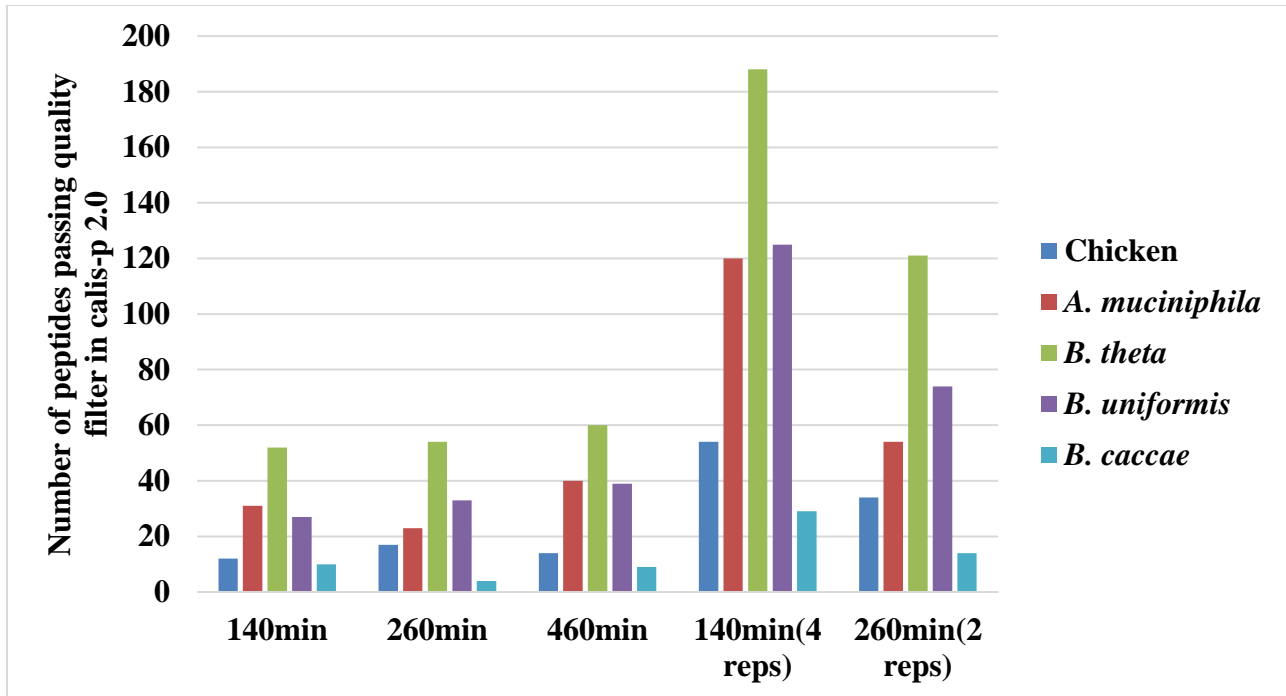
1. Formulation based on calculated values from the latest ingredient analysis information. Since nutrient composition of natural ingredients varies and some nutrient loss will occur due to manufacturing processes, analysis will differ accordingly.
2. Nutrients expressed as percent of ration except where otherwise indicated. Moisture content is assumed to be 10.0% for the purpose of calculations.
3. NDF = approximately cellulose, hemicellulose and lignin.
4. ADF = approximately cellulose and lignin.
5. Physiological Fuel Value (kcal/gm) = Sum of decimal fractions of protein, fat and carbohydrate (use Nitrogen Free Extract) x 4,9,4 kcal/gm respectively.

LabDiet
www.labdiet.com

Supplemental figure 4-8. Standard chow

Supplementary table S4-4. Results from verification of diet compositions by LC-MS/MS
Each row represents one diet replicate analyzed using the methods described in the main text (4.3.4 - 8).

Diet	Protein source	Fat source	Fiber source	starch	Most abundant protein (#PSMs)	Main organism (#proteins, 5%FDR)	Other relevant organisms (#proteins, 5% FDR)
Control diet	Egg whites	Corn oil	Corn fiber	Corn starch	Ovotransferrin (1474)	Gallus gallus (35)	Zea mays (5)
Fiber diet 1	Egg whites	Corn oil	Cellulose	Corn starch	Ovotransferrin (1578)	Gallus gallus (47)	Zea mays (13)
Fiber diet 2	Egg whites	Corn oil	Inulin	Corn starch	Ovotransferrin (1572)	Gallus gallus (47)	Zea mays (12)
Fat diet 1	Egg whites	Soybean oil	Corn fiber	Corn starch	Ovotransferrin (2296)	Gallus gallus (26)	Zea mays (2), glycine max (10)
Fat diet 2	Egg whites	Sunflower oil	Corn fiber	Corn starch	Ovotransferrin (1398)	Gallus gallus (37)	Zea mays(17), Helianthus annuus(2)
Protein diet 1	Casein	Corn oil	Corn fiber	Corn starch	Alpha-S1-Casein (1560)	Bos taurus (183)	Zea mays (41)
Protein diet 2	Soy protein	Corn oil	Corn fiber	Corn starch	Glycinin G4 (1260)	Glycine max (69)	Zea mays (32)



Supplemental figure 4-9. Running samples in 4 replicates yields enough peptides for SIF estimation.

Supplementary table S4-5. Proteinaceous biomass contribution, group 1

Diet	Mouse	MFR (%)	BOV (%)	BUN (%)	RIN (%)	BTH (%)	ERC (%)	AMC (%)	ECL (%)	BCC (%)	CSM (%)	BIN (%)	CAR (%)	FPR (%)
Casein	1	18.59	2.37	27.41	1.52	17.61	2.97	20.59	1.53	5.02	1.34	0.46	0.59	0.00
Casein	2	20.98	4.99	31.07	0.55	20.52	1.21	13.61	1.86	3.53	1.17	0.07	0.44	0.00
Casein	3	12.94	2.15	29.12	0.46	17.38	2.52	25.05	2.31	5.68	1.64	0.03	0.71	0.00
Casein	4	13.36	4.66	27.24	1.05	19.60	3.18	20.33	1.19	6.44	1.88	0.35	0.72	0.00
Casein	5	22.36	4.29	28.96	0.43	18.21	1.68	16.46	1.50	4.48	1.13	0.06	0.44	0.00
EggWhites	2	9.07	6.49	28.21	0.05	18.49	2.15	19.27	2.49	7.75	0.98	3.64	1.41	0.00
EggWhites	1	10.57	5.61	24.67	0.59	16.96	2.40	24.63	2.55	7.92	0.86	1.93	1.31	0.00
EggWhites	3	8.71	5.83	32.27	0.16	17.00	1.51	18.05	3.53	7.03	1.65	2.71	1.55	0.00
EggWhites	4	10.51	6.78	29.72	0.02	17.27	1.36	18.80	2.19	7.57	1.25	3.10	1.43	0.01
EggWhites	5	11.12	6.18	28.83	0.71	19.00	2.92	17.16	1.98	7.56	0.60	2.55	1.39	0.00
SC_end	1	27.88	19.85	12.43	4.50	11.07	9.51	7.37	2.25	1.90	1.85	0.93	0.43	0.00
SC_end	2	28.31	19.84	13.48	3.15	10.76	9.75	7.55	1.77	2.05	2.04	0.84	0.47	0.01
SC_end	3	24.91	16.18	14.04	7.39	12.08	8.49	7.17	2.52	3.91	1.96	0.84	0.46	0.04
SC_end	4	19.25	17.13	11.07	6.40	14.89	7.87	3.93	1.45	11.56	2.56	2.53	0.59	0.78
SC_end	5	29.84	15.12	14.17	8.15	8.25	9.08	8.24	2.07	2.46	1.24	0.89	0.47	0.00
SC_Start	1	28.63	18.82	12.65	9.14	8.71	8.13	7.75	2.09	1.52	1.48	0.61	0.48	0.01
SC_Start	2	31.56	19.85	13.02	6.80	6.53	7.47	7.87	1.77	2.44	1.35	0.75	0.59	0.01
SC_Start	3	28.70	19.65	12.31	6.65	10.00	4.48	8.79	2.99	3.18	1.60	0.83	0.81	0.00
SC_Start	4	28.95	22.60	12.54	7.69	11.02	1.73	7.66	2.64	2.63	1.36	0.62	0.53	0.03
SC_Start	5	38.12	18.73	10.98	1.86	7.49	5.68	7.63	3.35	3.05	1.84	0.70	0.57	0.00
SoyProtein	2	8.13	8.30	30.78	0.79	21.61	2.49	12.42	9.57	2.86	1.46	0.69	0.90	0.00
SoyProtein	3	7.45	7.48	24.52	0.80	14.85	2.48	21.00	14.81	3.57	1.09	0.98	0.96	0.00
SoyProtein	4	7.80	10.17	28.19	1.16	12.75	2.56	14.62	16.21	3.98	1.47	0.02	1.09	0.00
SoyProtein	5	7.95	8.69	28.14	0.58	13.20	2.63	26.60	5.90	2.99	1.04	0.82	1.44	0.00
SoyProtein	1	10.15	8.32	29.93	1.40	20.24	3.19	13.84	6.37	3.64	1.60	0.01	1.25	0.05

Supplementary table S4-6. Proteinaceous biomass contribution, group 2

Diet	Mouse	AM C (%)	BC C (%)	BIN (%)	BO V (%)	BT H (%)	BU N (%)	CA R (%)	CS M (%)	ECL (%)	ER C (%)	FP R (%)	MF R (%)	RIN (%)			
Corn fiber & Corn oil	12	13.9	5	7.49	3.51	1.48	23.5	29.1	0.92	0.72	10.8	8	1.91	0.00	6.35	0.04	
Corn fiber & Corn oil	11	17.5	0	6.37	1.61	3.85	16.9	30.1	1.33	0.39	15.0	4	1.10	0.00	5.63	0.01	
Corn fiber & Corn oil	10	17.0	9	8.43	1.89	4.47	21.0	29.5	0.86	0.56	9.01	1.40	0.00	0.00	5.73	0.02	
Corn fiber & Corn oil	9	16.1	2	10.1	1.67	5.39	19.0	31.2	0.82	0.31	3.60	2.29	0.00	0.00	9.22	0.05	
Corn fiber & Corn oil	8	17.0	9	8.43	1.89	4.47	21.0	29.5	0.86	0.56	9.01	1.40	0.00	0.00	5.73	0.02	
Corn fiber & Corn oil	7	23.6	4	8.37	2.45	4.81	18.0	26.2	0.84	0.24	5.70	1.59	0.02	0.00	8.08	0.01	
Cellulose	12	22.9	5	15.2	8	8.43	17.7	17.1	0.79	0.93	2.58	0.00	0.00	0.00	4.92	0.00	
Cellulose	11	21.6	1	13.7	6	9.51	16.8	20.4	1.47	1.41	2.78	0.22	0.00	0.00	4.67	0.00	
Cellulose	10	21.4	8	14.2	8	9.02	18.9	18.3	1.62	1.36	2.72	0.20	0.00	0.00	3.08	0.00	
Cellulose	9	22.2	7	13.0	0	9.24	18.2	17.0	1.08	0.61	3.23	0.02	0.00	0.00	5.43	0.00	
Cellulose	8	22.9	6	11.3	3	12.4	13.6	22.5	0.24	1.13	2.93	0.00	0.00	0.00	3.91	0.00	
Cellulose	7	20.3	4	14.6	4	9.07	16.8	19.7	1.14	0.79	2.27	0.00	0.00	0.00	7.09	0.00	
Inulin	12	15.1	7	25.2	5	5.01	16.8	16.8	1.84	0.30	9.01	0.27	0.00	0.00	11.1	5	0.00
Inulin	11	19.1	3	18.4	9	5.02	21.0	21.0	0.00	0.35	12.4	3	0.00	0.00	11.5	1	0.00
Inulin	10	13.1	6	21.4	2	5.86	11.3	24.4	1.40	0.36	7.14	0.23	0.00	0.00	8.67	0.00	
Inulin	9	14.5	0	20.9	2	5.89	10.4	31.6	3.49	0.25	3.54	0.00	0.00	0.00	4.59	0.00	
Inulin	8	12.8	8	18.2	3	4.00	25.0	25.0	2.45	0.83	3.54	0.00	0.00	0.00	19.9	0	0.00
Inulin	7	19.7	8	19.4	6	3.97	11.0	21.9	2.63	0.45	4.12	0.24	0.00	0.00	10.5	5	0.00
SC_end	12	4.36	2.99	0.88	1	8.32	19.0	11.5	0.37	1.28	2.06	7.65	0.00	0.00	37.0	5	4.45
SC_end	11	4.68	1.81	1.00	0	5.59	17.2	13.2	0.49	1.17	4.11	7.56	0.00	0.00	40.3	4	2.83
SC_end	10	6.24	1.85	0.11	9	7.20	19.5	11.4	0.55	1.25	2.54	7.99	0.00	0.00	36.0	1	5.18
SC_end	9	4.79	1.98	0.81	5	6.38	16.6	13.6	0.33	1.57	1.77	5.23	0.00	0.00	39.7	0	7.19
SC_end	8	6.14	2.45	0.52	4	6.05	16.8	11.6	0.48	1.16	2.75	8.30	0.00	0.00	35.3	9	8.25
SC_end	7	4.77	2.38	0.58	4	9.16	17.7	12.6	0.39	1.24	1.95	5.73	0.01	0.00	37.2	5	6.13
SC_start	12	4.59	1.93	0.70	6	8.60	16.5	11.5	0.58	1.54	1.62	3.86	0.00	0.00	40.0	3	10.5
SC_start	11	5.72	2.14	0.83	9	9.51	24.0	14.4	0.67	1.47	2.19	1.90	0.01	0.00	33.9	9	3.00
SC_start	10	5.82	2.04	0.47	1	7.30	20.6	11.2	0.56	1.44	2.36	2.37	0.00	0.00	37.7	9	7.95
SC_start	9	6.76	1.71	0.80	7	8.46	19.9	12.2	0.64	1.57	1.82	3.60	0.01	0.00	34.2	4	8.21
SC_start	8	6.78	2.57	0.52	0	6.66	19.3	13.6	0.45	1.51	1.50	5.32	0.02	0.00	31.1	3	10.6
SC_start	7	6.78	2.57	0.52	0	6.66	19.3	13.6	0.45	1.51	1.50	5.32	0.02	0.00	31.1	3	10.6

Supplementary table S4-6 (continued).

SC_start	7	7.08	2.92	0.73	21.3	9.31	7	0.46	1.57	1.19	3.43	0.00	31.8	3	9.91
	1	11.7				20.6	30.7								
Soybean oil	2	8	8.75	2.43	6.37	0	2	0.68	0.34	9.51	1.21	0.00	7.53	0.07	
	1	14.8				16.5	28.8			14.0					
Soybean oil	1	8	7.57	2.72	6.20	4	9	1.01	0.61	1	1.49	0.00	5.87	0.22	
	1	14.6				16.4	31.8			10.1					
Soybean oil	0	7	8.76	2.86	6.50	0	8	0.92	0.21	1	1.25	0.00	6.42	0.02	
		13.7				17.5	33.8								
Soybean oil	9	6	8.56	1.79	6.39	1	5	0.87	0.44	5.38	1.64	0.00	9.71	0.09	
		12.4				17.4	33.6								
Soybean oil	8	9	8.01	1.97	6.23	7	3	0.67	0.71	6.06	1.33	0.04	11.3	9	0.01
		18.1				15.6	31.2								
Soybean oil	7	9	9.54	1.17	5.19	1	7	0.80	0.37	5.51	1.72	0.20	10.4	0	0.04
	1	11.3				17.6	31.8			10.6					
Sunflower oil	2	5	8.87	4.42	6.08	1	4	0.56	0.38	4	1.40	0.00	6.74	0.10	
	1	14.3				15.4	32.2			11.9					
Sunflower oil	1	7	7.69	2.71	6.87	9	4	0.93	0.40	1	1.03	0.00	6.35	0.01	
	1	13.8				15.7	32.2			10.5					
Sunflower oil	0	7	7.45	3.42	7.20	1	4	0.76	0.66	6	1.67	0.00	6.42	0.04	
		11.1				19.1	31.9								
Sunflower oil	9	3	8.94	2.66	5.88	1	5	0.79	0.69	4.20	2.02	0.04	12.1	1	0.48
		11.2				18.9	31.2								
Sunflower oil	8	6	9.60	3.31	4.98	8	1	0.68	0.76	3.72	1.55	0.03	13.8	6	0.09
		12.4				18.1	31.0								
Sunflower oil	7	6	9.49	2.98	5.99	6	5	0.68	0.48	5.53	1.54	0.00	11.3	3	0.30

Supplementary table S4-7. Offset

GROUP 1						
Organism	Peptide File	#Peptides	#PSMs	Median Delta ¹³C (‰) (Protein-SIF)	Delta ¹³C (‰) (IRMS)	Offset (‰)
CHICKEN	CaseinEggwhites_Start_PSMs.txt	525	14721	-33.7	-17.19	-16.51
CHICKEN	CaseinEggwhites_End_PSMs.txt	351	8576	-33.2	-17.19	-16.01
COW	CaseinEggwhites_End_PSMs.txt	311	10112	-45.3	-27.81	-17.49
COW	CaseinEggwhites_Start_PSMs.txt	263	7160	-45.8	-27.81	-17.99
HOMO	MKH_Rep1_140min_600ng_Run4_PSMs.txt	201	7037	-36.3	-20.3	-16
HOMO	MKH_Rep2_140min_600ng_Run4_PSMs.txt	142	5792	-36	-20.3	-15.7
MEAN Offset						-16.62

Supplementary table S4-7 (continued).

GROUP 2						
Organism	Peptide File	#Peptides	#PSMs	Median Delta ¹³C (‰) (Protein- SIF)	Delta ¹³C (‰) (IRMS)	Offset (‰)
HOMO	MKH_140min_600ng_Rep1_Run5_PSMs.txt	157	7085	-37.2	-20.3	-16.9
HOMO	MKH_140min_600ng_Rep2_Run5_PSMs.txt	136	7337	-37.8	-20.3	-17.5
HOMO	MKH_140min_600ng_Rep3_Run5_PSMs.txt	141	7244	-37.6	-20.3	-17.3
HOMO	MKH_140min_600ng_Rep4_Run5_PSMs.txt	137	4584	-37	-20.3	-16.7
COW	CaseinEggwhites_140min_600ng_Rep4_Run5_PSMs.txt	85	3025	-42.7	-27.81	-14.89
COW	CaseinEggwhites_140min_600ng_Rep3_Run5_PSMs.txt	77	2900	-45.7	-27.81	-17.89
COW	CaseinEggwhites_140min_600ng_Rep1_Run5_PSMs.txt	90	2888	-46.6	-27.81	-18.79
COW	CaseinEggwhites_140min_600ng_Rep2_Run5_PSMs.txt	86	2701	-47.3	-27.81	-19.49
CHICKEN	CaseinEggwhites_140min_600ng_Rep3_Run5_PSMs.txt	80	2693	-34.7	-17.19	-17.51
CHICKEN	CaseinEggwhites_140min_600ng_Rep2_Run5_PSMs.txt	92	2545	-35.7	-17.19	-18.51
CHICKEN	CaseinEggwhites_140min_600ng_Rep4_Run5_PSMs.txt	75	2513	-32.2	-17.19	-15.01
CHICKEN	CaseinEggwhites_140min_600ng_Rep1_Run5_PSMs.txt	99	2379	-35.4	-17.19	-18.21
MEAN Offset						-17.39

Table S4-8. Stable isotope fingerprints (SIFs) of the organisms in Group 1
 All values that passed the 30 peptide threshold.

#Organism	Peptide File (Mouse#_Diet)	#Peptides	#PSMs	Median Delta ¹³ C (‰)
AMC	1_BAS_140min_600ng_Run4_PSMs.txt	343	964	-14.3
AMC	1_CAS_140min_600ng_Run4_PSMs.txt	308	844	-17.7
AMC	4_CAS_140min_600ng_Run4_PSMs.txt	202	697	-18.3
AMC	5_BAS_140min_600ng_Run4_PSMs.txt	186	480	-15.9
AMC	2_BAS_140min_600ng_Run4_PSMs.txt	175	564	-11.3
AMC	4_BAS_140min_600ng_Run4_PSMs.txt	164	551	-14.3
AMC	3_CAS_140min_600ng_Run4_PSMs.txt	152	666	-17
AMC	3_SOY_140min_600ng_Run4_PSMs.txt	109	411	-13.1
AMC	1_SC_end_140min_600ng_Run4_PSMs.txt	107	175	-26.1
AMC	5_CAS_140min_600ng_Run4_PSMs.txt	94	350	-16.8
AMC	4_SC_day0_140min_600ng_Run4_PSMs.txt	84	172	-19.8
AMC	3_BAS_140min_600ng_Run4_PSMs.txt	84	387	-11.8
AMC	3_SC_day0_140min_600ng_Run4_PSMs.txt	79	171	-19.3
AMC	1_SC_day0_140min_600ng_Run4_PSMs.txt	79	155	-20.9
AMC	1_SOY_140min_600ng_Run4-(1)_PSMs.txt	74	429	-21.5
AMC	5_SOY_140min_600ng_Run4_PSMs.txt	74	396	-11.1
AMC	5_SC_end_140min_600ng_Run4_PSMs.txt	69	143	-24.2
AMC	3_SC_end_140min_600ng_Run4_PSMs.txt	63	123	-20.5
AMC	2_SC_end_140min_600ng_Run4_PSMs.txt	57	132	-26.4
AMC	4_SC_end_140min_600ng_Run4_PSMs.txt	55	108	-19.7
AMC	4_SOY_140min_600ng_Run4_PSMs.txt	53	231	-14.6
AMC	2_SC_day0_140min_600ng_Run4_PSMs.txt	45	87	-13
AMC	2_CAS_140min_600ng_Run4_PSMs.txt	44	206	-17.2
AMC	2_SOY_140min_600ng_Run4_PSMs.txt	44	249	-18.4
AMC	5_SC_day0_140min_600ng_Run4_PSMs.txt	41	105	-20.4
BCC	1_BAS_140min_600ng_Run4_PSMs.txt	73	167	-12.5
BCC	5_BAS_140min_600ng_Run4_PSMs.txt	56	132	-16.1
BCC	2_BAS_140min_600ng_Run4_PSMs.txt	42	114	-17.9
BCC	4_BAS_140min_600ng_Run4_PSMs.txt	35	97	-17.1
BCC	1_CAS_140min_600ng_Run4_PSMs.txt	34	55	-22.9
BOV	1_SC_end_140min_600ng_Run4_PSMs.txt	155	342	-25.1
BOV	4_SC_day0_140min_600ng_Run4_PSMs.txt	125	367	-18.7
BOV	3_SC_day0_140min_600ng_Run4_PSMs.txt	113	295	-25.5
BOV	2_SC_day0_140min_600ng_Run4_PSMs.txt	92	278	-16.7
BOV	4_SC_end_140min_600ng_Run4_PSMs.txt	85	243	-28.3
BOV	1_SC_day0_140min_600ng_Run4_PSMs.txt	78	172	-22.2
BOV	2_SC_end_140min_600ng_Run4_PSMs.txt	71	157	-20.8
BOV	5_SC_day0_140min_600ng_Run4_PSMs.txt	67	201	-25.2
BOV	3_SC_end_140min_600ng_Run4_PSMs.txt	64	195	-20.5
BOV	5_SC_end_140min_600ng_Run4_PSMs.txt	51	112	-24.2
BTH	1_CAS_140min_600ng_Run4_PSMs.txt	225	691	-20.4

Table S4-8 (continued).

BTH	1_BAS_140min_600ng_Run4_PSMs.txt	206	599	-10.8
BTH	5_BAS_140min_600ng_Run4_PSMs.txt	197	584	-16.3
BTH	4_CAS_140min_600ng_Run4_PSMs.txt	164	573	-15.4
BTH	2_BAS_140min_600ng_Run4_PSMs.txt	157	475	-16
BTH	4_BAS_140min_600ng_Run4_PSMs.txt	117	483	-17.6
BTH	3_CAS_140min_600ng_Run4_PSMs.txt	109	607	-14.4
BTH	5_CAS_140min_600ng_Run4_PSMs.txt	107	599	-14.4
BTH	1_SOY_140min_600ng_Run4-(1)_PSMs.txt	98	806	-15.4
BTH	3_SOY_140min_600ng_Run4_PSMs.txt	68	331	-18.9
BTH	3_BAS_140min_600ng_Run4_PSMs.txt	62	334	-8.2
BTH	1_SC_end_140min_600ng_Run4_PSMs.txt	62	97	-22.4
BTH	2_SOY_140min_600ng_Run4_PSMs.txt	57	411	-12.5
BTH	3_SC_end_140min_600ng_Run4_PSMs.txt	48	97	-13.2
BTH	4_SC_day0_140min_600ng_Run4_PSMs.txt	44	75	-13
BTH	3_SC_day0_140min_600ng_Run4_PSMs.txt	40	83	-12.2
BTH	4_SC_end_140min_600ng_Run4_PSMs.txt	39	101	-14.9
BTH	5_SOY_140min_600ng_Run4_PSMs.txt	36	171	-18.8
BTH	2_SC_end_140min_600ng_Run4_PSMs.txt	36	59	-32.5
BUN	1_CAS_140min_600ng_Run4_PSMs.txt	394	1373	-18.2
BUN	5_BAS_140min_600ng_Run4_PSMs.txt	286	1067	-8.8
BUN	1_BAS_140min_600ng_Run4_PSMs.txt	280	911	-16.3
BUN	2_BAS_140min_600ng_Run4_PSMs.txt	228	1167	-14.1
BUN	4_BAS_140min_600ng_Run4_PSMs.txt	197	835	-16.4
BUN	4_CAS_140min_600ng_Run4_PSMs.txt	178	675	-21.9
BUN	5_CAS_140min_600ng_Run4_PSMs.txt	142	694	-17.4
BUN	3_CAS_140min_600ng_Run4_PSMs.txt	114	652	-19.3
BUN	3_SOY_140min_600ng_Run4_PSMs.txt	102	524	-9.8
BUN	3_BAS_140min_600ng_Run4_PSMs.txt	95	836	-12.8
BUN	4_SOY_140min_600ng_Run4_PSMs.txt	82	469	-13
BUN	1_SC_end_140min_600ng_Run4_PSMs.txt	80	164	-30.6
BUN	4_SC_day0_140min_600ng_Run4_PSMs.txt	70	158	-15.5
BUN	2_SC_end_140min_600ng_Run4_PSMs.txt	63	146	-15.3
BUN	1_SC_day0_140min_600ng_Run4_PSMs.txt	60	95	-18.9
BUN	3_SC_end_140min_600ng_Run4_PSMs.txt	60	110	-13.4
BUN	2_SOY_140min_600ng_Run4_PSMs.txt	59	480	-10.2
BUN	2_CAS_140min_600ng_Run4_PSMs.txt	58	434	-15
BUN	2_SC_day0_140min_600ng_Run4_PSMs.txt	53	149	-17.3
BUN	5_SC_end_140min_600ng_Run4_PSMs.txt	50	99	-23.9
BUN	4_SC_end_140min_600ng_Run4_PSMs.txt	48	101	-15.7
BUN	5_SOY_140min_600ng_Run4_PSMs.txt	46	194	-13.4
BUN	3_SC_day0_140min_600ng_Run4_PSMs.txt	39	91	-24.9
CHICKEN	2_BAS_140min_600ng_Run4_PSMs.txt	49	436	-13.3

Table S4-8 (continued).

CHICKEN	5_BAS_140min_600ng_Run4_PSMs.txt	45	304	-13.6
CHICKEN	1_BAS_140min_600ng_Run4_PSMs.txt	41	277	-16.8
ECL	3_SOY_140min_600ng_Run4_PSMs.txt	128	664	-15.1
ECL	4_SOY_140min_600ng_Run4_PSMs.txt	92	668	-17.8
ECL	3_SC_end_140min_600ng_Run4_PSMs.txt	45	83	-16.6
ECL	1_SOY_140min_600ng_Run4-(1)_PSMs.txt	42	357	-8.9
ECL	3_SC_day0_140min_600ng_Run4_PSMs.txt	39	99	-14.5
ECL	5_BAS_140min_600ng_Run4_PSMs.txt	38	113	-3.4
ECL	1_BAS_140min_600ng_Run4_PSMs.txt	38	163	-6.9
ECL	2_SOY_140min_600ng_Run4_PSMs.txt	32	281	-13.1
ECL	5_SC_day0_140min_600ng_Run4_PSMs.txt	31	117	-5.5
ECL	2_BAS_140min_600ng_Run4_PSMs.txt	31	154	-13.5
ERC	1_SC_end_140min_600ng_Run4_PSMs.txt	104	191	-27.9
ERC	4_SC_end_140min_600ng_Run4_PSMs.txt	68	160	-19.9
ERC	2_SC_end_140min_600ng_Run4_PSMs.txt	66	139	-23.4
ERC	3_SC_end_140min_600ng_Run4_PSMs.txt	62	120	-14.8
ERC	5_SC_end_140min_600ng_Run4_PSMs.txt	56	103	-21.9
ERC	1_SC_day0_140min_600ng_Run4_PSMs.txt	37	86	-4.7
<i>L. lactis</i>	1_CAS_140min_600ng_Run4_PSMs.txt	59	121	-27.2
<i>L. lactis</i>	4_CAS_140min_600ng_Run4_PSMs.txt	44	103	-28.3
MFR	1_SC_end_140min_600ng_Run4_PSMs.txt	610	1698	-19.9
MFR	4_SC_day0_140min_600ng_Run4_PSMs.txt	496	1488	-20.2
MFR	1_SC_day0_140min_600ng_Run4_PSMs.txt	386	1268	-21.5
MFR	4_SC_end_140min_600ng_Run4_PSMs.txt	386	1292	-23.1
MFR	3_SC_end_140min_600ng_Run4_PSMs.txt	361	1256	-21.7
MFR	2_SC_day0_140min_600ng_Run4_PSMs.txt	358	1337	-20.9
MFR	3_SC_day0_140min_600ng_Run4_PSMs.txt	358	1228	-17.9
MFR	5_SC_end_140min_600ng_Run4_PSMs.txt	340	1256	-21.5
MFR	2_SC_end_140min_600ng_Run4_PSMs.txt	336	1294	-17.9
MFR	5_SC_day0_140min_600ng_Run4_PSMs.txt	278	1405	-20.5
MFR	1_CAS_140min_600ng_Run4_PSMs.txt	242	609	-20.6
MFR	5_CAS_140min_600ng_Run4_PSMs.txt	142	568	-15.7
MFR	4_CAS_140min_600ng_Run4_PSMs.txt	110	311	-21.3
MFR	1_BAS_140min_600ng_Run4_PSMs.txt	95	212	-23.3
MFR	5_BAS_140min_600ng_Run4_PSMs.txt	83	222	-13.3
MFR	2_BAS_140min_600ng_Run4_PSMs.txt	69	177	-16.4
MFR	4_BAS_140min_600ng_Run4_PSMs.txt	64	158	-14.8
MFR	3_CAS_140min_600ng_Run4_PSMs.txt	60	226	-18.9
MFR	1_SOY_140min_600ng_Run4-(1)_PSMs.txt	46	262	-9.7
MFR	2_CAS_140min_600ng_Run4_PSMs.txt	34	134	-16.2
MFR	3_BAS_140min_600ng_Run4_PSMs.txt	31	142	-17.7
MFR	3_SOY_140min_600ng_Run4_PSMs.txt	31	158	-5.5

Table S4-8 (continued).

MOUSE	5_BAS_140min_600ng_Run4_PSMs.txt	668	5543	-12.3
MOUSE	1_BAS_140min_600ng_Run4_PSMs.txt	662	5337	-13.1
MOUSE	1_SC_end_140min_600ng_Run4_PSMs.txt	622	2526	-20.6
MOUSE	1_CAS_140min_600ng_Run4_PSMs.txt	612	3814	-18.2
MOUSE	4_SC_day0_140min_600ng_Run4_PSMs.txt	506	2499	-18
MOUSE	4_CAS_140min_600ng_Run4_PSMs.txt	487	3888	-17.2
MOUSE	2_SC_end_140min_600ng_Run4_PSMs.txt	478	3063	-18.2
MOUSE	2_BAS_140min_600ng_Run4_PSMs.txt	477	4576	-12.4
MOUSE	4_SC_end_140min_600ng_Run4_PSMs.txt	465	3224	-18.1
MOUSE	3_SC_end_140min_600ng_Run4_PSMs.txt	464	2963	-16.5
MOUSE	3_SC_day0_140min_600ng_Run4_PSMs.txt	459	2517	-18.7
MOUSE	1_SC_day0_140min_600ng_Run4_PSMs.txt	447	2350	-20.1
MOUSE	3_SOY_140min_600ng_Run4_PSMs.txt	427	4509	-15.6
MOUSE	4_BAS_140min_600ng_Run4_PSMs.txt	422	4397	-13.1
MOUSE	5_SC_end_140min_600ng_Run4_PSMs.txt	421	2845	-17.5
MOUSE	2_SC_day0_140min_600ng_Run4_PSMs.txt	392	2825	-19.4
MOUSE	5_SOY_140min_600ng_Run4_PSMs.txt	360	6938	-14.5
MOUSE	3_CAS_140min_600ng_Run4_PSMs.txt	351	4289	-16.2
MOUSE	4_SOY_140min_600ng_Run4_PSMs.txt	320	6307	-16.3
MOUSE	5_CAS_140min_600ng_Run4_PSMs.txt	305	3346	-16.4
MOUSE	5_SC_day0_140min_600ng_Run4_PSMs.txt	290	2599	-17.4
MOUSE	1_SOY_140min_600ng_Run4-(1)_PSMs.txt	282	4587	-15.2
MOUSE	2_SOY_140min_600ng_Run4_PSMs.txt	236	4510	-13.6
MOUSE	3_BAS_140min_600ng_Run4_PSMs.txt	225	4189	-12.3
MOUSE	2_CAS_140min_600ng_Run4_PSMs.txt	192	2714	-14
RIN	4_SC_day0_140min_600ng_Run4_PSMs.txt	58	125	-22.4
RIN	1_SC_day0_140min_600ng_Run4_PSMs.txt	55	107	-24.4
RIN	5_SC_end_140min_600ng_Run4_PSMs.txt	37	128	-19.8
RIN	4_SC_end_140min_600ng_Run4_PSMs.txt	33	94	-15.3
SOYBEAN	1_SC_day0_140min_600ng_Run4_PSMs.txt	80	613	-21.9
SOYBEAN	4_SC_day0_140min_600ng_Run4_PSMs.txt	69	481	-26.2
SOYBEAN	1_SC_end_140min_600ng_Run4_PSMs.txt	67	325	-20.8
SOYBEAN	3_SC_day0_140min_600ng_Run4_PSMs.txt	61	541	-23
SOYBEAN	2_SC_end_140min_600ng_Run4_PSMs.txt	57	444	-17.1
SOYBEAN	4_SC_end_140min_600ng_Run4_PSMs.txt	56	641	-19.7
SOYBEAN	2_SC_day0_140min_600ng_Run4_PSMs.txt	55	519	-23.4
SOYBEAN	3_SC_end_140min_600ng_Run4_PSMs.txt	39	286	-21.9
SOYBEAN	5_SC_end_140min_600ng_Run4_PSMs.txt	34	233	-18.5

Table S4-9. Stable isotope fingerprints (SIFs) of the organisms in Group 2

All values that passed the 30 peptide threshold.

#Organism	Peptide File (Mouse #_Diet)	#Peptides	#PSMs	Median Delta ¹³ C (‰)
AMC	11_SOYO_140min_600ng_Rep_PSMs.txt	259	1255	-12.6
AMC	7_BAS_140min_600ng_Rep_PSMs.txt	217	852	-10.9
AMC	10_CEL_140min_600ng_Rep_PSMs.txt	199	806	-13.7
AMC	7_CEL_140min_600ng_Rep_PSMs.txt	228	767	-14.8
AMC	11_CEL_140min_600ng_Rep_PSMs.txt	167	726	-16.9
AMC	10_BAS_140min_600ng_Rep_PSMs.txt	197	683	-9.9
AMC	9_CEL_140min_600ng_Rep_PSMs.txt	180	647	-13.1
AMC	9_BAS_140min_600ng_Rep_PSMs.txt	155	558	-6.6
AMC	10_SOYO_140min_600ng_Rep_PSMs.txt	126	556	-11.4
AMC	7_INU_140min_600ng_Rep_PSMs.txt	210	538	-16.3
AMC	7_SOYO_140min_600ng_Rep_PSMs.txt	118	527	-11.7
AMC	12_CEL_140min_600ng_Rep_PSMs.txt	141	503	-14.1
AMC	11_INU_140min_600ng_Rep_PSMs.txt	183	477	-16.6
AMC	12_INU_140min_600ng_Rep_PSMs.txt	124	473	-16.7
AMC	11_BAS_140min_600ng_Rep_PSMs.txt	96	469	-13.1
AMC	8_INU_140min_600ng_Rep_PSMs.txt	124	437	-9.4
AMC	8_BAS_140min_600ng_Rep_PSMs.txt	64	430	-10.9
AMC	7_SUNO_140min_600ng_Rep_PSMs.txt	82	414	-13.7
AMC	11_SUNO_140min_600ng_Rep_PSMs.txt	91	395	-20.7
AMC	10_INU_140min_600ng_Rep_PSMs.txt	104	390	-13
AMC	8_SOYO_140min_600ng_Rep_PSMs.txt	74	378	-26.3
AMC	10_SUNO_140min_600ng_Rep_PSMs.txt	113	358	-15.2
AMC	12_SUNO_140min_600ng_Rep_PSMs.txt	95	356	-18.3
AMC	9_SOYO_140min_600ng_Rep_PSMs.txt	104	350	-21.7
AMC	8_SUNO_140min_600ng_Rep_PSMs.txt	89	345	-14
AMC	12_SOYO_140min_600ng_Rep_PSMs.txt	103	318	-13
AMC	9_SUNO_140min_600ng_Rep_PSMs.txt	81	300	-16.4
AMC	9_INU_140min_600ng_Rep_PSMs.txt	41	275	-9.4
AMC	8_CEL_140min_600ng_Rep_PSMs.txt	49	222	-7.2
AMC	7_SC_day0_140min_600ng_Rep_PSMs.txt	80	161	-23.1
AMC	11_SC_day0_140min_600ng_Rep_PSMs.txt	54	143	-14.3
AMC	12_SC_day0_140min_600ng_Rep_PSMs.txt	46	135	-15.9
AMC	10_SC_end_140min_600ng_Rep_PSMs.txt	53	128	-25.2
AMC	8_SC_day0_140min_600ng_Rep_PSMs.txt	38	113	-19
AMC	9_SC_day0_140min_600ng_Rep_PSMs.txt	41	92	-29.5
AMC	9_SC_end_140min_600ng_Rep_PSMs.txt	31	77	-22.6
AMC	7_SC_end_140min_600ng_Rep_PSMs.txt	36	56	-17.4
BCC	10_INU_140min_600ng_Rep_PSMs.txt	164	595	-17.1
BCC	12_INU_140min_600ng_Rep_PSMs.txt	122	558	-19.2
BCC	8_INU_140min_600ng_Rep_PSMs.txt	126	413	-10.2
BCC	7_INU_140min_600ng_Rep_PSMs.txt	132	379	-18.6

Table S4-9 (continued).

BCC	11_CEL_140min_600ng_Rep_PSMs.txt	71	283	-10.2
BCC	9_INU_140min_600ng_Rep_PSMs.txt	51	278	-7.2
BCC	12_CEL_140min_600ng_Rep_PSMs.txt	57	264	-9.7
BCC	7_CEL_140min_600ng_Rep_PSMs.txt	76	259	-9.7
BCC	11_SOYO_140min_600ng_Rep- _PSMs.txt	66	241	-10
BCC	10_CEL_140min_600ng_Rep_PSMs.txt	70	217	-13.3
BCC	11_INU_140min_600ng_Rep_PSMs.txt	89	209	-17.6
BCC	7_SUNO_140min_600ng_Rep_PSMs.txt	45	197	-19.3
BCC	12_SOYO_140min_600ng_Rep_PSMs.txt	59	196	-14.6
BCC	12_SUNO_140min_600ng_Rep_PSMs.txt	53	195	-15.4
BCC	8_SUNO_140min_600ng_Rep_PSMs.txt	50	184	-11.3
BCC	9_SUNO_140min_600ng_Rep- _PSMs.txt	52	183	-14.5
BCC	8_SOYO_140min_600ng_Rep_PSMs.txt	41	165	-20.8
BCC	10_SOYO_140min_600ng_Rep_PSMs.txt	53	161	-18.8
BCC	11_SUNO_140min_600ng_Rep_PSMs.txt	33	143	-8.5
BCC	9_CEL_140min_600ng_Rep_PSMs.txt	46	138	-16.5
BCC	9_BAS_140min_600ng_Rep_PSMs.txt	52	134	-18
BCC	7_BAS_140min_600ng_Rep_PSMs.txt	41	115	-15.3
BCC	9_SOYO_140min_600ng_Rep_PSMs.txt	35	112	-12.4
BCC	10_BAS_140min_600ng_Rep_PSMs.txt	38	109	-4.9
BCC	10_SUNO_140min_600ng_Rep_PSMs.txt	39	100	-8.4
BIN	11_CEL_140min_600ng_Rep_PSMs.txt	56	248	-10.1
BIN	10_CEL_140min_600ng_Rep_PSMs.txt	64	189	-11.7
BIN	7_CEL_140min_600ng_Rep_PSMs.txt	57	185	-10.9
BIN	9_CEL_140min_600ng_Rep_PSMs.txt	45	133	-18.5
BOV	10_SC_end_140min_600ng_Rep_PSMs.txt	100	283	-20.7
BOV	11_SC_day0_140min_600ng_Rep_PSMs.txt	84	255	-20.3
BOV	9_SC_day0_140min_600ng_Rep_PSMs.txt	70	241	-23.7
BOV	7_SC_end_140min_600ng_Rep_PSMs.txt	112	239	-16.4
BOV	12_SC_end_140min_600ng_Rep_PSMs.txt	76	235	-15.7
BOV	7_SC_day0_140min_600ng_Rep_PSMs.txt	91	225	-19.4
BOV	12_SC_day0_140min_600ng_Rep_PSMs.txt	84	223	-19.1
BOV	10_SC_day0_140min_600ng_Rep_PSMs.txt	58	191	-26.2
BOV	9_SC_end_140min_600ng_Rep_PSMs.txt	57	184	-18.4
BOV	7_INU_140min_600ng_Rep_PSMs.txt	51	173	-17.5
BOV	8_SC_day0_140min_600ng_Rep_PSMs.txt	48	163	-29.3
BOV	8_SC_end_140min_600ng_Rep_PSMs.txt	46	134	-20.7
BOV	7_CEL_140min_600ng_Rep_PSMs.txt	31	113	-14.6
BOV	10_INU_140min_600ng_Rep_PSMs.txt	36	106	-5.3
BTH	11_SOYO_140min_600ng_Rep- _PSMs.txt	246	1523	-7.4
BTH	12_BAS_140min_600ng_Rep_PSMs.txt	215	873	-10.7
BTH	10_BAS_140min_600ng_Rep_PSMs.txt	206	792	-10.5

Table S4-9 (continued).

BTH	8_SUNO_140min_600ng_Rep_PSMs.txt	169	654	-14.8
BTH	12_SOYO_140min_600ng_Rep_PSMs.txt	163	654	-12.5
BTH	9_SOYO_140min_600ng_Rep_PSMs.txt	139	589	-10.7
BTH	8_SOYO_140min_600ng_Rep_PSMs.txt	99	547	-13.8
BTH	12_SUNO_140min_600ng_Rep_PSMs.txt	158	541	-10.3
BTH	9_SUNO_140min_600ng_Rep_PSMs.txt	130	536	-20
BTH	11_BAS_140min_600ng_Rep_PSMs.txt	90	474	-18.7
BTH	7_BAS_140min_600ng_Rep_PSMs.txt	135	458	-9.8
BTH	7_SUNO_140min_600ng_Rep_PSMs.txt	98	439	-14.1
BTH	10_SOYO_140min_600ng_Rep_PSMs.txt	131	422	-12.4
BTH	9_BAS_140min_600ng_Rep_PSMs.txt	137	417	-12.5
BTH	11_SUNO_140min_600ng_Rep_PSMs.txt	82	408	-18.4
BTH	10_CEL_140min_600ng_Rep_PSMs.txt	109	389	-12.3
BTH	8_BAS_140min_600ng_Rep_PSMs.txt	59	380	-17.9
BTH	12_CEL_140min_600ng_Rep_PSMs.txt	94	366	-13.8
BTH	7_SOYO_140min_600ng_Rep_PSMs.txt	84	360	-16.4
BTH	10_SUNO_140min_600ng_Rep_PSMs.txt	102	336	-16.4
BTH	7_CEL_140min_600ng_Rep_PSMs.txt	97	321	-10.8
BTH	9_CEL_140min_600ng_Rep_PSMs.txt	89	304	-10.4
BTH	10_INU_140min_600ng_Rep_PSMs.txt	37	91	-22.2
BTH	7_INU_140min_600ng_Rep_PSMs.txt	33	79	-15.3
BTH	7_SC_end_140min_600ng_Rep_PSMs.txt	55	72	-24.2
BTH	8_INU_140min_600ng_Rep_PSMs.txt	32	67	-21
BTH	12_SC_day0_140min_600ng_Rep_PSMs.txt	39	65	-2.8
BTH	7_SC_day0_140min_600ng_Rep_PSMs.txt	32	58	-20.2
BTH	11_INU_140min_600ng_Rep_PSMs.txt	31	54	-28.2
BUN	11_SOYO_140min_600ng_Rep_PSMs.txt	393	2649	-10.4
BUN	12_SUNO_140min_600ng_Rep_PSMs.txt	318	1811	-13.5
BUN	8_SUNO_140min_600ng_Rep_PSMs.txt	265	1631	-10.3
BUN	7_SUNO_140min_600ng_Rep_PSMs.txt	211	1583	-11.1
BUN	9_SOYO_140min_600ng_Rep_PSMs.txt	269	1530	-13
BUN	9_SUNO_140min_600ng_Rep_PSMs.txt	226	1349	-12.8
BUN	12_SOYO_140min_600ng_Rep_PSMs.txt	269	1335	-10.6
BUN	10_SUNO_140min_600ng_Rep_PSMs.txt	234	1238	-9.6
BUN	8_SOYO_140min_600ng_Rep_PSMs.txt	186	1214	-13.4
BUN	11_SUNO_140min_600ng_Rep_PSMs.txt	189	1181	-10
BUN	10_SOYO_140min_600ng_Rep_PSMs.txt	273	1171	-14.6
BUN	7_SOYO_140min_600ng_Rep_PSMs.txt	176	1115	-12
BUN	12_BAS_140min_600ng_Rep_PSMs.txt	223	1084	-13.2
BUN	9_BAS_140min_600ng_Rep_PSMs.txt	270	1060	-13.4
BUN	7_BAS_140min_600ng_Rep_PSMs.txt	214	866	-12
BUN	11_BAS_140min_600ng_Rep_PSMs.txt	142	853	-18.2
BUN	9_INU_140min_600ng_Rep_PSMs.txt	88	845	-10.9
BUN	10_BAS_140min_600ng_Rep_PSMs.txt	216	824	-10.5
BUN	10_INU_140min_600ng_Rep_PSMs.txt	171	743	-11.7

Table S4-9 (continued).

BUN	8_INU_140min_600ng_Rep_PSMs.txt	159	676	-16.1
BUN	11_CEL_140min_600ng_Rep_PSMs.txt	113	594	-11.8
BUN	8_BAS_140min_600ng_Rep_PSMs.txt	81	523	-9.7
BUN	7_CEL_140min_600ng_Rep_PSMs.txt	104	443	-12.6
BUN	7_INU_140min_600ng_Rep_PSMs.txt	130	441	-13.5
BUN	12_INU_140min_600ng_Rep_PSMs.txt	97	393	-15.2
BUN	11_INU_140min_600ng_Rep_PSMs.txt	93	320	-9.3
BUN	11_SC_day0_140min_600ng_Rep_PSMs.txt	75	280	-19.1
BUN	10_CEL_140min_600ng_Rep_PSMs.txt	90	276	-18.5
BUN	9_CEL_140min_600ng_Rep_PSMs.txt	68	225	-25.6
BUN	9_SC_day0_140min_600ng_Rep_PSMs.txt	50	182	-14.6
BUN	8_SC_day0_140min_600ng_Rep_PSMs.txt	36	177	-11.7
BUN	7_SC_end_140min_600ng_Rep_PSMs.txt	78	161	-13.2
BUN	12_CEL_140min_600ng_Rep_PSMs.txt	44	149	-11.9
BUN	10_SC_day0_140min_600ng_Rep_PSMs.txt	40	122	-20.4
BUN	12_SC_day0_140min_600ng_Rep_PSMs.txt	43	103	-11.3
BUN	9_SC_end_140min_600ng_Rep_PSMs.txt	38	96	-15.1
BUN	10_SC_end_140min_600ng_Rep_PSMs.txt	38	78	-30.7
BUN	7_SC_day0_140min_600ng_Rep_PSMs.txt	31	62	-30.5
CHICKEN	7_SUNO_140min_600ng_Rep_PSMs.txt	35	926	-12
CHICKEN	9_CEL_140min_600ng_Rep_PSMs.txt	45	902	-11.8
CHICKEN	10_CEL_140min_600ng_Rep_PSMs.txt	53	792	-10.1
CHICKEN	11_SUNO_140min_600ng_Rep_PSMs.txt	46	762	-10.7
CHICKEN	12_BAS_140min_600ng_Rep_PSMs.txt	55	738	-13
CHICKEN	9_BAS_140min_600ng_Rep_PSMs.txt	63	674	-10.7
CHICKEN	9_SUNO_140min_600ng_Rep_PSMs.txt	43	668	-10.7
CHICKEN	10_SUNO_140min_600ng_Rep_PSMs.txt	50	639	-13.9
CHICKEN	7_SOYO_140min_600ng_Rep_PSMs.txt	36	624	-12.8
CHICKEN	10_SOYO_140min_600ng_Rep_PSMs.txt	53	609	-13
CHICKEN	10_INU_140min_600ng_Rep_PSMs.txt	48	590	-10
CHICKEN	12_INU_140min_600ng_Rep_PSMs.txt	34	494	-11.1
CHICKEN	7_BAS_140min_600ng_Rep_PSMs.txt	46	488	-11.9
CHICKEN	12_SUNO_140min_600ng_Rep_PSMs.txt	47	486	-11.2
CHICKEN	10_BAS_140min_600ng_Rep_PSMs.txt	51	471	-11.1
ECL	11_SOYO_140min_600ng_Rep_PSMs.txt	279	2305	-9.4
ECL	11_BAS_140min_600ng_Rep_PSMs.txt	121	1279	-7.5
ECL	11_SUNO_140min_600ng_Rep_PSMs.txt	120	1194	-13.4
ECL	10_SOYO_140min_600ng_Rep_PSMs.txt	145	1006	-9.9
ECL	12_SUNO_140min_600ng_Rep_PSMs.txt	178	958	-7.7
ECL	12_INU_140min_600ng_Rep_PSMs.txt	119	955	-6.8
ECL	10_SUNO_140min_600ng_Rep_PSMs.txt	143	922	-11.5

Table S4-9 (continued).

ECL	12_BAS_140min_600ng_Rep_PSMs.txt	166	841	-9.7
ECL	7_SUNO_140min_600ng_Rep_PSMs.txt	52	822	-5.3
ECL	10_BAS_140min_600ng_Rep_PSMs.txt	151	658	-10.2
ECL	11_INU_140min_600ng_Rep_PSMs.txt	153	624	-12.1
ECL	12_SOYO_140min_600ng_Rep_PSMs.txt	126	587	-6.9
ECL	10_INU_140min_600ng_Rep_PSMs.txt	112	509	-8.6
ECL	8_BAS_140min_600ng_Rep_PSMs.txt	35	450	-6.7
ECL	7_BAS_140min_600ng_Rep_PSMs.txt	83	447	-6.1
ECL	7_SOYO_140min_600ng_Rep_PSMs.txt	56	414	-9.3
ECL	9_SOYO_140min_600ng_Rep_PSMs.txt	62	409	-12
ECL	8_SOYO_140min_600ng_Rep_PSMs.txt	50	322	-11.5
ECL	8_SUNO_140min_600ng_Rep_PSMs.txt	46	303	-9.8
ECL	7_INU_140min_600ng_Rep_PSMs.txt	72	238	-7.2
ECL	9_SUNO_140min_600ng_Rep- PSMs.txt	45	225	-5
ECL	11_CEL_140min_600ng_Rep_PSMs.txt	30	153	-10.2
ECL	10_CEL_140min_600ng_Rep_PSMs.txt	39	138	-10.2
ECL	9_CEL_140min_600ng_Rep_PSMs.txt	30	133	-2.8
ECL	9_BAS_140min_600ng_Rep_PSMs.txt	46	112	-14.7
ERC	10_SC_end_140min_600ng_Rep_PSMs.txt	40	116	-30.8
ERC	12_SC_end_140min_600ng_Rep_PSMs.txt	35	100	-18.9
MFR	12_SC_day0_140min_600ng_Rep_PSMs.txt	569	2655	-18.9
MFR	9_SC_end_140min_600ng_Rep_PSMs.txt	323	2184	-17.9
MFR	12_SC_end_140min_600ng_Rep_PSMs.txt	353	2050	-17.8
MFR	11_SC_day0_140min_600ng_Rep_PSMs.txt	459	1994	-17.4
MFR	7_SC_end_140min_600ng_Rep_PSMs.txt	568	1892	-20.8
MFR	10_SC_end_140min_600ng_Rep_PSMs.txt	356	1791	-18.4
MFR	11_SC_end_140min_600ng_Rep_PSMs.txt	184	1761	-18.2
MFR	7_SC_day0_140min_600ng_Rep_PSMs.txt	442	1734	-17.7
MFR	9_SC_day0_140min_600ng_Rep_PSMs.txt	359	1576	-18.4
MFR	10_SC_day0_140min_600ng_Rep_PSMs.txt	290	1486	-18.3
MFR	8_SC_end_140min_600ng_Rep_PSMs.txt	287	1464	-18
MFR	8_SC_day0_140min_600ng_Rep_PSMs.txt	274	1276	-18.5
MFR	8_INU_140min_600ng_Rep_PSMs.txt	174	722	-15.8
MFR	8_SUNO_140min_600ng_Rep_PSMs.txt	87	417	-13.9
MFR	8_BAS_140min_600ng_Rep_PSMs.txt	46	336	-12.8
MFR	7_SUNO_140min_600ng_Rep_PSMs.txt	70	288	-12.3
MFR	11_SOYO_140min_600ng_Rep- PSMs.txt	69	277	-14.7
MFR	7_SOYO_140min_600ng_Rep_PSMs.txt	52	277	-13.5
MFR	8_SOYO_140min_600ng_Rep_PSMs.txt	60	263	-13.3
MFR	9_SUNO_140min_600ng_Rep- PSMs.txt	70	255	-11.2
MFR	10_INU_140min_600ng_Rep_PSMs.txt	68	232	-17.2
MFR	7_INU_140min_600ng_Rep_PSMs.txt	90	227	-14.9

Table S4-9 (continued).

MFR	12_INU_140min_600ng_Rep_PSMs.txt	66	212	-24.4
MFR	9_SOYO_140min_600ng_Rep_PSMs.txt	64	210	-16.6
MFR	9_BAS_140min_600ng_Rep_PSMs.txt	65	202	-7.9
MFR	12_BAS_140min_600ng_Rep_PSMs.txt	46	181	-6.8
MFR	12_SUNO_140min_600ng_Rep_PSMs.txt	44	171	-8.1
MFR	11_INU_140min_600ng_Rep_PSMs.txt	73	165	-17.6
MFR	7_BAS_140min_600ng_Rep_PSMs.txt	61	162	-10.2
MFR	10_SUNO_140min_600ng_Rep_PSMs.txt	37	132	-4.6
MFR	12_SOYO_140min_600ng_Rep_PSMs.txt	45	130	-17.4
MFR	7_CEL_140min_600ng_Rep_PSMs.txt	43	108	-5.4
MFR	10_BAS_140min_600ng_Rep_PSMs.txt	35	108	-17
MFR	10_SOYO_140min_600ng_Rep_PSMs.txt	35	81	-11.8
MFR	9_CEL_140min_600ng_Rep_PSMs.txt	32	80	-11.4
MOUSE	11_SOYO_140min_600ng_Rep_PSMs.txt	698	13974	-12.3
MOUSE	7_CEL_140min_600ng_Rep_PSMs.txt	597	8863	-11.5
MOUSE	9_CEL_140min_600ng_Rep_PSMs.txt	542	8251	-11.7
MOUSE	7_SOYO_140min_600ng_Rep_PSMs.txt	454	7608	-10.9
MOUSE	11_INU_140min_600ng_Rep_PSMs.txt	706	7541	-13.9
MOUSE	7_BAS_140min_600ng_Rep_PSMs.txt	599	7468	-10
MOUSE	12_INU_140min_600ng_Rep_PSMs.txt	521	7389	-13
MOUSE	10_SOYO_140min_600ng_Rep_PSMs.txt	550	7330	-11.1
MOUSE	11_BAS_140min_600ng_Rep_PSMs.txt	404	7324	-11.3
MOUSE	9_SUNO_140min_600ng_Rep_PSMs.txt	439	7093	-9.8
MOUSE	7_SUNO_140min_600ng_Rep_PSMs.txt	431	7089	-10.8
MOUSE	10_CEL_140min_600ng_Rep_PSMs.txt	562	7025	-12.1
MOUSE	11_SUNO_140min_600ng_Rep_PSMs.txt	390	6778	-10.7
MOUSE	11_CEL_140min_600ng_Rep_PSMs.txt	472	6569	-10.4
MOUSE	8_SOYO_140min_600ng_Rep_PSMs.txt	397	6471	-11.3
MOUSE	8_CEL_140min_600ng_Rep_PSMs.txt	225	6418	-12.1
MOUSE	12_BAS_140min_600ng_Rep_PSMs.txt	514	6379	-10.8
MOUSE	8_SUNO_140min_600ng_Rep_PSMs.txt	500	6329	-11.1
MOUSE	10_SUNO_140min_600ng_Rep_PSMs.txt	469	6296	-12.2
MOUSE	10_INU_140min_600ng_Rep_PSMs.txt	540	6239	-10.4
MOUSE	12_SUNO_140min_600ng_Rep_PSMs.txt	489	6173	-11.3
MOUSE	9_BAS_140min_600ng_Rep_PSMs.txt	548	5849	-11.6
MOUSE	9_SOYO_140min_600ng_Rep_PSMs.txt	417	5844	-10
MOUSE	7_INU_140min_600ng_Rep_PSMs.txt	656	5766	-11.9
MOUSE	8_BAS_140min_600ng_Rep_PSMs.txt	277	5623	-11.2
MOUSE	12_SOYO_140min_600ng_Rep_PSMs.txt	480	5602	-11.9
MOUSE	9_INU_140min_600ng_Rep_PSMs.txt	195	5420	-11.5
MOUSE	12_CEL_140min_600ng_Rep_PSMs.txt	422	5333	-12.1
MOUSE	10_BAS_140min_600ng_Rep_PSMs.txt	469	4913	-11.7

Table S4-9 (continued).

MOUSE	8_INU_140min_600ng_Rep_PSMs.txt	476	4496	-11.1
MOUSE	12_SC_end_140min_600ng_Rep_PSMs.txt	410	3312	-19.2
MOUSE	8_SC_end_140min_600ng_Rep_PSMs.txt	441	3256	-17.2
MOUSE	10_SC_end_140min_600ng_Rep_PSMs.txt	419	3236	-15.7
MOUSE	10_SC_day0_140min_600ng_Rep_PSMs.txt	326	3127	-16.4
MOUSE	11_SC_end_140min_600ng_Rep_PSMs.txt	219	2979	-16.8
MOUSE	11_SC_day0_140min_600ng_Rep_PSMs.txt	399	2974	-17.2
MOUSE	9_SC_end_140min_600ng_Rep_PSMs.txt	350	2935	-16.1
MOUSE	9_SC_day0_140min_600ng_Rep_PSMs.txt	366	2924	-18
MOUSE	7_SC_day0_140min_600ng_Rep_PSMs.txt	473	2902	-18.1
MOUSE	8_SC_day0_140min_600ng_Rep_PSMs.txt	370	2750	-16.1
MOUSE	12_SC_day0_140min_600ng_Rep_PSMs.txt	447	2633	-18.8
MOUSE	7_SC_end_140min_600ng_Rep_PSMs.txt	382	1781	-17
RIN	12_SC_day0_140min_600ng_Rep_PSMs.txt	77	268	-20.3
RIN	7_SC_day0_140min_600ng_Rep_PSMs.txt	52	152	-21.5
RIN	8_SC_day0_140min_600ng_Rep_PSMs.txt	38	146	-7.8
RIN	7_SC_end_140min_600ng_Rep_PSMs.txt	41	69	-28.9
SOYBEAN	11_SC_end_140min_600ng_Rep_PSMs.txt	44	988	-21.9
SOYBEAN	11_SC_day0_140min_600ng_Rep_PSMs.txt	69	765	-19.7
SOYBEAN	8_SC_day0_140min_600ng_Rep_PSMs.txt	53	705	-23.2
SOYBEAN	8_SC_end_140min_600ng_Rep_PSMs.txt	51	586	-23.4
SOYBEAN	12_SC_day0_140min_600ng_Rep_PSMs.txt	56	564	-19.3
SOYBEAN	9_SC_day0_140min_600ng_Rep_PSMs.txt	40	474	-22
SOYBEAN	7_SC_day0_140min_600ng_Rep_PSMs.txt	53	401	-18.2
SOYBEAN	10_SC_end_140min_600ng_Rep_PSMs.txt	46	401	-21.4
SOYBEAN	12_SC_end_140min_600ng_Rep_PSMs.txt	44	382	-20
SOYBEAN	10_SC_day0_140min_600ng_Rep_PSMs.txt	39	331	-24.9
SOYBEAN	7_SC_end_140min_600ng_Rep_PSMs.txt	55	278	-21.3

Supplementary table S4-10. Differentially abundant *M. formatexigens* proteins between the soy protein and egg whites diet (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, S0=0.1, both sides, not paired)

Accession	Description	Directionality	Avg orgNS AF in Soy diet	Avg orgNS AF in Egg diet	Ratio	Stdev SOY (n = 5)	Stdev EGG (n = 5)
MFR_C6LG S7	ABC transporter, solute-binding protein	Up in EGG	0.2575 83	0.7564 56	2.9367 5	0.0910 78	0.2758 92
MFR_C6LA I8	ABC transporter, solute-binding protein	Up in EGG	0.1686 66	0.3104 28	1.8404 85	0.0209 3	0.0730 12
MFR_C6LE W0	Ethanolamine utilization protein EutQ	Up in EGG	0.0608 04	0.2639 78	4.3414 56	0.0143 48	0.1421 78
MFR_C6LD Z3	ABC transporter, ATP-binding protein	Up in EGG	0.0565 83	0.2120 99	3.7484 64	0.0115 21	0.0558 32
MFR_C6LD Z4	Branched-chain amino acid ABC transporter, permease protein	Up in EGG	0.0458 06	0.2105 29	4.5960 64	0.0218 49	0.0622 96
MFR_C6LB 27	S1 RNA binding domain protein	Up in EGG	0.0383 51	0.1844 26	4.8088 83	0.0289 19	0.0603 27
MFR_C6LL 31	Branched-chain amino acid ABC transporter, permease protein	Up in EGG	0.0343 93	0.0964 03	2.8029 42	0.0177 44	0.0169 52
MFR_C6LM 79	Glucose-1-phosphate adenylyltransferase	Up in EGG	0.0311 77	0.0951 78	3.0528 31	0.0159 06	0.0275 97
MFR_C6LC U0	Glucosamine-6-phosphate deaminase	Up in EGG	0.0242 98	0.1062 98	4.3747 62	0.0067 86	0.0533 56
MFR_C6LH D2	TRAP transporter solute receptor, DctP family	Up in EGG	0.0200 42	0.1342 06	6.6962 39	0.0050 32	0.0706 69
MFR_C6LH Q2	L-arabinose isomerase	Up in EGG	0.0184 88	0.1282 91	6.9392 03	0.0093 31	0.0563 91
MFR_C6LC T9	N-acetylglucosamine-6-phosphate deacetylase	Up in EGG	0.0180 27	0.0941 67	5.2236 14	0.0045 27	0.0549 87
MFR_C6LA I2	ABC transporter, solute-binding protein	Up in EGG	0.0173 05	0.4300 66	24.851 71	0.0048 33	0.2163 49
MFR_C6LII 1	Putative ABC transporter periplasmic binding protein BHWA1_00430	Up in EGG	0.0172 05	0.0890 2	5.1740 12	0.0048 05	0.0556 47
MFR_C6L8 Y8	ABC transporter, solute-binding protein	Up in EGG	0.0144 49	0.1911 75	13.230 98	0.0040 35	0.1228 9

Supplementary table S4-10 (continued).

MFR_C6LM 58	ABC transporter, solute-binding protein	Up in EGG	0.0141 74	0.0850 85	6.0029 3	0.0039 58	0.0530 17
MFR_C6L9I 6	Glucose-1-phosphate adenylyltransferase	Up in EGG	0.0133 78	0.0505 74	3.7804 79	0.0037 36	0.0257 49
MFR_C6LB 26	Uncharacterized protein	Up in EGG	0.0099 45	0.1451 14	14.591 77	0.0041 93	0.0547 54
MFR_C6LF 52	Sugar ABC transporter, periplasmic sugar-binding protein	Up in SOY	5.6346 34	1.5224 52	3.7010 27	2.2262 1	1.0224 7
MFR_C6LH E5	50S ribosomal protein L1	Up in SOY	1.0249 45	0.6796 13	1.5081 29	0.1036 86	0.1214 13
MFR_C6LH E4	50S ribosomal protein L11	Up in SOY	0.6340 27	0.3664 81	1.7300 42	0.2233 89	0.1161 56
MFR_C6LJ L3	UPF0145 protein BRYFOR_08850	Up in SOY	0.5856 77	0.1488 39	3.9349 8	0.5344 59	0.1485 77
MFR_C6LF 51	ABC transporter, ATP-binding protein	Up in SOY	0.4557 19	0.0957 36	4.7601 54	0.0764 97	0.1402 17
MFR_C6LJ L2	ABC transporter, ATP-binding protein	Up in SOY	0.4202 5	0.0361 63	11.621 09	0.1454 9	0.0259 19
MFR_C6LB 56	Uncharacterized protein	Up in SOY	0.3540 14	0.0099 3	35.649 23	0.1949 84	0.0017 02
MFR_C6LA A8	Hydrolase, alpha/beta domain protein	Up in SOY	0.3250 16	0.0272 31	11.935 51	0.1358 13	0.0170 11
MFR_C6LI4 3	Glutamate--ammonia ligase, catalytic domain protein	Up in SOY	0.1874 51	0.0099 82	18.778 4	0.0923 64	0.0060 33
MFR_C6LB X2	Translation initiation factor IF-3	Up in SOY	0.0774 04	0.0297 28	2.6037 18	0.0381 73	0.0103 15
MFR_C6LI G3	B12 binding domain protein	Up in SOY	0.0572 32	0.0193 93	2.9512 64	0.0140 75	0.0033 23
MFR_C6LL 77	Repeat protein	Up in SOY	0.0496 87	0.0166 34	2.9870 81	0.0273 09	0.0095 08
MFR_C6LE 08	Pyridine nucleotide-disulfide oxidoreductase	Up in SOY	0.0328 2	0.0104 49	3.1411 04	0.0175 91	0.0060 76
MFR_C6LB C3	Glutamate--ammonia ligase, catalytic domain protein	Up in SOY	0.0239 32	0.0058 48	4.0922 97	0.0144 8	0.0010 02
MFR_C6LH J8	Beta-galactosidase	Up in SOY	0.0199 15	0.0060 82	3.2746 44	0.0110 86	0.0010 42
MFR_C6LB A1	CTP synthase	Up in SOY	0.0179 55	0.0076 85	2.3365 6	0.0089 05	0.0013 17

Supplementary table S4-11. Differentially abundant *M. formatexigens* proteins between the soy protein and casein diet (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, S0=0.1, both sides, not paired)

Accession	Description	Directionality	Avg orgNS AF in Casein diet	Avg orgNS AF in Soy diet	Ratio	Stdev SOY (n = 5)	Stdev CAS (n = 5)
MFR_C6LDZ3	ABC transporter, ATP-binding protein	Up in CAS	0.0565 83	0.0115 21	4.7506 05	0.2688 03	0.0898 22
MFR_C6LCU0	Glucosamine-6-phosphate deaminase	Up in CAS	0.0242 98	0.0067 86	4.3909 43	0.1066 91	0.0321 64
MFR_C6LCT9	N-acetylglucosamine-6-phosphate deacetylase	Up in CAS	0.0180 27	0.0045 27	6.0857 26	0.1097 09	0.0196 68
MFR_C6LY8	ABC transporter, solute-binding protein	Up in CAS	0.0144 49	0.0040 35	7.5801 54	0.1095 26	0.0713 59
MFR_C6LM58	ABC transporter, solute-binding protein	Up in CAS	0.0141 74	0.0039 58	13.092 46	0.1855 71	0.1346 88
MFR_C6L9I6	Glucose-1-phosphate adenylyltransferase	Up in CAS	0.0133 78	0.0037 36	3.0838 03	0.0412 54	0.0283 21
MFR_C6LH55	50S ribosomal protein L24	Up in CAS	0.0679 66	0.0270 89	4.5230 5	0.3074 13	0.1505 47
MFR_C6L950	ABC transporter, solute-binding protein	Up in CAS	0.0551 65	0.0272 61	5.4878 85	0.3027 38	0.2308 62
MFR_C6LL96	ABC transporter, solute-binding protein	Up in CAS	0.0445 52	0.0211 38	4.2591 01	0.1897 53	0.1254 43
MFR_C6LDG5	Alpha-1,4 glucan phosphorylase	Up in CAS	0.0423 14	0.0313 78	5.1475 99	0.2178 14	0.0431 08
MFR_C6L9A5	Putative N-acetylmannosamine-6-phosphate 2-epimerase	Up in CAS	0.0256 6	0.0071 66	8.2052 81	0.2105 44	0.1035 65
MFR_C6L9F6	Phosphoribosylformylglycinamide synthase	Up in CAS	0.0216 15	0.0069 82	3.2308 16	0.0698 35	0.0462 11
MFR_C6LJR6	4Fe-4S binding domain protein	Up in CAS	0.0211 92	0.0058 4	3.2579 98	0.0690 45	0.0430 91
MFR_C6LFC3	Uncharacterized protein	Up in CAS	0.0187 72	0.0047 14	3.6167 26	0.0678 94	0.0463 85
MFR_C6L9Q8	Leucine--tRNA ligase	Up in CAS	0.0159 17	0.0098 58	3.2425 95	0.0516 13	0.0240 28
MFR_C6LM62	Family 4 glycosyl hydrolase	Up in CAS	0.0154 52	0.0038 8	10.947 4	0.1691 58	0.1277 98

Supplementary table S4-11 (continued).

MFR_C6LIL3	Alpha-1,4 glucan phosphorylase	Up in CAS	0.0153 92	0.0099 85	3.8240 55	0.0588 59	0.0316 62
MFR_C6LF07	Glycosyl hydrolase family 3 N-terminal domain protein	Up in CAS	0.0139 28	0.0063 54	4.8548 22	0.0676 19	0.0447 6
MFR_C6LHD3	Glycosyl hydrolase, family 31	Up in CAS	0.0085 36	0.0035 99	12.672 17	0.1081 71	0.1410 38
MFR_C6LDZ4	Branched-chain amino acid ABC transporter, permease protein	Up in SOY	0.0458 06	0.0218 49	2.0588 43	0.2050 14	0.0222 49
MFR_C6LAI2	ABC transporter, solute-binding protein	Up in SOY	0.0173 05	0.0048 33	9.3358 57	0.0085 74	0.0018 54
MFR_C6LII1	Putative ABC transporter periplasmic binding protein BHWA1_00430	Up in SOY	0.0172 05	0.0048 05	9.3358 57	0.0085 25	0.0018 43
MFR_C6LF52	Sugar ABC transporter, periplasmic sugar-binding protein	Up in SOY	5.6346 34	2.2262 1	33.507 85	0.8848 89	0.1681 59
MFR_C6LHE5	50S ribosomal protein L1	Up in SOY	1.0249 45	0.1036 86	10.190 89	0.5717 58	0.1005 75
MFR_C6LJL3	UPF0145 protein BRYFOR_08850	Up in SOY	0.5856 77	0.5344 59	27.940 39	0.0352 96	0.0209 62
MFR_C6LF51	ABC transporter, ATP-binding protein	Up in SOY	0.4557 19	0.0764 97	16.449 9	0.0355 57	0.0277 03
MFR_C6LJL2	ABC transporter, ATP-binding protein	Up in SOY	0.4202 5	0.1454 9	36.300 71	0.0171 66	0.0115 77
MFR_C6LB56	Uncharacterized protein	Up in SOY	0.3540 14	0.1949 84	96.349 25	0.0085 79	0.0036 74
MFR_C6LA8	Hydrolase, alpha/beta domain protein	Up in SOY	0.3250 16	0.1358 13	24.100 42	0.0207 63	0.0134 86
MFR_C6LIG3	B12 binding domain protein	Up in SOY	0.0572 32	0.0140 75	10.319 52	0.0164	0.0055 46
MFR_C6LE08	Pyridine nucleotide-disulfide oxidoreductase	Up in SOY	0.0328 2	0.0175 91	7.3825 34	0.0085 69	0.0044 46
MFR_C6LH48	30S ribosomal protein S19	Up in SOY	0.7811 03	0.2193 16	4.8488 34	0.2775 82	0.1610 91
MFR_C6L8S5	Arabinose-binding protein	Up in SOY	0.4160 16	0.1257 45	4.8476 85	0.1935 88	0.0858 17
MFR_C6L9A7	ABC transporter, solute-binding protein	Up in SOY	0.3787 56	0.1178 23	1.8780 67	1.0434 67	0.2016 73

Supplementary table S4-11 (continued).

MFR_C6LI3 8	Uncharacterized protein	Up in SOY	0.3737 05	0.0425 88	4.7765 24	0.1741 28	0.0782 38
MFR_C6LJB 4	Oxidoreductase, short chain dehydrogenase/reductase family protein	Up in SOY	0.2203 94	0.0984 65	8.0656 83	0.0415 49	0.0273 25
MFR_C6LK U1	Oxidoreductase, short chain dehydrogenase/reductase family protein	Up in SOY	0.2163 74	0.0988 39	7.9185 75	0.0415 49	0.0273 25
MFR_C6LM B6	Glutamate dehydrogenase	Up in SOY	0.1569 48	0.0656 57	2.7370 1	0.4725 48	0.0573 43
MFR_C6LG Z6	Ribosome-binding ATPase YchF	Up in SOY	0.1417 8	0.0401 88	8.3121 81	0.0747 09	0.0170 57
MFR_C6LF B3	ABC transporter, solute-binding protein	Up in SOY	0.0818 5	0.0414	6.8508 65	0.0132 17	0.0119 47
MFR_C6LB8 2	ABC transporter, solute-binding protein	Up in SOY	0.0792 71	0.0290 2	5.2049 1	0.0335 09	0.0152 3
MFR_C6LL5 9	Trigger factor	Up in SOY	0.0606 36	0.0308 1	1.2071 75	0.1714 32	0.0502 29
MFR_C6LIZ 0	Enolase	Up in SOY	0.0575 85	0.0165 03	3.0171 96	0.0236 53	0.0190 86
MFR_C6LCJ 1	Hydrolyase, tartrate alpha subunit/fumarate domain protein, Fe-S type	Up in SOY	0.0510 68	0.0570 94	1.8260 24	0.1453 82	0.0279 67
MFR_C6LE C5	ATP-dependent zinc metalloprotease FtsH	Up in SOY	0.0411 65	0.0179 97	4.2091 22	0.0131 75	0.0097 8
MFR_C6LH U4	ATP-dependent zinc metalloprotease FtsH	Up in SOY	0.0402 02	0.0175 76	4.2091 22	0.0128 67	0.0095 51
MFR_C6LK M1	Hsp90 protein	Up in SOY	0.0332 45	0.0129 22	2.7301 67	0.0841 22	0.0121 77
MFR_C6LH 12	NTP_transferase domain-containing protein	Up in SOY	0.0298 35	0.0159 38	1.0546 15	0.1040 1	0.0282 9
MFR_C6LIC 2	Sortase B signal domain, QVPTGV class	Up in SOY	0.0276 81	0.0239 23	10.340 92	0.0044 73	0.0026 77
MFR_C6LCJ 3	2Fe-2S iron-sulfur cluster binding domain protein	Up in SOY	0.0265 76	0.0074 22	1.3644 32	0.0992 75	0.0194 78
MFR_C6LE X2	Acetaldehyde dehydrogenase (Acetylating)	Up in SOY	0.0222 24	0.0119 81	1.4753 42	0.0611	0.0150 64

Supplementary table S4-12. Differentially abundant *B. thetaiotaomicron* proteins between the inulin and cellulose diet (Student's T-test corrected for multiple hypothesis testing with a permutation-based FDR of 5%, S0=0.1, both sides, not paired)

Accession	Description	Directionality	Avg orgNSAF in Cellulose diet	Avg orgNSAF in Inulin diet	Ratio	Stdev CEL (n = 6)	Stdev INU (n = 6)
BTH_Q8A6C7	Uncharacterized protein	Up in CEL	0.155303957	0.00777429	19.977	0.1367578	0.0003997
BTH_Q8AAC2	Glutamine synthetase	Up in INU	0.002409737	0.0401692	16.67	0.0002398	0.0472335
BTH_Q8A6W5	DUF4960 domain-containing protein	Up in INU	0.005204267	0.07313143	14.052	0.0029525	0.0284925
BTH_Q8A1G2	Starch-binding protein SusD	Up in CEL	0.085229425	0.00684725	12.447	0.0210696	0.0022567
BTH_Q8A6W6	Glycoside hydrolase family 32	Up in INU	0.007220661	0.08868421	12.282	0.0051855	0.0267605
BTH_Q89ZL9	SusD homolog	Up in CEL	0.063060747	0.00519682	12.134	0.0156086	0.0030566
BTH_Q8A6W9	Fructokinase	Up in INU	0.019584009	0.21457038	10.956	0.012055	0.126454
BTH_Q8A6W7	Levanase (2,6-beta-D-fructofuranosidase)	Up in INU	0.029893082	0.28036961	9.3791	0.0137851	0.0537511
BTH_Q8A6W4	SusD homolog	Up in INU	0.024169773	0.22557141	9.3328	0.0124344	0.0713585
BTH_Q8A0I2	SusC homolog	Up in CEL	0.027180317	0.00322034	8.4402	0.0094937	0.0012889
BTH_Q89YL2	Uncharacterized protein	Up in CEL	0.115150673	0.01391935	8.2727	0.0360752	0.0057816
BTH_Q8A0I3	SusD homolog	Up in CEL	0.065703795	0.00857968	7.6581	0.0175983	0.007144
BTH_Q8A0G5	Uncharacterized protein	Up in INU	0.004359053	0.03253201	7.4631	0.0004337	0.030094
BTH_Q8A9Y2	Alpha-glucosidase	Up in CEL	0.026166468	0.00353377	7.4047	0.0062933	0.0001817
BTH_Q8A2C3	Uncharacterized protein	Up in CEL	0.032518755	0.00484086	6.7176	0.007943	0.001588
BTH_Q8A8W9	Putative secreted endoglycosidase	Up in CEL	0.680618155	0.1026481	6.6306	0.1428556	0.0601285

Supplementary table S4-12 (continued).

BTH_Q89ZS 7	Putative cell surface protein	Up in CEL	0.0613293 95	0.009413 44	6.5151	0.033 3235	0.00535 8
BTH_Q8A1V 7	Putative zinc protease	Up in CEL	0.0268079 03	0.004131 9	6.488	0.003 9615	0.00195 68
BTH_Q8A7E 8	Putative Pirin family protein	Up in INU	0.0186319 85	0.112618 45	6.0444	0.017 5126	0.07023 4
BTH_G8JZT0	Outer membrane protein SusE	Up in CEL	0.0370111 78	0.006127 03	6.0406	0.016 2056	0.00031 5
BTH_Q89ZS 0	Uncharacterized protein	Up in CEL	0.0908494 58	0.015809 03	5.7467	0.012 202	0.01486 18
BTH_Q8A88 7	Uncharacterized protein	Up in CEL	0.3937562 85	0.068709 3	5.7308	0.160 4079	0.05703 57
BTH_Q8A8R 2	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase 1	Up in CEL	0.0886769 95	0.015480 18	5.7284	0.038 678	0.00939 7
BTH_Q89ZN 9	Beta-hexosaminidase	Up in CEL	0.0340780 68	0.006230 35	5.4697	0.010 433	0.00325 08
BTH_Q89ZS 8	SusC homolog	Up in CEL	0.0199654 36	0.003694 69	5.4038	0.007 4185	0.00205 11
BTH_Q89ZM 0	SusC homolog	Up in CEL	0.0344306 42	0.006391 17	5.3872	0.006 4152	0.00317 51
BTH_Q89ZZ 4	Fimbrillin_C domain-containing protein	Up in CEL	0.0277104 27	0.005566 1	4.9784	0.013 2435	0.00028 62
BTH_Q8A0N 7	SusD homolog	Up in CEL	0.6104406 72	0.127745 48	4.7786	0.120 7354	0.04471 44
BTH_Q8A7M 2	Uncharacterized protein	Up in CEL	0.0576628 79	0.012076 93	4.7746	0.018 1751	0.00773 5
BTH_Q8A8 W7	Coagulation factor 5/8 type, C-terminal	Up in CEL	0.2075641 44	0.04405	4.712	0.055 1235	0.02088 22
BTH_Q8A5H 0	SusD homolog	Up in CEL	0.1737296 93	0.037810 74	4.5947	0.047 392	0.01359 4
BTH_Q89ZT 1	Putative chitobiase	Up in CEL	0.0999802 95	0.021804 35	4.5853	0.025 6162	0.01310 38
BTH_Q8A55 5	O-acetylhomoserine (Thiol)-lyase	Up in INU	0.0116038 94	0.051955 71	4.4774	0.008 2851	0.03361 84
BTH_Q8A6L 6	2-isopropylmalate synthase	Up in INU	0.0078934 71	0.035308 3	4.4731	0.006 4	0.02378 56

Supplementary table S4-12 (continued).

BTH_Q8A4B7	Putative cation efflux transporter	Up in CEL	0.0100122 45	0.002253 95	4.4421	0.004 6509	0.00011 59
BTH_Q8A9T0	Aspartate decarboxylase AsdA	Up in INU	0.0079892 2	0.035429 83	4.4347	0.005 8422	0.02531 85
BTH_Q8AA A8	Carbamoyl-phosphate synthase (glutamine-hydrolyzing)	Up in INU	0.0124619 34	0.054989 16	4.4126	0.003 4494	0.02209 81
BTH_Q89ZT0	MscS mechanosensitive ion channel	Up in CEL	0.0482419 52	0.011099 38	4.3464	0.016 1483	0.00412 26
BTH_Q8A6W3	SusC homolog	Up in INU	0.0280818 95	0.121346 51	4.3212	0.008 8613	0.02789 08
BTH_Q8AAZ1	Cytidine/deoxycytidylate deaminase	Up in CEL	0.0920808 84	0.021318 46	4.3193	0.037 2206	0.00858 94
BTH_Q8A5H1	SusC homolog	Up in CEL	0.2015051 7	0.047731 96	4.2216	0.049 6571	0.01758 86
BTH_Q89ZG5	Transglutaminase-like protein	Up in CEL	0.0221094 06	0.005314 93	4.1599	0.009 9014	0.00279 83
BTH_Q8A2J3	Glucosylceramidase	Up in INU	0.0035417 31	0.014702 7	4.1513	0.000 3524	0.01028 68
BTH_G8JZS6	Outer membrane protein SusF	Up in CEL	0.0201763 55	0.004888 99	4.1269	0.010 1922	0.00025 14
BTH_Q89ZS9	SusD homolog	Up in CEL	0.0312441 79	0.007611 73	4.1047	0.007 3651	0.00444 64
BTH_Q89ZX0	Alpha-1,3-galactosidase B	Up in CEL	0.0158135 72	0.003855 54	4.1015	0.004 9008	0.00019 82
BTH_Q8A600	Choloylglycine hydrolase	Up in INU	0.0247267 91	0.101136 71	4.0902	0.006 2608	0.02506 41
BTH_Q8A6K3	Putative alpha-1,2-mannosidase	Up in CEL	0.0595360 05	0.014759 33	4.0338	0.015 7044	0.00978 27
BTH_Q8A0N6	Uncharacterized protein	Up in CEL	1.0258794	0.260437 42	3.9391	0.357 3189	0.08596 85
BTH_Q8A2R4	SusD homolog	Up in CEL	0.0179690 03	0.004586 38	3.9179	0.003 9135	0.00023 58
BTH_Q8A1U1	Periplasmic beta-glucosidase	Up in INU	0.0058272 44	0.021634 45	3.7126	0.002 7385	0.00756 2
BTH_Q8A2L1	Putative alpha-glucosidase	Up in CEL	0.0187082 35	0.005117 86	3.6555	0.010 2616	0.00204 84

Supplementary table S4-12 (continued).

BTH_Q8A9U 7	ATP synthase subunit alpha	Up in INU	0.0118871 32	0.043172 19	3.6318	0.006 1366	0.02771 27
BTH_Q8A2C 2	Homeodomain-like protein	Up in CEL	0.1096396 39	0.030356 03	3.6118	0.040 1654	0.01878 67
BTH_Q8A1D 9	Uncharacterized protein	Up in CEL	0.0277562 11	0.007807 19	3.5552	0.006 492	0.00472 72
BTH_Q8A8X 0	SusD homolog	Up in CEL	0.8479845 46	0.238569 49	3.5545	0.196 3432	0.06070 45
BTH_Q8A4N 9	SusC homolog	Up in CEL	0.0411975 01	0.011734 8	3.5107	0.012 4804	0.01103 08
BTH_Q8A0H 4	Alpha-2- macroglobulin-like protein	Up in INU	0.0012816 33	0.004484 5	3.4991	0.000 816	0.00213 73
BTH_Q89ZR 8	Pantothenate synthetase	Up in INU	0.0123627 38	0.042823 92	3.464	0.005 8654	0.02218 17
BTH_Q89ZD 0	DUF1080 domain- containing protein	Up in CEL	3.1556916 25	0.917341 43	3.44	0.979 5907	0.44629 11
BTH_Q8A6H 3	7-alpha- hydroxysteroid dehydrogenase	Up in INU	0.0094339 86	0.031696 11	3.3598	0.003 554	0.01786 06
BTH_Q8A10 3	Beta-N- hexosaminidase, glycosyl hydrolase family 20	Up in CEL	0.1000243 88	0.029872 44	3.3484	0.014 7224	0.01492 83
BTH_Q8A0N 4	Endo-beta-N- acetylglucosaminidase F1	Up in CEL	0.6640868 12	0.199200 77	3.3338	0.112 8037	0.07620 3
BTH_Q89ZY 1	APH domain- containing protein	Up in CEL	0.0470768 95	0.014202 38	3.3147	0.017 6698	0.00805 97
BTH_Q8A5I1	GTPase HflX	Up in INU	0.0099594 18	0.032655 82	3.2789	0.004 425	0.01947 99
BTH_Q8A32 6	Putative flagellar motor protein MotB	Up in CEL	0.0842089 23	0.025819 76	3.2614	0.032 9331	0.01467 84
BTH_Q8A7T 4	SusC homolog	Up in INU	0.0113906 77	0.037123 61	3.2591	0.006 067	0.00672 59
BTH_Q8A3Y 7	Putative TonB- dependent receptor	Up in CEL	0.0371692 4	0.011469 64	3.2407	0.012 355	0.00503 37
BTH_Q89ZB 9	Pyridoxal kinase	Up in INU	0.0060575 81	0.019387 15	3.2005	0.000 6027	0.00869 37

Supplementary table S4-12 (continued).

BTH_Q8A5T 1	Probable secreted glycosyl hydrolase	Up in CEL	2.8591723 79	0.897528 2	3.1856	0.824 1209	0.38538 63
BTH_Q8A1I7	Six-hairpin glycosidase	Up in INU	0.0047131 77	0.014849 25	3.1506	0.002 1013	0.00437 08
BTH_Q8A0N 5	Putative patatin-like protein	Up in CEL	1.2105008 82	0.384447 08	3.1487	0.142 2928	0.06008 95
BTH_Q8A0L 6	Uncharacterized protein	Up in CEL	0.2312538 6	0.073958 33	3.1268	0.105 1628	0.02492 62
BTH_Q8A6T 7	Aminopeptidase	Up in CEL	0.2346965 89	0.075933 99	3.0908	0.094 9996	0.05405 49
BTH_Q8A8C 4	Protein GrpE	Up in INU	0.0196014 89	0.059453 05	3.0331	0.015 3845	0.03283 07
BTH_Q8A84 1	Putative dihydropyrimidine dehydrogenase [NADP+]	Up in CEL	0.1101043 22	0.036477 61	3.0184	0.019 5507	0.02097 72
BTH_Q8A2T 2	Uncharacterized protein	Up in INU	0.0243677 07	0.073252 33	3.0061	0.024 825	0.04020 65
BTH_Q8AB0 6	Outer membrane protein oprM	Up in INU	0.0038439 79	0.011448 34	2.9783	0.000 3825	0.00695 26
BTH_Q8A8 W8	Putative patatin-like protein	Up in CEL	0.6804068 11	0.229641 7	2.9629	0.106 3908	0.06986 38
BTH_Q8A6I 4	Aconitate hydratase	Up in INU	0.0056003 29	0.016493 58	2.9451	0.001 1221	0.00705 49
BTH_Q8A7X 1	1-deoxy-D-xylulose-5- phosphate synthase	Up in CEL	0.0141024 02	0.004841 62	2.9127	0.006 194	0.00171 21
BTH_Q8A1U 0	SusD homolog	Up in INU	0.0497654 83	0.141515 88	2.8437	0.012 5674	0.04722 48
BTH_Q8A3V 3	Alpha-galactosidase	Up in INU	0.0049141 36	0.013802 28	2.8087	0.003 4897	0.00149
BTH_Q8A2B 3	Nitroreductase	Up in INU	0.0142167 41	0.039896 98	2.8063	0.005 5457	0.00904 63
BTH_Q8A6B 5	NADP-dependent malate dehydrogenase	Up in INU	0.0551694 22	0.152279 82	2.7602	0.023 8606	0.03801 89
BTH_Q8AB W3	Fido domain- containing protein	Up in INU	0.0041724 87	0.011501 75	2.7566	0.001 47	0.00590 6
BTH_Q8AB W9	Uncharacterized protein	Up in CEL	0.4007969 68	0.146476 79	2.7362	0.154 2607	0.02486 15

Supplementary table S4-12 (continued).

BTH_Q8A9Z 2	NADP(H) oxidoreductase	Up in CEL	0.0633071 85	0.023153 41	2.7342	0.018 1156	0.01654
BTH_Q8A79 9	Beta-galactosidase	Up in INU	0.0037362 53	0.010161 62	2.7197	0.001 6752	0.00244 99
BTH_Q8A7A 6	SusC homolog	Up in CEL	0.0577791 41	0.021271 07	2.7163	0.005 0089	0.00808 84
BTH_Q8A01 9	Uncharacterized protein	Up in CEL	0.0437103 38	0.016130 33	2.7098	0.017 457	0.00082 93
BTH_Q89ZW 7	Putative TPR domain protein	Up in CEL	0.1425113 38	0.052663 74	2.7061	0.027 6466	0.01651 88
BTH_Q8AA8 8	LysM-repeat protein	Up in INU	0.0029875 82	0.007983 46	2.6722	0.000 2973	0.00377 7
BTH_Q8A9M 2	Xylose isomerase	Up in CEL	0.4963064 66	0.187708 17	2.644	0.173 2309	0.06840 38
BTH_Q8A1T 9	SusC homolog	Up in INU	0.0249793 89	0.066017 06	2.6429	0.008 76	0.02349 05
BTH_Q8A23 6	Putative dehydrogenase	Up in CEL	0.0546810 05	0.020722 48	2.6387	0.017 0122	0.01892 3
BTH_Q8A7A 5	SusD homolog	Up in CEL	0.0924108 93	0.035110 05	2.632	0.012 7661	0.01297 34
BTH_Q8A7H 3	Peptidase M1A and M12B	Up in CEL	0.0083477 91	0.003233 27	2.5818	0.004 0647	0.00114 33
BTH_Q8A50 9	OMP_b-brl domain- containing protein	Up in INU	0.2062315 14	0.531675 94	2.5781	0.115 8085	0.18705 19
BTH_Q8AA W1	L-arabinose isomerase	Up in CEL	0.0665424 36	0.025886 15	2.5706	0.014 8268	0.01044 08
BTH_Q8A8X 1	SusC homolog	Up in CEL	0.6213043 61	0.242078 97	2.5665	0.128 8598	0.06456 84
BTH_Q8A6N 9	Putative alanyl dipeptidyl peptidase	Up in CEL	0.0757133 86	0.030080 19	2.5171	0.022 3251	0.00858 19
BTH_Q8A1C 4	Uncharacterized protein	Up in CEL	0.0956724 77	0.038044 4	2.5148	0.035 9652	0.02166 49
BTH_Q8A9N 0	Glutamine synthetase I	Up in CEL	0.0210234 47	0.008466 18	2.4832	0.009 4699	0.00370 52
BTH_Q8A8L 1	Na(+)-translocating NADH-quinone reductase subunit E	Up in CEL	0.0332066 71	0.013617 05	2.4386	0.010 7938	0.00481 52

Supplementary table S4-12 (continued).

BTH_Q8A76 5	2,3- bisphosphoglycerate- dependent phosphoglycerate mutase 2	Up in INU	0.0070834 61	0.017239 96	2.4338	0.000 7048	0.00448 22
BTH_Q8A0E 0	SusC homolog	Up in CEL	0.0100259 71	0.004126 83	2.4295	0.005 0342	0.00292 2
BTH_Q8A2G 9	Acyl-[acyl-carrier- protein]-UDP-N- acetylglucosamine O- acyltransferase	Up in CEL	0.0310782 06	0.012865 36	2.4156	0.008 935	0.00792 72
BTH_Q8A6K 0	Purine nucleoside phosphorylase	Up in INU	0.0222954 71	0.053341 74	2.3925	0.012 112	0.02096 78
BTH_Q8A5H 5	Outer membrane protein Omp121	Up in CEL	0.6464250 17	0.273408 52	2.3643	0.145 1471	0.06102 87
BTH_Q8A1L 2	Putative modulator of DNA gyrase	Up in CEL	0.0174536 87	0.007468 27	2.337	0.008 7222	0.00262 5
BTH_Q8A9E 8	SusC homolog	Up in CEL	0.0259259 98	0.011125 65	2.3303	0.004 8105	0.00440 85
BTH_Q8A8L 4	Phosphoserine aminotransferase	Up in INU	0.6705784 87	1.552475 47	2.3151	0.142 814	0.31145 75
BTH_Q8A6A 6	Meso- diaminopimelate D- dehydrogenase	Up in INU	0.2905435 91	0.663525 63	2.2837	0.044 8308	0.20657 66
BTH_Q8A1G 1	TonB-dependent receptor SusC	Up in CEL	0.0546733 4	0.024021 16	2.276	0.013 6674	0.00522 55
BTH_Q8A76 6	Fructose-bisphosphate aldolase class I	Up in INU	0.0865758 39	0.196279 45	2.2671	0.024 4188	0.04034 52
BTH_Q8AA1 9	OMP_b-brl_2 domain- containing protein	Up in CEL	0.7564280 02	0.333832 05	2.2659	0.212 8648	0.12955 29
BTH_Q8A0N 8	SusC homolog	Up in CEL	0.5405167 73	0.247486 38	2.184	0.096 636	0.03838 77
BTH_Q8A5H 6	Putative lipoprotein	Up in CEL	0.8069937 17	0.373634 97	2.1598	0.141 5843	0.08710 81
BTH_Q89YZ 6	Aminomethyltransfera se	Up in CEL	0.0282468 35	0.013092 58	2.1575	0.007 6714	0.00906 3
BTH_Q8AAF 9	Beta-hexosaminidase	Up in CEL	0.0331109 59	0.015384 97	2.1522	0.004 1971	0.00475 74
BTH_Q8A8X 5	SusC homolog	Up in CEL	0.0366905 56	0.017183 78	2.1352	0.006 8271	0.01453 94

Supplementary table S4-12 (continued).

BTH_Q8A86 2	Serine protease	Up in INU	0.0308688 3	0.065273 35	2.1145	0.008 7855	0.01618 76
BTH_Q8A35 2	SusC homolog	Up in INU	0.0096095 02	0.020012 65	2.0826	0.003 2509	0.00418 06
BTH_Q8A4M 8	Glutaminase	Up in INU	0.0561503 68	0.116555 95	2.0758	0.017 7592	0.04479 02
BTH_Q8A2F 9	SusC homolog	Up in INU	0.0099276 36	0.020240 7	2.0388	0.003 3585	0.00429 63
BTH_Q8AAZ 6	Putative hemin receptor	Up in CEL	0.0851411 52	0.041900 93	2.032	0.014 9548	0.00919 62
BTH_Q8A7Y 3	Bacterial outer membrane protein	Up in INU	0.0865777 26	0.175363 32	2.0255	0.012 989	0.01112 75
BTH_Q8AA1 0	Superoxide dismutase	Up in INU	0.7293242 74	1.447095 83	1.9842	0.144 171	0.47291 03
BTH_Q8A2K 8	SusC homolog	Up in INU	0.0102504 06	0.020260 43	1.9765	0.003 1844	0.00430 05
BTH_Q8A5H 7	Uncharacterized protein	Up in CEL	0.5340114 94	0.272636 5	1.9587	0.106 13	0.09081 49
BTH_Q8A7S 4	Propionyl-CoA carboxylase beta chain	Up in INU	0.0251858	0.049206 24	1.9537	0.006 7126	0.01469 61
BTH_Q8A87 4	Uncharacterized protein	Up in INU	0.0062739 23	0.011709 17	1.8663	0.000 6242	0.00411 57
BTH_Q89ZV 9	Peptidyl-dipeptidase	Up in INU	0.0025496 35	0.004758 44	1.8663	0.000 2537	0.00167 25
BTH_Q8A7J8	Myo-inositol-1- phosphate synthase	Up in INU	0.4936014 75	0.899054 39	1.8214	0.054 7862	0.13710 92
BTH_Q8A29 2	Uncharacterized protein	Up in INU	0.0042843 98	0.007767 44	1.813	0.001 3452	0.00252 64
BTH_Q8A52 9	Aminotransferase	Up in INU	0.0658566 71	0.118826 05	1.8043	0.019 8871	0.03334 25
BTH_Q8A6X 8	Pyruvate-flavodoxin oxidoreductase	Up in INU	0.6367799 54	1.125885 71	1.7681	0.123 8095	0.23288 66
BTH_Q8A61 5	Isocitrate dehydrogenase [NADP]	Up in INU	0.0271277	0.046749 43	1.7233	0.007 463	0.00850 39
BTH_Q89ZR 4	Seryl-tRNA synthetase	Up in INU	0.0758737 02	0.130583 91	1.7211	0.015 4755	0.03357

Supplementary table S4-12 (continued).

BTH_Q8A03 3	Enoyl-[acyl-carrier-protein] reductase [NADH]	Up in INU	0.0650568 4	0.106812 4	1.6418	0.009 8973	0.02716 65
BTH_Q8A48 0	30S ribosomal protein S19	Up in INU	0.2269065 67	0.368695 94	1.6249	0.022 0738	0.09149 76
BTH_Q8A91 0	Putative lipoprotein	Up in INU	0.0066794 62	0.010769 38	1.6123	0.000 6646	0.00380 82
BTH_Q8A84 5	Thiol peroxidase	Up in INU	0.7017305 88	1.113977 91	1.5875	0.176 9289	0.18069 41
BTH_Q89Y W6	Chaperone protein DnaK	Up in INU	0.3141986 92	0.477956 06	1.5212	0.052 9797	0.07523

SCRIPTS**# Statistics on relative abundance plots:**

```

one.way <- aov(Abundance ~ diet, data = Biomass.OrganismName)
Tukey <- TukeyHSD(one.way, ordered = TRUE)
TK <- (Tukey)
TK_data <- as.data.frame(TK[1])
write.csv(TK_data, "TK_data.csv")
#In Excel: Add heading to first column
TK_data <- read.csv("~/TK_data.csv")
if(!require(rcompanion)){install.packages("rcompanion")}
library(rcompanion)
cldList(p-value ~ Comparison, data = TK_data.csv)
Letters <- cldList(p-value ~ Comparison, data = TK_data)
Letters_data <- as.data.frame(Letters)
write.csv(Letters_data, "Letters.csv")

```

Statistics on SIFs – Example for *M. formatexigens*

```

MFR_stats <- pairwise.t.test(MFR$delta, MFR$diet, p.adjust.method = "BH")
MFR_Stats_data <- as.data.frame(MFR_stats[1-12])
write.csv(MFR_Stats_data, "MFR_SIF_STATS.csv")

```

APPENDIX C – OTHER CONTRIBUTIONS TO SCIENCE

1. An experimental and statistical method to compare metagenomics measurement bias between control samples and natural microbial communities

Michael R. McLaren¹, **Angie L. Mordant**¹, David S. Clausen, Amy D. Willis, Manuel Kleiner¹,
Benjamin J. Callahan¹

Author affiliations:

1: North Carolina State University

2: University of Washington

Status: follow-up experiment in progress

Metagenomics-based microbiome measurements have protocol-specific biases that cause some taxa to be overestimated while others go virtually undetected. Industrial, academic, and governmental organizations have developed microbiome control materials for characterizing the bias of a given protocol. But it is unknown whether the bias in such control measurements accurately reflects the bias in measurements of natural samples. We describe a new experimental methodology for such assessments and apply it to test whether bias is consistent between in vitro mixtures of bacterial species and fecal samples from mice in which these same species are living residents. The primary difficulty facing such assessments is that a protocol's bias cannot be measured in natural samples, as their true composition is unknown. We surmount this problem by instead measuring the differential bias between different protocols used on the same samples, and determining how this differential bias varies across sample types. Our method provides a means to validate microbiome calibration controls, an essential step to establishing the feasibility of calibration as a path to truly quantitative microbiome measurements.

**abstract written by Michael R. McLaren*

Contributions: Prepared microbial mock community samples and performed DNA extractions

2. Growth and proteomic shifts associated with vitamin B1 and precursor use in the microalgae *Ostreococcus lucimarinus*

Nathaniel P. Curtis¹, **Angie L. Mordant**¹, Manuel Kleiner¹, Ryan Paerl¹

Author affiliations:

1: North Carolina State University

Status: manuscript in preparation

Thiamin (vitamin B1) is an essential cofactor required for enzymes in multiple metabolic pathways. Despite its vital role, ubiquitous phytoplankton that are key primary producers and prey for zooplankton cannot synthesize B1 de novo and survive on exogenous picomolar supplies of this vitamin in the oligotrophic ocean. Such auxotrophy is diverse, with some picoeukaryotes taking up various pyrimidine and thiazole precursor moieties as well as the intact vitamin. Growth experiments and comparative proteomics were employed to test if *Ostreococcus lucimarinus* CCE9901 exhibits varied responses when growing on exogenous B1 or precursors. CCE9901 reached higher maximum cell yields growing on precursors versus B1, but maximum growth rates were comparable. CCE9901 proteomes of B1- or precursor-grown cells were largely similar (74 proteins of 1543 detected showed significant differences), yet select proteomic responses were linked to multiple cellular functions, including lipid biosynthesis, B-vitamin metabolism, and vacuolar/vesicle storage and transport. Similarly, the abundances of the membrane transporter SSSP (OSTLU_24399) and a putative pyrimidine salvage enzyme similar to YlmB (OSTLU_29911) differed in abundance between treatments. These results show that B1 and its precursors not only influence cell numbers, but they also implicate a role in the formation of metabolites which may be shared within microbial communities or transferred to higher trophic levels. These responses putatively influence community structure, biogeochemistry, and food webs in marine environments.

**abstract written by Nathaniel P. Curtis*

Contributions: Optimized protein extraction methods, extracted and prepared samples for proteomics, analyzed samples by LC-MS/MS, and some data analyses.

3. Optimization of In Vitro Growth Conditions for the Cultivation of Anaerobic Gut Strains

Abigail Korenek¹, **Angie L. Mordant**¹, Manuel Kleiner¹

Author affiliations:

1: North Carolina State University

Status: Experiments in progress

Human gut isolates used to study the microbiome have very specific growth requirements for nutrients, temperature, and oxygen concentration. Extremely oxygen-sensitive species derived from the gut, such as *Faecalibacterium prausnitzii* and *Marvinbryantia formatexigens*, remain understudied due to the challenging nature of their cultivation; thus, it is important to optimize methods for their growth. While anaerobic benches are the gold standard, they are also time, space, and cost-prohibitive to use; therefore, efficient and accessible methods for benchtop cultivation are critical for advancing research into the gut microbiome. We developed a reliable method for the benchtop cultivation of extremely oxygen-sensitive species and assessed the impact of *Escherichia coli* on maintaining strict anaerobiosis.

**abstract written by Abigail Korenek*

Contributions: Conceptualization, training and supervision of Abigail