LIPID DIETS AFFECT THE HOST-MICROBE DYNAMIC IN THE GUT

by

Njiati Abulizi

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

January 2017

© Njiati Abulizi, 2017
The undersigned certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis entitled:

**LIPID DIETS AFFECT THE HOST-MICROBE DYNAMIC IN THE GUT**

submitted by __________ Nijiati Abulizi _______________ in partial fulfilment of the requirements of the degree of _______ Master of Sciences _______.

________________________________________
Dr. Deanna Gibson
**Supervisor, Professor** (please print name and faculty/school above the line)

________________________________________
Dr. Michael Deyholos
**Supervisory Committee Member, Professor** (please print name and faculty/school in the line above)

________________________________________
Dr. John Klironomos
**Supervisory Committee Member, Professor** (please print name and faculty/school in the line above)

________________________________________
Dr. Wesley Zandberg
**University Examiner, Professor** (please print name and faculty/school in the line above)

________________________________________
**External Examiner, Professor** (please print name and university in the line above)

January 2017

(Date Submitted to Grad Studies)

Additional Committee Members include:

________________________________________
(please print name and faculty/school in the line above)

________________________________________
(please print name and faculty/school in the line above)
Abstract

In the gut, there is a complex relationship formed between the host and the bacteria which is further influenced by dietary antigens. The dynamics of this tripartite relationship is for the most part unknown. An imbalance between harmful and protective gut bacteria, termed dysbiosis, has been associated with high fat diets. Dysbiosis has been linked to several inflammatory conditions, such as inflammatory bowel disease and diabetes. Whether different types of fatty acids have similar effects is not fully known. This is important because in Canada, while saturated fatty (SFA) consumption has remained the same, total fat containing n-6 polyunsaturated fatty acid (PUFA) has increased by 54%. To understand the host-microbe dynamic in the gut in response to different lipid diets, we combined 16S rRNA metagenomic sequencing of the microbiome, computational metagenomic prediction of microbiota function and mass spectrometry-based relative quantification of the bacterial and host metaproteome of the colon. We exposed 3 week old C57BL/6 mice to isonitrogenous and isocaloric diets composed of 40% energy from either anhydrous milk fat, corn oil, olive oil or a low fat diet for 5 weeks and then collected their small and large intestinal tissues for analysis. Overall, the corn oil diet rich in n-6 PUFA resulted in a microbiome that showed enhanced virulence associated with increased host inflammation, oxidative stress and increased barrier dysfunction evident by a reduction in protective mucin2 proteins and increase in inflammatory mucin13 proteins. While the milk fat diet rich in SFA resulted in a host-microbe relationship that promoted inflammation, there was also a compensatory protective response evident by the increased tissue repair proteins. In contrast, the olive oil diet rich in monounsaturated fatty acids (MUFA) resulted in increased digestive proteins. We conclude that various lipids uniquely alter the host-microbe dynamic in the gut. Overall, n-6 PUFA increases the potential for pathobiont survival and invasion in an inflamed, oxidized and damaged gut while SFA promotes tissue repair and MUFA enhances metabolism. These results have the potential to guide evidence-based nutrition recommendations to inflammatory bowel disease patients who suffer from malnutrition yet are currently advised to eat low fat diets.
Preface

This research presented in this thesis was divided into two experimental sections. I analyzed a metagenomics sequencing data generated by the genomic sequencing center in Vancouver from the amplicon libraries created in our laboratory by previous trainees (Dr. Carol Chan and Kirsty Brown) and wrote and edited a manuscript on the results. Short chain fatty acid analysis was performed by Sandeep Gill under the guidance of Dr. Sanjoy Ghosh. I was fully involved in conducting experiments, analyzing the data, writing and editing the manuscript of the second section where the proteomics experiment was performed with help from Candice Quin and then sequenced at the Proteomics Center in Victoria. All animal experiments were approved by the UBC Animal Care Committee (certificate # A11-0367); Biosafety (certificate # B13-0122) and performed at the Centre for Disease Modeling (CDM) in Vancouver, BC Canada by our laboratory technician Ben Dai with the supervision of Dr. Deanna Gibson.

The two sections of this work have been presented as two separate poster presentations entitled “The effects of lipid diets on the gut microbiome” and “The effects of lipid diets on the gut microbiome and the host” at the biology graduate symposium, UBC Okanagan over the last two years. The work presented in this thesis is submitted for publication in Cell Host & Microbe.
# Table of Contents

Examination Committee ........................................................................................................ ii
Abstract ................................................................................................................................... iii
Preface ...................................................................................................................................... iv
Table of Contents .................................................................................................................... v
List of Tables .......................................................................................................................... viii
List of Figures ........................................................................................................................ ix
List of Symbols, Abbreviations .............................................................................................. x
Acknowledgements ................................................................................................................ xiv

## Chapter 1: Introduction ..................................................................................................... 1

1.1 Host-microbe interaction in the gut ............................................................................. 1
1.2 Anatomy of intestine and microbial interaction ......................................................... 3

### 1.2.1 The mucosal structure and the epithelial cells ................................................. 3

1.3 Types of lipid diets and its effect on the host and microbiome .............................. 5

### 1.3.1 n-6 polyunsaturated fatty acids ................................................................... 6

### 1.3.2 Monounsaturated fatty acids ......................................................................... 7

### 1.3.3 Saturated fatty acids ....................................................................................... 8

1.4 Tools to study the microbiome ............................................................................... 8

### 1.4.1 16S RNA based metagenomics and data analysis ........................................ 8

### 1.4.2 Shotgun metagenomics ..................................................................................... 10

### 1.4.3 Prediction based metagenomics and data analysis ......................................... 10

### 1.4.4 Metaproteomics ............................................................................................... 11

### 1.4.5 Gel-free mass spectrometry-based metaproteomics .................................... 12

### 1.4.6 Label-based (mass spectrometry) metaproteomics ........................................ 12

### 1.4.7 iTRAQ metaproteomics and data analysis ...................................................... 13

## Chapter 2: Overall rationale, hypothesis and objectives ............................................. 15

Objective 1 ............................................................................................................................ 15
Objective 2 ............................................................................................................................ 15
Chapter 3: Experimental Methods

3.1 Animals

3.2 Diets (Performed by Ben Dai)

3.3 Tissue collections (Performed by Dr. Carol Chan)

3.4 Bacterial DNA extraction and amplicon preparation (Performed by Dr. Carol Chan and Kirsty Brown)

3.5 High throughput amplicon based metagenomics sequencing analysis

3.5.1 Alpha diversity and beta diversity

3.5.2 Compositional analysis

3.5.3 Computer prediction based bacterial metagenomics

3.6 High throughput mass spectrometry based metaproteomics

3.6.1 Protein extraction (Performed with Candice Quin)

3.6.2 iTRAQ experiment (Performed at The University of Victoria Proteome Center)

3.6.3 iTRAQ mass spectrometry data analysis

Chapter 4: Results

4.1 Objective 1: Effects of the lipid diets on the metagenome of the microbiome

4.1.1 The type of lipid diet affects the microbial phyla ratio

4.1.2 The type of lipid diet differentially affects the alpha diversity of microbial communities both in the colon and in the ileum

4.1.3 The type of lipid diet differentially affects the beta diversity of microbial communities in the colon but not in the ileum

4.1.4 The type of lipid diet promotes the abundance of different bacterial taxa

4.1.5 Lipid diets confer core functionality to each microbial community

4.2 Objective 2: Gut proteomics

4.2.1 Microbial functions were differentially affected by dietary lipid

4.2.2 Effects of lipid diets on the metaproteome of the host tissue

4.2.3 Corn oil exposure impaired the production of proteins involved in intestinal integrity

4.2.4 Corn oil exposure increased cell turnover events

4.2.5 Milk fat diet was associated with higher collagen and tissue repair activity
4.2.6 Olive oil consumption was associated with higher metabolism and energy production ................................................................. 47
4.2.7 Corn oil diet altered epithelial adherens junctions......................... 49

Chapter 5: Discussion ........................................................................ 52
  5.1 Microbial metagenomics............................................................... 52
  5.2 Microbial proteome.................................................................... 54
  5.3 Host proteome............................................................................ 57

Chapter 6: Conclusion ....................................................................... 60
  6.1 Limitations of this study ............................................................. 60
  6.2 Future work................................................................................ 61

References ............................................................................................ 62

Appendices ........................................................................................... 70
  Appendix A. Statistical analysis of the phyla ratio data ...................... 70
  Appendix B. Statistical analysis of the Gram-positive and Gram negative bacteria abundance comparison......................................................... 71
  Appendix C. Statistical analysis of alpha and beta diversity .................. 72
List of Tables

Table 1. Mucosal and submucosal protein identification summary using both bacterial and mouse proteome databases ................................................................. 33
Table 2. Summary of the identified bacterial proteome (Firmicutes and Bacteroidetes combined) in the mucosal and submucosal samples .................................................. 35
Table 3. Proteins that contributed to the epithelial remodeling processes ................................................. 51
Table A.1. Kruskal-Wallis rank sum test p-values computed by comparing the Firmicutes to Bacteroidetes ratio across lipid diets in the colon and the ileum ..................... 70
Table A.2. Firmicutes to Bacteroidetes ratio pairwise group comparison p-values ....................... 70
Table B.1. Kruskal-Wallis rank sum test p-values computed by comparing the Gram-positive and Gram-negative bacteria abundance across lipid diets both in the colon and the ileum ................................................................. 71
Table B.2. Gram-positive and Gram-negative bacteria abundance pairwise group comparison p-values ................................................................. 71
Table C.1. Kruskal-Wallis rank sum test p-values computed by comparing the three alpha diversity index values across lipid diets both in the colon and the ileum .......... 72
Table C.2. Alpha diversity indexes pairwise group comparison p-values ............................................. 72
Table C.3. PERMOANOVA statistics on weighted and unweighted UniFrac distances for the colon and the ileum ........................................................................... 72
## List of Figures

Figure 1. Phyla abundance and ratio plots among different lipid diets and anatomical regions. ................................................................. 23

Figure 2. Relative abundance of Gram-negative and Gram-positive bacteria across lipid diets. .................................................................. 24

Figure 3. Alpha diversity of the microbial communities across lipid diets. ................................................................. 25

Figure 4. The effect of lipid diets on the beta diversity of the gut microbiota. ................................................................. 27

Figure 5. The effect of lipid diets on the gut microbial taxa. ................................................................................................. 29

Figure 6. The effect of lipid diets on the predicted microbial functions. ............................................................................... 31

Figure 7. Protein level fold change comparison plots of the bacterial proteins across lipid diets. .................................................................. 36

Figure 8. Gene ontology term pie chart summary for the mucosal and submucosal proteome. ................................................................. 38

Figure 9. Protein level fold change comparison plots of the mucosal proteins across lipid diets. .................................................................. 39

Figure 10. Protein level fold change comparison plots of the mucosal proteins among lipid diets. .................................................................. 41

Figure 11. Protein level fold change comparison plots of the submucosal proteins among lipid diets. ................................................................. 43

Figure 12. Protein level fold change comparison plots of the mucosal proteins among lipid diets. ................................................................. 44

Figure 13. Protein level fold change comparison plots of the submucosal proteins among lipid diets. ................................................................. 46

Figure 14. Protein level fold change comparison plots of the mucosal proteins among lipid diets. ................................................................. 47

Figure 15. Protein level fold change comparison plots of the submucosal proteins among lipid diets. ................................................................. 48

Figure 16. Effects of lipid diets on the mucosal and submucosal proteome shown by the Ingenuity pathway comparative analysis. ................................................................. 50
List of Symbols, Abbreviations

AA Arachidonic acid
AF Animal fat
AMF Anhydrous milk fat
BCA Bicinchoninic acid
Bp Base pair
CD Crohn`s Disease
COG Clusters of Orthologous Groups
COX Cyclooxygenases
CSS Cumulative sum scaling
CRP C reactive protein
CHD Coronary heart disease
DHA Docosahexaenoic acid
DSS Dextran sodium sulfate
DGLA Dihomo-γ –linolenic acid
EDTA Ethylenediaminetetraacetic acid
EPA Eicosapentaenoic acid
FA Fatty acids
FABP Fatty acid binding protein
GI Gastrointestinal
HEPES 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid
HDL High density lipoprotein
HFD(s) High fat diet(s)
HPLC High performance liquid chromatography
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAnN</td>
<td>The HMP (Human microbiome project) Unified metabolic analysis network</td>
</tr>
<tr>
<td>IAP</td>
<td>Intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>IECs</td>
<td>Intestinal epithelial cells</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity pathway analysis</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>Isobaric tags for relative and absolute quantitation</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>KO</td>
<td>Knock out</td>
</tr>
<tr>
<td>LA</td>
<td>Linoleic acid</td>
</tr>
<tr>
<td>LEfSe</td>
<td>Linear discriminant analysis Effect Size</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LDL</td>
<td>Low density lipoprotein</td>
</tr>
<tr>
<td>LOX</td>
<td>Lipoxygenase</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MAMPS</td>
<td>Microbe-associated molecular patterns</td>
</tr>
<tr>
<td>MUC2/MUC13</td>
<td>Mucin2/Mucin13</td>
</tr>
<tr>
<td>MMTS</td>
<td>Methyl methanethiosulfonate</td>
</tr>
<tr>
<td>MUFA</td>
<td>Monounsaturated fatty acids</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>NCBI</td>
<td>National center for biotechnology information</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear factor kappa-light-chain-enhancer of activated B cells</td>
</tr>
<tr>
<td>OA</td>
<td>Oleic acid</td>
</tr>
<tr>
<td>OTU</td>
<td>Operational taxonomic unit</td>
</tr>
<tr>
<td>PA</td>
<td>Palmitic acid</td>
</tr>
<tr>
<td>PCoA</td>
<td>Principle coordinate analysis</td>
</tr>
<tr>
<td>PGE2</td>
<td>Prostaglandin E2</td>
</tr>
<tr>
<td>PICRUSt</td>
<td>Phylogenetic Investigation of Communities by Reconstruction of Unobserved States</td>
</tr>
<tr>
<td>PRRs</td>
<td>Pattern recognition receptors</td>
</tr>
<tr>
<td>PUFA</td>
<td>Polyunsaturated fatty acids</td>
</tr>
<tr>
<td>QIIME</td>
<td>Quantitative Insights into Microbial Ecology</td>
</tr>
<tr>
<td>SAP</td>
<td>Serum amyloid P component</td>
</tr>
<tr>
<td>SCFA</td>
<td>Short chain fatty acids</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SFA</td>
<td>Saturated fatty acids</td>
</tr>
<tr>
<td>SiRNA</td>
<td>Small interfering Ribonucleic Acid</td>
</tr>
<tr>
<td>T2D</td>
<td>Type 2 diabetes</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid cycle</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris (2-carboxyethyl) phosphine hydrochloride</td>
</tr>
<tr>
<td>TEAB</td>
<td>Triethylammonium bicarbonate</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptors</td>
</tr>
<tr>
<td>TMT</td>
<td>Tandem mass tags</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>UC</td>
<td>Ulcerative colitis</td>
</tr>
<tr>
<td>UniFrac</td>
<td>A distance metric</td>
</tr>
<tr>
<td>2D-DIGE</td>
<td>2 dimensional difference gel electrophoresis</td>
</tr>
<tr>
<td>2D-Gel</td>
<td>2 dimensional gel</td>
</tr>
<tr>
<td>16S RNA</td>
<td>16S Ribosomal RNA</td>
</tr>
</tbody>
</table>
Acknowledgements

I owe particular thanks to Dr. Deanna Gibson, whose guidance and investment helped me complete this work and made other achievements over the last two years possible. I also owe thanks to my committee members, Dr. John Klironomos, and Dr. Michael Deyholos who gave me better insight into research by providing their expertise and valuable time and energy. I thank my parents and my partner, who have supported me throughout my years of education, both emotionally and financially. Finally, offer my enduring gratitude to the faculty, staff and my fellow students at the UBCO, who have been of tremendous help and inspiration to continue my work in this field.

I acknowledge funding support from: UBCO Graduate Fellowship Awards, Golden Key Graduate Scholarship and British Columbia Proteomics Network Training Fund over the last two years. This work was supported from operating grants from Crohn’s and colitis Canada and NSERC to my supervisor, Dr. Deanna Gibson.
Chapter 1: Introduction

1.1 Host-microbe interaction in the gut

The diverse community of microorganisms, including Eukarya (1), Archaea (2), bacteria and viruses (3) that inhabit our body cavities are collectively referred to as microbiota. The highly complex bacterial community in the human gastrointestinal (GI) tract is composed of two major phyla: Firmicutes and Bacteroidetes. In addition to these two major phyla, there are members of Proteobacteria, Actinobacteria, Fusobacteria, Spirochaetes, Verrucomicrobia and Lentisphaerae (4). Bacteroidetes (including Bacteroides) and Firmicutes (including Clostridium, Bacillus and Lactobacillus) which make up about 90% of the bacteria together whereas the Proteobacteria (Escherichia) and Actinobacteria (Bifidobacterium) are typically under 5% of the total population and the remaining phyla are present in minor abundance (5).

The bacterial community is not evenly distributed along the GI tract. They are less abundant in the stomach ($10^3$-$10^4$ bacteria per gram of content) (6, 7) with low bacterial diversity (8), but more abundant in jejunum ($10^4$ bacteria per gram of content) and in ileum ($10^8$ bacteria per gram of content). Bacterial numbers can increase up to $10^8$ bacteria per gram of content in the large intestine (7, 9). Moreover, bacterial composition in the fecal samples can differ significantly from that of the colonic mucosa (10, 11). Microbiota in the large intestine appear to be relatively stable over time (12, 13) and more diverse compared to microbiota in the small intestine (10).

These intestinal microbiota and their host are in a mutualistic and homeostatic relationship (7, 14, 15). The microbial community gains nutrients in the intestinal tract from the host diet and they are capable of fermenting molecules such as cellulose, pectin, glucans that the human enzymes cannot fully digest (16). The major microbial fermentation products include short chain fatty acids (SCFA). Microbiota produce SCFAs as their necessary waste products to balance the increased reducing equivalents in the gut (17). The host gains these fermented products (10% of our energy requirement) such as acetate (C2), propionate (C3) and butyrate (C4) (18), as well as vitamin K and B12 from the microbiota (6, 7, 15). SCFA have been shown to exert beneficial effects on the health, such as body weight, insulin sensitivity and glucose homeostasis. Moreover, consuming higher fiber is associated with
reduced risk of inflammatory bowel disease, colon cancer, diabetes and cardiovascular disease (19). Among the dominant bacterial phyla, bacteria from the Bacteroidetes phyla produce high amounts of SCFAs including acetate and propionate, while bacteria from the Firmicutes phyla produce high amounts of butyrate (major energy source for colonocytes) (20).

Besides being in a mutualistic relationship with the host, the human intestinal microbiota is also associated with human diseases. The balance between the gut microbiota and the host immune system drives a healthy state. Healthy microbiota continually stimulates the host immune system for its better development (15). This continued stimulation is achieved directly through host receptors such as Microbe-associated molecular patterns (MAMPs), Pattern recognition receptors (PRRs) or indirectly by production of bacterial metabolites (14, 15). Dysbiosis occurs when there is an imbalance between protective and harmful bacteria resulting in a deficient host-microbial mutualism. Symbiotic and protective bacteria are adapted to harvest nutrients from the host diet such as Bacteroides thetaiotaomicron which has a variety of saccharolytic enzymes allowing them to breakdown complex polysaccharides. However, pathogenic bacteria such as Salmonella and Shigella are poor competitors with symbionts for the nutrients from the host diet because they have limited capacity for hydrolysis and saccharide uptake. Thus pathogenic bacteria evolved to invade host tissue and use them as a nutrient source (14). Pathogenic bacteria can produce proteins that bind to host cells to illicit different host responses and cause diseases. The mechanism that bacteria utilize include the production of capsules (high molecular weight polysaccharides), proteinaceous toxins (exotoxins), including A-B toxins, proteolytic toxins and pore forming toxins, adhesins and proteins that are involved in invasions (21).

Several human disease states are associated with dysbiosis such as inflammatory bowel disease (IBD), irritable bowel syndrome, and colorectal cancer (22). It has been observed that changes in bacterial metabolic pathway expression and reduced bacterial diversity are correlated with obesity (23). Specifically, studies found decreased Bacteroidetes phyla and increased Firmicutes (23, 24). This shift is further associated with obesity and increased low grade inflammation. The high capacity to harvest energy is related to the increased Firmicutes phyla and their capacity to break down polysaccharides not digestible by the host. Studies use mice that are reared in the absence of microbes, called germ-free
mice, also support the role of microbiota in obesity development whereby germ-free mice orally gavaged with stool from an obese animal will be obese in contrast to the being lean if gavaged with stool from a lean animal (25). This suggests that obesity phenotype is not only caused by dietary calories, but microbiota is also involved.

Microbiota functions have also been linked to type 2 diabetes (T2D) (26). T2D patient’s microbiota is reported to have increased xenobiotic degradation, branched chain amino acid transport, membrane transport of carbohydrates and sulphate reduction. However, the levels of butyrate biosynthesis, bacterial chemotaxis, flagellar assembly, vitamin and cofactor metabolism are found to be reduced in the T2D patients’ microbiota compared to controls (27). It has also been shown that bacterial defence against oxidative stress was also stimulated in these patients. Therefore, there is a strong connection between microbial changes and corresponding physiological changes. Thus, understanding the factors that can contribute to the intestinal microbial community composition is essential to prevent and treat human diseases.

1.2 Anatomy of intestine and microbial interaction

The GI tract is a complex organ system that is made up of the mouth, esophagus, stomach, small intestine, large intestine, rectum and anus. Histologically, the GI tract can be divided into the mucosa, submucosa, muscularis, and the serosa. Mucosa is the first layer of the GI tract surrounding the lumen (location of food antigens and microbiota) and it can further be subdivided into epithelium, lamina propria, and muscularis mucosa. Thus, the mucosa presents the first line of barrier against the external milieu. Submucosa, on the other hand, is a dense layer of connective tissue that holds the enteric nervous system, the blood vessels and lymphatic system.

1.2.1 The mucosal structure and the epithelial cells

The intestinal epithelium is the largest mucosal surface made up of single layer of columnar cells with a surface area of ~400 m² (28). The mucosal epithelial cells form barrier surfaces that separate the external environment from the internal host cells. Mucous is present between the epithelial surfaces and the external environment and forms a continuous layer (29). Large glycoproteins called mucins make up the mucous layer, which further
contains other intestinal proteins such as immunoglobulins, growth factors and defensins. It is estimated in humans that 10 L of mucous is secreted per day (29).

The mucous layer can act as lubricant and allows easier passage of food in the intestine. It can function as a selective barrier permitting low molecular weight molecules to reach the epithelial cells. The mucous layer also protects the epithelial cells from physical, chemical (drugs, toxins and heavy metals) and biological (bacteria, viruses and parasites) damage. Moreover, mucus can serve as growth substrates, adhesion sites and protection for microbial cells (30, 31).

Two different sub-layers of mucus are present in the colon (32). While the outer layer is non adherent and mostly soluble and can act as lubricant, the inner non soluble layer adheres to the epithelium and acts as selective barrier (33). The loosely attached outer layer of mucus contains high concentrations of bacteria, whereas the firmly attached inner layer contains hardly any bacteria. In both layers, MUC2 (mucin 2) is the main structural glycoprotein, but other specialized MUC proteins are present in each layer (34, 35). The protective role of the dominant mucin MUC2 was demonstrated in genetically engineered mice. These mice spontaneously developed intestinal inflammation in the absence of the MUC2 glycoproteins (36, 37).

Based on the mucin structures, mucins are divided into membrane-bound and secretory mucins. Membrane bound mucins (e.g. MUC13) are synthesized by the epithelial cells, whereas secretory mucins (e.g. MUC2) (high molecular weight) are secreted by specialized cells such as goblet cells and can form a viscous gel. The disulfide bridges are formed by cysteine rich sequences of secretory mucins to make filamentous multimers (38, 39). However, membrane bound mucin proteins do not form covalent multimers but they are thought to act as cell surface receptors which sense external conditions and translate sensed information into cellular responses such as apoptosis, cellular differentiation or cell proliferation (40). Recently it has been shown that the epithelial inflammation is differentially regulated by the MUC1 and MUC13 proteins in inflammatory conditions. Using siRNA silencing techniques, the authors showed that the chemokine secretion increased in the absence of MUC1, conversely the chemokine secretion decreased in the absence of MUC13 when stimulated with tumor necrosis factor alpha (TNF-alpha) (41).
Mucins have extremely diverse carbohydrate structures, thus they can provide binding sites for bacteria as adhesion sites. For example, *Lactobacillus johnsonni* protein GroEL aggregates *Helicobacter pylori* by biding to mucins, facilitating pathogen clearance (42). Pathogens use mucin carbohydrate structure as attachment sites for access to the epithelial cells (43). Moreover, GI tract pathologies such as Crohn’s disease (CD) and ulcerative colitis (UC) feature altered mucous layer integrity. In UC, a thinner mucous layer was found, whereas in CD the mucus thickness was higher than normal (44). In another study of intestinal infection induced by intestinal nematode *Nippostrongylus brasiliensis*, not only is MUC2 production increased but the mucin carbohydrate structure also changed during infection (45, 46).

The epithelial layer is the next barrier after the secretory and membrane-bound mucosa. The most of the cells surrounding the intestinal lumen are absorptive enterocytes, and other specialized intestinal epithelial cells (IEC) are present for a diversity of functions. These cells include enteroendocrine cells (regulates digestive functions through hormones), goblet cells (secretes mucins) and Paneth cells (secretes antimicrobial peptides). These cells play an important role in separating microbial contact with the underlying epithelial surface and immune cells by forming the intestinal barrier adjoined with numerous adherens and junctional proteins. Thus these specialized cells provide diverse functions to protect the host from infectious or inflammatory environmental stimuli (28).

### 1.3 Types of lipid diets and its effect on the host and microbiome

Diet is a recognized factor that can influence both the host and the microbiota. High fat diets are implicated in the induction of bacterial imbalance in the gut. For example, it has been shown that high fat diets can increase Firmicutes phyla and reduce Bacteroidetes phyla in the gut and this change has been shown to have causal relationship to the obesity phenotype (47). This was demonstrated in an experiment that the total body fat increased in the germ free lean mice when they are colonized by obese mice microbiota (48).

Lipids change microbial community composition by direct antibiotic-like actions, or indirectly by inducing bile salts during the degradation of dietary fats after their consumption. Bile salts are produced by the liver and function during digestion as follows: lipids are first broken into droplets by the mixing action of the stomach. Next, pancreatic
lipase hydrolyzes triacylglycerol to fatty acids and 2-monoacylglycerol in the intestine. Bile salts (amphipathic molecules) then emulsify the hydrolyzed products and form micelles. Finally the small intestine absorbs these micelles (49).

The interactions between the bacteria, the lipids, and the bile salts have been investigated in some studies. In mice, the transport of the dietary fat, emulsification and absorption and peroxidation are affected by the intestinal bacteria (50). Some bacteria have bile salt hydrolysis activity, thus they can survive the harsh environment created by these salts (51). Lithocholic acid and deoxycholic acid, also known as secondary bile acids, can be produced by bacterial modification (52). These biologically highly active molecules are associated with the development of gallstones and colorectal cancer (53). In CD patients, the bile resistance capacity in Firmicutes was shown to be reduced significantly based on the metagenome data (51).

However, studies that focused on a high fat diet did not differentiate the types of dietary fats in their experiments. Considering that different types of fatty acids can have different effects on physiological responses, it seems likely that they might also have different effects on dysbiosis. In our lab, we showed that n-6 PUFA exacerbated murine colitis through changes in the microbiome (54, 55). Moreover, in another study in mice, saturated, but not polyunsaturated, dietary fat promoted the growth of intestinal deltaproteobacteria that elicited a proinflammatory response and increased incidence of colitis in genetically susceptible mice (56). While various lipids may induce changes to the microbiota, a full characterization of the lipid-induced changes in the microbiome (the microbes and their associated functions) has not been elucidated.

1.3.1 n-6 polyunsaturated fatty acids

Linoleic acid is the most abundant n-6 polyunsaturated fatty acid (n-6 PUFA) in common vegetable seeds oil including corn oil, soybean and safflower oil. Linoleic acid (LA; 18:2n6), with two double bonds, is an important component of plant and animal cell membranes. n-6 PUFA help maintain membrane integrity, fluidity, and correct membrane protein function. Humans cannot synthesize n-6 PUFA and they must be obtained from food sources, therefore, they are called essential fatty acids (57). PUFA are metabolized into more highly unsaturated molecules to carry out other functions. For example, LA is first
converted into γ-linolenic acid (GLA, 18:3n-6) and dihomo-γ-linolenic acid (DGLA, 20:3n-6), then these molecules are further converted into arachidonic acid (AA) (20:4n-6). Cyclooxygenases (COX) and lipoxygenases (LOX) can convert AA to potent forms of the prostaglandins, thromboxanes, and the leukotrienes. These short lived molecules are called “eicosanoids” and they are involved in inflammatory conditions such as atherosclerosis, obesity, and IBD (58).

There is conflicting evidence with regards to the effects of consuming n-6 PUFA on health. A meta-analysis summarized randomized controlled trials and found that n-6 PUFA reduced coronary heart disease (59). However, some studies showed that n-6 PUFA increased coronary heart disease risk (60). Furthermore, other systemic reviews found no association between n-6 PUFA with inflammatory markers or coronary risks (61). These studies focused on the cardiovascular diseases, however, the effects of n-6 PUFA on the intestinal diseases are not clear.

Recently, our lab has shown that corn oil rich in n-6 fatty acids induced exacerbated colitis during acute colitis (62). Our lab showed that corn oil diets fed to mice had augmented colitis evident from the increased immune cell infiltration, intestinal damage and prostaglandin E2 expression found in the gut tissues as well as Gram-negative bacteria translocated across the intestinal mucosae (63). In a separate study, our lab observed bacterial overgrowth, increased macrophage and neutrophil infiltration and weight gain associated with dysbiosis in the corn oil diets (54). Moreover, when comparing several high fat diets, bacteria infiltrated across the intestinal barrier only in the corn oil diet (54). However, the underlying mechanism for the increased permeability of the intestine under the n-6 PUFA rich corn oil diets is not well characterized.

1.3.2 Monounsaturated fatty acids

Oleic acid, an 18:1 monounsaturated fatty acid (MUFA), is the main component of olive oil, a major dietary lipid in the Mediterranean diet. It is distinct from other fatty acids because it only has one double bond in its structure. MUFA consumption is associated with several health benefits. Epidemiological evidence shows that MUFAs have beneficial effects on the risk of heart disease. A systematic review and meta-analysis involving 13800 patients and 23340 controls in 19 observational studies indicated that olive oil consumption is
inversely correlated with cancer prevalence (64). Moreover, Mediterranean diet intake is associated with reduced diabetes, metabolic syndrome and related micro-inflammation (65), and IBD (66). Our lab has shown that olive oil is protective in the gut during injury due to the induction of intestinal alkaline phosphatase which detoxifies lipopolysaccharides (LPS) from Gram-negative bacteria (62). Overall, these studies suggest that MUFAs may be beneficial during inflammatory diseases but studies are needed to establish its effects on a molecular level.

1.3.3 Saturated fatty acids

Saturated fatty acids (SFA), which have no double bond, are found in milk and dairy products. Saturated fatty acids gained a “bad fat” reputation since Ancel Keys’s Seven Country study (67), which correlated total cholesterol with coronary heart disease. Moreover, a positive correlation was made between saturated fat consumption and death rates. Since then dietary guidelines advised to reduce saturated fat consumption to reduce cardiovascular diseases. While the global consumption of SFA has been in line with recommended low consumption levels (68), inflammatory diseases are on the rise and cardiovascular diseases are still the number one cause of death. Indeed, growing amount of evidence is suggesting that dietary SFA is not associated with adverse health effects, or is weakly associated when other major contributing factors may not be taken into consideration. (69). Furthermore, our lab has shown that during intestinal inflammation, milk fat promoted beneficial physiological effects compared to a corn oil diet (62). However, the interaction among SFA, microbiota and the host responses are not well characterized.

1.4 Tools to study the microbiome

1.4.1 16S RNA based metagenomics and data analysis

Next generation metagenomics consists of two general approaches. The first method is 16S ribosomal RNA (rRNA) gene sequencing. The second metagenomic approach identifies community taxa by shotgun sequencing of the DNA and comparing it to reference genome databases. In the 16S RNA gene sequencing approach, DNA is first extracted from samples containing multiple microbial species. Then the 16S rRNA genes of the microbial DNA is amplified and sequenced to determine the microbial taxa present in the samples (70).
Two major sequencing methods include pyrosequencing and Illumina sequencing, each having their own advantages and disadvantages. In pyrosequencing, light is produced when a pyrophosphate that is released upon the addition of nucleotides by DNA polymerase participates in the release of photons in an enzyme catalyzed reaction. The amount of light produced is proportional to the numbers of nucleotides. Specifically, fragmented DNA is first ligated to adaptors and these ligation products are annealed to beads that have complimentary sequences to the adapter sequences, which allows unique DNA fragment and bead linkage. DNA fragments are amplified by emulsion polymerase chain reaction (PCR) into millions of copies on the surface of each bead. These beads are then put into a well of picotitre plate (only one bead can fit into each well). Solutions containing nucleotides are added to the plate together with enzymes while being imaged. In pyrosequencing, consecutive nucleotides (AAAAA), also known as homopolymers, are susceptible to inaccurate sequencing results due to deletions and insertions. However, this technique generated sequences have the longest sequence read length out of all second generation sequencing methods. In the Illumina technique, the adaptor ligated DNA fragments are hybridized on a flow cell where a “bridge” amplification reaction (localized PCR reaction) is used on the surface of the flow cell to amplify and generate clusters. To add only one nucleotide at a time to the growing sequence, a mixture of 3’OH blocked nucleotides with fluorescent labels as well as the primers are added to the flow cell. A microscope is placed on top of the flow cell, so when light is emitted, the detector records the wavelength and intensity corresponding each added nucleotide. Billions of reads are generated in the Illumina sequencing. The major drawback of this technique is substitution but not deletions and insertions. Therefore, although Illumina is the most frequently used technique with less deletion and insertion errors and pyrosequencing is no longer supported and is not used any more, pyrosequencing provides a larger read length for increased sequence identification.

After sequencing, software programs, such as Quantitative Insights Into Microbial Ecology (QIIME) are used for subsequent analysis. Sequences that are highly similar are grouped into Operational Taxonomic Units (OTUs), and the OTU sequences are then searched for 16S databases such as Greengenes to determine their identities. The community can then be characterized by the presence of different OTUs, or the relative abundances of
OTUs (72). Finally, an association is made between microbial community changes and corresponding phenotypes or disease states.

1.4.2 Shotguns metagenomics

The shotgun metagenomics approach is more expensive than the 16S rRNA region sequencing but provides higher sequence coverage targeting all regions of the genome, not just the 16S rRNA region of the genome in the amplicon sequencing (72). The sequences can be compared to the functional databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) or the Clusters of Orthologous Groups (COG) to determine the functional capabilities of the bacterial community. Therefore, relative abundance of functional pathways can be compared different conditions.

Despite a deeper sequencing depth than 16S rRNA region sequencing, shotgun metagenomic experiments only discover the functional potential of the microbial populations and they are not a good representation of actual expressed (transcribed and translated) functions. Thus, they are not as popular as conducting 16S rRNA amplicon based metagenomic experiments. However, software programs called Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (73) can reconstruct a metagenomic profile of the microbial population from 16S rRNA data. This software package predicts metagenome functional content from marker genes such as 16S rRNA genes. Overall, this approach is more economical and time saving than shotgun genome sequencing as it circumvents the need to sequence the entire genome by predicting the whole genome by only using the 16S region of the genome.

1.4.3 Prediction based metagenomics and data analysis

Langille et al. (2013) demonstrated that PICRUSt outperforms shallow metagenomic sequencing: PICRUSt predictions achieved a maximum accuracy with only 105 16S sequences assigned to OTUs per sample (Spearman r= 0.82, P<0.001); This same accuracy is achieved by at least 15,000 annotated or 72,650 raw read metagenomic sequences per sample. Therefore, a typical 16S rRNA gene sequencing is adequate to predict metagenomics functions. However, this study also showed the following weaknesses in PICRUSt: 1) viral or eukaryotic metagenomes are not predicted, because PICRUSt uses bacterial 16S rRNA data as input data, 2) if the primer used does not amplify an organism, then that organism’s
contribution to the metagenome is not included/predicted, 3) only the already known gene families included in KEGG database are predicted and 4) accuracy for any sample depends on the availability of appropriate reference databases.

Metagenomics approaches may reveal species (identity and abundance) or genes that may be important in certain diseases. However, they do not provide evidence for the actual involvement or function of these species or genes (gene transcription and translation). For example, in one comprehensive study of the fecal intestinal metaproteome of humans, the identified peptides were taxonomically classified and the results were compared to phylogenetic analysis based on 16S rRNA sequence diversity (74). It appeared that proteins belonging to Actinobacteria were overrepresented, suggesting high bacteria activity of this phylum, including *Bifidobacteria*. This supports the idea that species abundance alone cannot be used as an indicator of microbiome functions, since Actinobacteria exists in lower abundance in the gut. In another study, 2D-DIGE-based (difference gel electrophoresis) technique discovered an imbalance of intestinal bacterial functions in CD patients (75). Many proteins from the *Bacteroides* sp. were overexpressed in this study, whereas Firmicutes and Prevotella proteins were under-represented. The overrepresented proteins linked to functions enabling the pathobionts to colonize the mucous layer and invade the host barrier. More importantly, some of the TonB dependent receptors and other uncharacterized proteins on the cell surface of the *Bacteroides* sp. increased in CD patients independent of the abundance of the corresponding bacterial population (75). This also suggests that changes in microbial functionality cannot be directly inferred from the bacterial compositional changes. To understand the ecological changes in terms of function, identifying actively involved proteins would be more targeted and satisfactory. Quantitative metaproteomics and transcriptomics are good solutions for this problem. However, while proteomic experiments show the final translated products, transcriptomic experiments show the expression levels of the transcripts that are not always well correlated to the translated proteome due to post-transcriptional regulation.

1.4.4 Metaproteomics

Quantitative metaproteomics identifies and quantifies all the proteins in a sample using mass spectrometry methods. Then, microbial functions can be directly studied from the pool of determined proteins in a sample. This is because proteins determine cellular structure
and biological function. Quantitative proteomic experiments are divided into gel-based (2D gels) methods and gel-free mass spectrometry based methods. In 2D gel-based proteomics, proteins that are not detected are never analyzed thus they are completely ignored. The yield of 2D electrophoresis is rather moderate (20-40%) (76). Moreover, protein losses are not homogenous and this effect may be much greater for non-soluble proteins such as membrane proteins (77, 78). Thus, 2D gel-based proteomics has poor ability to analyze membrane proteins. Gel-based techniques are also inefficient in characterizing hydrophobic, large (>200kDa) and low abundant proteins (79, 80).

1.4.5 Gel-free mass spectrometry-based metaproteomics

Gel-free mass spectrometry-based proteomics on the other hand, depending on the question being asked, can identify and quantify individual peptides or broadly compare nearly the entire proteome. Mass spectrometry based proteomics include absolute quantification (absolute amount or concentration of the proteins within a sample) and relative quantification (amounts of proteins or whole proteomes between samples) of the proteins. Relative quantification is further divided into label-based quantification and label free quantification. In label free quantification, peptides and proteins in their natural states are compared in consecutive experiments. Thus gel-free proteomic technique is largely dependent on highly reproducible sample handling, liquid chromatography separation and mass spectrometry measurements. One major advantage of gel-free proteomic technique is that unlimited number of samples without any chemical, metabolic and enzymatic modification can be quantified, thus, it is lower in cost. Still, two major disadvantages exist: the work flow for gel-free proteomic technique is less robust than label-based approaches, and it requires higher reproducibility (81, 82).

1.4.6 Label-based (mass spectrometry) metaproteomics

The aforementioned limitations led to the development of the label-based methods including metabolic labeling, chemical labeling and enzymatic labeling. Metabolic labeling such as SILAC is restricted to small organisms and cell cultures, while chemical labeling such as isobaric tags for relative and absolute quantitation (iTRAQ) or tandem mass tag (TMT) can quantify all kinds of protein samples (83, 84). In label-based methods, the addition of the labels shifts the mass values in the mass spectrum, thus, the ratio of the peak intensities of the differentially labeled proteins (same proteins from different samples) yields
the relative abundance ratio of the protein species concerned. Thus up to 6 samples in TMT and up to 8 samples in iTRAQ experiments can be compared in a single mass spectrum. This method yields higher accuracy of quantitative measurements compared to label-free techniques.

Steps in labeling approaches are similar to the label-free approaches and they include the following major steps (85). First, proteins are extracted from samples and subsequently reduced, alkylated and digested. Second, digested peptides from each sample are labeled with isobaric tags and subsequently mixed to prepare a pool of differentially labeled proteins. Third, samples are prepared by liquid chromatography (LC) and analyzed by mass spectrometry (MS). Fourth, data analysis is performed by identifying and quantifying peptides/proteins followed by conducting statistical analysis on the quantified peptides/proteins. Proteins are quantified frequently by spectral counting of identified proteins after the mass spectrometry analysis.

In the protein labeling approach, different protein samples can be pooled together once the proteins are labeled and this combined mixture is processed before being analyzed by a LC-MS/MS experiment. Major disadvantages of the labeling method are the high cost of the labeling reagents and addition of sample handling steps compared to label-free methods. iTRAQ and TMT are two good examples of this type of labeling method (86).

1.4.7 iTRAQ metaproteomics and data analysis

iTRAQ uses isobaric chemical labels and each contains an amine reactive moiety, a balancing moiety and a reporter moiety (the reporter moieties have the same chemical composition but different molecular weight because of different isotope composition). Different reagents have the same molecular weight and once labeled, they produce identical mass shifts. Different samples are labeled with different isobaric tags and mixed together. Identical peptides from samples co-elute and are detected as a single precursor ion. When fragmented, the reporter ions are released and their intensities are proportional to the relative abundances of the labeled peptide from different samples (83). TMT is similar to iTRAQ but it uses tandem mass tags and it compares up to 6 samples. TMT tags are similar to iTRAQ tags in that they are also made up of a reactive amine moiety, a balancing moiety and a
reporter moiety. These are fragmented during MS/MS and their intensity is used to calculate relative amounts between samples (84).

A single iTRAQ experiment compares up to 8 samples. In iTRAQ experiments, experimental and technical variation can account for about 23% and 11% iTRAQ of the data variability. Preferably, in a single iTRAQ experiment, at least two biological replicates are measured (87), because biological variation explains the highest variation (25%). If the sample size is bigger, all the available samples are pooled to create a reference and this pooled sample is used as one sample in each of the 8-plex iTRAQ set (88). This pooled sample can be used for cross-comparison between the ratios of multiple iTRAQ sets (88). Sample pooling is a reasonable approach because of the following: 1) cDNA microarray applications commonly use sample pooling to reduce biological variation and pooled samples are contained in more than 15% of Gene expression database 2) sample pooling is also used when community effect is measured with higher precision in fold change 3) iTRAQ reagents and mass spectrometry time are expensive.

In conclusion, label-free relative quantitative metaproteomics and label-based relative quantitative metaproteomics are well-suited methods for the relative quantification of the proteins in highly complex microbiota samples. Label-free quantitative proteomics can be used for an unlimited number of sample quantifications. However, this suffers from low accuracy from highly reproducible sample handling issues. Therefore, the iTRAQ method is suitable for determining both the absolute and relative abundances of the proteins for up to 8 samples in a single run or even more samples with the use of a pooled reference sample.
Chapter 2: Overall rationale, hypothesis and objectives

**Rationale:** The effect of dietary lipids on the host-microbe dynamic in the gut is unknown. Our lab previously showed that dietary lipid type, and not total calories from fat, influenced physiological and immunological responses in the gut. Specifically, olive oil protected against colitis, while corn oil and milk fat resulted in a similarly exacerbated colitis. However, the milk fat group also resulted in several protective compensatory responses for inflammation in the gut unlike corn oil. Moreover, only the corn oil diet rich in n-6 PUFA resulted in bacterial infiltration across the intestinal barrier. Based on this data, our overall hypothesis is that the host-microbe interaction will be uniquely affected by exposure to each type of dietary lipid. Specifically, *we hypothesize that corn oil diets, rich in n-6 PUFA, promote a pathogenic microbiome associated with intestinal barrier damage.*

**Approach:** To understand the tripartite relationship between lipid diet, gut bacteria and the host, we fed mice a 40% (by energy) isocaloric and isonitrogenous diet composed of either corn oil, olive oil or milk fat for 5 weeks post-weaning. The gut tissues were collected for metagenomics and metaproteomic analysis.

**Objective 1:** To determine the effects of lipid diets on the metagenome of microbiome.

**Objective 2:** To determine the effects of lipid diets on the metaproteome of the colon (host and microbial).
Chapter 3: Experimental Methods

3.1 Animals

C57BL/6 male and female mice (Jackson Laboratories, Bar Harbor, Maine) were maintained at the Center for Disease Modeling at the University of British Columbia (UBC), Vancouver, Canada. The animal room was temperature controlled (22+/-2 °C) with a 12-hour light/dark cycle and fed with respective diets ad libitum with free access to water under a specific pathogen-free condition. Food intake and blood glucose was monitored weekly as described previously [89]. The protocols used were approved by the Animal Care Committee of UBC and in direct accordance with guidelines drafted by the Canadian Council on the Use of Laboratory Animals.

3.2 Diets (Performed by Ben Dai)

Three week old C57BL/6 mice (total n=48, n=8 each diet) were fed irradiated isocaloric, isonitrogenous diets for 5 weeks. The high fat diets were composed with lipids at 40% energy whereas the chow control was composed with 9% energy from fat as previously reported. High fat diets contained 20% w/w of olive oil, corn oil or anhydrous milk fat prepared by blending dietary oils to a basal diet mix as previously reported.

3.3 Tissue collections (Performed by Dr. Carol Chan)

Mice were anaesthetized with isoflurane and euthanized by cervical dislocation. Ileum and distal colon (with the luminal content and stool removed) were collected, snap frozen in liquid nitrogen and stored at -80°C in 1 cm pieces and then used for either metagenomics or proteomic experiments.

3.4 Bacterial DNA extraction and amplicon preparation (Performed by Dr. Carol Chan and Kirsty Brown)

Frozen ileum and colon pieces were homogenized (60Hz for 1 minute) using stainless steel beads in Mixer Mill MM 400 (Retsch). Bacterial genomic DNA was extracted with QIAamp® DNA Stool Mini Kit (Qiagen) according to the manufacturer’s instructions. DNA concentration and purity were checked with Nanodrop 2000 (Thermo Scientific). Primers were used to amplify the 16S rRNA gene. The same linker primer (CCTACGGGNGGCWGCAG) and reverse primer (CTATGCGCCTTGCCAGCCCGCTAGGACTACHVGGGTATCTAATCC) were used for
all samples while specific barcode sequences (TACTCTCGTG, TAGAGACGAG, TCGTGCCTCG, ACATAACGCTG, TGTACTACTC, ACGACTACAG, CGTAGACTAG, TACGAGTATG, TCTCTATGCG, TGATACGTCT, CATAGTAGTG, CGAGAGATAC, ATACGACGTA, TCACGTACTA, CGTCTAGTAC, TCTACGTAGC, ACGAGTGCGT, ACGCTCGACA, AGACGCCTC, AGCACTGTAG, ATCAGACACG, ATATCGCGAG, CGTGTCTCTA, CTCGCGTGTC) were used to identify each sample. EconoTaq® PLUS 2X Master Mix (Lucigen Corporation, Middleton, USA) was used for the amplification: the 25 µL PCR mixture contained 1X EconoTaq master mix, 1 µM forward and reverse primers and 50-200ng template genomic DNA isolated. The samples were amplified with a PCR protocol of initial denaturation at 94°C for 2 min, 30 cycles of denaturation, annealing and extension at 94 °C for 30 s, 58 °C for 30 s and 72 °C for 60 s, then a final extension at 72 °C for 10 min. The PCR product was run on a 1% agarose gel with the band at 464 bp gen purified with QIAquick® Gel Extraction Kit (Qiagen). PCR amplicons concentration was normalized with SequlPrepTM Normalization Plate (96) Kit (Invitrogen). Library preparation, emPCR amplification and picotitre plate pyrosequencing using titanium chemistry was carried out by Vancouver Prostate Centre, UBC and Vancouver General Hospital Centre of Excellence in accordance with Roche/454 Life Sciences protocol on the 454 GS FLX+ System.

3.5 High throughput amplicon based metagenomics sequencing analysis

Sequence reads were analyzed using the quantitative insights in microbial ecology (QIIME) pipeline. Raw sequence (sff) files were first demultiplexed and chimeras were removed using usearch61. Using the quality filtered sequences, operational taxonomic units (OTUs) were defined using usearch61 with a similarity threshold of 0.97. Representative sequences were aligned with PyNAST (90) using the Greengenes Core Set as a template and taxonomy was assigned using the ribosomal databased project with the Greengenes database (version 13.8). For subsequent analysis, only the OTU tables with no PyNAST failures were used for subsequent analysis.

3.5.1 Alpha diversity and beta diversity

R package phyloseq (91) was used for both alpha diversity and beta diversity analysis. For alpha diversity analysis we first pruned OTUs that were not present in at least one sample. Then we plotted Observed, Chao1, Shannon and Simpson’s diversities using boxplots. Before conducting beta diversity analysis, the OTU matrix was normalized to
account for uneven column (sample) sums. Thus, OTU tables were first normalized using metagenomeSeq’s CSS (cumulative sum scaling) method. (92) The CSS transformed OTU table was not recommended for presence/absence matrix analysis such as unweighted UniFrac. However, other transformation techniques such as rarefied and relative abundance transformation yielded the same results for the unweighted UniFrac analysis. Therefore, CSS transformed data was used for both the weighted and unweighted UniFrac analysis.

3.5.2 Compositional analysis

The linear discriminant analysis (LDA) effect size (LEfSe) (93) method was used to determine the differentially abundant OTUs. LEfSe first determines a list of features that are differential among conditions of interest with statistical and biological significance, ranking them according to the effect size. Then it constructs a mapping of the differences to taxonomic trees. Additionally, it produces a histogram visualizing the raw data within the specified problem structure for each relevant feature.

3.5.3 Computer prediction based bacterial metagenomics

We used PICRUSt (73) to predict the metagenome of the microbiome using 16RNA dataset. PICRUST uses closed reference OTUs, since it can only predict the metagenome of the known KEGG functions. Therefore, we filtered closed reference OTUs from the original open reference OTU table. This produced an OTU table with ~2000 OTUs. Then the OTUs were first normalized for 16S gene copy numbers and then PICRUSt prediction algorithm was used to predict the metagenome of the community. The predicted metagenomic files were then submitted for further analysis to HUMAnN (94). HUMAnN takes gene abundances as input and produces gene and pathway summaries as outputs. The summary output files were then analyzed using LEfSe to determine the differential pathways among conditions.

3.6 High throughput mass spectrometry based metaproteomics

3.6.1 Protein extraction (Performed with Candice Quin)

Frozen colon pieces were scraped to separate the mucosa from the submucosa following a similar protocol as previously described (95). We separated the mucosa from the submucosa because the majority of the bacterial proteins should be present at the mucosal surface and we wanted to understand the dynamics between the host and the microbes at the
mucosal surface. Since we removed the luminal content, any proteins seen in this layer are due to a true interaction at this surface. The submucosal and the mucosal samples were separately put into lysis buffer made up of 25 mM HEPES solution (pH=7.5) with 1 protease inhibitor (bestatin, AEBSF, EDTA, pepstatin, and E-64), 7 M urea, 2 M thiourea, 4% CHAPS (six biological replicates were pooled into one tube for each diet group). Then we performed homogenization with bead beating. Insoluble materials were removed by centrifugation. Then the supernatant was transferred to clean tubes and mixed with approximately 15 ml pre-chilled acetone. The proteins were precipitated overnight (4 °C) and precipitated proteins were pelleted by centrifugation. Acetone was decanted and precipitated protein pellets were washed twice with acetone.

3.6.2 iTRAQ experiment (Performed at The University of Victoria Proteome Center)

The samples were submitted for iTRAQ labeling, high pH reversed phase fractionation and LC-MS analysis at the University of Victoria Genome BC Proteome Center located at the Vancouver Island Technology Park.

Supernatant from each sample was used to determine the concentration in a BCA assay (Sigma, ON, Canada). The volume containing 100 µg of protein was acetone precipitated in 10 volumes of ice-cold acetone overnight at -20 °C. The precipitated samples were centrifuged and the acetone was removed. The protein pellets were re-solubilized in 30 µL, 0.5 M TEAB and 3 µL, 2 % SDS. Samples were rehydrated for 4h at 4 °C. 2 µL, 50 mM TCEP was added and samples were incubated for 1 h at 60 °C and then allowed to cool to room temperature. 1 µL, 200 mM MMTS was added and samples were incubated at room temperature for 10 minutes.

Eighty micrograms of trypsin (Promega, Sequencing Grade Modified Trypsin) was re-suspended in 250 µL, 100 mM TEAB and 120 µL (10µg) was added to each sample. Samples were incubated overnight at 37 °C. The digested samples were dried by vacuum centrifugation (Savant Instruments, Holbrook, NY) and then 30 µL, 0.5 M TEAB was added to each dried sample followed by the addition of 50uL isopropanol. The iTRAQ label (AB Sciex, ON, Canada) was added to each sample and then incubated at room temperature for two hours. The labeled peptides were pooled and vacuum centrifuged until the final volume was approximately 100µL.
An Agilent 1290 HPLC (Agilent, CA, USA) was equipped with an XBridge C18 BEH300 (Waters, MA, USA) 250 mm X 4.6 mm, 5 µm, 300 A HPLC column. Buffer A was 10 mM Ammonium hydroxide (pH=10). Buffer B was 80% Acetonitrile, 10 mM Ammonium hydroxide (pH=10). The flow rate was set to 0.75 ml/min. Samples were brought up to 1.8 mL with buffer A and injected onto the column. The column was allowed to equilibrate for 5 min in buffer A before a gradient was applied; 5-45% B in 75 minutes. Fractions were collected every minute for 96 minutes. The HPLC fractions were then reduced in volume by lyophilization and concatenated into 24 fractions by combining every 24th fraction (e.g. fractions 1, 25, 49, and 73 were combined). C18 StageTip concentrated samples were analyzed by reversed phase nanoflow HPLC with nano-electrospray ionization using a LTQ-Orbitrap Velos Pro mass spectrometer operated in positive ion mode with a was 2 hours reverse phase gradient per HPLC fraction.

Each sample was rehydrated to 15µl (2% Acetonitrile, 3% Formic acid). A five micro-litre aliquot of sample was separated by on-line reversed phase liquid chromatography using a Thermo Scientific EASY-nanoLC II system with a reversed-phase pre-column Magic C-18AQ (100 µm I.D., 2.5 cm length, and an in-house prepared reversed-phase nano-analytical column packed with Magic C-18AQ (75 µm I.D., 15 cm length, 5 µm, 100 Å, Michrom BioResources Inc, Auburn, CA), at a flow rate of 300 nl/min. The chromatography system was coupled on-line to an LTQ Orbitrap Velos Pro mass spectrometer equipped with a Nanospray Flex source (Thermo Fisher Scientific). Solvents used were A: 2% Acetonitrile, 0.1% Formic acid; B: 90% Acetonitrile, 0.1% Formic acid. After a 249 bar (~ 5µL) pre-column equilibration and 249 bar (~ 8µL) nanocolumn equilibration, samples were separated using a 120 minute gradient (0 min: 5 % B; 100 min: 40% B; 5 min: 80% B; 2 min: 100% ; 13 min: 100% B).

The parameters for the LTQ Orbitrap Velos Pro (Thermo Fisher Scientific, Bremen, Germany) were as follows: Nano-electrospray ion source with spray voltage 2.458kV; capillary temperature 250 °C; Survey MS1 scan m/z range 400-1800 profile mode; resolution 30,000 @ m/z 400 with AGC target 1E6; and one microscan with maximum inject time 500 ms. A Siloxane m/z 445.120024 lock mass was used for internal calibration. Other settings were: Preview mode for FTMS master scans: on; injection waveforms: on; monoisotopic precursor selection: on; rejection of charge state: 1. The experiment was acquired as a top 15
five data dependent analysis of the most abundant ions with charge states of 2-4, exceeding 20,000 counts, being selected for HCD FT MS/MS fragmentation (FTMSMS scans 2-16) and detection in centroid mode. Dynamic exclusion settings were: repeat count: 1; repeat duration: 10 seconds; exclusion list size: 500; exclusion duration: 10 seconds with a 10ppm mass window. The FT HCD settings were: resolution 7,500 @ m/z 400 in centroid mode with AGC target 1E5, 1.6amu isolation width, and normalized collision energy 45%, 0.1msec activation time.

3.6.3 iTRAQ mass spectrometry data analysis

All data was analyzed using Proteome Discoverer 1.4 and MASCOT v2.4 software. Spectrum Selection was used to generate peak lists of the HCD spectra (parameters: activation type: HCD; s/n cut-off: 1.5; total intensity threshold: 0; minimum peak count: 1; precursor mass: 350-5000 Da). The peak lists were submitted to an in-house database search using Mascot 2.4 (Matrix Science), and were searched against the following databases Uniprot-Mouse (43,908 sequences; 19,909,825 residues), Bacteroidetes (unknown version, 11363 entries) (only "20160330_REC-1563_Fr_1_24 Bacteroidetes") assuming the digestion enzyme trypsin and Firmicutes (unknown version, 17039 entries) (only "20160330_REC-1563_Fr_1_24 Firmicutes") also assuming trypsin. The data were also searched for Proteobacteria and Actinobacteria with SearchGUI. Search parameters were: precursor tolerance 10 ppm; MS/MS tolerance 15mmu (for FT MS/MS HCD data); enzyme trypsin; 1 missed cleavage; instrument type FT-ICR ESI; fixed modification: Methylthio (C), iTRAQ8plex (K), and iTRAQ8plex (N-term); variable modifications: Oxidation (M), Deamidated (NQ), iTRAQ8plex (Y). The HCD Percolator settings were: Max delta Cn 0.05; Target FDR (false discovery rate) strict 0.01, Target FDR relaxed 0.05 with validation based on q-Value. Proteome Discoverer result files were then analyzed using Scaffold for statistical validation of protein identifications.

Scaffold (version Scaffold_4.6.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least two identified peptides for the host proteome and at least one identified peptide for the bacterial proteome. Protein probabilities were
assigned by the Protein Prophet algorithm (96). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Normalization was performed iteratively (across samples and spectra) on intensities, as described earlier (97). Medians were used for averaging. Spectra data were log-transformed, pruned of those matched to multiple proteins and those missing a reference value, and weighted by an adaptive intensity weighting algorithm. Differentially expressed proteins were determined by applying Permutation Test with unadjusted significance level p < 0.05 corrected by Benjamini-Hochberg.
Chapter 4: Results

4.1 Objective 1: Effects of the lipid diets on the metagenome of the microbiome

4.1.1 The type of lipid diet affected the microbial phyla ratio

Several studies have found that high fat diets alter the microbial ecosystem causing dysbiosis in mice and in humans. Some studies found that the ratio of Firmicutes to Bacteroidetes phyla was increased in high fat diets. However, the effects of different lipid diets were not well understood. Thus, we have examined the effects of high fat diets composed of either olive oil, corn oil or milk fat on the ileum and colon microbiota using 16S rRNA-based 454 pyrosequencing. In our study, we found that corn oil and milk fat diet group significantly increased the Firmicutes to Bacteroidetes ratio (Figure 1, Tables A.1-A.2), whereas olive oil group did not increase this ratio significantly compared to the low fat diet group. This effect was more pronounced in the colon compared to the ileum (Figure 1 C-D).

Figure 1. Phyla abundance and ratio plots among different lipid diets and anatomical regions. (A) Abundance of bacterial phyla among different lipid diets in the ileum. (B) Abundance of bacterial phyla among different lipid diets in the colon. (C) Comparison of the log abundance of the Firmicutes to Bacteroidetes ratio among experimental diet groups in the ileum. (D) Comparison of the log abundance of the Firmicutes to Bacteroidetes ratio among experimental diet groups in the ileum. The y-axis of the box plot indicates the log of the abundance of the Firmicutes divided by the abundance of Bacteroidetes.
The Firmicutes to Bacteroidetes ratio increased in all high fat diets, and this suggests that the balance of Gram-positive and Gram-negative bacteria were altered. To confirm this, we further compared the relative abundance of Gram-positive and Gram-negative bacteria. Figure 2 shows that the abundance of the Gram-negative bacteria are the largest in the low fat diets both in the ileum (Figure 2A) and in the colon (Figure 2B), whereas the abundance of Gram-positive bacteria are elevated in all high fat diets (Figure 2 C-D), especially in the milk fat and the corn oil diet (Tables B.1-B.2).

Figure 2. Relative abundance of Gram-negative and Gram-positive bacteria across lipid diets. (A) Ileum Gram-negative bacterial abundance comparison among lipid diets. (B) Colon Gram-negative bacterial abundance comparison among lipid diets. (C) Ileum Gram-positive bacterial abundance comparison among lipid diets. (D) Colon Gram-positive bacterial abundance comparison among lipid diets. Plots were generated by Bugbase online analysis application.
4.1.2 The type of lipid diet differentially affected the alpha diversity of microbial communities both in the colon and in the ileum

To analyze alpha diversity, we looked at two separate components of species diversity; number of species present (species richness) and their relative abundances (dominance or evenness). We found differential effects of diets on alpha diversity within the colon and ileum with the colon showing increased microbial diversity in mice fed corn oil and milk fat diets (Figure 3B), while in the ileum, all high fat diets tended to show a decrease in diversity although no significant effects were observed (Figure 3A, Tables C.1-C.2). Specifically, in the colon (Figure 3B), both the observed (the OTUs that were observed at least once in each sample) and Chao1 index (estimate the total number of rare species present in a community) estimated highest species richness for corn oil diet followed by milk fat diet. Olive oil and low fat diet, on the other hand, had similar species richness compared to each other but reduced richness compared to corn oil and milk fat diets. In the ileum (Figure 3A), however, both the observed and Chao 1 diversities decreased in all high fat diet groups.

![Figure 3. Alpha diversity of the microbial communities across lipid diets. A) Alpha diversity of ileum microbiota in lipid diets. Observed, Chao1, Shannon, Simpsons and Inverse Simpsons indexes are used as alpha diversity. B) Alpha diversity of colon microbiota in lipid diets: anhydrous milk fat, corn oil, low fat and olive oil diet groups. Observed, Chao1, Shannon, Simpsons and Inverse Simpsons indexes are used as alpha diversity. Figure color code: red (milk fat), blue (corn oil), purple (olive oil) and green (low fat).](image)

We used two of the most widely used indices including the Shannon (or Shannon-Wiener) index and Simpson’s index. When both the richness and the evenness of the community increase, the Shannon index increases. The Shannon index showed that the milk
fat diet exhibited both high richness and evenness compared to low fat and olive oil diets. Richness and evenness are confounded in the Shannon index, thus, we used a direct estimate of species richness (Chao1) combined with a measure of dominance or evenness (Simpsons index).

Simpson’s index is computed as the probability of any two individuals picked randomly from a sample belonging to the same species. When Simpsons index increases, diversity (evenness) decreases. Simpsons indexes showed that the low fat diet promoted the highest evenness compared to all high fat diets, while decreased evenness was observed for all high fat diet groups in the colon (Figure 3B). However, the corn oil diet showed bigger variation in alpha diversity (Shannon) because it has the highest species richness but lower species evenness in the colon. In the ileum, Shannon diversity combined with Simpsons index showed that all high fat diet groups exhibited increased evenness compared to low fat diet, while there were no significant changes observed among milk fat, corn oil and olive oil diets.

Overall, the type of fatty acid present in the diet was found to have differential effects on the alpha diversity of the bacterial communities. This effect was more pronounced in the colon than in the ileum, thus bacterial communities in the colon seemed to be susceptible to change compared to the ileum bacterial communities. Moreover, milk fat and corn oil diets increased diversity in the colon, whereas olive oil diet had reduced diversity similar to that of the low fat chow diet.

4.1.3 The type of lipid diet differentially affected the beta diversity of microbial communities in the colon but not in the ileum

To examine how different types of microbes were distributed among dietary groups, we generated principal coordinate analysis (PCoA) plots based on weighted and unweighted UniFrac distances (98). Unweighted UniFrac is most useful when communities differ primarily by their presence or absence (e.g., diet, pH). On the other hand, weighted UniFrac is a quantitative method that is useful when the relative taxon abundance results in the community changes (e.g., particular set of taxa is more abundant because of some abundant nutrient supply). Therefore, we used both weighted and unweighted UniFrac to describe the beta diversity among the four diet groups.
PCoA plot of unweighted UniFrac distance for colon samples (Figure 4B, Table C.3) shows four clusters corresponding to the four diet groups of interest. This shows the presence of differential taxa in each diet in the colon, showing that bacterial communities that live in each diet differ significantly. However, similar to alpha diversity, ileum was resistant to change in beta diversity when exposed to different lipid diets (Figure 4 A-C). Overall, all high fat diets had similar bacterial taxa compared to each other, but they had significantly different bacterial taxa compared to the low fat diet in the ileum.

Figure 4. The effect of lipid diets on the beta diversity of the gut microbiota. A) PCoA plots of the unweighted UniFrac distance of ileum microbiota samples from milk fat (n=8), corn oil (n=8), low fat (n=7) and olive oil (n=8) diet fed mice. B) PCoA plots of the unweighted UniFrac distance of colon microbiota samples from milk fat (n=8), corn oil (n=8), low fat (n=8) and olive oil (n=5) diet fed mice. C) PCoA plots of the weighted UniFrac distance of ileum microbiota samples from milk fat (n=8), corn oil (n=8), low fat (n=7) and olive oil (n=8) diet fed mice. D) PCoA plots of the weighted UniFrac distance of colon microbiota samples from milk fat (n=8), corn oil (n=8), low fat (n=8) and olive oil (n=5) diet fed mice. The first two PCs from the PCoA are plotted. Areas are constructed by linking all the individuals within each diet group for visualizations of clusters.
PCoA plot of weighted UniFrac distance for colon samples (Figure 4D) showed that there were three distinct clusters. These three clusters corresponded to low fat diet, olive oil diet and milk fat/corn oil combined (overlapping) diet. Milk fat and corn oil diet cluster together, showing that the communities do not differ in relative taxon abundance. In the ileum (Figure 4C), however, there are only two distinct clusters. All the high fat diet groups clustered together distinct from the low fat diet, showing that high fat diets did not promote differential relative taxon abundance in the ileum.

PCoA produces a set of uncorrelated axes to summarize the variability in the data set seen in the magnitude of each axis. In our data, weighted UniFrac for the colon showed that x axis (56.8%) and y axis (8%) captured most of the variability that can be explained by the different types of diets. Hence diet is the major factor that drives the composition of the microbiota in the colon. The same is similar for the ileum with the x axis 33.6% and y axis representing 25.6% variability explained by different diets. Therefore, approximately 60% of the variability of the data was explained by diet alone both in the ileum and the colon. Overall, dietary lipids strongly predict the type of microbes that are present in the gut.

4.1.4 The type of lipid diet promoted the abundance of different bacterial taxa

Since unweighted UniFrac showed four distinct clusters based on our four diets, we next used LEfSe to determine specific bacteria taxa that were differentially abundant. Both colon and ileum showed differentially abundant bacteria (LEfSe Kruskal-Wallis test, P<0.05) in each lipid diet. In the colon (Figure 5B), we found an elevation in unclassified Lachnospiraceae (Firmicutes), Coriobacteriaceae (Adlercreutzia) from the Actinobacteria phyla in low fat chow diet. Olive oil diet was enriched with Firmicutes with abundant unclassified Clostridia, unclassified Peptostreptococcaceae, unclassified Ruminococcaceae, Dorea and unclassified Mogibacterium. Linear discriminate analysis also detected increased Firmicutes such as Lactococcus from the Streptococcaceae family and unclassified Erysipelotrichales in the milk fat diet, whereas high Coprococcus (Lachnospiraceae), Turicibacter were found abundant in the corn oil diet. Overall, lipid diets promoted the growth of Firmicutes phyla. However, different types of Firmicutes, family and genus level, become abundant depending on the type of fatty acids they were in.
Figure 5. The effect of lipid diets on the gut microbial taxa. A) Differentially abundant microbial clades in the ileum microbiota from milk fat (n=8), corn oil (n=8), low fat (n=8) and olive oil (n=5) diet fed mice. B) Differentially abundant microbial clades in the colon microbiota from milk fat (n=8), corn oil (n=8), low fat (n=7) and olive oil (n=8) diet fed mice. Cladogram represents taxonomic representation of statistically and biologically consistent differences among lipid diet groups. In cladograms, significant differences are represented in the color of the most abundant class. Yellow circles represent non-significant microbial clades. The most strict (all-to-all) version of LEfSe was used with Kruskal-Wallis test (P<0.05) and only these results are shown in the figure. LDA score threshold was set at default value 2.
Despite that we did not find distinct clusters in the unweighted UniFrac in the ileum, LEfSe detected abundant unclassified Lachnospiraceae, unclassified Firmicutes and Anaeroplasmataceae from the Tenericutes phyla in low fat diet (Figure 5A). Milk fat fed mice had more abundant *Lactobacillus* from the Firmicutes phyla, whereas corn oil promoted an increased abundance of Firmicutes such as *Coprococcus* and unclassified Peptostreptococcaceae. Lastly, Firmicutes such as Streptococcaceae, unclassified Clostridiaeae and unclassified Erysipelotrichales were found to be abundant in the olive oil diet. All in all, high fat diets again promoted the growth of Firmicutes phyla with differing species depending on the lipid type they are in. However, Tenericutes phylum was suppressed in all high fat diets.

4.1.5 Lipid diets conferred core functionality to each microbial community

Since microbial compositions change according to type of lipid diets, we next investigated how fatty acids differentially affect the functionality of the microbiota. In order to investigate the functional changes in microbiota, we first predicted metagenomics functional content from 16S rRNA genes. We started with obtaining close-reference OTUs. This step reduced the OTU numbers from open-reference OTUs (~ 6000) to closed-reference OTUs (~2000). Then the closed-reference OTUs were normalized by dividing each OTU by known/predicted 16S copy number abundance. Finally, the final metagenome functions were created by multiplying each normalized OTU abundance by each predicted functional trait abundance to produce a table of functions by samples. Therefore, only the known OTUs (~2000) that were picked by closed-reference OTU picking method were used in the metagenomic prediction. We further performed linear discriminate analysis (LDA) on this predicted metagenome among lipid diets both in the ileum and colon as shown in Figure 6.

Low fat diet fed mice microbiota in the colon was enriched with functions like vitamin and cofactor biosynthesis (biotin, vitamin B6, folate, lipoic acid, riboflavin, nicotinamide and nicotinamide, pantothenate and CoA biosynthesis, thiamine) and higher amino acid (D_glutamate and D_glutamine, histidine, alanine, aspartate and glutamate, glycine, serine and threonine) and protein export, digestion and absorption, carbohydrate metabolism (TCA cycle) and fatty acid biosynthesis (Figure 6B). Similar functions were also observed in the ileum such as vitamin and cofactor biosynthesis, amino acid metabolism, and protein export, digestion and absorption and fatty acid metabolism (Figure 6A).
Figure 6. The effect of lipid diets on the predicted microbial functions. A) Differentially abundant microbial pathways predicted from PICRUSt in the colon microbiota from milk fat (n=8), corn oil (n=8), low fat (n=8) and olive oil (n=5) diet fed mice. B) Differentially abundant microbial pathways predicted from PICRUSt in the colon microbiota from milk fat (n=8), corn oil (n=8), low fat (n=8) and olive oil (n=5) diet fed mice. Statistically and biologically differentially abundant pathways among the four diet group shown as histograms of the LDA scores. The most strict (all-to-all) version of LEfSe was used with Kruskal-Wallis test (P<0.05). LDA score threshold was set at default value 2.
Olive oil fed mice microbiota had increased potential for altered short chain fatty acid metabolism, such as butanoate metabolism and propanoate metabolism as well as amino acid (arginine, proline, phenylalanine) metabolism. Some xenobiotic degradations (nitrotoluene, limonene and pinene, naphthalene, polycyclic aromatic hydrocarbons) were also abundant in olive oil diet in the colon (Figure 6B). Similar to the colon olive oil diet, increased arginine and proline metabolism and biosynthesis of other secondary metabolites were detected in the ileum (Figure 6A).

Corn oil fed mice microbiota functions could be characterized by increased flagella assembly, bacterial chemotaxis, ABC transporters, lipid metabolism (biosynthesis of unsaturated fatty acids, sphingolipid metabolism), glutathione metabolism, and epithelial cell signaling in H. pylori infection (Figure 6B). The ileum microbiota function (Figure 6A) was similar to the colon microbiota in that it was also enriched with flagella assembly. Moreover, the ileum corn oil fed mice microbiota has the potential for staphylococcus aureus infection, xenobiotic degradation (limonene, pinene, naphthalene, aminobenzoate) and carbohydrate metabolism (fructose and mannose, starch and sucrose).

Lastly, colon milk fat fed mice microbiota had the highest potential for sugar (starch, sucrose, fructose, mannose and galactose) metabolism, glycolysis and gluconeogenesis, energy metabolism (sulfur relay system), fatty acid metabolism (fatty acid metabolism, alpha_linolenic acid metabolism) and xenobiotic degradation (styrene, dioxin, xylene, caprolactam, benzoate) in the colon (Figure 6B). The ileum also has enriched energy metabolism (sulfur relay system), xenobiotic degradation (nitrotoluene, styrene, dioxin) (Figure 6A).
### 4.2 Objective 2: Gut proteomics

Since the ileum appeared to be more resistant to diet-induced dysbiosis compared to the colon, we proceeded with the proteomic analysis using only the colon samples. To understand the dynamic between lipid diets, gut bacteria and the host in the gut, we used iTRAQ shotgun metaproteomic technique for both the mucosa and submucosa. Precipitated proteins from both fractions were sent for proteomic analysis at the University of Victoria Genome BC Proteome Center. Table 1 presents the number of proteins identified in each layer of the gut from the host and the microbiome.

| Table 1. Mucosal and submucosal protein identification summary using both bacterial and mouse proteome databases. (A) Mucosal proteome with two minimum peptide (B) Mucosal proteome with one minimum peptide (C) Submucosal proteome with two minimum peptide (D) Submucosal proteome with one minimum peptide. |
|---|---|---|---|---|
| **(A) Mucosa, minimum peptide number 2** | **Protein threshold 95%, 0% decoy FDR (Bacteria and Mouse)** | **Number of identified proteins** | **Number of identified spectra** | **Total number of spectra** | **Spectrum identification rate** |
| **Database** | | | | |
| Firmicutes | 9 | 29 | 224804 | 0.01% |
| Bacteroidetes | 16 | 75 | 224804 | 0.03% |
| Mouse | 1842 | 29684 | 224804 | 13% |
| **(B) Mucosa, minimum peptide number 1** | **Protein threshold 95%, 65.5% decoy FDR(Bacteria), 0.1% decoy FDR(Mouse)** | **Number of identified proteins** | **Number of identified spectra** | **Total number of spectra** | **Spectrum identification rate** |
| **Database** | | | | |
| Firmicutes | 258 | 309 | 224804 | 0.14% |
| Bacteroidetes | 97 | 166 | 224804 | 0.07% |
| Mouse | 2731 | 31230 | 224804 | 14% |
| **(C) Submucosa, minimum peptide number 2** | **Protein threshold 95%, 0% decoy FDR (Bacteria and Mouse)** | **Number of identified proteins** | **Number of identified spectra** | **Total number of spectra** | **Spectrum identification rate** |
| **Database** | | | | |
| Firmicutes | 7 | 27 | 219790 | 0.01% |
| Bacteroidetes | 15 | 56 | 219790 | 0.03% |
| Mouse | 2079 | 29890 | 219790 | 14% |
| **(D) Submucosa, minimum peptide number 1** | **Protein threshold 95%, 53.7% decoy FDR (Bacteria),0.9% decoy FDR(Mouse)** | **Number of identified proteins** | **Number of identified spectra** | **Total number of spectra** | **Spectrum identification rate** |
| **Database** | | | | |
| Firmicutes | 47 | 73 | 219790 | 0.03% |
| Bacteroidetes | 86 | 142 | 219790 | 0.07% |
| Mouse | 3082 | 31462 | 219790 | 14% |
4.2.1 Microbial functions were differentially affected by dietary lipid

We searched the raw dataset against the bacterial proteome to identify bacterial functions and to determine if the PICRUSt obtained functional differences were actually present. We used both the mucosal samples and submucosal tissue samples for database searching. Because Bacteroidetes and Firmicutes are the two major phyla in the gut, raw data was searched for Bacteroidetes and Firmicutes databases and the results were combined and shown in table 2. We also searched the data for other databases such as Proteobacteria, Actinobacteria and Verrucomicrobia (data not shown). A search for Bacteroidetes and Firmicutes databases identified 25 proteins (0% FDR) with 95% confidence with two minimum peptides. When the minimum peptide threshold was decreased to one minimum peptide, the identified protein number increased to a total of 355 proteins (FDR 65.5%, percent of decoy protein) in the mucosal samples with 31% uncharacterized or unknown protein peptides. When the same analysis was done on the submucosal samples, 22 proteins (0% FDR) were identified with two minimum peptides, while 133 proteins (FDR 53.7%) containing 24% unknown and uncharacterized proteins were identified with one minimum peptide (Table 1). High confidence proteins with at least two minimum peptides include ATP synthases, some uncharacterized proteins, the chaperone protein DnaK, isocitrate dehydrogenase and adenosylhomocysteinas in both in the mucosal and submucosal data. Among these, relative intensities of the ATP synthase and the chaperone protein DnaK were the highest in the corn oil diet compared to the olive oil and the milk fat diets (Table 2). The same high confidence proteins were identified when searched for Proteobacteria, Actinobacteria and Verrucomicrobia databases (data not shown).

Considering a previous high impact paper that used one minimum peptide to identify proteins in a mass spectrometry based on a fecal bacterial proteomic experiment (99), we analyzed the data using proteins identified with one peptide. Since using one peptide decreases overall confidence we further analyzed proteins that had more than one-fold change difference when compared among diets. In the corn oil diet, we found high chaperone DnaK protein (Figure 7A) and elongation factor Tu (Figure 7B). DnaK proteins have strong immunogenic properties both at the level of the innate and adaptive immune system (100). DnaK proteins have strong immunogenic properties both at the level of the innate and adaptive immune system (100). Elongation factor Tu is recognized as a new bacterial
virulence factor that helps the pathogen in complement escape and also in extracellular matrix degradation (101).

**Table 2. Summary of the identified bacterial proteome (Firmicutes and Bacteroidetes combined) in the mucosal and submucosal samples.** Proteins that had higher than 0.5 fold change values were placed in their respective dietary group when compared to the pooled reference channel. Proteins in red colour indicate that these proteins were identified with at least two minimum peptides. Proteins in black colour indicate that these proteins were identified with one minimum peptide.

<table>
<thead>
<tr>
<th></th>
<th><strong>Mucosa</strong></th>
<th><strong>Submucosa</strong></th>
</tr>
</thead>
</table>
| **Olive oil** | DNA helicase                      | Adenosylhomocysteinase **  
Phosphoglycerate kinase  
Ruminococcaceae protein  
Uncharacterized Firmicutes protein  
Unknown Bacteroides peptide  |
|             | ATP dependant zinc metalloprotease  
Phosphoglycerate kinase  
Ruminococcaceae protein  
Uncharacterized Firmicutes protein  
Unknown Bacteroides peptide  | Transaldolase  
Phosphoglycerate kinase  
Probable cytosol amino peptidase  
2,3-bisphosphoglycerate-dependant phosphoglycerate  
NADP dependent glyceraldehyde-3-phosphate dehydrogenase  
Uncharacterized Bacteroidetes protein  |
| **Milk fat** | S-layer domain protein  
Unknown Bacteroides peptide  
Uncharacterized Firmicutes protein  | Subtilisin-like serine protease  
Methyltransferase  
NADH dehydrogenase  
ABC transporter like protein  
Uncharacterized Firmicutes protein  |
| **Corn oil** | Chaperone protein DnaK **  
Cluster of ATP synthase **  
Putative GTPase  
Starch binding protein  
Pyruvate dehydrogenase  
60kDa chaperonin  
50S ribosomal protein L9  
Polyamine aminopropyltransferase  
Malate dehydrogenase  
Ubiquitin family proteins  
Acetyl-CoA acetyltransferase  
Elongation factor Tu  
GDP-mannose 4,6-dehydrogenase  
TonB receptor protein  
Uncharacterized Faecalibacterium prausnitzii  | Glutamine-fructose-6-phosphate aminotransferase  
Uncharacterized Firmicutes protein  |

We also found increased energy production indicated by higher levels of ATP synthase (Figure 7D), starch binding protein, acetyl-CoA acetyltransferase, malate dehydrogenase, pyruvate dehydrogenase (Table 2). TonB receptor protein (Figure 7C), another highly immunogenic protein that executes high-affinity binding and energy-dependent uptake of substrates into the periplasmic space (102) (i.e. iron) was also elevated in the corn oil diet. Overall, this suggests that microbial uptake of nutrients, energy production as well as pathogenicity of the microbiota are enhanced in the corn oil diet.
Figure 7. Protein level fold change comparison plots of the bacterial proteins across lipid diets. (A) Chaperone protein DnaK, (B) Elongation factor Tu, (C) TonB dependent receptor, (D) Cluster of ATP synthase, (E) Pyruvate dehydrogenase, (F) Faecalibacterium prausnitzii uncharacterized protein, (G) S-layer domain protein, (H) Subtilisin-like protease, (I) NADH dehydrogenase, (J) Transaldolase, (K) Phosphoglycerate kinase, (L) Glyceraldehyde 3-phosphate dehydrogenase (NADP). The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein. The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
In the milk fat diet group, we found increased levels of S layer domain proteins (Figure 7G). S-layer proteins protect bacteria from other bacteria or predators, provides adhesion sites. We also found increased uncharacterized Bacteroidetes and Firmicutes proteins from the mucosal samples (Table 2). Moreover, elevated subtilisin-like serine proteases (Figure 7H) and methyltransferase proteins were found from the submucosal data (Table 2). Serine proteases have been shown to protect against colitis by decreasing intestinal permeability(103). Overall, this suggests that microbial protection, attachment, and protease activity are enhanced in the milk fat diet.

Finally, olive oil promoted bacterial proteins involved in pentose phosphate pathways and glycolysis, such as transaldolases, phosphoglycerate kinase, phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenases (Figure 7 J-L). Some uncharacterized Firmicutes and Bacteroidetes, ATP dependent zinc metalloprotease proteins were also found to have higher intensities in the olive oil diet compared to the other diet groups (Table 2). This suggests that olive oil promotes the expression of microbial fermentation pathway associated proteins.

Overall, we found more bacterial protein hits in the mucosal samples compared to the submucosal samples and that the high confidence proteins were similar in both layers. However, we found more bacterial hits and their differential intensities among lipid diets when the peptide per protein threshold was lowered. Overall, we found that lipid diets differentially affected microbial protein abundance, thus had differential functions.

4.2.2 Effects of lipid diets on the metaproteome of the host tissue

Although we found fewer bacterial proteins, this iTRAQ experiment yielded 2079 high confidence mouse proteins with at least two minimum peptides from the submucosal data and 1842 high confidence mouse proteins with at least two minimum peptides from the mucosal data with 0% decoy FDR (Table 1). Therefore, host proteins dominated both the mucosal and submucosal samples compared to the bacterial proteins. Moreover, the host protein numbers are consistent with similar mass spectrometry based studies published earlier (96). To establish how the colon and mucosal proteome differ from each other, we performed GO term functional analysis on the obtained data set shown in Figure 8.
Figure 8. Gene ontology term pie chart summary for the mucosal and submucosal proteome. (A) GO term annotation pie chart for biological processes for mucosal samples. (B) GO term annotation pie chart for cellular components for mucosal samples. (C) GO term annotation for molecular functions for mucosal samples. (D) GO term annotation pie chart for biological processes for submucosal samples (E) GO term annotation pie chart for cellular components for submucosal samples. (F) GO term annotation for molecular functions for submucosal samples.
Figure 8 (A, B, C) represent GO term functional analysis for biological processes, cellular component and molecular functions from the mucosal samples respectively, whereas Figure 8 (D, E, F) represent GO term functional analysis for biological processes, cellular component and molecular functions from the submucosal samples. The GO term analysis shows that the colon and the mucosal samples share similar proteomic profiles. More mouse proteins were identified in the submucosal samples (2079) than the mucosal samples (1842). Although the submucosal and the mucosal samples share similar higher-ranking functional categories, we found differences at the protein level. For example, we found no MUC13 proteins with the set threshold (two minimum peptides) in the submucosal proteome data. However, both the MUC2 and MUC13 were found with more than two minimum peptides in the mucosal data. Interestingly, more proteins with unknown functions (>50%) are present in the mucosal samples (327) compared to the submucosal samples (110).

4.2.3 Corn oil exposure impaired the production of proteins involved in intestinal integrity

Compared to the submucosal proteome, the mucosa derived proteome contained more mucosal structure proteins such as mucin proteins. Thus, quantification results were more robust than the colon-derived proteome, since we did not find any MUC13 proteins at minimum two peptide threshold in the submucosal data. We found that MUC 2 (Figure 9A) protein was reduced in the corn oil diet, whereas MUC 13 (Figure 9B) was increased in the corn oil and milk fat diet.

Figure 9. Protein level fold change comparison plots of the mucosal proteins across lipid diets. (A) Mucin2 (0.5 fold lower than the pool), (B) MUC13 (0.5 fold higher than the pool), (C) Tight junction protein ZO-1. The total protein pool was used as a reference for all the diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
All high fat diets had the tendency to reduce MUC2 intensity. However, this effect was the largest for the corn oil diet (0.5 fold lower than the pool) compared to other diet groups. On the other hand, MUC13 had the highest intensity in the corn oil and milk fat diets, but not in the olive oil diet group. These results are important because it has been shown that MUC2 deficiency can lead to spontaneous colitis (35). Finally, we found that tight junction proteins (Figure 9C) were similarly increased compared to the pooled reference, however, the median intensity of the tight junction protein in the corn oil diet was slightly lower than the olive oil and milk fat diet groups compared to the pooled reference.

4.2.4 Corn oil exposure increased cell turnover events

In the mucosal samples derived proteins, we found 488 differentially abundant proteins (compared to the pooled reference). In order to identify proteins that were highly affected in each lipid diet, we first looked at proteins that were at least 0.5 fold change higher in corn oil diet compared to the pooled reference channel. Proteins listed in Figure 10 were differentially expressed in the colonic mucosa with at least 0.5 fold difference compared to the pooled reference.

Mucosal pentraxin, annexin 3, caspase 14, death domain protein, aldehyde dehydrogenase (Figure 10 A-E) proteins were abundant in the corn oil diet compared to the other diet groups. These proteins are associated with cell death and turnover events. Proteins that are involved in oxidative stress responses such as peroxiredoxin 6 and peroxiredoxin 5 (Figure 10 G-H) were also highly represented in the corn oil diet. A heightened immune responses were present in the corn oil diet, for example, serum amyloid A1 (acute phase molecule), increased complement component 1, CD 166 antigen (Figure 10 I-K). Among these serum amyloid A protein is shown to be associated with insulin resistance (104). Lastly, proteins that are important in the regulation of calcium ions such as S100 calcium biding proteins (Figure 10L) were abundant in the corn oil diet group as well.
Figure 10. Protein level fold change comparison plots of the mucosal proteins among lipid diets. Protein log2 fold change values that exceed at least 0.5 fold in the corn oil diet compared to the pooled reference channel. (A) Mucosal pentraxin, (B) Annexin A3, (C) Caspase 14, (D) Fas-associated death domain protein, (E) Aldehyde dehydrogenase, (F) Fatty acid binding protein 4, (G) Peroxiredoxin 6, (H) Peroxiredoxin 5, (I) Serum amyloid A, (J) Complement component 1, (K) CD166 antigen, (L) Protein S100-G. The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
Next we analyzed the submucosa derived proteome and also ranked the proteins with at least 0.5 fold change compared to the pooled reference. This analysis also showed similar responses found in the mucosal samples. This includes the heightened cell turn over event associated proteins, mucosal pentraxin, caspase 14 (Figure 11 A-B), and increased oxidative stress response shown by increased carbonic anhydrase 3 (Figure 11C). Moreover, fibronectin (Figure 11 D), a molecule that is associated with extra cellular matrix formation and wound healing (105), was decreased in the corn oil diet.

We found increased fatty acid binding protein 4 (Figure 11E), a protein that is associated with the development of insulin resistance. Other increased proteins in the corn oil diet were cluster of histone molecules and ribosomal proteins (Figure 11 G-I), including the core histone macro H2A.1, 40S ribosomal protein and histone H4 molecules. H4 molecule are shown to be major molecules contributing to death in sepsis (106). Other elevated proteins in the corn oil diet include increased fascin, immunoglobulin molecules and alpha1 anti-trypsin proteins (Figure 11 J-L). Overall, both the mucosa and colon derived proteome show similar responses to the increased consumption of n-6 PUFA diets with increased cell death and turn over events.as well as increased pro-inflammatory and oxidative stress responses as shown in Figure 11.
Figure 11. Protein level fold change comparison plots of the submucosal proteins among lipid diets. Protein log2 fold change values that exceed at least 0.5 fold in the corn oil diet compared to the pooled reference channel. (A) Mucosal pentraxin, (B) Caspase 14, (C) Carbonic anhydrase 3, (D) Fibronectin, (E) Fatty acid binding protein 4, (F) Platelet glycoprotein, (G) Core histone macro H2A.1, (H) 40S ribosomal protein S11, (I) Histone H4, (J) Fascin, (K) Ig Kappa chain C, (L) Alpha-1 anti-trypsin. The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
4.2.5 Milk fat diet was associated with higher collagen and tissue repair activity

Next we ranked the intensities of proteome fold change values in the milk fat group. Mucosa associated proteome revealed the proteins that were at least 0.5 fold different in the milk fat group compared to the pooled reference channel shown in Figure 12.

Figure 12. Protein level fold change comparison plots of the mucosal proteins among lipid diets. Protein log2 fold change values that exceed at least 0.5 fold in the milk fat diet compared to the pooled reference channel. (A) 60S ribosomal protein L13, (B) 40S ribosomal protein S23, (C) Core histone macro H2A.1, (D) Core histone macro H2B, (E) Heterochromatin protein 1 binding protein, (F) Histone H1.1, (G) Leucine-rich HEV glycoprotein, (H) Serine protease inhibitor A3N, (I) Four and a half LIM-domains protein. The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
In the mucosal proteome, we found increased mouse 40S and 60S ribosomal proteins (Figure 12 A-B), increased histone associated proteins such core histone macro H2A.1 (Figure 12 C-D) which can function as antimicrobial peptide (107, 108), increased heterochromatin binding protein (Figure 12E) that are involved in transcriptional regulation as well as increased histone linker proteins such as histone H1.1 (Figure 12F). We also found leucine rich glycoproteins (Figure 12G). Leucine rich glycoproteins can bind to extracellular matrix proteins such as fibronectin, collagen and laminin. It can also bind to TGF-beta capturing humoral factors to modulate cell adhesion locally (109). Other proteins that are increased in the milk fat diet were serine protease inhibitor A3N and four and a half LIM domains protein (Figure 12 H-I).

In the submucosal proteome, we found that collagen proteins such as collagen alpha-1, collagen alpha-2 and collagen 6A3 molecules were increased in the milk fat diet (Figure 13 A-C). These molecules are major components of extracellular matrix and involved in all stages of wound healing processes. Anterior gradient protein 2 (Figure 13D), a protein important in maintaining the epithelial barrier function (110), was increased in the milk fat diet. However, acid glycoproteins (Figure 13E), acute phase molecules expressed in response to injury or infectious stimuli, were found to be abundant in the milk fat diet. Thrombosis associated protein fibrinogen (Figure 13F) and metabolism regulator protein glucagon (Figure 13G) were also elevated in the milk fat diet. Enzymes such as carnitine o-palmitoyltransferase, corticosteroid 11 beta dehydrogenase proteins (Figure 13 H-I) had higher intensities in the milk fat diet as well. Interestingly, hydroxymethyl glutaryl-CoA synthase (Figure 13J), a protein whose expression can be stimulated by histone deacetylase inhibitors (butyrate) (111) was also increased. Other proteins heightened in the milk fat diet include serine peptidase inhibitor and pyridoxal dependent decarboxylase containing protein 1 (Figure 13 K-L). Therefore, although some pro-inflammatory responses were elevated in the milk fat diet, compensatory anti-inflammatory responses were also increased in the milk fat diet.
Figure 13. Protein level fold change comparison plots of the submucosal proteins among lipid diets. Protein log2 fold change values that exceed at least 0.5 fold in the milk fat diet compared to the pooled reference channel. (A) Collagen alpha 1 (VI), (B) Collagen alpha 2 (VI), (C) Collagen 6A3, (D) Anterior gradient protein 2, (E) Cluster of alpha-1 acid glycoprotein, (F) Fibrinogen, (G) Glucagon, (H) Carnitine O-palmitoyltransferase, (I) Corticosteroid 11-beta dehydrogenase, (J) Hydroxymethyl glutaryl-CoA synthase, mitochondria, (K) Serine (Cysteine) peptidase inhibitor, clade A, (L) Pyridoxal dependent decarboxylase containing protein 1. The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
4.2.6 Olive oil consumption was associated with higher metabolism and energy production

Finally, we ranked the high intensity proteins from the mucosal proteome in the olive oil diet group. The mucosal proteome was enriched with higher collagen, including collagen 6A3, collagen alpha 1 (VI) and alpha 2 (VI) (Figure 14 A-C), pancreatic amylase and PDZ and LIM domain proteins in the olive oil diet (Figure 14 D-F).

![Figure 14. Protein level fold change comparison plots of the mucosal proteins among lipid diets.](image)

In the submucosal proteome, we found increased protease molecules such as carboxypeptidase A1, carboxypeptidase B1, chymotrypsin-like-elastase family 1, chymotrypsin-like-elastase family 2A (Figure 15 A-E). Although proteases were increased in the olive oil diet, protease inhibitor proteins such as serpin protein (Figure 15F) was also increased. Lipid metabolism associated proteins and other pancreatic enzymes such as bile salt activated lipase, pancreatic triacylglycerol lipase, pancreatic alpha-amylase proteins (Figure 15 G-I) were also abundant in the olive oil diet. Other increased proteins in the olive oil diet were protein disulfide isomerase, ER resident protein and protein Fcgbp (Figure 15 K-L).
Figure 15. Protein level fold change comparison plots of the submucosal proteins among lipid diets. Protein log2 fold change values that exceed at least 0.5 fold in the olive oil diet compared to the pooled reference channel. (A) Carboxypeptidase A1, (B) Carboxypeptidase B1, (C) Chymotrypsinogen, (D) Chymotrypsin-like-elastase family member 1, (E) Chymotrypsin-like-elastase family member 2A, (F) Serpin, (G) Pancreatic alpha amylase, (H) Pancreatic triacylglycerol lipase, (I) Bile salt activated lipase, (J) Protein disulfide isomerase, (K) ER resident protein, (L) Protein Fcgbp. The total protein pool is used as a reference for all diets. Protein level chart shows the overall peptide label information for each protein (at least two peptides per protein). The box plot shows the relative median value and the range for the log2 fold change derived from the relative log2 fold change for each peptide from the selected protein. The whiskers are lines extending from the interquartile range representing the extreme data points or possible outliers of peptide intensities.
Overall, we found that compared to the olive oil and milk fat diet groups, most of the highly expressed proteins in the corn oil diet showed similar differential abundance both in the mucosal and submucosal proteins. These included mucosal pentraxin, carbonic anhydrase, caspases and oxidative stress response molecules. This shows that both the colon and mucosa derived proteins exhibited similar biological responses and that the effects of diets were consistent. Thus, mucosal and colonic proteome data are complementary and can be used to validate each other. Therefore, the corn oil diet is associated with higher cellular death and inflammatory responses, whereas olive oil and milk fat diets, although they have some oxidative stress induced injury, have higher levels of healing associated proteins or processes in place.

4.2.7 Corn oil diet altered epithelial adherens junctions

To understand higher ranking response pathways in response to different lipid diets, the iTRAQ results were further evaluated using ingenuity pathway analysis (IPA; www.ingenuity.com). IPA software identifies the most significant canonical pathways, biological functions, and networks in a genetic or proteomic data set (112). The data set that contained the differentially expressed proteins identified in the iTRAQ experiment was saved as ‘fold change’ values and uploaded into IPA. No expression value cutoff was selected. To identify the cellular mechanisms most related to diet effects, we first conducted a core analysis which analyzed datasets in terms of pathways, upstream regulators, molecular networks. We then used comparison analysis to understand which biological processes and pathways are more relevant in each lipid diet.

Overall, we detected more differential changes in the submucosal proteome than the mucosa-associated proteins. In the submucosal proteome, corn oil diet showed a distinct proteomic profile from milk fat and olive oil proteomic profile. After generating the pathway comparison heat map, we ranked the effects of each diet. When we ranked the effects by corn oil in descending order, both the mucosal and submucosal data showed that the highly affected pathway in the corn oil diet was the epithelial tissue remodeling pathway (Figure 16 A-B). Another pathway that was heightened in the corn oil diet was NRF2 mediated oxidative stress. Although all high fat diets were associated with increased oxidative stress levels, the corn oil diet had the strongest expression of oxidative stress related proteins, indicated by the dark orange color in the heat map (Figure 16). Other pathways that were
elevated in the corn oil diet compared to the olive oil and milk fat diets include Cdc42 signaling, calcium signaling, fMLP signaling in neutrophils, actin nucleation by ARP/WASP complex and Rac signaling.

Figure 16. Effects of lipid diets on the mucosal and submucosal proteome shown by the Ingenuity pathway comparative analysis. A) Submucosal proteome comparative analysis among lipid diets. B) Mucosal proteome comparative analysis among lipid diets. Orange color indicates the higher activation score, whereas blue color indicates a lower activation score.

We further analyzed the epithelial cell remodeling pathway, because epithelial cell remodeling is the most affected pathway. The identified molecules that contributed to the epithelial cell remodeling included actinin, cadherin, IQ motif containing GTPase activating protein, RAB5, RAB7 and vinculin proteins. Among these, expression of cadherin, RAB5, RAB7 (Table 3) are more abundant in the corn oil diet compared to the other diets.
Table 3. Proteins that contributed to the epithelial remodeling processes. Log2 fold change values are compared among lipid diets where values are normalized to the pooled samples.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Low fat</th>
<th>Corn oil</th>
<th>Milk fat</th>
<th>Olive oil</th>
<th>Location</th>
<th>Biomarker Application(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinin alpha 1</td>
<td>0.34</td>
<td>-0.33</td>
<td>0.05</td>
<td>0.11</td>
<td>Cytoplasm</td>
<td></td>
</tr>
<tr>
<td>Cadherine 1</td>
<td>-0.48</td>
<td>0.38**</td>
<td>0.07</td>
<td>-0.04</td>
<td>Membrane</td>
<td>diagnosis, disease progression, efficacy, prognosis</td>
</tr>
<tr>
<td>Catenin alpha 1</td>
<td>0.43</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>Membrane</td>
<td>diagnosis</td>
</tr>
<tr>
<td>IQ motif containing GTPase activating protein</td>
<td>0.20</td>
<td>0.13</td>
<td>0.09</td>
<td>-0.15</td>
<td>Cytoplasm</td>
<td></td>
</tr>
<tr>
<td>RAB5A, member RAS oncogene family name</td>
<td>-0.19</td>
<td>0.1**</td>
<td>-0.04</td>
<td>-0.02</td>
<td>Cytoplasm</td>
<td></td>
</tr>
<tr>
<td>RAB7A, member RAS oncogene family</td>
<td>-0.53</td>
<td>0.16**</td>
<td>0.02</td>
<td>0.04</td>
<td>Cytoplasm</td>
<td></td>
</tr>
<tr>
<td>Vinculin</td>
<td>0.60</td>
<td>-0.35</td>
<td>-0.12</td>
<td>0.12</td>
<td>Membrane</td>
<td>Unspecified application</td>
</tr>
</tbody>
</table>

Finally, IPA also indicated that calcium signaling was also relatively over represented in the corn oil diet (Figure 16). This was consistent with the individual proteins that were found abundant in the corn oil diet group. Most of the highly abundant proteins were either dependent on calcium levels or they regulate calcium levels directly to some degree. For example, mucosal pentraxin expression is modulated by calcium levels and proteins such as S100G bind to calcium ions directly. Overall, calcium dependent pathways were affected in the n-6 fatty acid rich diet group compared to the other diet groups.
Chapter 5: Discussion

5.1 Microbial metagenomics

Trillions of microbial organisms coevolved with humans and it has been recognized that they are intimately related with human health. Many factors can contribute to microbial community changes in the gut, such as diet. Although, over the past few years, studies have shown that high fat diets change the gut ecosystem, the effects of different lipid types on bacterial taxa and their functional responses are not well understood. To understand the bacterial compositional and functional changes in response to different lipid types, we analyzed the effects of dietary lipids on the colon and ileum microbiota. Although containing similar calorie load, olive oil, milk fat and corn oil diets changed the microbial composition distinctively. Overall, olive oil diet had similar alpha diversity compared to low fat diet. In contrast, corn oil diet and milk fat diet had increased alpha diversity compared to low fat diet.

The current hypothesis for microbial composition states that a low microbial diversity and high Firmicutes-to-Bacteroidetes ratio is associated with a high-calorie diet contributing to obesity (24, 47). This hypothesis links the Western diet with poorer health and greater disease risk to gut microbiota. Greater diversity of the gut microbiota is linked to better health outcomes assuming that a diverse gut microbiota is resistant to change and would be beneficial for health. Our study results are consistent with the observation that high fat diets promote Firmicutes-to-Bacteroidetes ratio. Indeed, all high fat diets except olive oil promoted the abundance of the Firmicutes phyla. Firmicutes are mostly Gram-positive (113) and butyrate is their primary metabolic end product (20). Conversely, Bacteroidetes are dominated by Gram-negative bacteria (114). It has been shown that a high fat diet is associated with increased endotoxemia in the gut (115). Bacterial endotoxins or LPS are major glycolipid components of the Gram-negative bacteria and if absorbed into circulation, would induce inflammation. This inflammation, known as endotoxemia, is increased by emulsified lipids, and therefore high fat diet is also associated with endotoxemia. Therefore, in order to decrease the endotoxin absorption, the host might selectively configure the composition of the gut bacteria in the high fat diet condition, favoring the selection of the Gram-positive bacteria over the Gram-negative bacteria. Thus, the growth or the diversification of the Gram-negative bacteria might be suppressed by the host under high fat conditions. Indeed, in our study, the metagenomics prediction using PICRUSt algorithm
revealed that a low fat diet had the highest abundance of LPS biosynthesis genes corresponding with the high levels of Bacteroidetes. This is also consistent with the relative abundance of the Gram-negative bacteria which increased in the low fat diet group. In reverse, the high fat diets had reduced Bacteroidetes with lower LPS biosynthesis and increased the Gram-positive bacteria corresponding with the dominance of the Firmicutes phyla in the high fat diet.

Firmicutes are mostly Gram-positive and are the major producers of butyrate SCFA known to have beneficial effects on the host including cancer prevention, glycemic control and balanced immunity (116). On the other hand, the Bacteroidetes phylum mainly produce acetate and propionate (117). Our study showed that levels of all SCFA (data not shown), particularly acetate and propionate, were higher in the low fat chow group corresponding with the dominance of the Bacteroidetes phyla. While SCFA production was suppressed in the corn oil and olive oil group, the milk fat group had similar levels of acetate and propionate production as the low fat control although this did not correspond with the decreased abundance of the Bacteroidetes phyla. However, the presence of SCFA did parallel with the increased energy and lipid metabolism associated genes in the milk fat group suggesting that the taxa containing these genes were enriched in this group. Genomic studies show that Bacteroidetes have more polysaccharide-degrading enzymes but they have less ABC and PTS transporters compared to Firmicutes (118), suggesting Bacteroidetes do not efficiently take up substrates for fermentation. However, Firmicutes efficiently take up nutrients such as acetate by their ABC transporters and produce butyrate as a fermentation by-product. Indeed, acetate produced by Bacteroidetes was harvested by Firmicutes to produce butyrate (118). While more studies are needed, in our study it was possible that the milk fat group resisted loss of SCFAs since the increase in Firmicutes like *Lactococcus* and *Ruminococcus* spp., major SCFA producers (119, 120), were able to utilize the acetate produced by the Bacteroidetes and harvested this into butyrate unlike the other high fat diets. While the olive oil group had increased genetic pathways involved in SCFA production, this did not translate into secreted metabolites suggesting that other mechanisms might be involved.

It is evident from this study that not only was the colonic microbiome susceptible to the type of lipid consumed. Each lipid diet promoted a specific genera of bacteria within the
Firmicutes phyla, indicating that different Firmicutes species would prefer the presence of different lipid substrates. Not only did the taxa differ among lipid diets, corresponding microbial functions were also different. For example, the corn oil diet microbiome had increased motility potential whereas milk fat diet had increased energy metabolism with increased carbohydrate and lipid metabolism.

This study provided evidence for high bacterial diversity in the case of corn oil and milk fat diet. However, the differential composition and predicted functions of the gut microbiota did not seem to link to better health outcomes with corn oil promoting bacteria with high motility potential. Therefore, our data did not support the diversity hypothesis. For example, the high fat diet represented extreme conditions compared to the low fat diet. The diversity of the low fat diet and olive oil similarly displayed lower bacterial alpha diversity, compared to the high fat diet. Evidences show that a better health outcome is associated with higher olive oil consumption (62). However, milk fat and corn oil diet increased the diversity of microbiota and there is evidence showing that corn oil and milk fat diets can promote inflammation compared to the olive oil diet (62). Therefore, diversity alone may not be a predictor for a better health in all circumstances.

5.2 Microbial proteome

To determine the true expressed functions of mucosal microbiota, we conducted an iTRAQ experiment on the mucosal and submucosal proteome. Compared to the host proteome, we were able to identify fewer proteins with higher confidence. Identified higher confidence proteins include ATP synthases, homocysteinases and DnaK chaperones. However, when we dropped the peptide per protein threshold to one, more proteins with lower confidence were identified. We continued our bacterial proteome analysis with at least one peptide per protein threshold, because one study published in Nature that investigated fecal microbial proteome had the same lower peptide per protein scenario and this was accepted in the literature (99). Reducing the threshold increased the identified protein numbers to around 300. However, most of these proteins turned out to be uncharacterized or unknown and the FDR increased up to 60% in the mucosal proteome when one minimum peptide is considered. In contrast, we found around 2000 high confidence (0% FDR) host proteins both in the mucosal and submucosal samples. These together suggested that both the mucosa and submucosa were dominated by host proteome and identification of the bacterial
proteins suffered from abundance issues. Although this proteomic experiment was not able to identify the expected level of bacterial proteome, we found more identified bacterial proteins in the mucosal samples compared to the submucosal samples. This further suggests that it is necessary to separate the mucosa from the submucosa to avoid as much host proteome as possible.

The identified bacterial proteins with known functions provided more functional information. Some proteins were not characterized, so we are unable to include them in this analysis. After analyzing the known proteins, we found that corn oil diet increased bacterial chaperone and chaperonin (with at least two minimum peptides) molecules such as DnaK. This is a chaperone molecule that is increased against oxidative or heat induced stress (121). We also found that there was an increase in ATP synthesis in the corn oil diet. Interestingly, there is a positive correlation between ATP synthesis and DnaK molecule activation (121). This might indicate that corn oil induced stress can stimulate bacterial defense against oxidative stress. In addition, DnaK is highly immunogenic and is present on the cell surface (75), therefore, the increased intensity of this molecule in the corn oil diet can induce host inflammation. This finding of increased host inflammation further suggests that corn oil diet increases bacterial virulence, which also supports PICRUSt results. Although we did not find bacterial flagella proteins as predicted by PICRUSt, bacterial virulence factors were increased in the corn oil diet. For example, PICRUSt predicted that glutathione metabolism was higher in the corn oil diet. An increase in the glutathione level is also indicative of increased protection against oxidative stress, thus increasing the potential virulence and ability to survive and stay attached to the mucosa. Other molecules that were also higher in the corn oil diet included TonB dependent receptors. TonB is a cell envelope protein and used in nutrient uptake and environmental sensing (122). The increase in its expression might indicate increased need for carbon substrates or other micronutrients to assist metabolism. The beta-barrel of TonB receptors can also act as an outer membrane virulence factor (123). Overall, bacterial virulence factors might be promoted in the corn oil diet.

Unlike corn oil, we found increased S layer proteins in the milk fat diet. Organisms with S layer protein are ubiquitous as it is one of the most abundant cellular proteins. Multiple functions of this protein have been discovered, including providing adhesion sites for bacteria, protecting cells from bacterial parasites and, providing the ability to grow in
extreme conditions (124). Both the beneficial and virulent activities of this protein have been reported. It has been shown that innate immunity is stimulated by the S-layer proteins of the Lactobacillus helveticus MIMLh5 (125). Hence, milk fat might induce the expression of bacterial proteins that have an anti-inflammatory effect on the host. S-layer proteins can also contribute to virulence when they are expressed on the surface of pathogens (124). They can confer the bacteria the ability to resist the host complement system. Other proteins that were elevated in the milk fat diet were the serine proteases. Serine proteases have been shown to decrease permeability of the epithelia towards high molecular molecules and ions (103). The increase in the serine protease activity might suggest that bacterial secretion of proteases is affected by lipid diets.

Lastly, olive oil promoted bacterial pentose phosphate pathway related proteins, such as transaldolases. Bacteria metabolize carbohydrates through glycolysis (6C sugars) or the pentose phosphate pathway (5C sugars) and converts simple carbohydrates into phosphoenolpyruvate (PEP) (126). PEP is then fermented into metabolites such as SCFAs. We also found increased glyceraldehyde-3-phosphate expression in the olive oil microbiota, suggesting that there were increased reducing equivalents produced in the form of NADH (17). In order to get rid of excessive reducing power, bacteria use fermentation pathways to produce SCFA (17). Therefore, SCFA production might be promoted in the olive oil microbiota. This observation also supports the PICRUSt prediction that SCFA metabolism (increased propionate and butyrate metabolism) is elevated in the olive oil diet. However, SCFA analysis does not support this observation suggesting that other mechanisms involved in the depletion of SCFAs might be at play.

Overall, lipid diets had differential effects on the functionality of the microbiota. Moreover, the bacterial proteome data was mostly consistent with PICRUSt predicted microbial functions. Although we were not able to compare all functions with PICRUSt predictions due to relatively fewer identified and uncharacterized proteins, the functions predicted by PICRUSt fall into the larger functional categories such as virulence and major phyla characteristics (e.g. increased LPS biosynthesis corresponding to increased Bacteroidetes in the low fat diet). Therefore, the prediction-based metagenomics could provide important information about the system being studied.
5.3 Host proteome

This proteomics experiment successfully discovered around 2000 proteins, both in the mucosal and submucosal samples. Among these, we found around 950 differentially abundant proteins in the submucosa samples and 450 in the mucosal samples. Mucosal proteome data revealed that the corn oil diet reduced the abundance of MUC2 compared to olive oil and milk fat diets, suggesting that mucosal barrier function might be weakened in the corn oil diet. On the other hand, MUC13 has the highest intensity in the corn oil and milk fat diets, but not in the olive oil and low fat diet. This is consistent with the earlier observations that increased MUC13 is associated with heightened immune responses (41) and that both milk fat and corn oil diets exhibited robust inflammatory responses to acute colitis (62). Overall, corn oil diet might impair intestinal barrier function by reducing MUC2 production. MUC2 is secreted by goblet cells, thus goblet cell function might be affected by different lipid types or indirectly through microbial changes.

Mucosal pentraxin molecules and carbonic anhydrase molecules were found to be the most abundant in the corn oil diet. A recent in vivo nutrigenomic study shows that mucosal pentraxin and carbonic anhydrase molecules are associated with epithelial proliferation and differentiation and this process is affected by the levels of calcium and iron ions (127). The authors of this paper proposed that mucosal pentraxin can be a colon-specific marker for diet induced stress. There is evidence to show that mucosal pentraxin molecules recognize and clear dead host cells (128). Mucosal pentraxin molecules are expressed largely in the colon and are strongly regulated by iron and calcium ions. Carbonic anhydrase is another molecule expressed specifically in the colonic epithelium and its function is also associated with cell death and epithelial proliferation (129).

Other interesting molecules that are differentially abundant in the corn oil diet include fatty acid binding protein (FABP4) and calcium-activated chloride channel proteins. FABP4 binds to both long chain fatty acids and the increased expression of FABP4 is positively associated with cytokine production and positive regulation of inflammatory responses (130). Interestingly, Lactobacillus can affect the expression of FABP4 in the Paneth cells (131). Several important functions of calcium-activated chloride protein include: mediating calcium-activated chloride conductance, goblet cell metaplasia and mucus hypersecretion, regulating mucus production and/or secretion by goblet cells and regulating tissue
inflammation in the innate immune response (132). The higher expression of this molecule might be to secrete more mucin proteins to compensate for the lower relative abundance of the MUC2 proteins. Overall, these results together show that corn oil diet induces stress and increases cell turnover processes such as cell death and proliferation of the mucosal cells compared to olive oil and milk fat diet.

Adhesive interaction associated molecules and calcium metabolism related proteins were also affected in the corn oil diet, namely fascins and S100 proteins respectively. Fascins are associated with actin bundles of cell-surface projections and they are involved in cell motility. It has been proposed that upregulation of this protein could be important in the invasive phenotype of the colonic carcinoma (133). S100 protein acts as intracellular Ca\textsuperscript{2+} sensors and extracellular factors. Many human cancers feature dysregulated expression of S100 family proteins. Earlier evidence indicates that S100 proteins actively contribute to proliferation, metastasis, angiogenesis and immune invasion processes. These molecules are used for cancer diagnosis and that many S100 inhibitors are being discovered and developed for cancer treatment (134). Therefore, cells in the corn oil diet might have heightened motility and oncogenic properties compared to the other lipid diets.

IPA analysis showed host responses were similar in olive oil diet and milk fat diet. In contrast, corn oil diet showed differences in several pathways. One of the most abundant pathways was the epithelial adherens junctions remodeling pathway that functions to mediate cell-cell adhesion. Disassembly of these structures is associated with lack of cell contacts and increased motility potential. The adherens junctions transmits extracellular stimuli and modulates the epithelial cell dynamics. The transmembrane protein cadherins such as E-cadherins are important cell-cell adhesion molecules in epithelia (135). The extracellular domain of E-cadherin binds to the neighboring cells’ extracellular E-cadherin domains by the stabilization of Ca\textsuperscript{2+} (calcium) ions. Normally, E-cadherin proteins are endocytosed and recycled to new cell contact sites. However, extracellular stimuli such as depleted Ca\textsuperscript{2+} levels and high oxidative stress levels increase this process and make cells more loosely attached. In our study, Rab family GTPases such as Rab5 and Rab7 molecules were abundant in the corn oil diet. These mediate the intracellular trafficking of E-cadherin for the formation of endosomes (135). This evidence combined suggests that there might be an increase in the endocytosis of adherens junctions molecules and decrease in their recycling to sites of new
cell-cell contacts. Thus, cell-cell attachment is weakened and cells might gain migratory potential in the corn oil diet compared to the other diet.

Based on the proteomic data shown here, over consumption of n-6 PUFA rich diet might weaken the adherens junction, probably by increasing endocytosis of cadherin molecules through altering extracellular calcium concentration or by increasing oncogenic molecules such as Rab family GTPases to subject cadherin molecules for destruction. Indeed, one study that investigated the effects of polyunsaturated fatty acids on calcium response showed that only the n-6 PUFA diet increased cytosolic Ca\textsuperscript{2+} concentration probably by interacting with ion channels (136). This effect was not found for saturated or monounsaturated fatty acids. Moreover, another study showed that high ratio of n-6 to n-3 PUFA consumption is correlated with lower bone mineral hip bone density both in males and females (137). All this evidence together suggests that n-6 PUFA might be involved in destabilizing cadherin mediated adherens junctions probably by affecting extracellular calcium levels.

Although we found higher oxidative stress levels in all high fat diets, unlike the corn oil diet, the milk fat diet was enriched with higher levels of collagen molecules. These molecules are important in the wound healing processes. Hence, milk fat diet might be better equipped with compensatory mechanisms for oxidative stress induced injury. Most strikingly, we found at least an eight-fold change increase in protein intensities involved in digestion and metabolism, including lipid and carbohydrate metabolism in the olive oil diet. Among many pancreatic enzymes, we found many proteases such as chymotrypsin like elastases, chymotrypsinogen, carboxypeptidases. This suggests that olive oil may enhance digestion and metabolism. The increased presence of these enzymes might directly affect bacterial communities because some proteases have direct killing activities. Indeed, the elevated levels of these enzymes correspond with the decreased microbial alpha diversity in the olive oil diet. It has also been suggested that some proteases might also have proteolytic effect on the host barrier proteomes. However, in the olive oil diet, we found increased serpin molecules which act as protease inhibitors. Therefore, the gut in the olive oil diet might also be protected against proteolytic enzyme (bacterial and host) induced injury. Overall, each lipid diet affected the host-microbe dynamic in unique ways.
Chapter 6: Conclusion

The data presented in this work show that lipid diets interact both with the host and the microbiota and have differential effects on them. Specifically, olive oil diet induced changes in the microbial community that had the potential to produce short chain fatty acids. Olive oil might also directly increase host metabolism by elevating host digestive enzymes. Microbial glycolysis and pentose phosphate pathway proteins were also promoted in the olive oil diet providing evidence for PICRUST predicted functions. Milk fat diet increased diversity of microbiota in the gut whose functions can be characterized by increased potential in carbohydrate and lipid metabolism. The host proteomic data showed that milk fat might also promote host repair associated proteins as well as bacterial defense, attachment and protease activities. Lastly, a corn oil diet increased the microbial diversity in the gut that was associated with increased potential microbial virulence such as increased microbial motility, and capacity to withstand oxidative stress. We also found that corn oil promoted virulence associated proteins namely DnaK chaperones, TonB dependent receptors and the elongation factor Tu. Host proteome data further indicated that cell death and turn over events, and oxidative stress associated proteins increased in the corn oil diet. Moreover, protective barrier protein MUC2 decreased while inflammatory MUC13 increased in the corn oil diet. Overall, diets rich in n-6 PUFA might impair intestinal barrier function over time. In addition, our data suggests that corn oil may affect cadherin mediated cell-cell adhesion by affecting Ca^{2+} regulation and increasing cellular endocytosis of cadherin molecules. We conclude that increased n-6 PUFA in the diet may be a risk factor for the development of a leaky gut. More in vivo studies are needed to investigate if diets rich in n-6 PUFA, such as corn oil, have direct effects on the intestinal barrier through calcium mediated pathways.

6.1 Limitations of this study

There are several limitations to this study outline below.

1. This study investigated the potential functional differences of microbiota among lipid diets by prediction-based methods. However, only the known bacterial functions were predicted, because the prediction was based on the already known bacterial database. This severely limited our understanding of the microbial proteome since close to 70% of the microbiome was unknown.
2. One of the goals of this study was to determine how lipid diets affect the bacterial proteome. However, the iTRAQ based proteomic experiment identified more host proteome than the bacterial proteome. This is because host proteome dominated the scraped mucosal samples. Although previously reported to be successful (95), the scraping technique was not sufficient for isolating bacterial proteome from the host.

3. Compared to the identified host proteome, the bacterial proteome coverage was much lower, thus, we were not able to confirm all the predicted functions by PICRUST. We lowered the confidence level of the microbial proteome to one peptide per protein threshold to include any bacterial protein peptides in the study. This yielded more bacterial proteome, however, the false discovery rate increased up to 60%. Therefore, the bacterial proteins identified need further validation by other methods.

4. Although we were able to identify changes due to lipid diets, we still cannot rule out the effects of lipid soluble bioactive molecules in this study.

5. Another limitation of this study was that the individual biological variations cannot be quantified due to sample pooling.

6. The identification rate of the host proteome was only 14%. This low identification rate suggests that the existing database coverage is not adequate or the post translational modification of the peptides play an important role in affecting the identification process.

### 6.2 Future work

More sophisticated mucosal bacterial isolation techniques should be developed for mucosa associated proteomic and genomic studies. The databases especially the gut specific microbial databases (proteome) should be created/curated to facilitate the search and identification processes. Software functions that can analyze microbial proteome interactions should be integrated into the currently existing analysis software. This study found a large amount of host and bacterial proteins that were affected in the lipid diets. This information should be validated to further uncover the specific interactions between the diet, bacteria and the host.
References

42. G. Bergonzelli et al., GroEL of Lactobacillus johnsonii La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen Helicobacter pylori. *Infect Immun* **74**, 425-434 (2006).


108. N. Sathyan et al., Characterization of Histone H2A Derived Antimicrobial Peptides, Harriottins, from Sicklefin Chimaera Neoharriotta pinnata (Schnakenbeck, 1931) and Its Evolutionary Divergence with respect to CO1 and Histone H2A. *ISRN Mol Biol* 2013, 930216 (2013).
115. S. Pendyala, J. Walker, P. Holt, A high-fat diet is associated with endotoxemia that originates from the gut. *Gastroenterology* 142, 1100-1101.e1102 (2012).
120. V. C. Antharam et al., Intestinal dysbiosis and depletion of butyrogenic bacteria in Clostridium difficile infection and nosocomial diarrhea. J Clin Microbiol 51, 2884-2892 (2013).
Appendices

Appendix A. Statistical analysis of the phyla ratio data

Table A.1. Kruskal-Wallis rank sum test p-values computed by comparing the Firmicutes to Bacteroidetes ratio across lipid diets in the colon and the ileum.

<table>
<thead>
<tr>
<th>Region</th>
<th>Colon</th>
<th>Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteroidetes to Firmicutes ratio</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table A.2. Firmicutes to Bacteroidetes ratio pairwise group comparison p-values. FDR corrected pairwise comparison p values using Dunn’s-test for multiple comparisons of independent samples. P value adjustment method: Benjamini–Hochberg.

<table>
<thead>
<tr>
<th>Region</th>
<th>Colon</th>
<th>Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk fat~Corn oil</td>
<td>0.592</td>
<td>0.363</td>
</tr>
<tr>
<td>Milk fat~Low fat</td>
<td>0.0004</td>
<td>0.012</td>
</tr>
<tr>
<td>Milk fat~Olive oil</td>
<td>0.289</td>
<td>0.317</td>
</tr>
<tr>
<td>Corn oil~Low fat</td>
<td>0.006</td>
<td>0.001</td>
</tr>
<tr>
<td>Corn oil~Olive oil</td>
<td>0.592</td>
<td>0.050</td>
</tr>
<tr>
<td>Low fat~Olive oil</td>
<td>0.279</td>
<td>0.344</td>
</tr>
</tbody>
</table>
Appendix B. Statistical analysis of the Gram-positive and Gram negative bacteria abundance comparison

Table B.1. Kruskal-Wallis rank sum test p-values computed by comparing the Gram-positive and Gram-negative bacteria abundance across lipid diets both in the colon and in the ileum.

<table>
<thead>
<tr>
<th>Region</th>
<th>Colon</th>
<th>Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gram-positive</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>Gram-negative</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table B.2. Gram-positive and Gram-negative bacteria abundance pairwise group comparison p-values. FDR corrected pairwise comparison p-values generated by using the Mann-Whitney-Wilcoxon test when comparing Gram-positive to Gram-negative bacteria both in the colon and in the ileum. P value adjustment method: Benjamini–Hochberg.

<table>
<thead>
<tr>
<th></th>
<th>Colon Gram-positive</th>
<th>Colon Gram-negative</th>
<th>Ileum Gram-positive</th>
<th>Ileum Gram-negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk fat~Corn oil</td>
<td>0.130</td>
<td>0.130</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>Milk fat~Low fat</td>
<td>0.001</td>
<td>0.001</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Milk fat~Olive oil</td>
<td>0.068</td>
<td>0.068</td>
<td>0.128</td>
<td>0.128</td>
</tr>
<tr>
<td>Corn oil~Low fat</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>Corn oil~Olive oil</td>
<td>0.130</td>
<td>0.130</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td>Low fat~Olive oil</td>
<td>0.022</td>
<td>0.022</td>
<td>0.128</td>
<td>0.128</td>
</tr>
</tbody>
</table>
Appendix C. Statistical analysis of alpha and beta diversity

Table C.1. Kruskal-Wallis rank sum test p-values computed by comparing the three alpha diversity index values across lipid diets both in the colon and in the ileum.

<table>
<thead>
<tr>
<th>Region</th>
<th>Colon</th>
<th>Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chao1</td>
<td>0.001</td>
<td>0.156</td>
</tr>
<tr>
<td>Shannon</td>
<td>0.003</td>
<td>0.058</td>
</tr>
<tr>
<td>Simpsons</td>
<td>0.120</td>
<td>0.076</td>
</tr>
</tbody>
</table>

Table C.2. Alpha diversity indexes pairwise group comparison p-values. FDR corrected pairwise comparison p-values using Dunn’s-test for multiple comparisons of independent samples. P value adjustment method: Benjamini–Hochberg.

<table>
<thead>
<tr>
<th>Region</th>
<th>Colon</th>
<th>Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha diversity indexes</td>
<td>Chao1</td>
<td>Shannon</td>
</tr>
<tr>
<td>Milk fat~Corn oil</td>
<td>0.061</td>
<td>0.790</td>
</tr>
<tr>
<td>Milk fat~Low fat</td>
<td>0.029</td>
<td>0.019</td>
</tr>
<tr>
<td>Milk fat~Olive oil</td>
<td>0.009</td>
<td>0.021</td>
</tr>
<tr>
<td>Corn oil~Low fat</td>
<td>0.009</td>
<td>0.023</td>
</tr>
<tr>
<td>Corn oil~Olive oil</td>
<td>0.003</td>
<td>0.035</td>
</tr>
<tr>
<td>Low fat~Olive oil</td>
<td>0.613</td>
<td>0.790</td>
</tr>
</tbody>
</table>

Table C.3. PERMOANOVA statistics on weighted and unweighted UniFrac distances for the colon and the ileum. Number of permutations: 999.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>R squared value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon (unweighted UniFrac)</td>
<td>0.326</td>
<td>0.001</td>
</tr>
<tr>
<td>Colon (weighted UniFrac)</td>
<td>0.629</td>
<td>0.001</td>
</tr>
<tr>
<td>Ileum (unweighted UniFrac)</td>
<td>0.193</td>
<td>0.001</td>
</tr>
<tr>
<td>Ileum (weighted UniFrac)</td>
<td>0.286</td>
<td>0.001</td>
</tr>
</tbody>
</table>