ASSOCIATIONS OF GUT MICROBIOTA PATTERNS WITH COLORECTAL POLYP PREVALENCE AND DIETARY FIBER INTAKE

By

TYLER CULPEPPER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013
To my family members, immediate and extended, who have always provided their unconditional love, and the mentors, peers, pupils, and friends who never stopped believing I would arrive at this point.
ACKNOWLEDGMENTS

I owe thanks to many people, whose assistance was indispensable in completing this project. First, I wish to thank my advisor, Dr. Volker Mai, for his guidance and patience throughout my degree program, and for allowing me the opportunity and providing me with the resources to complete my project. I appreciate all of his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. I am thankful for the excellent example he has provided as a successful scientist.

I also thank Dr. Bobbi Langkamp-Henken for always being available to discuss results, provide advice, and most importantly, for providing motivation and positive reinforcement through some of the most difficult periods of my project. I thank my other committee members, Dr. Claudio Gonzalez, Dr. Nematollah Keyhani, and Dr. Joseph Larkin III, for always taking time to meet with me and provide constructive criticism and recommendations to improve the quality of my project. Additionally, I thank Dr. Eric Triplett and the faculty and staff of the Microbiology and Cell Science for their support and assistance. In particular, I thank my fellow graduate students who provided much needed social stimulation and served as an outlet during difficult times.

Much of the data generated in this thesis would not have been possible without the support and generous contribution of past and present members of the Mai lab, especially Dr. Maria Ukhanova for her indispensable help with experiments and stress management and Dr. Vinata Vedam-Mai for her extensive help with coordinating the polyp study. Her help and kind words always encouraged me to continue. I would like to express my appreciation for the help of Dr. Yijun Sun and Dr. Xiaoyu Wang for their invaluable help with DNA sequencing analysis and figure preparation. I would also like
to thank previous undergraduates Madeline Bost, Fleta Netter, and Adeeb Rohani for their generous contributions of time and effort at the bench. Outside of the lab, I would like to thank the entire Henken and Dahl lab research groups for their countless hours working on our clinical trials. I also thank the National Cancer Institute, the Matsutani company, Corn Products, and General Mills for funding this research.

Lastly, I would like to thank my family for all their encouragement. I thank my parents, who ensured my emotional stability and understood my absence in many familial affairs throughout this challenging process.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>10</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>14</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>18</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 FOREWORD</td>
<td>20</td>
</tr>
<tr>
<td>2 REVIEW OF LITERATURE</td>
<td>23</td>
</tr>
<tr>
<td>Background</td>
<td>23</td>
</tr>
<tr>
<td>Normal Gut Physiology</td>
<td>24</td>
</tr>
<tr>
<td>Normal Commensal Gut Microbiota</td>
<td>25</td>
</tr>
<tr>
<td>Colorectal Carcinogenesis</td>
<td>28</td>
</tr>
<tr>
<td>Methylation</td>
<td>34</td>
</tr>
<tr>
<td>Risk Factors for CRC</td>
<td>37</td>
</tr>
<tr>
<td>Mucosal Immunology</td>
<td>42</td>
</tr>
<tr>
<td>Microbiota, Aberrant Immune Function and IBD</td>
<td>44</td>
</tr>
<tr>
<td>Evidence for Correlations of Microbiota with CRC</td>
<td>47</td>
</tr>
<tr>
<td>Means of Influencing Gut Microbiota Composition</td>
<td>52</td>
</tr>
<tr>
<td>Conclusions</td>
<td>55</td>
</tr>
<tr>
<td>3 MATERIALS AND METHODS</td>
<td>56</td>
</tr>
<tr>
<td>Stool Sample Collections</td>
<td>56</td>
</tr>
<tr>
<td>Biopsy Collections</td>
<td>56</td>
</tr>
<tr>
<td>Microbiota Analysis</td>
<td>56</td>
</tr>
<tr>
<td>Denaturing Gradient Gel Electrophoresis (DGGE) Analysis</td>
<td>57</td>
</tr>
<tr>
<td>Quantitative PCR</td>
<td>58</td>
</tr>
<tr>
<td>454-based 16S rDNA Sequencing</td>
<td>59</td>
</tr>
<tr>
<td>Metagenomic Sequencing</td>
<td>60</td>
</tr>
<tr>
<td>Fecal pH</td>
<td>61</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>61</td>
</tr>
<tr>
<td>4 GUT MICROBIOTA PATTERNS ASSOCIATED WITH COLORECTAL POLYP PREVALENCE</td>
<td>63</td>
</tr>
</tbody>
</table>
Differences in Diet and Intestinal Microflora: Potential Associations with Increased CRC Rates in African-Americans ................................................................. 63
   Introduction ........................................................................................................... 63
   Study Design ......................................................................................................... 63
   Results ..................................................................................................................... 65
      Study population demographics and dietary habits .............................................. 65
      Fecal microbiota community diversity ............................................................... 66
      Quantification of targeted bacteria ..................................................................... 66
      454 pyrosequencing analysis of 16S rDNA ......................................................... 67
      Metagenomic shotgun sequencing analysis ....................................................... 71
   Discussion ............................................................................................................... 71

5 EFFECTS OF SPECIFIC COMPLEX CARBOHYDRATES ON FECAL MICROBIOTA COMMUNITY COMPOSITION ................................................................. 104

Study #1: Resistant Maltodextrin Increases Bifidobacteria Counts in Healthy Males .................................................................................................................. 104
   Introduction ........................................................................................................... 104
   Study Design ......................................................................................................... 105
   Results ..................................................................................................................... 106
      Fecal microbiota community diversity ............................................................... 106
      Bifidobacteria qPCR ......................................................................................... 106
      454 pyrosequencing analysis of rDNA ............................................................... 107
   Conclusion ............................................................................................................. 108

Study #2: The Addition of Whole Grains to the Diets of Middle-school Children: Effects on Fecal Microbiota Community Structure ............................................... 108
   Introduction ........................................................................................................... 108
   Study Design ......................................................................................................... 110
      Subject recruitment ............................................................................................. 110
      Experimental protocol ...................................................................................... 111
      Grain-based foods and administration protocol ................................................. 111
   Results ..................................................................................................................... 112
      Fecal pH ............................................................................................................. 112
      Fecal microbiota community diversity ............................................................... 112
      Quantification of targeted bacteria .................................................................... 112
      454 pyrosequencing of 16S rRNA analysis ......................................................... 113
   Conclusion ............................................................................................................. 113

Study #3: The Effects of Galactooligosaccharides on Fecal Microbiota Composition in 3.1 Undergraduate Students and 3.2 Aged Adults .............................................. 114
   Introduction ........................................................................................................... 114
   Protocol 1: Galactooligosaccharides Supplementation in Healthy University Students .................................................................................................................. 115
      Study Design ......................................................................................................... 115
      Subject recruitment ............................................................................................. 115
      Experimental protocol ...................................................................................... 116
      GOS administration protocol ........................................................................... 117
   Results ..................................................................................................................... 117
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fecal microbiota community diversity</td>
<td>117</td>
</tr>
<tr>
<td>Quantification of targeted bacteria</td>
<td>118</td>
</tr>
<tr>
<td>454 pyrosequencing analysis of 16S rDNA</td>
<td>118</td>
</tr>
<tr>
<td>Protocol 2: Galactooligosaccharides Supplementation in Healthy Aged Adults</td>
<td>119</td>
</tr>
<tr>
<td>Study Design</td>
<td>119</td>
</tr>
<tr>
<td>Subject recruitment</td>
<td>119</td>
</tr>
<tr>
<td>Experimental protocol</td>
<td>120</td>
</tr>
<tr>
<td>GOS administration protocol</td>
<td>121</td>
</tr>
<tr>
<td>Results</td>
<td>121</td>
</tr>
<tr>
<td>Fecal pH</td>
<td>121</td>
</tr>
<tr>
<td>Fecal microbiota community composition</td>
<td>121</td>
</tr>
<tr>
<td>Quantification of targeted bacteria</td>
<td>121</td>
</tr>
<tr>
<td>454 pyrosequencing analysis of 16S rDNA</td>
<td>122</td>
</tr>
<tr>
<td>Conclusion</td>
<td>123</td>
</tr>
<tr>
<td>6 DISCUSSION AND CONCLUSION</td>
<td>154</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>187</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Disease-specific predisposition genes implicated in CRC.</td>
<td>31</td>
</tr>
<tr>
<td>4-1</td>
<td>Demographic and lifestyle data by case status.</td>
<td>78</td>
</tr>
<tr>
<td>4-2</td>
<td>Dietary differences between case groups and racial groups.</td>
<td>80</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Genetic alterations and the progression of colorectal cancer</td>
<td>29</td>
</tr>
<tr>
<td>4-1</td>
<td>Diversity indices in stool communities between cases and controls. Error bars represent the standard deviation. Total n = 114 (57 cases, 57 controls).</td>
<td>81</td>
</tr>
<tr>
<td>4-2</td>
<td>qPCR results showing no significant differences between demographic groups for lactic acid bacteria and bifidobacteria in stools</td>
<td>82</td>
</tr>
<tr>
<td>4-3</td>
<td>qPCR results showing no significant differences between demographic groups for lactic acid bacteria and bifidobacteria in biopsies</td>
<td>83</td>
</tr>
<tr>
<td>4-4</td>
<td>qPCR results showing no significant differences between BMI groups for lactic acid bacteria and bifidobacteria in stools (A) and biopsies (B)</td>
<td>84</td>
</tr>
<tr>
<td>4-5</td>
<td>qPCR results showing abundances for lactic acid bacteria and bifidobacteria between racial groups in stools(A) and biopsies (B)</td>
<td>85</td>
</tr>
<tr>
<td>4-6</td>
<td>Rarefaction curves for stool and biopsy communities based on Chao1</td>
<td>86</td>
</tr>
<tr>
<td>4-7</td>
<td>Principal coordinate analysis (PCoA) based on UNIFRAC showing differences between stool and biopsy microbiota</td>
<td>87</td>
</tr>
<tr>
<td>4-8</td>
<td>Relative abundance of bacteria phyla between cases and controls in stools (A) and biopsies (B)</td>
<td>88</td>
</tr>
<tr>
<td>4-9</td>
<td>Heat map showing the distribution of the 20 most significantly differing individual OTUs between all cases and controls in stools at the 95% similarity level</td>
<td>89</td>
</tr>
<tr>
<td>4-11</td>
<td>Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases defined conventionally and controls in stools at the 95% similarity level</td>
<td>91</td>
</tr>
<tr>
<td>4-12</td>
<td>Heat map showing the distribution of the 20 most significantly differing individual OTUs between all cases and controls in stools at the 98% similarity level</td>
<td>92</td>
</tr>
<tr>
<td>4-13</td>
<td>Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases and controls in stools at the 98% similarity level</td>
<td>93</td>
</tr>
<tr>
<td>4-14</td>
<td>Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases defined by convention and controls in stools at the 98% similarity level</td>
<td>94</td>
</tr>
</tbody>
</table>
4-15 Heat map showing the distribution of all significantly differing individual OTUs between all cases and controls in biopsies at the 95% similarity level .................................. 95

4-16 Heat map showing the distribution of all significantly differing individual OTUs between high-risk cases and controls in biopsies at the 95% similarity level ........................................................................................................ 96

4-17 Heat map showing the distribution of all significantly differing individual OTUs between all cases and controls in biopsies at the 98% similarity level .......... 97

4-18 Heat map showing the distribution of all significantly differing individual OTUs between high-risk cases and controls in biopsies at the 98% similarity level .... 98

4-19 Heat map showing the distribution of selected individual OTUs that significantly differed in stools between racial groups at the 95% similarity level ........................................................................................................ 99

4-20 Heat map showing the distribution of selected individual OTUs that significantly differed in stools between racial groups at the 98% similarity level ........................................................................................................ 100

4-21 Heat map based on prevalent OTUs in either case or control groups at the 92% similarity level ........................................................................................................ 101

4-22 Number of shotgun sequence reads with closest matches to gamma-proteobacteria in the case and the control pools .................................................. 102

4-23 Number of shotgun sequence reads with closest matches within bacterial metabolic pathways in the case and the control pools .................................................. 103

5-1 DGGE gel showing the enrichment of a particular band as a result of RM treatment ........................................................................................................ 126

5-2 qPCR results showing abundances in Bifidobacteria genome equivalents between treatment groups over time ........................................................................ 127

5-3 Rarefaction curves for RM treatment based on Chao1, which is a measure of estimated diversity if sequenced to completion ........................................................................ 128

5-4 Unifrac-based PCoA comparing RM baseline, RM post-treatment, and placebo treatment periods ........................................................................................................ 129

5-5 Heat map showing the distribution of selected individual OTUs that significantly differed in stools between time points at the 95% similarity level . 130

5-6 Differences in stool pH between WG and RG treatment groups over time ...... 131
Diversity indices in stool communities between WG and RG treatment groups over time .................................................. 132

qPCR results showing abundances of Bifidobacteria and Lactic acid bacteria (LAB) genome equivalents between WG and RG treatment groups over time. 133

Rarefaction curves for WG treatment based on Chao1 .................................. 134

Unifrac-based PCoA comparing WG and RG treatment periods .............. 135

Relative abundances of bacterial phyla in stools in WG-treated subjects .... 136

Heat map showing the distribution of individual OTUs at the 98% similarity level from subjects within the WG treatment group that significantly differed in stools between time and treatment groups ................................................................. 137

Shannon diversity indices in stool communities between treatment groups over time in the GOS student study population .................................................. 138

Inverse Simpson diversity indices in stool communities between GOS treatment groups over time in the student study population ...................... 139

qPCR results showing no significant differences in Bifidobacteria genome equivalents between GOS treatment groups over time in the student study population ................................................................. 140

qPCR results showing no significant differences in Lactic acid bacteria (LAB) genome equivalents between GOS treatment groups over time in the student study population ................................................................. 141

Relative abundances of bacterial phyla in stools in GOS-treated subjects ...... 142

Rarefaction curves for GOS treatment in the student study population based on Chao1 ................................................................. 143

Heat map showing the distribution of individual OTUs that significantly differed in stools between time and GOS treatment groups in the student study population ................................................................. 144

Measurements of stool pH between GOS treatment groups over time in the aged adult study population ................................................................. 145

Diversity indices in stool communities between GOS treatment groups over time in the aged adult study population ................................................................. 146

qPCR results showing no significant differences in Bifidobacteria and Lactic acid bacteria (LAB) genome equivalents between GOS treatment groups over time in the aged adult study population ................................................................. 147
5-23 qPCR results showing differences in Bifidobacteria genome equivalent proportions between GOS treatment groups over time in the aged adult study population .................................................. 148

5-24 The predicted proportion of bifidobacteria in fecal samples after two weeks of supplementation by the proportion of bifidobacteria in baseline samples for healthy older adults ≥65 y (A) or <65 y of age (B) .................................................. 149

5-25 Relative abundances of bacterial phyla in stools in placebo-treated subjects and GOS-treated subjects post-treatment in the aged adult study population.. 150

5-26 Rarefaction curves for GOS treatment in the aged-adult study population based on Chao1 .................................................................................................................. 151

5-27 Unifrac-based PCoA comparing GOS and placebo treatment periods in the aged adult study population.......................................................................................... 152

5-28 Heat map showing the distribution of individual OTUs that significantly differed in stools between collection periods in the GOS-treated group in the aged adult study population.................................................................................. 153
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>µg</td>
<td>Microgram</td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
</tr>
<tr>
<td>µM</td>
<td>Micromolar</td>
</tr>
<tr>
<td>B. fragilis</td>
<td>Bacteroides fragilis</td>
</tr>
<tr>
<td>B. lactis</td>
<td>Bifidobacterium lactis</td>
</tr>
<tr>
<td>B. longum</td>
<td>Bifidobacterium longum</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>AA</td>
<td>African-American</td>
</tr>
<tr>
<td>ACF</td>
<td>Aberrant crypt foci</td>
</tr>
<tr>
<td>AOM</td>
<td>Azoxymethane</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the ROC curve</td>
</tr>
<tr>
<td>CA</td>
<td>Caucasian-American</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-dependent kinase</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylator phenotype</td>
</tr>
<tr>
<td>COX</td>
<td>Cyclooxygenase</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DF</td>
<td>Dietary fiber</td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing gradient electrophoresis</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EPIC</td>
<td>European Prospective Investigation into Cancer</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td><em>E. faecalis</em></td>
<td><em>Enterococcus faecalis</em></td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
</tr>
<tr>
<td>g</td>
<td>Gram</td>
</tr>
<tr>
<td>GALT</td>
<td>Gut-associated lymphoid tissue</td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td>GOS</td>
<td>Galactooligosaccharides</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>IgA</td>
<td>Immunoglobin A</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>L</td>
<td>Liter</td>
</tr>
<tr>
<td><em>L. acidophilus</em></td>
<td><em>Lactobacillus acidophilus</em></td>
</tr>
<tr>
<td><em>L. rhamnosus</em></td>
<td><em>Lactobacillus rhamnosus</em></td>
</tr>
<tr>
<td>LAB</td>
<td>Lactic acid bacteria</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase pairs</td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>MG-RAST</td>
<td>Metagenomics Rapid Annotation using Subsystems Technology</td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>mm</td>
<td>Millimeter</td>
</tr>
<tr>
<td>mM</td>
<td>Milimolar</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatch repair</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
</tbody>
</table>
ng  Nanogram
NK  Natural killer
NKT  Natural killer T
OTU  Operational taxonomic unit
PBS  Phosphate buffered saline
PCoA  Principle coordinate analysis
PCR  Polymerase chain reaction
PSA  Polysaccharide A
qPCR  Quantitative PCR
RDP  Ribosomal database project
ROC  Receiver operator characteristic
rRNA  Ribosomal ribonucleic acid
SEM  Standard error of the means
SFB  Segmented filamentous bacteria
*S. bovis*  *Streptococcus bovis*
*S. gallolyticus*  *Steptococcus gallolyticus*
SPF  Specific pathogen free
T1D  Type 1 diabetes
TAE  Tris-acetate-EDTA
TGF  Transforming Growth Factor
T_H  T Helper
TLR  Toll-like receptor
TNF  Tumor necrosis factor
T_{reg}  Regulatory T cell
Tris  Tris(hydroxymethyl)aminomethane
<table>
<thead>
<tr>
<th>TST</th>
<th>Thiosulfate sulfur transferase</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>v / v</td>
<td>Volume to volume</td>
</tr>
<tr>
<td>w / v</td>
<td>Weight to volume</td>
</tr>
</tbody>
</table>
ASSOCIATIONS OF GUT MICROBIOTA PATTERNS WITH COLORECTAL POLYP PREVALENCE AND DIETARY FIBER INTAKE

By

Tyler Culpepper

August 2013

Chair: Volker Mai
Major: Microbiology and Cell Science

Colorectal cancers (CRCs) remain among the most frequently detected cancers. Higher CRC prevalence and mortality are observed in African Americans (AAs). One of the unique characteristics of the colon environment is the presence of a diverse and metabolically active commensal microbiota. Sequencing technologies allow for an in depth analysis of associations between microbiota composition and various diseases. This work tested the hypothesis that gut microbiota composition is associated with colorectal polyp presence. Furthermore, the hypothesis that differences in microbiota composition are associated with the increased CRC risk observed in AAs was tested. Data on dietary and medical history, a fecal sample, and multiple colon biopsy samples were collected from 126 subjects undergoing a screening colonoscopy. A 16S rDNA sequencing-based microbiota analysis in 30 individuals presenting with at least one polyp and 30 matched controls revealed that bacterial signatures differed between groups. The most significant of these differences were detected in subjects with high risk polyps, strengthening the correlation with carcinogenesis. Furthermore, using a discriminant analysis a microbial pattern predictive of polyp status was detected. These data are consistent with the hypothesis that microbiota is associated with CRC risk. We
detected differences in specific bacterial signature sequences but not in overall diversity between AAs and Caucasian Americans (CAs).

A parallel goal was to determine the effects of fibrous dietary substrates on gut microbiota composition. Because epidemiological evidence suggests that dietary fiber intake decreases CRC risk, the effects of various fibrous substrates on gut microbiota composition were determined. First, resistant maltodextrin (RM) was shown to increase some butyrate-producing bacteria and also bifidobacteria, both of which have previously been implied in protection against CRC development. Second, whole grains (WG), which contain a mixture of dietary fibers, were shown to increase lactic acid bacteria (LAB), which are thought to promote gut health. Lastly, galactooligosaccharides (GOS) increased the proportions of bifidobacteria in aged adults but not young adults. Though no fibrous substrate changed overall gut microbiota composition markedly, all substrates affected the prevalence of several operational taxonomic units (OTUs). Each fiber studied may be used as means to influence gut microbiota composition.
Colorectal cancers are commonly diagnosed and are a leading cause of cancer-related death. In the United States in 2009, 136,717 new cases were reported with 51,848 deaths (http://apps.nccd.cdc.gov/uscs/toptencancers.aspx). Current screening techniques, including the colonoscopy procedure, are invasive, while others, including the fecal occult blood test, are non-specific. Patients undergoing a colonoscopy risk perforation of the large bowel. Furthermore, compensation for the team of healthcare providers needed for this procedure increases the economic burden on the patient. Therefore, the development of a non-invasive specific screening test is needed. Though many risk factors have been identified, a large proportion of cases are unexplained. It is thought that 65% of CRC arise from environmental factors; however, the exact nature of these environmental factors is not fully understood (69). Understanding the etiology of CRC will allow for the design of more effective screening and intervention regimens.

The work presented here is based on the hypothesis that resident members of the normal gut microbiota may contribute to CRC development. While many gut organisms are shared among most individuals, a unique intestinal environment exists within each individual due to the variation in abundance and distribution of these organisms. Individuals belonging to the same demographic groups or individuals who share particular diseases may display similarities in gut microbiota composition based on the particular disease present (26, 61, 204, 211, 226, 227). Thus, the gut microbiota composition may be an environmental factor that contributes to CRC development. The hypothesis will be tested by determining the differences between the gut microbiota communities of individuals with polyps (a preneoplastic lesion that is a marker for CRC
risk) and individuals without polyps. Selecting individuals with early lesions, such as polyps as opposed to fully blown cancer, is important because cancers can change the gut environment (particularly immune surveillance, which likely has a strong impact on microbiota). Furthermore, cancerous tissue often stimulates angiogenesis and increases blood flow to the affected area. This can change the concentration of nutrients available to the microbiota, and thus affect the abilities of particular populations of organisms to survive and/or grow. Cancers can also cause blockages and likewise affect the amount of nutrients available. However, a caveat of studying polyps is that not all polyps develop into cancer. Roughly 12% of adenomatous polyps become cancerous and less than 5% lead to invasive cancer (213).

AAs suffer from an increased burden of CRC. This association led to the hypothesis that differences in the microbiota composition of AAs contribute to increased risk for CRC development. The work presented here determined both the differences between the gut microbiota community composition of subjects with polyps and polyp-free subjects and the differences between AAs and CAs. The differences between racial groups did not relate to the differences found between subjects with polyps and polyp-free subjects; however, the development of a stool-based test for aberrant microbiota patterns based on the differences between subjects with polyps and poly-free subjects would offer a non-invasive risk assessment tool that would increase earlier detection and benefit all racial groups. Thus, diagnosis and treatment efforts could begin earlier and reduce mortality from this disease.

Beyond the polyp study, other studies were aimed at modifying microbiota compositions towards a beneficial composition. Evidence in the literature that suggests
lower CRC risk for individuals that consume higher amounts of dietary fiber led to the hypothesis that potentially beneficial members of the resident gut microbiota interact positively with the fibrous substrates that escape host digestion and reach the colon (22, 143, 205). These interactions, which may include stimulation of growth of potentially beneficial microbes or an increase in certain metabolic activities, may lead to a gut environment that is less conducive to CRC development. Such favorable interactions may include the production of butyrate, which may contribute to proper cell cycle regulation, and modulation of the immune system away from an inflammatory state. Several studies involving various fibrous substrates showed that microbiota is changed as a result of fiber supplementation. The community compositions resulting from such change were compared with the community composition of polyp-free subjects and showed that while each substrate changed gut microbiota community composition, no substrate was able to shape communities such that they displayed any remarkable similarity to communities in polyp-free patients.

The discovery of aberrant gut microbiota patterns in subjects with polyps would provide one possible mechanism to the development of polyps and would allow for the development of the screening method based on such microbiota patterns. Coupled with results associating the composition of gut microbiota communities in subjects without polyps to gut microbiota communities after fiber supplementation, this work would also demonstrate means of influencing gut microbiota composition. Combined, these results would contribute to earlier detection and increased prevention, and would contribute to the current knowledge of CRC development, resulting in a decrease in CRC morbidity and mortality.
CHAPTER 2
REVIEW OF LITERATURE

Background

Colorectal cancers remain among the most frequently observed and fatal malignancies worldwide (http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/index.php). Mutations in many genes resulting from genetic predispositions or various environmental exposures have been associated with increased CRC risk (31, 62, 167, 200, 232, 234). Dietary habits and the resulting differences in gut microbiota have been suggested to contribute to the increase CRC risk (45, 116, 137, 138, 155, 202). Chronic inflammation and obesity, which both can affect gut microbiota composition, have also been linked to CRC (165, 231, 257). A large body of evidence supports the idea that altering gut microbiota can change host physiology (88, 194, 195, 257). Abnormal activation of immune responses towards commensal microbes is thought to contribute to inflammatory bowel diseases (IBD) that are known to increase CRC risk (231). Many studies also support the idea that changes in host diet, especially the addition of fermentable dietary fiber (DF), can modulate the gut microbiota towards a more protective composition (150, 151).

Establishing a clear link between gut microbiota and colorectal carcinogenesis would be crucial for the development of novel prevention approaches that could be directed at both early detection of aberrant microbiota patterns indicative of increased CRC risk and shaping microbiota towards a low-risk composition by dietary means. Emerging evidence has shown that changes in diet can quickly affect microbiota composition (77). To provide a clear understanding on the development of CRC, aspects of normal colorectal physiology and changes occurring during CRC...
development will be reviewed first. As various links between CRC, inflammation, and gut microbiota have been established, this review will then describe in some detail host immunity, especially at the mucosal gut interface, as it relates to CRC. Finally, the current evidence for associations between CRC and gut microbiota will be discussed.

**Normal Gut Physiology**

The gastrointestinal (GI) tract consists of several accessory glands, including salivary glands, the pancreas, and the liver, which are linked to the alimentary canal which stretches from the mouth to the anus. Beginning at the most proximal site and continuing distally, these luminal organs consist of the mouth, oropharynx, esophagus, stomach, small intestine and large intestine. The small intestine is further divided into the duodenum, jejunum, ileum, while the large intestine is further divided into the cecum, ascending colon, transverse colon, descending colon, and rectum. Together, the organs of the GI system serve to digest and absorb nutrients, electrolytes, and fluids to maintain metabolic homeostasis (21).

The luminal organs along the length of the GI tract beyond the mouth are characterized by the presence of several layers which include, beginning from the most apical layer, the mucosa (which consists of the epithelium and the lamina propria), muscularis mucosae, submucosa, muscularis externa, and serosa. In the healthy gut, commensal microbes are limited to the luminal contents where they interact primarily with the epithelium and lamina propria (21).

Both the small and large intestines have increased surface areas due to the three structural features; (1) macroscopic semi lunar folds known as folds of Kerckring, (2) finger-like projections known as villi (found only in the small intestine) surrounded by inwardly juxtaposed structures known as crypts of Lieberkuhn (found in both the small
and large intestine), and (3) microvilli present on the membranes of epithelial cells. Intestinal crypts are composed mainly of absorptive epithelial cells and goblet cells, the latter of which functions in mucin secretion. Intestinal crypts throughout the entire gut are characterized by the presence of stem/progenitor cells and enteric endocrine cells at their bases (21). Progenitor cells multiply at the base of crypts and migrate up the crypt-villous axis, where within at most a few days after reaching the luminal border they are sloughed off as exfoliated colonocytes. This process results in a constant turnover of the large intestinal epithelium (21). Maintenance of structural integrity of the epithelium layer is crucial to avoid leakage of luminal contents. Epithelial permeability is regulated through paracellular resistance, which is dependent upon highly regulated tight junctions between epithelial cells (21). The colonic epithelium observed in mouse models provides three distinct layers of protection against luminal microbes. A 150 µM viscous mucus coating separated into a diffuse outer layer accessible to bacteria and a dense inner layer resistant to bacterial colonization exists atop a layer of intestinal epithelial cells (120). Mucus primarily consists of heavily O-glycosylated proteins known as mucins. Mucins form oligomers which retain water in the glycoprotein matrix (157).

**Normal Commensal Gut Microbiota**

The colon harbors up to one hundred trillion bacteria that constitute the commensal human gut microbiota (203). The human microbiome is estimated to contain up to 100-fold more genes than the human genome (111). Microbial communities differ by anatomical site along the colon and their location either within the lumen (fecal samples) or attached to or penetrating the mucin layer (mucosa) (71). While the stomach and duodenum harbor roughly $10^1 - 10^3$ organisms per milliliter of luminal
contents, the small intestine and the large intestine contain an estimated $10^4 - 10^7$ and
$10^{11} - 10^{12}$ organisms, respectively (171).

While many of the gut microbes represent commensal organisms that live in a
beneficial balance within the protected host environment, others can become
pathogenic (111). The exploration of the microbiome has generated the concept of
humans as “super organisms” in which the combined microbial gene pool contributes
functions that the host could not accomplish alone (87). While microbiota composition
differs considerably between individuals, important microbial functions appear more
conserved (57). Various functions have been attributed to the microbiota including
digestion of complex carbohydrates (such as those in DF), synthesis of vitamins, and
modification by deconjugation, dehydrogenation, and dehydroxylation of primary bile
acids (5, 86, 163).

Through extensive efforts associated with the Human Microbiome Project and
related projects, large advances have been made in our understanding of microbiota
variation over time and between individuals. These projects have shown that microbiota
varies greatly between body sites, and to a lesser extent within body sites between
individuals. The gut microbiota is dominated by Firmicutes, Bacteroidetes,
Actinobacteria, and Proteobacteria at the phylum level (113). Variation increases at
lower taxonomic levels, with thousands of operational taxonomic units (OTUs) detected
between individuals (229). However, some studies suggest the presence of a microbial
core, or group of organisms found in the majority of individuals (229, 236). Based on the
analysis of a large metagenomic data set, the existence of three distinct enterotypes of
microbial colonization pattern has been suggested (11, 88).
The integrity of tight junctions is compromised in many autoimmune disorders and leads to a condition known as leaky gut syndrome (78). This condition promotes bacterial translocation from the gut into circulation, which results in systemic immune responses and increases the likelihood of sepsis (32). It is unclear as to whether potentially harmful bacteria first breach the epithelial barrier and generate inflammation which then leads to leaky gut syndrome or if the syndrome develops first facilitating the subsequent translocation of bacteria. Dysbiosis of intestinal microbiota has been associated with autoimmune diseases including eczema and IBD (92, 227). Likewise, many other non-autoimmune conditions including type 2 diabetes, obesity, and celiac disease have been linked with aberrant microbiota patterns (65, 136, 237). Recent work has linked distortions in normal gut microbiota establishment in preterm infants with necrotizing enterocolitis, a disease characterized by extensive intestinal inflammation (152).

Despite the recently accumulated wealth of evidence for correlations between various disease states and microbiota composition, the evidence for any causal links is sparse. While observational studies allow for the initial establishment of correlations, showing causality requires microbiota analysis in multiple prospectively collected samples. Without such evidence it is not at all obvious if differences in microbiota observed before and after diagnosis or between cases suffering from the respective disease and matched controls are due to the disease or contributed to its development. The large variation in gut microbiota composition between individuals and within individuals over time indicates the requirement for a large sample size. A nested case/control design, with a prospective collection of a series of microbiota samples from
populations at increased risk for the disease of interest offers the most efficient study design. Such studies will be especially difficult for chronic diseases that develop over many years or decades, such as CRC. However, without such studies it is unlikely that causal contributions of the microbiota can be determined. While animal models have served well to prove the general principle that microbes can contribute to carcinogenesis, the specific findings of microbiota contributions cannot likely be extrapolated to humans.

**Colorectal Carcinogenesis**

Colorectal cancer is thought to initiate in crypt stem cells that form aberrant crypt foci (ACF), regions of cell hyper proliferation at the base of intestinal crypts (24). ACF are found more often in rodents with intestinal polyps and are thought to initiate polyp development (228). Polyps, precancerous legions characterized by hyper proliferation and lack of differentiation that vary in their genetic makeup and potential to progress to CRC, can develop into either hyperplastic polyps or adenomas. Although earlier only adenomas were considered as precancerous, evidence suggests hyperplastic polyps may also be a threat (117). Polyps then develop into cancerous legions known as carcinomas classified into four stages (83). Polyp development is displayed in Figure 2-1 from Volgelstein, et al.,(243).
Figure 2-1. Genetic alterations and the progression of colorectal cancer.
Polyps of interest for studying CRC development because 95% of CRCs arise from polyps (115). This risk factor weighs more heavily than lifestyle risk factors and is therefore a desirable marker for prospective cohort studies; however, approximately 5% of polyps develop into cancer. Though such low incidence is not ideal for CRC development, it is important to study this environment to determine causal associations. Studies conducted in CRC patients cannot discern whether environmental changes contributed to CRC development or if CRC contributed to the environmental changes.

Like other cancers, CRC arises from abnormalities in proliferation, differentiation, or apoptosis due to either genetic or epigenetic factors that can be inherited or acquired from environmental exposures. Table 2-1 from Kilpivaara and Aaltonen shows many genes commonly associated with CRC (126). Furthermore, mutations of *KRAS* and *BRAF* genes, a GTPase and serine/threonine kinase, respectively, which activate the mitogen-activated protein kinase (MAPK) pathway, are found abundantly in colorectal cancers (31, 63). *PI3KCA* encodes the catalytic subunit of phosphatidylinositol 3-kinase (PIK3), which is important for cell proliferation and survival, and has been found to be mutated in 32% of colorectal cancers (199, 200).
<table>
<thead>
<tr>
<th>High-penetrance colon cancer syndrome</th>
<th>Predisposition gene(s)</th>
<th>Gene function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch syndrome</td>
<td>mutL homolog 1 (MLH1)</td>
<td>DNA mismatch repair</td>
</tr>
<tr>
<td></td>
<td>mutS homolog 2 (MSH2)</td>
<td>DNA mismatch repair</td>
</tr>
<tr>
<td></td>
<td>mutS homolog 6 (MSH6)</td>
<td>DNA mismatch repair</td>
</tr>
<tr>
<td></td>
<td>postmeiotic segregation increased 2 (PMS2)</td>
<td>DNA mismatch repair</td>
</tr>
<tr>
<td>Familial adenomatous polyposis</td>
<td>adenomatous polyposis coli (APC)</td>
<td>Wnt signaling</td>
</tr>
<tr>
<td>Peutz-Jeghers syndrome</td>
<td>serine/threonine kinase 11 (LKB1/STK11)</td>
<td>Controlling the activity of adenosine monophosphate– activated protein kinase (AMPK) family members</td>
</tr>
<tr>
<td>Juvenile polyposis</td>
<td>SMAD family member 4 (SMAD4)</td>
<td>Signal transduction of the transforming growth factor–β (TGF-β) superfamily and bone morphogenetic proteins (BMPs) TGF-β signaling</td>
</tr>
<tr>
<td>MYH-associated polyposis</td>
<td>mutY homolog (MUTYH)</td>
<td>DNA base excision repair</td>
</tr>
<tr>
<td>Colorectal cancer and familial tooth agenesis</td>
<td>axin 2 (AXIN2)</td>
<td>β-catenin/Wnt signaling</td>
</tr>
<tr>
<td>Polymerase proofreading– associated polyposis</td>
<td>polymerase delta 1, catalytic subunit (POLD1)</td>
<td>Catalytic and proofreading subunit of DNA polymerase δ1</td>
</tr>
<tr>
<td></td>
<td>polymerase epsilon, catalytic subunit (POLE)</td>
<td>Catalytic subunit of DNA polymerase ε</td>
</tr>
</tbody>
</table>
The activation of the Wnt signaling pathway, important for embryonic development and cell cycle regulation, is often thought to be the initiating event in CRC. Various mutations and deletions in the adenomatous polyposis coli (APC) gene have been detected in CRC models. One study showed that roughly 80% of colorectal adenomas possessed APC mutations, while another study in CRC cell lines showed that about 80% of the cell lines studied had truncated forms of APC. (118, 219). The condition in which these mutations are inherited is known as familial adenomatous polyposis coli (FAP). FAP patients often display up to thousands of polyps along the length of their colons (168). This gene acts upstream of many cell cycle regulation proteins through its interaction with β-catenin, a transcription factor to which it binds and targets for degradation (154). Colorectal tumors with intact APC have been shown to have activating β-catenin mutations, demonstrating the ability of either type of mutation to induce hyper proliferation (167). The expression of the transcription factor c-myc, a known oncogene, is induced by β-catenin (102). Nuclear expression of β-catenin has been correlated with c-myc expression, both of which have been correlated with adenoma size (30). The interactions of c-Myc with many genes, including cell cycle regulatory genes, have been demonstrated; however, many of these interactions are not clearly defined (60). c-Myc has been shown to induce the expression of cyclin D, a driver of cell cycle progression, and its activating counterpart, cyclin-dependent kinase 4 (CDK4). Cell culture experiments have demonstrated the dependence of CDK4 expression on c-Myc and an increase in cellular proliferation correlated with CDK4 production (104). Furthermore, c-Myc has been shown to inhibit the expression of p21, a CDK inhibitor (53). Additionally, loss of the CDK inhibitor p27 is associated with
decreased stage 3 CRC survival (17). These interactions demonstrate an oncogenic mechanism by which mutations in the Wnt signaling pathway lead to CRC.

Mutations in cell cycle regulatory proteins are often detected in CRC. Mutations in the cell-cycle arrest gene p53 have been detected in 50% of colorectal cancers (234). Wild-type p53 transfected into CRC cell lines has been shown to inhibit cell proliferation with up to 10-fold greater efficiency than mutant p53 (14). A population based analysis revealed that p53 mutations are more common in the distal colon than the proximal colon and are less likely to occur with KRAS mutations or microsatellite instability (MSI) (198). Mutation of p53 is thought to be a late transformation event as it is rarely detected in adenomas but can be found abundantly in carcinomas (15). However, mouse models in which human gastrin was overexpressed demonstrated an increased number of ACF in p53 mutants, which suggests p53 mutation is an early event (185). These data may conflict due to the difference between models or other genetic factors and demonstrate the involvement of many independent genetic abnormalities in CRC development.

Genes involved in growth factor pathways involving prostaglandins are also important for CRC development. Transforming growth factor β (TGF-β) receptors are commonly mutated in CRC (95). Furthermore, mutations of downstream MAD-related proteins in the TGF-β signaling pathway have likewise been observed in CRC (75). Additionally, mutations in epidermal growth factor, involved in the MAPK/ERK pathway, have also been associated with CRC (251).

The activation of prostaglandin signaling, notably with COX-2, is also associated with CRC. Prostaglandin E2 contributes to CRC development through activation of PI3K
and subsequently β-catenin (43). The effect of COX-2 is ameliorated in regular aspirin users (44). Studies conducted in CRC cell lines have also demonstrated an increase in COX-2 expression (48). Furthermore, a decrease in the COX-2 inhibitor 15-PGDH, hydroxyprostaglandin dehydrogenase, has been observed in CRC and adenomas (258). Mutations in these CRC-associated genes can affect cell turnover, mucin production, secretion of antimicrobial peptides, thus changing luminal conditions affecting gut microbiota composition and activities.

Enzymes involved in DNA repair are often mutated or silenced in CRC. Germ-line mutations in the base excision repair gene MUTYH have been observed in CRC (4). Additionally, germ-line mutations in DNA mismatch repair genes hMLH1 and hMSH2 lead to a condition known as hereditary nonpolyposis colon cancer, or Lynch syndrome, and lead to an increased risk of CRC (25). This condition can also develop as a result of promoter hypermethylation described in detail below. Adherent E. coli have been shown to decrease expression levels of DNA mismatch repair proteins MLH1 and MSH2 (149, 149).

**Methylation**

Epigenetic modifications have been shown to occur in response to environmental changes and thus might also correlate with activities of the gut microbiota, especially in the gut epithelium. DNA methyl transferase enzymes, particularly DNMT1, are responsible for CpG methylation through a reaction with the methyl donor S-adenosylmethionine (187, 188). CpG islands, frequently located in the promoters of eukaryotic genes, are usually unmethylated. In contrast, CpG dinucleotides not associated with CpG islands are frequently methylated. (23). The number of CpG islands in the human genome has been estimated to be roughly 45,000 (7). Promoter methylation is generally
correlated with gene silencing, and reversing methylation using 5'-deoxyazacytidine has been shown to restore gene expression (121, 159).

According to the Knudson two-hit hypothesis, once both alleles of a tumor suppressor gene have become either methylated in their promoter regions or mutated, malignant transformation of the cell may begin (131). Hypermethylation of the promoters of both alleles is common in sporadic colon cancers; however, it is less common in familiar cancers as the germ line mutation is the first hit. (76). Loss of function in tumor suppressor genes due to promoter methylation is observed frequently (121).

Several mechanisms exist by which methylation inhibits gene expression, the first of which is through mutation. Methylation of cytosine residues promotes cytosine-to-thymidine point mutations via deamination at the 4' location of the cytosine ring (191). The remaining mechanisms depend upon protein interactions with the methylated cytosine. Direct binding of methyl-cytosine-binding transcription repressors can inhibit the formation of the open promoter complex (214). Additionally, methylated DNA can recruit histone deacetylases, the activity of which subsequently leads to transcriptional repression, as deacetylation of histones also inhibits the open promoter complex (206). Histone deacetylase inhibitors alone are not able to restore the expression of genes with methylated promoters, but work synergistically with demethylating agents such as 5'-deoxyazacytidine to restore gene expression (39). This suggests that methylation, not deacetylation, is the dominant suppressive epigenetic modification.

Promoter hypermethylation of several genes has been associated with colorectal carcinogenesis. Inactivation of hMLH1 via promoter methylation leads to MSI, and is associated with sporadic CRC (164). Reversal of hMLH1 promoter methylation has
been shown to restore expression of the HMLH1 protein and subsequently mismatch repair (MMR) gene capacity in MMR-deficient cell lines (103). CpG island methylator phenotype (CIMP), which has been defined as a condition in which three or more CpG island loci in colorectal tumors are methylated, has been associated with hMLH1 methylation and MSI, and is thought to contribute to MSI development. Though CpG island methylation has been found in normal tissue and is associated with increasing age, some CpG island methylation has been found only in cancerous tissues independent of age (235). Though a molecular mechanism by which CIMP develops has not yet been elucidated, Ogino et al have developed a panel of eight genetic markers that effectively predict CIMP status (174).

Similarly, the promoters of several other genes have been shown to be highly methylated in CRC, including tumor suppressor gene p16, serine proteinase inhibitor tissue factor pathway inhibitor-2, Netrin-1 receptor UNC5C, and E3 ubiquitin ligase HACE1 (94, 106-108). Burri et al reported tumor-specific methylation of both p16 and p14, with p16 methylation found primarily in poorly differentiated proximal colon adenocarcinomas and p14 methylation found primarily in well differentiated distal colon adenocarcinomas (37). The occurrence O-6-methylguanine-DNA methyltransferase (MGMT) promoter methylation is significantly increased in tumors and adjacent normal tissues of sporadic CRC subjects (210).

Hypomethylation of genes, particularly growth-related genes, also contributes to the development of CRC. Sharrard et al found that the oncogene c-myc was more frequently hypomethylated in adenocarcinomas, metastatic deposits, adenomatous polyps, and hyperplastic polyps compared to normal mucosal tissue (209). H-RAS has
also been shown to be hypomethylated in colonic tumors (80). Studies of potential associations between microbiota composition and methylation status are currently ongoing but have not yet been reported in the literature. Once such an association is established the next step would be to determine if changes in microbiota composition, through targeted dietary interventions that might include prebiotics, probiotics or synbiotics, can reverse methylation pattern in the intestinal epithelium.

**Risk Factors for CRC**

For average-risk individuals aged 50 years or older, the American Cancer Society currently recommends one of four routine CRC screening options: (1) fecal occult blood testing every year, (2) flexible sigmoidoscopy every 5 years, (3) double contrast barium enema every five years, or (4) colonoscopy every 10 years (220).

Common factors influencing CRC risk include physical activity, diet, family history, obesity, smoking, and aspirin use (232). Many of the dietary factors associated with CRC that are discussed below have been or likely could be linked to microbiota composition (151).

Obesity has been positively associated with CRC risk (165). Data suggest associations between microbiota and obesity, although causality is still not well established (236, 237). Similarly, smoking has been shown to have a positive dose-dependent association with CRC risk (46). Inverse associations between CRC and both anti-inflammatory medications and physical activity have been established (52). A case-control study conducted by Sandler et al found that daily aspirin use significantly reduced the occurrence of adenomas in patients who previously had CRC (201). Rofecoxib, another cyclooxygenase inhibitor, has likewise been shown to decrease the
occurrence of adenomas (184). Again, both exposures are likely to affect microbiota composition.

The evidence regarding associations of dietary fiber intake with CRC risk is not fully consistent. A strong association between DF intake and CRC would be consistent with a role of microbiota in CRC as DF can change microbiota composition (150). The analysis of several prospective cohort studies showed that there was an inverse relationship between CRC and dietary fiber intake using an age-adjusted model, but no significant relationship was seen in fully adjusted models (179). The National Institute of Health – AARP Diet and Health Study found that a lower consumption of meat and potatoes and a higher consumption of fruits and vegetables was associated with lower CRC risk (82). Data from this study also showed that both whole grain intake and intake of fiber from whole grains were significantly associated with a decrease in CRC risk (205, 205). Larsson et al likewise observed that the consumption of whole grains was associated with decreased CRC risk (137). However, Terry et al showed that the consumption of larger amounts of fruit and vegetables was significantly associated with decreased CRC risk while fiber consumption from cereal was not (230). Several studies, including the Netherlands Cohort Study on Diet and Cancer and the Tennessee Colorectal Polyp Study, likewise found a significant association between fruit and vegetable intake and decreased CRC risk (189, 244, 252). The Women’s Health Study showed no significant associations of fiber intake from fruits and vegetables with CRC risk, but showed a significant association of legume fiber intake with decreased CRC risk (143). Data from the European Prospective Investigation into Cancer and Nutrition (EPIC) study showed that a higher consumption of dietary fiber was associated with
lower CRC risk, but the food source of the fiber was not important (22). Methodological differences may account for this discrepancy, as one study found that fiber intake in both an age-adjusted model and multivariable models was significantly associated with decreased CRC risk when using data obtained from food diaries, but results using data from food frequency questionnaires provided by the same subjects were not significant (59). Confounding variables may also account for discrepancies as suggested by other prospective cohort studies (i.e., EPIC, the Nurses’ Health Study, and the Health Professional’s follow-up Study) in which no significant associations between dietary fiber intake and CRC risk were found (160, 161). The evidence for associations between fruit, vegetable, and whole grain intake and CRC are not much stronger although there are some supportive data (218).

Other dietary habits associated with CRC risk include consumption of red meat, fish, vitamins, minerals, and fat. A meta-analysis of several prospective studies showed an increase in CRC risk associated with red and processed meat consumption (138). A prospective cohort study conducted by Norat et al showed that increased red meat consumption was significantly associated with increased CRC risk while consumption of fish was significantly associated with decreased risk (170). The method and degree of preparation have been also associated with CRC risk. Butler et al found that consumption of pan-fried and well-done red meat had the largest associations with CRC risk (38). Calcium and milk consumption were both found to significantly reduce CRC risk in a meta-analysis spanning several cohort studies (51). Another meta-analysis conducted by Sanjoaquin et al showed that dietary folate is significantly inversely associated with CRC risk (202). A positive association between CRC risk and energy
intake was found in yet another meta-analysis conducted by Howe, et al, but no significant associations were found regarding total fat and cholesterol (112). However, data from the Finnish Mobile Clinic Health Examination Survey showed significant positive associations between cholesterol intake and CRC risk, but not other fats (116). Additionally, a study conducted at the National Naval Medical Center showed that increased oleic acid intake was associated with increased CRC risk (155). Gut microbiota composition may play a role in these dietary associations, as the metabolic activities and by products of the microbiota (discussed below) are affected by the substrates which they receive. Many of these dietary associations are not strongly correlated with case status, and this may be because of differences in gut microbiota community composition.

Some studies show significant dietary associations with CRC risk only in specific regions of the colon. For instance, there was a significant association between high fruit and vegetable intake and low distal CRC risk only (133). Another study showed that rectal cancer was inversely associated with fruit, vegetable, and whole grain intake, and was positively associated with refined grain intake (218). Associations between CRC and red meat intake also appear to vary by location. Chao et al showed that both distal colon and rectal cancer risk were reduced within subjects consuming lower amounts of red meat (45). However, English et al showed only an increase in rectal cancer, but not colon cancer, associated with higher consumption of red meat (74). The data suggesting that some risk factors vary by anatomic location suggests that microbiota might behave in a similar fashion as a risk factor, which would indicate a benefit for stratifying studies of associations between CRC and microbiota by location.
Some data suggest that race and gender may contribute to dietary associations with CRC. African-Americans suffer from an increased incidence and mortality of CRC in comparison to Caucasian-Americans and males suffer higher rates within each racial category (American Cancer Society: Statistics for 2012, http://www.cancer.org/docroot/stt/stt_0.asp?from=fast). The observed racial differences may be due to variations in the diets between groups contributing to differences in microbiota (151). We showed that AAs consumed significantly more heterocyclic amines, potentially carcinogenic products of food preparation discussed later, than did CAs. Gender differences were observed in a study conducted by Deneo-Pellegrini et al which showed that only males experienced a protective effect of plant food intake against CRC (66). Furthermore, data from the Japan Collaborative Cohort Study showed a decrease in CRC risk associated with an increase in fiber intake, but this association was stronger in males (247). Distal, but not proximal, CRC risk was reduced by increasing calcium intake in both men and women (217). Similarly, data from the Shanghai Women’s Health Study showed a significant reduction of CRC risk associated with elevated calcium intake, but no significant associations were seen with fiber or vitamins (212). However, data from the Women’s Health Initiative trial conducted in post menopausal women found no significant associations between calcium and vitamin D intake and CRC risk, though another trial did find a significant association of vitamin D intake with decreased CRC risk (93, 246). These data did, however, show a significant reduction in CRC risk in subjects treated with estrogen and progesterone (50). Nilsen et al reported a positive association between blood glucose and diabetes with CRC risk in women, while in men the only significant association with increased CRC risk was low
physical activity (169). Similarly, a high glycemic load has been associated with CRC risk in women (109).

Cellular and molecular effects of dietary intake have also been documented. Slattery et al have shown that high vegetable, whole grain, and fiber intake are associated with a lower incidence of p53 (a cell cycle regulatory protein) mutations, while only high vegetable and fiber intake are associated with a lower incidence of KRAS (an oncogenic growth factor) mutations (217). Fish and olive oil added to HT-29 and Caco-2 cell cultures increased both apoptosis and cell differentiation and decreased COX-2 expression and subsequently Bcl-2 (an apoptosis regulatory protein) expression; however, only cell differentiation was induced by oleic and linoleic acids. Fish oil also significantly diminished cell proliferation (145).

Animal studies provide better mechanistic insight to dietary associations with CRC. Dietary supplementation of either black or navy beans upon administration of azoxymethane (AOM), an experimental colon carcinogen, in rats led to significantly reduced incidence and multiplicity of adenocarcinomas (101). Mai, et al, have shown earlier, that in APC\textsuperscript{Min} mice, diet induced differences in microbiota composition correlate with numbers of intestinal polyps. Unique bacterial signatures were also detected in mice with no or few polyps which suggested that the gut environment between mice with low verses high polyp burden differed (150).

**Mucosal Immunology**

Colonic inflammation and CRC are closely associated (231). Therefore, it is important to understand the interactions between commensal gut microbiota and the mucosal immune system. Upon initial colonization, the human GI tract is continuously exposed to an abundance of diverse microbes. Through mechanisms of tolerance, and
possibly ignorance as long as microbial contents remain confined to the lumen, the immune system facilitates a balance that allows for co-existence between microbes and host. The epithelium provides a physical barrier against microbes while the lamina propria contains the immune cells that maintain a tolerogenic response to these microbes. In the lamina propria, microbes that breach the epithelium are eliminated by the action of various immune cells.

In addition to providing a physical barrier, epithelial cells produce RegIIIγ, and antibacterial lectin, in response to pathogen-associated molecular patterns (PAMPs) via the MyD88-dependent toll-like receptor (TLR) pathway (239). The production of RegIIIγ may help compensate for the lack of a dense inner mucus layer. Paneth cells within the small intestine produce antimicrobial peptides called α-defensins which are capable of altering the bacterial composition of the lumen, whereas the action of RegIIIγ is localized to the mucus layer (197, 240).

The intestinal lamina propria contains gut-associated lymphoid tissue (GALT) which consists of Peyer’s patches in the small intestine, lymphoid aggregates in the appendix and colon, and diffuse immune cells throughout the lamina propria. Macrophages produce few pro-inflammatory cytokines in response to inflammatory stimuli but have the ability to induce an anti-inflammatory effect via the production of interleukin-10 (IL-10) (67, 221). Dendritic cells (DCs) can project dendrites into the intestinal lumen and present microbial antigens to lymphocytes accompanied by IL-17 secretion to induce antibacterial immunity (49, 67). Upon stimulation by DCs, B cells become antibody-producing plasma cells which then home to the lamina propria where they secrete immunoglobin A (IgA) into the intestinal lumen via transcytosis (166). While
no systemic antibody response is induced, bacteria opsonized with IgA cannot cross the epithelial barrier (148). However, antibodies against commensal bacteria have been detected in serum of healthy donors (98).

Peyer’s patches, found only in the small intestine and most abundantly in the distal ileum, participate largely in the crosstalk between the gut microbiota and the host (242). Specialized enterocytes known as M-cells reside within the epithelium surrounding Peyer’s patches. These cells are capable of transcytosis and are important for delivering antigens to DCs and lymphocytes within Peyer’s patches. M-cells transport free antigens, including bacteria, viruses, protozoa, and macromolecules, in addition to IgA-coated antigens from the intestinal lumen, to DCs and other immune cells within the GALT. Analogous lymphoid aggregates which function similarly are found in the colon.

Several T helper (T\(\text{H}\)) lymphocyte subsets, which include inflammatory T\(\text{H}1\) and T\(\text{H}17\) cells, humoral immunity-inducing T\(\text{H}2\) cells, and anti-inflammatory T regulatory cells (T\(\text{regs}\)), are stimulated by gut microbiota. In mice, segmented filamentous bacteria (SFB), or Candidatus Savagella, have been shown sufficient to induce signaling of all four T helper subtypes, similar to responses to conventional microbiota (85, 197). Several clostridial strains from the clusters IV and XIVa, which contain many important butyrate producers discussed later, were shown to induce a T\(\text{reg}\) phenotype both in the lamina propria and systemically (12). Bacteroides fragilis induces IL-10 production though a polysaccharide A-dependent response which prevents the expansion of T\(\text{H}17\) cells (194).

**Microbiota, Aberrant Immune Function and IBD**

IBD is a group of inflammatory conditions of the intestine associated with increased risk of developing CRC and is thought to arise due to an aberrant immune
response to gut microbiota in genetically susceptible hosts (257). Patients with Crohn’s disease exhibit a higher prevalence of adherent-invasive *E. coli*, similar to CRC patients discussed later (61). Further evidence from murine models demonstrated interplay between gut microbiota and IBD in genetically susceptible hosts, as germ-free IL-10 knockout mice did not develop disease; however, specific pathogen-free (SPF)-colonized counterparts rapidly developed disease (208). Another study showed that germ-free mice compared with SPF-colonized mice possessed larger amounts of intestinal natural killer T (NKT) cells and displayed higher levels of oxazolone-induced IBD and allergic asthma (177). Conflicting data may be the result of different genetic or IBD induction models between experiments.

Additional indirect evidence for a contribution of microbiota comes from the detection in IBD cases of mutations in genes important for gut microbiota immune signaling. Mutations in *NOD2*, whose product binds to bacterial lipopolysaccharides and interacts with nuclear transcription factor NF-κβ, are associated with Crohn’s disease in humans (175). Increased production of IL-23, secreted largely by dendritic cells, has been shown in both mice and humans to be associated with active IBD, likely via downstream effects on IL-17 production (73, 110). Mutations in the IL-10 receptor have been shown in humans to be associated with IBD (89). An increase in both *Treg* numbers and IL-10 expression have been observed in response to PSA from *B. fragilis* in a TLR 2 dependent manner (195). These results together suggest that some members of the microbiota are important for immune homeostasis. Both the addition of potentially beneficial bacteria and the removal of potentially harmful bacteria have been demonstrated to improve disease state. Supplementation with the probiotic mixture
VSL#3 has been shown to significantly improve disease in pouchitis patients, while Crohn’s patients treated with the antibiotic metronidazole have shown similar improvement (88, 225).

Systemic immunity and autoimmune diseases other than inflammatory bowel diseases have been linked to gut microbiota. SFB promote arthritis and experimental autoimmune encephalitis development, a condition in animal models analogous to human multiple sclerosis (142, 253). PSA-producing *B. fragilis* induces the expansion of circulating CD4+ T cells (156). MyD88, Myeloid differentiation primary response gene (88), a common adaptor protein in multiple TRL signaling cascades, is important in the development of type 1 diabetes (T1D). *MyD88*−/− germ-free mice develop robust T1D, but this effect is attenuated in SPF colonized mice (249). Though many positive associations between bacterial groups and disease development have been demonstrated, some comportments of the gut microbiota have been shown to have a protective effect against autoimmunity. Clostridial strains belonging mainly to Clostridium clusters IV and XIVa have been shown to increase Treg numbers in both the lamina propria and systemically and furthermore reduce the intensity of colitis in mice (12).

The mechanisms by which inflammation induces CRC have been studied extensively. Many mechanisms involve increased activation of the transcription factor NF-κβ. Greten *et al* demonstrated a decrease in tumor incidence in mice lacking IkB kinase (IKKβ), an important activator of NF-κβ (96). In humans, Sakamoto *et al* observed constitutive activation of NF-κβ in 40% of tumor tissue and 67% of CRC cell lines (196). Tumor tissue examined in parallel with autologous normal tissue in 28
patients showed an increase in nuclear NF-κβ (132). NF-κβ induces Wnt signaling by interacting with β-catenin and subsequently up regulates downstream oncogenes including c-myc (123). IL-6 and TNF-α, both NF-κβ-dependent cytokines, are important for the development of colitis-associated CRC in the mouse model (97, 181). A cross-sectional study conducted by Kim et al demonstrated that CRC patients have significantly higher levels of serum IL-6 and TNF-α (128).

Evidence for Correlations of Microbiota with CRC

The discovery that Helicobacter pylori is causally associated with non-cardia stomach cancers spurred interest in the potential contribution of other infectious agents in intestinal cancers (180). Increased diversity in the Clostridium coccoides and Clostridium leptum groups has been observed in CRC patients compared with healthy controls; however these groups of organisms contain many known butyrate producers, whose abundance are thought to be inversely associated with CRC (204). Superoxide-producing Enterococcus faecalis, a commensal gut organism, has been shown to induce DNA damage in cell lines and rat colonic tissue (114). Other bacterial species or groups have been studied more extensively, and both positive associations – most notable with E. coli, B. fragilis, and S. gallolyticus – have been found in addition to inverse associations, most notably with Bifidobacteria and butyrate-producing bacteria (26, 176, 207, 215, 226). Though not many viral associations have been made with CRC, one study showed a positive association with a mutant JC virus strain possessing a deletion in the transcriptional control region (190).

Proteobacteria have also been shown to positively associate with CRC, as they were found to be more abundant in patients with adenomas compared with normal controls (211). Swidsinski et al reported an increase in intracellular mucosal carriage of
*E. coli* in patients with adenomas compared to healthy controls (226). Additionally, adherent *E. coli* have been shown to decrease expression levels of DNA mismatch repair proteins MLH1 and MSH2 (149). AOM-treated IL-10 knockout mice exhibit greater tumor multiplicity and invasion when colonized with *E. coli* compared to colonization with *E. faecalis* (10). Particular strains of *E. coli*, which posses the *pks* pathogenicity island responsible for the production of the bacterial toxin colibactin, have been shown to induce DNA damage in mammalian cells lines and lead to chromosomal instability (58). In an AOM-treated IL-10 knock-out mouse model, monocolonization with this strain promoted carcinoma formation. When the *pks* island was deleted, tumor multiplicity and invasion decreased (10). Using subtractive hybridization to remove common sequences between a non-pathogenic K-12 *E. coli* strain and a colon cancer mucosal isolate, Bronowski* et al* identified several pathogenicity islands present in 30-40% of isolates from CRC patients that were previously detected only in uropathogenic *E. coli* strains (33).

*Streptococcus bovis* has been frequently associated with colorectal cancer, but data are conflicting. Fecal carriage of *S. bovis* has been reported to be significantly higher in carcinoma patients compared to healthy controls (129). However, Potter* et al* reported no significant difference in the stool carriage of colorectal cancer patients and healthy controls (182). This group also reported no significant associations of *S. bovis* with malignant colon tissue using culture techniques, but a study using quantitative PCR (qPCR) detected significantly more *S. bovis* DNA in tumor tissue compared to normal tissue in the same host (1). Rats treated with AOM and either live bacterial cells or cell wall antigens via oral gavage displayed a higher incidence of hyper-proliferative ACF
This study also showed an increase in polyamines associated with an increase in ACF after *S. bovis* administration (72). Bacterial cell wall antigens have also been shown to increase the expression of COX-2 *in vitro*, which, as previously stated, is associated with CRC development (20).

Durante-Mangoni *et al* reported a significant association between the prevalence of *S. bovis* endocarditis and increasing age (70). This suggests that *S. bovis* associations with CRC may be confounded by age, as age is a well-established risk factor for CRC. Other discrepancies between these studies may be attributed to small sample sizes, strain-level differences among *S. bovis* species, or the use of animal models. A recent meta-analysis conducted by Boleij *et al* described a specific biotype, *Streptococcus gallolyticus* (*S. bovis* Biotype 1), that is more closely associated with CRC (26). Further studies on this strain may provide more consistent results.

Enterotoxin-producing *B. fragilis* has been suggested to contribute to intestinal carcinogenesis through the effects of enterotoxin on e-cadherin (233, 254). Nuclear localization of β-catenin has been observed after treating cell lines with *B. fragilis* enterotoxin, subsequently leading to increased expression of *c-myc* and cellular proliferation (255). Recent data from the same group suggest a Th17-dependent mechanism for tumorogenesis (256). *B. fragilis* was also determined by Wang *et al* to be enriched in the guts of CRC patients compared to healthy controls.

Another mechanism by which *B. fragilis* may contribute to CRC development is through the metabolism of heterocyclic amines. Heterocyclic amines are generated by cooking meat well-done, and their production and mutagenicity vary based on the preparation method and type of meat (216). Few studies have examined associations
between gut microbiota and heterocyclic amines; however, B. fragilis cultured in meat extract resulted in a 2-fold increase in extract mutagenicity when applied to rat liver homogenates (130). These results suggest that the microbiota may have an influence on the genotoxicity of nitrogenous compounds.

*Bacteroides intestinalis*, among other members of the normal gut microbiota have been shown to produce secondary bile acids (84). Furthermore, colonic mucosal proliferation has been positively associated with secondary bile acid levels (173). *In vitro* studies have demonstrated the ability of secondary bile acids to up regulate NF-κβ expression and induce DNA damage through the generation of reactive oxygen species (119). Bacterial deconjugation of glucuronic acid derivatives via β-glucuronidase is an important factor for CRC development. Carcinogens are detoxified in the liver via conjugation with glucuronic acid and are subsequently excreted in the bile. β-glucuronidase activity has been shown to be increased in CRC patients compared to healthy controls (127).

Sulfate-reducing bacteria reside in the gut as part of the normal microbiota. Though sulfate-reducing bacteria to the author’s knowledge have not been associated with CRC, the production of hydrogen sulfide has been implicated in CRC development. Hydrogen sulfide was determined to be more abundant in subjects who had previously undergone surgery for CRC and developed new neoplasia compared to healthy controls (125). Hydrogen sulfide was also shown to induce DNA damage mediated by oxidative free radicals. Additionally, COX-2 expression was increased as a result of hydrogen sulfide treatment (13). These observations represent two possible mechanisms of the carcinogenic potential of hydrogen sulfide. Furthermore, thiosulfate sulfurtransferase
(TST), and important enzyme for detoxifying hydrogen sulfide, was found to be decreased in CRC tissue. *In vitro* studies showed the abundance of TST was increased in HT-29 cells when cultured with butyrate, suggesting a possible mechanism by which butyrate inhibits CRC development (186).

Butyrate-producing bacteria are thought to have a protective effect against CRC. The phylogenetic distribution of butyrate-producing bacteria lies mostly within *Clostridium* clusters IV and XIVa, also known as the *Clostridium leptum* and *coccoides* groups, respectively (16). Fermentation substrates are obtained from DFs, including inulin and resistant starches that escape host digestion (140, 158). The production of butyrate is primarily accomplished through the reaction of acetate with butyryl CoA via the butyrl-CoA: acetate CoA-transferase enzyme, but may also be accomplished by replacing CoA with a phosphate group via phosphotransbutyrylase followed by subsequent dephosphorylation via butyrate kinase, which has been shown to have both kinase and phosphotase activity (146).

In rodent models, administration of both butyrate-producing bacteria and butyrate metabolic precursors have been shown to increase colonic levels of butyrate and decrease precancerous lesions (124, 176). Okhawara *et al* also observed increased spleenic NK cells and decreased β-glucuronidase activities in fecal cultures, both of which are features associated with protection against CRC.

Other protective effects of butyrate against CRC have been proposed to arise through various mechanisms. One such mechanism involves the promotion of apoptosis through the Wnt signaling pathway (28). A linear relationship has been demonstrated between Wnt signaling and apoptosis in CRC cell lines exposed to butyrate (139).
Bordonaro et al propose that these data – which conflict with numerous experiments showing an increase in CRC risk as a result of increase Wnt signaling – represent a situation in which low levels of signaling result in controlled self-renewal, moderate levels promote uncontrolled proliferation, and high levels induce apoptosis. This idea was described as the “just right hypothesis” (27). Members of the gut microbiota that produce butyrate may contribute to cell cycle regulation via this mechanism. Further studies are needed to confirm this hypothesis.

Other mechanisms by which butyrate may prevent CRC development are through epigenetic modification. Butyrate is a known histone deacetylase inhibitor and has been shown to promote DNA demethylation of pluripotency-associated genes in fibroblasts (153). Another study on CRC cell lines demonstrated the ability of butyrate to promote differentiation, though this was attributed to the interaction of butyrate with signaling cascade proteins (178). Butyrate has also been shown in tumor cell lines to both increase and decrease genomic methylation, depending on the cell line (55).

**Means of Influencing Gut Microbiota Composition**

Probiotics have been defined by the World Health Organization and Food and Agriculture Organization of the United Nations as “Live microorganisms which when administered in adequate amounts confer a health benefit on the host”. Probiotics, when administered with prebiotics – defined as selectively fermented ingredients that allow specific changes, both in the composition and/or activity in the gastrointestinal microflora, that confer benefits upon host well-being and health – are referred to collectively as synbiotics (192). Clinical trials using synbiotics have shown promising results in reducing CRC risk. A study in which polypectomized patients were administered oligofructose-enriched inulin, *Lactobacillus rhamnosus*, and
*Bifidobacterium lactis* showed a reduction in fecal water genotoxicity (183).

*Bifidobacterium longum* and *Lactobacillus acidophilus* have also been shown to decrease fecal water genotoxicity (172). Other studies in which cell lines were incubated *in vitro* with lactic acid bacteria and/or prebiotics demonstrated a decrease in fecal water genotoxicity (36). Clinical data are lacking regarding the prevention of CRC lesion development as a function of probiotic and synbiotic use as these data would require large prospective cohort studies over long periods of time.

Animal models provide evidence for the mechanisms by which probiotics and prebiotics reduce CRC risk. *B. longum* has been shown in rats to decrease tumor incidence, AOM-induced cell proliferation, the activity of ornithine decarboxylase – an important enzyme for the polyamine synthesis – and the expression of ras-p21 oncoprotein (215). In the same model it was shown that both *L. rhamnosus* GG and *B. lactis* in combination with inulin and oligosaccharides prevented the AOM-induced suppression of NK cell-like activity, stimulated to production of IL-10, and suppressed lymphocyte proliferation in Peyer’s patches (193). A significant reduction in tumor incidence and an increase in cecal propionate and butyrate have been demonstrated using the prebiotic alone and in combination with the probiotic (81). *B. lactis*, when co-administered to rats with *L. acidophilus* and resistant starch, has been shown to increase the acute apoptotic response to a genotoxic carcinogen and is associated with lower pH and coliform counts (141).

Galactooligosaccharides (GOS) are prebiotic dietary fibers consisting of 2 to 8 unit galactose chains with a terminal glucose. Studies in humans examining direct effects of GOS on CRC are lacking, but many studies have shown the ability of GOS to enrich
groups of colonic bacteria negatively associated with CRC. Davis et al demonstrated a
dose-dependent increase in various species of *Bifidobacterium* as a result of GOS
supplementation (63). A dose-dependent effect was also observed in a study in which
the GOS were produced using a probiotic *Bifidobacterium* strain, and this effect was
stronger than that of another commercially available GOS supplement (68). Another
study demonstrated a bifidogenic effect of not only GOS, but also with
fructooligosaccharides and resistant starch (29). van Dokkum et al showed that GOS
increased levels of fecal acetate and decreased β-glucuronidase activity, which
suggests a mechanism by which GOS are associated with a decrease in CRC risk
(241). Studies conducted by Gopal et al both *in vitro* and *in vivo* demonstrated the ability
of GOS to stimulate the growth of both lactobacilli and bifidobacteria. This effect on
lactobacilli was greater in subjects who began treatment with lower baseline levels of
these organisms (90, 91). Another study demonstrated not only the bifidogenic effect of
GOS, but also its ability to decrease *E. coli* and *Salmonella enterica* serotype
Typhimurium adhesion to HT-29 cells (238). GOS were also shown in this study to
increase lactobacilli, certain butyrate-producing bacteria, and to decrease *E. coli.*
Though an increase in bifidobacteria was seen in the placebo group, a larger increase
was seen in the GOS-treated group that was significantly higher than the increase in the
placebo group. The same study showed an increase in both NK cell activity and anti-
inflammatory cytokine production while pro-inflammatory cytokines were reduced (245).
Though mechanistic studies are difficult to conduct in humans, a rat AOM model
showed that aberrant crypt multiplicity, but not numbers of ACF, was reduced in rats
consuming relatively high amounts of GOS (250).
Conclusions

Multiple lines of evidence support the notion that gut microbiota can contribute to colorectal carcinogenesis. Diet, particularly dietary fiber intake, which has been associated with reduced CRC risk, affects microbiota composition. Thus, changes in microbiota might represent important mechanisms through which diet reduces CRC risk. Various bacteria have been linked with experimental carcinogenesis in animal models or correlated with CRC in human observational studies. However, causality has not been established in humans. Interventional studies that would introduce potential pathogens in humans are infeasible for ethical reasons. If these bacteria are causally associated with CRC in humans, they should be detected at early stages in which polyps have not fully developed into cancer. Multiple microbiota-based studies suggest differences in mucosa associated and luminal bacteria in subjects with CRC. Various pro- and prebiotic interventions aimed at modifying gut microbiota toward a more beneficial composition have been successful in reducing CRC risk markers, gut inflammation, and preneoplastic lesions. Although no single microbial agent has yet been shown to be causally linked to CRC, contributions of the gut microbiota to colorectal carcinogenesis are evident. It can be expected that further advances in the field can soon be translated into the development of microbiota-based CRC screening and prevention regimens. A more detailed understanding of how microbiota can be manipulated also might point toward novel means to minimize the detrimental effects of surgery, radiation, and chemotherapy treatment on gut health. Due to the established uniqueness of each individual’s microbiota, personalized microbiota manipulations, possibly based on the underlying dominant enterotype, will likely offer the best chance for success.
CHAPTER 3
MATERIALS AND METHODS

Stool Sample Collections

Subjects collected their own stool samples which were placed on ice and delivered to the lab within four hours of defecation. The samples were processed immediately upon arrival. Samples were homogenized by kneading in a strong plastic bag and aliquoted for long term storage at -80°C and short term storage pending molecular assays at -20°C.

Biopsy Collections

During the colonoscopy, the attending physician collected multiple biopsies from the ascending colon. If the ascending colon could not be reached, the biopsy was taken from the most proximal location. One biopsy sample was placed into a tube of RNA later (Ambion, # AM7024) and fixed overnight at 4°C. Samples in RNA later were then stored at -20°C pending analysis. The remaining biopsy samples for each patient were divided and fixed in either formalin and Carnoy’s solution (10% glacial acetic acid [MP Biomedicals, #193829], 30% chloroform [Acros, #67-66-2], 60% 200 proof ethanol [Acros, #64-17-5]) and fixed overnight at 4°C. Biopsy samples were then transferred to 70% ethanol and submitted to the University of Florida College of Dentistry Molecular Pathology Core lab for microscope slide preparation. Biopsy samples were embedded in paraffin blocks. Four slides in total were prepared – 2 unstained slides and 2 slides stained with Hematoxylin and Eosin.

Microbiota Analysis

A pea-sized amount of each stool sample was washed twice in 1 mL sterile phosphate buffered saline (PBS). PBS was prepared with 8.00 g/L sodium chloride
(NaCl) (Fisher Scientific, #S671-500), 2.68 g/L potassium chloride (Fisher Scientific, #BP366-500), 14.20 g/L sodium phosphate dibasic (MP Biomedicals, #194846), and 2.40 g potassium phosphate monobasic (Fisher Scientific, #S374-500) adjusted to pH = 7.4 with hydrochloric acid (Acros, #7647-01-0). Bacterial genomic DNA was isolated from fecal samples using the QIAamp DNA Stool kit (Qiagen, #51504) per manufacturer instructions with an added 3-minute bead beating step (3450 oscillations/min) after cell lysis (Zirconia/Silica beads, BioSpec Products, #110791012; Mini-BeadBeater, Biospec Products, #607), which efficiently lyses most bacterial cells with little bias (162). Heat-resistant 2 mL microcentrifuge tubes were used during the bead beating step to prevent melting or warping of the tubes (Shrstedt, #D-51588). Kit buffers were reconstituted using 100% ethanol (Acros, #64-17-5). To detach bacteria from biopsy samples, tissues were first subjected to vigorous agitation in 500 mL PBS using 1 mm glass beads. Biopsy tissues were removed and stored for later analyses. Bacterial genomic DNA was extracted from the PBS wash via the same method used for stool extractions.

**Denaturing Gradient Gel Electrophoresis (DGGE) Analysis**

DGGE profiles were generated as an initial quality control for DNA extraction and as a crude tool for judging diversity in each sample. A 457bp fragment from the V6 to V8 region of the bacterial 16S rDNA was amplified with primers U968-GC (5’ CGC CCG GGG CGC GCC CCG GGC GGG GCG GGG GCA CGG GGG GAA CGC GAA GAA CCT TAC), and L1401 (5’ GCG TGT GTA CAA GAC CC) as described by Zoetendal et al. (259). PCR reagents were obtained from Qiagen (#201205). Reactions included 1 mM magnesium chloride (Qiagen, #201205) and 4% formamide (Acros, #75-12-7). 15-20 µL of amplified DNA mixed with an equal volume of loading dye was added to each lane. Loading dye (Bio-Rad, DCode Kit) was diluted to 10% in glycerol (70%, Fisher
Scientific, #G33-500) and sterile deionized water (20%). DGGE profiling was performed on an 8% (w/v) acrylamide (Acyrlamide/Bis solution, 37.5:1, Bio-Rad, #161-0148) gel with a gradient ranging from 40% at the top to 50% at the bottom. 100% denaturing conditions were defined as 7 M urea (Bio-Rad, #161-0730) and 40% formamide (Amresco, #0606-500). Polymerization was achieved using ammonium persulfate (0.09% w/v, Bio-Rad, #161-0700) and N,N,N',N'-tetra-methyl-ethalenediamine (0.09% v/v, Bio-Rad, #161-0800). Gels were run at 60°C for 16 hours at 65V in 0.5 X TAE buffer (Bio-Rad, #161-0773) and stained with SYBR Gold (Life Technologies, #S11494). Images of the stained gels were scanned with Quantity One software (Biorad) and analyzed with Diversity Database software (Biorad).

**Quantitative PCR**

qPCR analysis was performed in duplicate using a QuantiTect SYBR Green PCR kit (Qiagen, Cat. No. 204143) on a Stratagene MX3000P (La Jolla, CA). Reactions were performed in a final volume of 12.5 µl with 10 ng DNA template 0.2 µM of each primer (Invitrogen). The following primers and annealing temperatures were used: 1) eubacteria (V3 F: 5’-CCTACGGGAGGCAGCAG-3’; R:5’-ATTACCGCGCTGCTGG-3’, 56°C); 2) bifidobacteria (F: 5’-TCGCGT(C/T)GGTGTGAAAG-3’; R: 5’-CCACATCCAGC(A/G)TCCAC-3’, 58°C); lactic acid bacteria (F: 5’-AGCAGTGGGAATCTTCCA-3’; R: 5’-ATTYCACCGCTACACATG-3’, 58°C). Standard curves for quantification were constructed using genomic DNA extracted from *Bifidobacterium adolescentis* (ATC15703D) for bifidobacteria and VSL #3 mix (Sigma-Tau Pharmaceuticals, Inc.) for LAB and eubacteria.
Units are expressed as genome equivalents (1 genome equivalent = mass of 1 genome) per 10 ng DNA, which are derived as follows from the average mass of bifidobacteria and lactic acid bacteria genomes (2.3 Mb):

$$\frac{2,300,000 \text{bp}}{\text{1bp}} \times \frac{660 \text{ g/mol}}{\text{mol}} \times \frac{\text{1mol}}{6.02 \times 10^{23} \text{ genomes}} = 2.52 \times 10^{-15} \text{ g / genome}$$

Input mass of DNA for standard curves is converted to genome equivalents using the multiplicity constant $3.97 \times 10^5$ derived below:

$$\frac{1\text{ genome}}{2.52 \times 10^{-15} \text{ g}} \times \frac{1\text{ g}}{10^5 \text{ ng}} = 3.97 \times 10^5 \text{ genomes/ ng}$$

**454-based 16S rDNA Sequencing**

16S rRNA amplicons were generated using a published barcoded primer set (100). The PCR reactions included an initial melting step at 95°C for 2 minutes and 30 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 68°C for 1 minute. Reactions included 1 mM magnesium chloride (Qiagen, #201205) and 4% formamide (Acros, #75-12-7). Sequences with a quality score of less than 20 or with a length of less than 150 nucleotides were removed from the analysis. Sequences were initially analyzed using the Ribosomal Database Project (RDP) pyrosequencing pipeline that included features to calculate diversity indices and rarefaction curves. Sequence reads were binned using ESPRIT into OTUs at various similarity levels (223). The QIIME package was used to calculate i) Chao rarefaction diversity, which estimates how many OTU are present in a sample, and ii) UniFrac distances (40, 147), that allow for a comparison of the distribution of OTU among samples.
Metagenomic Sequencing

Bacterial DNA was isolated using the GNOME DNA Kit (MP Biomedicals, #2010-400) per manufacturer’s instructions with the following changes: After protein digestion, a ten minute bead-beating step (apparatus and materials described above) was used to lyse remaining intact bacterial cells. 15 mg of Polypvinylpolypyrrolidone (PVPP) (Fluka analytical, #9003-39-8) was added to each sample. Samples were mixed and centrifuged for 3 minutes (Eppendorf 5804R; all centrifugation steps were performed at 20,000 g). The supernatants were removed and the pellets were washed three times in TNAP buffer (50 mM Tris (Fisher Scientific, #BP152-500), pH 8, 20mM EDTA (Fisher Scientific, #O2793-500) pH 8, 100 mM NaCl, 1% PVPP). Supernatants were pooled. DNA was pelleted from 750 µL of pooled supernatant using 1 mL isopropanol (Acros, #67-63-0) for 30 minutes at -20°C. The pellet was dissolved in 400 µL water and the procedure resumed according to the manufacturer’s instructions. Ethanol precipitation occurred at -80°C for one hour.

Bacterial genomic DNA sheared randomly via sonication from eight cases with high-risk polyps was pooled in equal molar amounts as was DNA from eight controls. Both DNA pools were then submitted to 454 pyrosequencing using one half plate for each DNA pool. Sequence reads were deposited and analyzed in MG-RAST using the GenBank reference database with 50% identity and a minimum alignment length of 40 bases. 233,141 sequence reads with an average of 180 bases per read were obtained for control subjects and 228,814 sequence reads with an average of 176 bases were obtained for case subjects.
Fecal pH

Approximately 1 g of homogenized stool was added to pre weighed 15 mL conical tubes. Exact stool weight was obtained on a digital scale and a 1:10 w/v dilution with sodium chloride (0.155M) was agitated with 4 to 5 glass beads (1 mm) to break up particles. Samples were centrifuged (Sorval legend RT+) for 5 minutes at 3716 rcf. Supernatants were measured using a glass electrode and Ultrabasic pH meter (Denver instruments).

Statistical Analysis

For DGGE profile analysis, background was subtracted from each lane and the profiles were subjected to Gauss modeling. The similarity matrix was calculated based on the Dice correlation coefficient. Phylogenetic trees were generated using Ward’s clustering method (Ward, UPMA). Shannon diversity indices were calculated for each sample using the total number of bands per sample and their relative intensities.

For qPCR assays, genome equivalents were used as the unit of comparison in an attempt to correct for differences in genome size and copy numbers of 16S rRNA genes. The Student's t-test was used to calculate differences in genome equivalents between cases and controls either for raw or log transformed data when the data were normally distributed. The Mann-Whitney U test was used on non-normally distributed data.

For analyzing differences in the prevalence of OTUs, a z score was obtained using a chi-square based test. For exploratory purposes we used a p<0.05, we corrected for multiple analyses by using a P<0.01 and/or FPR rate of 5% (α=0.05). For generating heat maps the most significantly different OTUs were chosen based upon Z score. In the case of more that more than 20 OTUs were significantly enriched, only the 20 with
the lowest p-values were included in the heat map. The QIIME package was used to calculate p-values for differences in UniFrac distances.
CHAPTER 4
GUT MICROBIOTA PATTERNS ASSOCIATED WITH COLORECTAL POLYP PREVALENCE

Differences in Diet and Intestinal Microflora: Potential Associations with Increased CRC Rates in African-Americans

Introduction

The work presented first is based on the hypothesis that distortions in gut microbiota composition are associated with colorectal polyp prevalence. Adults aged 40 years or older who were scheduled to undergo a screening colonoscopy were recruited. It was important to recruit subjects in this age group because the risk of developing CRC and the incidence of polyps increase with age. Furthermore, it is necessary to use subjects only referred for screening as other underlying diseases may be associated with microbiota composition. Microbiota was analyzed using DGGE, qPCR, 16S rRNA pyrosequencing, and metagenomic shotgun sequencing.

Study Design

Volunteers scheduled to undergo a screening colonoscopy were recruited from university hospital-associated endoscopy clinics in Baltimore, MD and Gainesville, FL. The study was approved by Institutional Review Boards at University of Maryland and the University of Florida. Upon consent to participate, subjects provided dietary and lifestyle information via a demographic questionnaire, food frequency questionnaire (NIH Block 98.2), and the National Institute of Health meats-module questionnaire. A 4-day food record reflecting the four days of food and fluid intake prior to stool sample collection was collected by the subjects. Each subject gave written informed consent, and all study procedures were in accordance with the ethical standards of the University...
of Maryland and the University of Florida Institutional Review Board. Inclusion criteria included the following:

- Referred for a screening colonoscopy
- Age of 40 years or older
- Able to provide informed consent for study procedures
- Able to provide information on dietary history and other demographic factors (e.g., family history)
- African American of Caucasian American

Exclusion criteria included the following:

- Known colorectal cancer
- Known or suspected inflammatory bowel disease
- Acute or chronic diarrhea
- Short gut syndrome
- Prior resection of small intestine or ileocecal valve
- Known small or large bowel obstruction
- Hospitalization within the past 4 weeks
- Current hospitalization for 3 days or longer
- Use of systemic antibiotics (oral, intravenous, or intramuscular) within the past 4 weeks
- Diagnosis of malignancy (other than skin cancer within the past five years)
- Chemotherapy treatment within the past three years

Colonoscopy and pathology reports were obtained to determine the presence of polyps, their sizes, and locations. Pathology reports were obtained to determine polyp histological classification. Non-neoplastic polyps, those commonly thought to be without
malignant potential, include hyperplastic polyps and lymphoid aggregates. Neoplastic polyps, or polyps with malignant potential, are mainly adenomas (54). Subjects with polyps were further classified into high risk or low risk groups based on polyp number and size. Specifically, high risk subjects were defined as those with total polyp diameters equal to 10 mm or greater from one or multiple polyps, since this allowed for a nearly equal distribution of subjects between low-risk and high-risk groups. Conventionally, high risk subjects have been defined as those with at least one polyp with a diameter of 10mm or greater (222). Because in this study only few subjects fit this criterion, decreasing statistical power, the broader criterion was used. 157 subjects consented to the study protocol. Case status was reported for 126 subjects. Any subject displaying at least one polyp was considered to be a case subject. Subjects without reported case status were excluded.

Results

Study population demographics and dietary habits

115 subjects completed demographic questionnaires (Table 4-1). Control subjects reported significantly more exercise than high-risk subjects and were less likely to have had polyps removed at previous colonoscopies than both high-risk case subjects and all case subjects. No statistically significant differences in dietary habits were detected between all cases and controls, but high-risk cases consumed significantly less dietary fiber, carbohydrates, fruit, copper, magnesium, potassium, vitamin E and vitamin C than controls (Table 4-2).

AAs consumed smaller amounts of both dietary and supplemental vitamins and minerals compared to CAs. AAs also consumed smaller amounts of total dietary fiber,
and specifically fiber from grains, vitamin E, magnesium, and potassium. Likewise, AAs ate fewer servings of fruits and vegetables per day.

**Fecal microbiota community diversity**

DNA was extracted from all available stool and biopsy samples. Initial DGGE profiling confirmed that DNA extractions were successful for all samples and that the samples contained a diverse microbiota community. No differences were detected in the prevalence of specific DGGE bands between cases and controls, or between the two racial groups in neither stool nor biopsy samples (data not shown). Similarly, no significant differences between case groups were detected for the DGGE-based Shannon and Simpson diversity indices (Figure 4-1). Compared to stool samples, biopsy samples exhibited much less diversity (data not shown).

**Quantification of targeted bacteria**

A qPCR analysis was performed to compare the amounts of lactic acid bacteria and bifidobacteria, which are generally thought of as beneficial members of the gut microbiota. No statistically significant differences were detected between cases or controls in both stool (Figure 4-2) and biopsy (Figure 4-3) samples, even after stratifying by body mass index (BMI) or between racial groups. When stratified by BMI independent of case status, no significant differences in either bifidobacteria of lactic acid bacteria abundances were observed in either stool or biopsy samples (Figure 4-4). However, there was a significantly higher abundance of bifidobacteria in stools samples from AA subjects compared to CA subjects independent of case status, but this was not seen in biopsy samples (Figure 4-5). Furthermore, in biopsy samples the abundance of lactic acid bacteria was significantly greater than the abundance of bifidobacteria, while in stools samples bifidobacteria were significantly more abundant than lactic acid.
bacteria. These results further demonstrate the differences between stool and biopsy microbiota communities.

**454 pyrosequencing analysis of 16S rDNA**

16S rRNA-based 454 pyrosequencing was used to analyze 30 case and 30 control samples from stools, and 16 case and 16 control samples from biopsies. After removal of low quality and short reads, a total of 209850 sequence reads with an average length of 216 nucleotides were retained from stools and 42150 sequence reads with an average length of 211 nucleotides were retained from biopsies. Less DNA was sequenced for biopsy samples because of the lower diversity compared to stool samples based on DGGE analysis. ESPRIT was used to bin sequences into a total of 1927 OTUs at the 95% similarity level and 6321 OTUs at the 98% similarity level. OTUs were blasted against the RDP database to obtain taxonomic identities.

UniFrac, a tool for comparing microbial community diversity, was used to determine if bacterial communities in stool samples differed from those in biopsy samples (99, 223). The unweighted approach was used, which considers only the absence or presence of each OTU, as opposed to the weighted approach, which also considers OTU abundance and can be strongly influenced by a few abundant OTUs. Chao1-based rarefaction curves did not significantly differ between case groups, suggesting that overall diversity is not significantly affected by case status. (Figure 4-6). As community structures in these two sets of samples clearly differed (Figure 4-7a) stool microbiota were analyzed separately from biopsy-adherent microbiota. When cases were compared to controls, no differences in UniFrac diversity were detected in either stool (Figure 4-7b) or biopsy samples (Figure 4-7c). There also were no detectable differences in UniFrac diversity between high-risk subjects and controls (data
not shown) or between AAs and CAs (Figure 4-7d). No significant differences were detected in phylum abundance between cases and controls in both stool and biopsy samples (Figure 4-8).

In both stool and biopsy samples more OTUs were significantly more prevalent in control rather than case samples, suggesting that a lack of potentially “beneficial” bacteria, rather than the presence of a pathogen, correlates with polyp prevalence. To maximize our ability to detect differences, control samples were compared with all cases, high-risk cases as defined in this study (the sum of polyp diameters equaling or exceeding 10mm), and high-risk cases as defined by convention (presence of a single polyp with a diameter of 10mm or greater). Figure 4-9 shows the 20 most significantly affected OTUs by case status between all cases and controls. Figure 4-10 represents the same analysis between only high-risk cases (defined in this study) and controls. Figure 4-11 shows the analysis with high-risk cases redefined by convention. The majority of the OTUs detected in these analyses matched within the class Clostridia.

Comparisons at the 98% similarity level revealed similar results in that most significantly affected OTUs belonged to the class Clostridia. Figure 4-12 shows the 20 most significantly affected OTUs by case status between all cases and controls. Figure 4-13 represents the same analysis between only high-risk cases (defined in this study) and controls. Figure 4-14 shows the analysis with high-risk cases redefined by convention.

For biopsy samples, many OTUs that were significantly affected by case status again matched within the class Clostridia; however, some OTUs matched within other phyla including Bacteroidetes and Proteobacteria. The comparison between controls
and all cases at the 95% similarity level is represented in Figure 4-15. Figure 4-16 shows a similar analyses considering only high-risk cases defined in this study. Figure 4-17 shows the analysis between controls and all cases at the 98% similarity level. Controls and high-risk cases are compared at the 98% similarity level in Figure 4-18. No comparison between controls and high-risk subjects defined conventionally was made for biopsy samples because only two case subjects belonged to that category.

Differences in OTU prevalence appeared stronger (based on lower p-values) when high-risk polyp cases were compared with controls, giving stronger evidence for those OTUs to truly be associated with CRC risk. The abundance of OTUs corresponding to butyrate-producing bacterial genera as described by Antharam, et al, were compared in both stool and biopsy samples. None of the comparisons (controls verses all cases, controls verses high-risk cases, and controls verses conventionally defined high-risk cases) showed any significant differences in the presence of butyrate-producing bacteria (data not shown)(8). Various OTUs differed in prevalence between AAs and CAs. These OTUs, similar to those found to be significantly enriched or depleted based on case status, matched mainly to members of the Clostridia class; however, these particular OTUs were different than the significant OTUs observed between case groups. Therefore, although there are differences in gut microbiota composition between race groups, these differences cannot be attributed to increased burden of CRC in AAs. Figure 4-19 displays those OTUs significantly affected by race at the 95% similarity level. Figure 4-20 displays the OTUs significantly affected by race at the 98% similarity level.
Because not a single OTU could effectively discriminate between cases and controls, a discriminant analysis approach was applied to develop a fecal microbiota pattern that would allow for such a distinction. At the 92% similarity level, a microbiota signature profile was detected based on 27 OTU's that effectively separated cases from controls (Figure 4-21a). The 92% similarity level is optimal for discriminant analysis accuracy(224). When the similarity level is lower, sequences are grouped into large clusters where discriminating and non-discriminating OTUs may be combined. If the similarity level is higher, the sequencing depth is not great enough to obtain accurate estimates of microbial composition profiles (54, 224). The formula for the discriminant analysis is as follows, where $X_1$ through $X_{15}$ represent the number of reads for each OTU, respectively, and $Y$ is the calculated risk assessment. In the formula below the OTUs correspond as follows: Bacteroides eggerthii ($X_1$), Sutterella stercoricanis ($X_2$), Parasutterella excrementihominis ($X_3$), Oscillibacter valericigenes ($X_4$), Alistipes massiliensis ($X_5$), Lactobacillus manihotivorans ($X_6$), Clostridium methylpentosum ($X_7$), Ruminococcus bromii ($X_8$), Clostridium sp. 14774 ($X_9$), Succinispira mobilis($X_{10}$), Lutispora thermophila($X_{11}$), Oscillibacter valericigenes($X_{12}$), human intestinal firmicute CO35 ($X_{13}$), Ruminococcus sp. CJ60 ($X_{14}$), Succinivibrio dextrinosolvens ($X_{15}$)

$$Y = 0.5206x_1 - 0.1281x_2 - 0.3423x_3 - 0.3169x_4 + 0.1982x_5 - 0.1714x_6$$
$$- 0.2007x_7 - 0.09999x_8 - 0.0824x_9 - 0.0572x_{10} + 0.0739x_{11} - 0.2362x_{12}$$
$$- 0.0716x_{13} + 0.2046x_{14} + 0.2079x_{15} + 7.7193$$

A positive value for $Y$ indicates a case subject and a negative value indicates a control subject. The corresponding receiver operator characteristic (ROC) curve yields
an area under the ROC curve (AUC) value of 0.81, which indicates the
sensitivity/specificity for predicting cases status based on microbiota profile (Figure 4-21b).

**Metagenomic shotgun sequencing analysis**

To expand the analysis beyond the limitations of the 16S rRNA-based approach, a pilot metagenomic analysis was performed using shotgun sequencing on the 454 pyrosequencing platform. For this analysis equimolar amounts of genomic DNA were pooled from 8 controls and 8 high-risk cases. A total of 233,141 and 228,814 were generated, respectively. The two datasets were analyzed using the Metagenomics Rapid Annotation using Subsystems Technology (MG-RAST) platform with the Greengenes database. An enrichment of sequences with closest matches to *Klebsiella*, *Shigella*, and *Citrobacter* was detected in cases (Figure 4-22). *Klebsiella* was also more prevalent in cases when the analysis was limited to 16S rRNA sequences that were detected in the shotgun datasets. Bacterial starch and sucrose metabolic functions were more highly represented in cases while fructose and galactose metabolic functions were more highly represented in controls (Figure 4-23).

**Discussion**

The results of this study indicate that there were differences in both lifestyle, including exercise and diet habits, and gut microbiota composition between healthy subjects and subjects with polyps. Exercise has been associated with reduced polyp development (52). Approximately half of control subjects, compared to only a fifth of subjects with polyps, exercised regularly ($p < 0.05$). Furthermore, high-risk subjects consumed significantly less dietary fiber and fruit, which likewise parallels results found in the literature (82, 189, 244, 252).
The apparent differences in bacterial phyla abundance did not reach significance, likely because of the large amount of variation between subjects (particularly those with a difference in magnitude of 10% or more). Ranges of bacterial abundance in the two predominating bacterial phyla, Bacteroidetes and Firmicutes, of 32 to 96% and 0.3 to 61%, respectively, in stools and 30 to 82% and 4 to 58%, respectively, in biopsies, were detected. The extensive interpersonal variation in this study has been previously observed others(113). Although not significant, the Firmicutes:Bacteroidetes ratio was higher in stools of control subjects. These results parallel those presented by chen, et al., conducted in CRC patients (47). Furthermore, the abundance of Fusobacterium was greater, though not significantly, in cases verses controls in both stool and biopsy samples. *Fusobacterium nucleatum* have been reported to associate with colorectal tumor tissue (42).

Differences between case and control subjects were detectable at the OTU level in both stool and biopsy samples. Some OTUs were more prevalent in case subjects while others were more prevalent in control subjects, which suggests that some OTUs may have pathogenic potential while others may be protective. When only high-risk case subjects were considered, many of the previously associated OTUs again reached significance in both stool and biopsy communities. OTUs that reach significance in multiple analyses can be considered with more confidence as they are less likely to be spurious results. In this study, the combined analysis of all cases and only high-risk cases verses controls was used to help eliminate OTUs that may occur by chance. A common problem when dealing with many variables is the detection of significant differences based on chance. With 6321 OTUs in the data set and an alpha level of
0.05, one would expect to find 316 significantly different OTUs by chance. One may reduce spurious results by lowering the threshold of significance or using different statistical tests, such as those based on the t distribution verses the chi squared distribution; however, there is currently no consensus regarding the appropriate statistical tests or analysis method for sequencing data. The false discovery rate can be used to assign a metric (q-value) to a range of p-values based on the distribution of p-values, where lower p-values are assigned a lower value. The q-value is generally represented by the height of the p-value distribution where the tail becomes flat. P-values assigned q-values that are lower than this threshold are assumed to be true positives.

While most OTUs associated with polyp presence were detected in both the analyses considering all cases and only high-risk cases analyses, several OTUs only reached significance in the comparison between high-risk cases and controls. These results suggest that some associations do not arise until after a certain level of disease progression. OTUs that reach significance only in the low risk population may contribute to polyp development via a “hit and run” mechanism, in which a pathogen initiates a pathogenic physiology which progresses after the pathogen has been eliminated.

In stools samples, most significantly different OTUs match within the phylum Firmicutes, particularly within the Classes Clostridia and Bacilli. Many of these OTUs are known butyrate producers, but while most butyrate producers were associated with controls, few were associated with cases. Therefore, they cannot be firmly linked with protection against polyps. While some significantly different OTUs in biopsy samples also matched within these groups, there were more matches to organisms in the phyla
Bacteroidetes, Proteobacteria, and Fusobacteria compared with stools. These OTUs did not correspond with organisms discussed from the literature in Chapter 2, with the exception of one OTU which was associated with case subjects in biopsies at the 98% similarity level (26, 176, 207, 215, 226). This association is consistent with the results from Swidsinski, et al. that demonstrate increase mucosal carriage of *E. coli* in CRC patients (226). Though some OTUs did match to streptococci groups, they associated in some analyses with case subjects and in others with control subjects; therefore results implicating *S. gallolyticus* in CRC were not supported in this study (26). The most likely reason for these inconsistencies is that this work examined differences between subjects with polyps and healthy controls verses subjects with CRC and healthy controls. Because only roughly 5% of subjects with polyps develop CRC, it is not surprising that these results did not parallel those in the CRC literature; however, this particular risk factor is optimal for this prospective cohort study because roughly 95% of CRCs develop from polyps (115). Future studies should examine a much larger samples size and follow the subjects for several years to determine which subjects develop CRC.

Across the comparisons of individual OTUs stratified by sample site and risk, there were OTUs belonging to the groups Oscillospiraceae, unidentified Firmicutes, unknown butyrate-producing bacteria, and *Alistipes* that consistently associated with control subjects. These results confirm previous findings in which OTUs matching to these groups associated with healthy subjects verses subjects with CRC (248). The inverse association of these OTUs with polyp incidence and CRC suggest that these groups
may be protective against CRC development. These OTUs may provide such a benefit through the production of butyrate (176).

Some OTUs showed contradictory results. For instance, at the 98% similarity level, some were associated positively with polyp prevalence in biopsy samples but inversely in stool samples suggesting that some OTUs may be harmful when in direct contact with the mucosa but beneficial when present in the lumen. At the 95% similarity level, there were no shared OTUs between stool and biopsy communities that associated with case status.

There are both advantages and disadvantages to analyzing stool and biopsy communities. The advantage of analyzing stool samples is that they can be obtained non-invasively; however, analyzing stool samples is disadvantageous in that they contain luminal organisms and not those closely associated with the epithelium, where polyps develop. This work has shown a clear distinction between luminal and mucosa adherent microbiota communities. The organisms in direct contact with the epithelium may have a stronger influence on epithelial cell physiology via stimulation of TLRs on epithelial cells or activation of dendritic cells, and thus may represent better candidates for future studies. However, the biopsies in this study were not obtained under physiological conditions. Patients preparing for colonoscopies must undergo a cleansing prep which can remove mucosally associated microorganisms. Therefore, in this study stool samples may be more appropriate in that only they can be used for developing a non-invasive screening method. Furthermore, stools samples are more representative of normal physiological conditions, as they were collected before patients began their cleansing prep.
The comparison between racial groups at the OTU level revealed that many OTUs are associated with racial group; however, most of these racial group-associated OTUs did not correlate with polyp prevalence or with organisms previously associated with CRC. There is no clear evidence based on OTU distribution that the increased incidence of CRC in AAs is associated with the differences in gut microbiota composition of AAs. While the abundance of bifidobacteria, determined by qPCR, was higher in AAs, there was no association between bifidobacteria counts and case status. These results suggest that the increased CRC rate in AAs verses CAs is not microbiota dependent. Further studies examining genetic predispositions to CRC between races may provide better insight.

Upon determining differences by case status at the OTU level, a predictive model based on OTU abundance was able to discriminate cases and controls. Several of these OTUs match to groups previously found to be inversely associated with CRC (*Alistipes*, *Oscillobacter*, unknown Firmicutes) (248). This model was verified using the leave one out cross validation method, in which a model is constructed with a training data set and tested with a validating datum. The training data set includes all but one of the experimental data, while the validating datum is the one excluded. This validation is conducted for each individual datum of the experimental data set and the results are averaged to produce a single model. This model could potentially be used clinically as a preventative screening method. Further studies with deeper sequencing technologies and larger samples sizes are needed to improve sensitivity and specificity.

Differences by case status in stools were also seen using an exploratory shotgun sequencing approach, though the bacterial groups associated with polyp prevalence
were different than those detected via 16S sequencing. However, the enrichment of E. coli in the stools of case subjects compliments the significant association of E. coli in biopsies with high-risk polyp incidence. The differences in methodologies may account for the ability to detect significant associations with luminal E. coli in this analysis. Metagenomic sequencing may provide better results in that there is no PCR bias, whereas 16S sequencing relies upon primer sets that may not be truly universal. Polymorphisms in the generally conserved regions of 16S gene, bacterial DNA methylation patterns, or DNA binding proteins carried throughout the DNA extraction procedure may influence primer affinity. Furthermore, metagenomic sequencing allows for functional assignment verses strictly taxonomic assignment. However, due to the severely limited sample size and the pooling of samples in this analysis, further metagenomic sequencing using more samples without pooling is needed to confirm results.

In conclusion, there are discernible differences in gut microbiota composition between AAs and CAs, but these differences are not consistent with the differences detected based on case status. Additionally, it was shown that there are large differences between stool and biopsy communities; however, the analysis of biopsy communities is compromised because all subjects were required to perform a colon cleansing prep prior to their colonoscopy. Therefore, the stool-based analysis may better represent normal physiological conditions. Finally, these results indicate that a stool-based test based at the OTU level is feasible for estimating polyp presence. This work may have implications in earlier detection of CRC and thus lower the morbidity and mortality of this disease.
Table 4-1. Demographic and lifestyle data by case status.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Study population (N = 115)</th>
<th>Controls (n = 61)</th>
<th>All cases (n = 54)</th>
<th>High-risk cases (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age</td>
<td>60 ± 9</td>
<td>60 ± 9</td>
<td>59 ± 9</td>
<td>62 ± 10</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>77 (67)</td>
<td>45 (74)</td>
<td>32 (59)</td>
<td>12 (50)</td>
</tr>
<tr>
<td>Male</td>
<td>38 (33)</td>
<td>16 (26)</td>
<td>22 (41)</td>
<td>12 (50)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-Americans</td>
<td>30 (26)</td>
<td>18 (29.5)</td>
<td>12 (22)</td>
<td>4 (17)</td>
</tr>
<tr>
<td>Caucasian-Americans</td>
<td>82 (71)</td>
<td>40 (65.5)</td>
<td>42 (78)</td>
<td>20 (83)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (3)</td>
<td>3 (5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean BMI</td>
<td>29 ± 8</td>
<td>30 ± 10</td>
<td>29 ± 6</td>
<td>28 ± 5</td>
</tr>
<tr>
<td>Healthy weight</td>
<td>34 (30)</td>
<td>18 (29)</td>
<td>15 (28)</td>
<td>5 (21)</td>
</tr>
<tr>
<td>Overweight</td>
<td>43 (37)</td>
<td>22 (36)</td>
<td>22 (40)</td>
<td>13 (54)</td>
</tr>
<tr>
<td>Obese</td>
<td>37 (32)</td>
<td>21 (35)</td>
<td>16 (29)</td>
<td>5 (21)</td>
</tr>
<tr>
<td>BMI Not reported</td>
<td>1 (1)</td>
<td>0</td>
<td>1 (3)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>13 (11)</td>
<td>6 (10)</td>
<td>7 (13)</td>
<td>3 (13)</td>
</tr>
<tr>
<td>High school diploma</td>
<td>36 (31)</td>
<td>18 (30)</td>
<td>18 (33)</td>
<td>8 (33)</td>
</tr>
<tr>
<td>Some college</td>
<td>63 (55)</td>
<td>35 (57)</td>
<td>28 (52)</td>
<td>12 (50)</td>
</tr>
<tr>
<td>Education not reported</td>
<td>3 (3)</td>
<td>2 (3)</td>
<td>1 (2)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Income per year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 30,000</td>
<td>37 (32)</td>
<td>19 (31)</td>
<td>18 (33)</td>
<td>8 (33)</td>
</tr>
<tr>
<td>30,000 to 50,000</td>
<td>30 (26)</td>
<td>17 (28)</td>
<td>13 (24)</td>
<td>5 (21)</td>
</tr>
<tr>
<td>Greater than 50,000</td>
<td>46 (40)</td>
<td>24 (39)</td>
<td>22 (41)</td>
<td>10 (42)</td>
</tr>
<tr>
<td>Education not reported</td>
<td>2 (2)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Previously detected polyps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyps detected and removed</td>
<td>42 (37)</td>
<td>11 (18)</td>
<td>31*** (57)</td>
<td>16** (67)</td>
</tr>
<tr>
<td>Polyps not detected</td>
<td>67 (58)</td>
<td>49 (80)</td>
<td>18 (33)</td>
<td>6 (25)</td>
</tr>
<tr>
<td>Polyps not reported</td>
<td>6 (5)</td>
<td>1 (2)</td>
<td>5 (10)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Exercise (45 minute activities)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 3 times per week</td>
<td>47 (41)</td>
<td>30 (49)</td>
<td>17 (31.5)</td>
<td>6 *(25)</td>
</tr>
<tr>
<td>Fewer than 3 times per week</td>
<td>64 (56)</td>
<td>31 (51)</td>
<td>33 (61)</td>
<td>16 (67)</td>
</tr>
<tr>
<td>Exercise not reported</td>
<td>4 (3)</td>
<td>0</td>
<td>4 (7.5)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Anti-inflammatory use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>51 (44.5)</td>
<td>30 (49)</td>
<td>21 (39)</td>
<td>11 (46)</td>
</tr>
<tr>
<td>Weekly</td>
<td>16 (14)</td>
<td>8 (13)</td>
<td>8 (15)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Rarely</td>
<td>37 (32)</td>
<td>18 (29.5)</td>
<td>19 (35)</td>
<td>9 (38)</td>
</tr>
<tr>
<td>Never</td>
<td>7 (6)</td>
<td>4 (6.5)</td>
<td>3 (5.5)</td>
<td>0</td>
</tr>
<tr>
<td>Anti-inflammatory use not reported</td>
<td>4 (3.5)</td>
<td>1 (2)</td>
<td>3 (5.5)</td>
<td>2 (8)</td>
</tr>
</tbody>
</table>
Table 4-1. Continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Study population (N = 115)</th>
<th>Controls (n = 61)</th>
<th>All cases (n = 54)</th>
<th>High-risk cases (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laxative use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>2 (1.5)</td>
<td>2 (3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Weekly</td>
<td>7 (6)</td>
<td>5 (8)</td>
<td>2 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Rarely</td>
<td>42 (37)</td>
<td>18 (30)</td>
<td>24 (44)</td>
<td>10 (42)</td>
</tr>
<tr>
<td>Never</td>
<td>61 (53)</td>
<td>36 (59)</td>
<td>25 (46)</td>
<td>12 (50)</td>
</tr>
<tr>
<td>Laxative use not reported</td>
<td>3 (2.5)</td>
<td>0</td>
<td>3 (6)</td>
<td>1 (4)</td>
</tr>
</tbody>
</table>

Parentheses indicate the percentage of the responding study population with the given characteristic. BMI is defined as weight in kg divided by height in meters squared. * = $p < 0.05$, ** = $p < 0.001$, *** = $p < 0.0001$ compared to the control group. ± indicates the standard deviation.
Table 4-2. Dietary differences between case groups and racial groups.

<table>
<thead>
<tr>
<th>Dietary variable</th>
<th>Controls (n = 48)</th>
<th>Cases (n = 41)</th>
<th>High-risk cases (n = 23)</th>
<th>AA (n = 25)</th>
<th>CA (n = 63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein (g)</td>
<td>58.8 ± 31.3</td>
<td>59.0 ± 32.4</td>
<td>51.3 ± 20.3</td>
<td>51.4 ± 33.1</td>
<td>61.1 ± 30.5</td>
</tr>
<tr>
<td>Total fat (g)</td>
<td>64.1 ± 34.7</td>
<td>69.4 ± 41.8</td>
<td>58.2 ± 24.2</td>
<td>57.4 ± 38.3</td>
<td>69.7 ± 37.8</td>
</tr>
<tr>
<td>Saturated fat (g)</td>
<td>18.6 ± 10.4</td>
<td>20.7 ± 13.3</td>
<td>17.3 ± 7.9</td>
<td>16.3 ± 11.7</td>
<td>20.7 ± 11.8</td>
</tr>
<tr>
<td>Carbohydrates (g)</td>
<td>192.3 ± 96.6</td>
<td>175.1 ± 86.6</td>
<td>146.1* ± 53.3</td>
<td>188.4 ± 114.9</td>
<td>182.1 ± 83.0</td>
</tr>
<tr>
<td>Fiber (g)</td>
<td>16.4 ± 9.2</td>
<td>12.8 ± 7.1</td>
<td>10.4* ± 4.4</td>
<td>11.9 ± 7.7</td>
<td>15.9* ± 8.6</td>
</tr>
<tr>
<td>Fiber from fruits and vegetables (g)</td>
<td>7.5 ± 5.2</td>
<td>5.7 ± 4.1</td>
<td>4.6* ± 2.8</td>
<td>4.9 ± 4.0</td>
<td>7.4 ± 4.9</td>
</tr>
<tr>
<td>Daily fruit servings</td>
<td>1.3 ± 0.9</td>
<td>1.1 ± 0.9</td>
<td>0.8* ± 0.7</td>
<td>0.9 ± 0.7</td>
<td>1.3* ± 1.0</td>
</tr>
<tr>
<td>Vitamin C (mg)</td>
<td>94.7 ± 50.2</td>
<td>79.6 ± 64.9</td>
<td>52.8** ± 33.5</td>
<td>77.9 ± 59.8</td>
<td>57.7 ± 57.2</td>
</tr>
<tr>
<td>Vitamin E (α-tocopherol eq)</td>
<td>8.9 ± 4.9</td>
<td>8.3 ± 4.7</td>
<td>6.6* ± 2.9</td>
<td>6.8 ± 5.0</td>
<td>9.3*** ± 4.6</td>
</tr>
<tr>
<td>Folate (mg)</td>
<td>334.6 ± 178.6</td>
<td>282.7 ± 130.6</td>
<td>248.5* ± 114.5</td>
<td>286.4 ± 194.6</td>
<td>318.7 ± 145.0</td>
</tr>
<tr>
<td>Magnesium (mg)</td>
<td>278.7 ± 137.7</td>
<td>239.5 ± 114.7</td>
<td>201.7* ± 70.8</td>
<td>212.6 ± 127.9</td>
<td>279.2*** ± 125.6</td>
</tr>
<tr>
<td>Potassium (mg)</td>
<td>2539.4 ± 1167.7</td>
<td>2315.4 ± 1221.8</td>
<td>1919.2* ± 627.4</td>
<td>1779.8 ± 1144.2</td>
<td>2634.6*** ± 1169.3</td>
</tr>
<tr>
<td>Copper (mg)</td>
<td>1.3 ± 0.8</td>
<td>1.1 ± 0.6</td>
<td>0.9* ± 0.5</td>
<td>1.1 ± 0.6</td>
<td>1.3 ± 0.7</td>
</tr>
</tbody>
</table>

Data are presented as means plus or minus the standard deviation. One subject did not provide information on race. * = $p < 0.05$, *** = $p < 0.0001$ compared to the control group. * = $p < 0.05$ compared to AA group.
Figure 4-1. Diversity indices in stool communities between cases and controls. Error bars represent the standard deviation. Total n = 114 (57 cases, 57 controls).
Figure 4-2. qPCR results showing no significant differences between demographic groups for lactic acid bacteria and bifidobacteria in stools. Error bars represent the standard deviation. BMI is defined as weight in kg divided by height in meters squared. All subjects for whom demographic data were provided are included. Eight subjects did not provide BMI data. Two subjects did not provide race data. Total (n = 111; 55 controls, 56 cases [29 high-risk cases]), BMI < 25 (n = 30; 15 controls, 15 cases [7 high-risk cases]), BMI 25 – 30 (n = 40; 12 controls, 22 cases [14 high-risk cases]), BMI > 30 (n = 33; 18 controls, 15 cases [5 high-risk cases]), AA (n = 25, 14 controls, 11 cases [6 high-risk cases]), CA (n = 84; 39 controls, 45 cases [23 high-risk cases])
Figure 4-3. qPCR results showing no significant differences between demographic groups for lactic acid bacteria and bifidobacteria in biopsies. Error bars represent the standard deviation. BMI is defined as weight in kg divided by height in meters squared. All subjects for whom demographic data were provided are included. Two subjects did not provide BMI data. Total (n = 78; 35 controls, 43 cases [22 high-risk cases]), BMI < 25 (n = 19; 8 controls, 11 cases [6 high-risk cases]), BMI 25 – 30 (n = 34; 13 controls, 21 cases [13 high-risk cases]), BMI > 30 (n = 23; 14 controls, 9 cases [2 high-risk cases]), AA (n = 10, 6 controls, 4 cases [2 high-risk cases]), CA (n = 68; 29 controls, 39 cases [20 high-risk cases])
Figure 4-4. qPCR results showing no significant differences between BMI groups for lactic acid bacteria and bifidobacteria in stools (A) and biopsies (B). Error bars represent the standard deviation. BMI is defined as weight in kg divided by height in meters squared. All subjects that provided BMI data are included. Stools (total n = 103; n = 30 BMI < 25, n = 40 BMI 25–30, n = 33 BMI > 30). Biopsies (total n = 76; n = 19 BMI < 25, n = 34 BMI 25–30, n = 23 BMI > 30).
Figure 4-5. qPCR results showing abundances for lactic acid bacteria and bifidobacteria between racial groups in stools (A) and biopsies (B). Error bars represent the standard deviation. All subjects for whom racial data were provided are included. * = $p < 0.05$. AA: $n = 10$; CA: $n = 68$. 
Figure 4-6. Rarefaction curves for stool and biopsy communities based on Chao1, which is a measure of estimated diversity if sequenced to completion. Stools (total n = 60; 30 cases and 30 controls); Biopsies (total n = 32; 16 cases and 16 controls).
Figure 4-7. Principal coordinate analysis (PCoA) based on UNIFRAC showing differences between stool and biopsy microbiota (A), microbiota by cases status in stool (B) and biopsy samples (C), and microbiota by racial group (D). Stools (total n = 60; 30 cases and 30 controls); Biopsies (total n = 32; 16 cases and 16 controls); Race (total n = 55; 18 AA and 37 CA).
Figure 4-8. Relative abundance of bacteria phyla between cases and controls in stools (A) and biopsies (B).
Figure 4-9. Heat map showing the distribution of the 20 most significantly differing individual OTUs between all cases and controls in stools at the 95% similarity level. Each column represents data from a single subject. Rows represent individual OTUs. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtain from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, mcrb = mixed culture rumen bacterium.
Figure 4-10. Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases and controls in stools at the 95% similarity level. Each column represents data from a single subject. Rows represent individual OTUs. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, mcrb = mixed culture rumen bacterium.
Figure 4.11. Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases defined conventionally and controls in stools at the 95% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, mcrb = mixed culture rumen bacterium.
Figure 4-12. Heat map showing the distribution of the 20 most significantly differing individual OTUs between all cases and controls in stools at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, bpb = butyrate-producing bacterium.
Figure 4-13. Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases and controls in stools at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtain from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, bpb = butyrate-producing bacterium.
Figure 4-14. Heat map showing the distribution of the 20 most significantly differing individual OTUs between high-risk cases defined by convention and controls in stools at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtain from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. rb = rumen bacterium, bpb = butyrate-producing bacterium.
Figure 4-15. Heat map showing the distribution of all significantly differing individual OTUs between all cases and controls in biopsies at the 95% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. bpβ = butyrate-producing bacterium.
Figure 4-16. Heat map showing the distribution of all significantly differing individual OTUs between high-risk cases and controls in biopsies at the 95% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. uec = unidentified Eubacterium clone, bpb = butyrate-producing bacterium.
Figure 4-17. Heat map showing the distribution of all significantly differing individual OTUs between all cases and controls in biopsies at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. hif = human intestinal firmicute.
Figure 4-18. Heat map showing the distribution of all significantly differing individual OTUs between high-risk cases and controls in biopsies at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. hif = human intestinal firmicute.
Figure 4-19. Heat map showing the distribution of the 20 most significantly differing individual OTUs in stools between racial groups at the 95% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. sfb = swine fecal bacterium, hif = human intestinal firmicute.
Figure 4-20. Heat map showing the distribution of the 20 most significantly differing individual OTUs in stools between racial groups at the 98% similarity level. Each column represents data from a single subject. Rows represent the OTU. Row numbers represent the OTU ID. Percentages represent the percent identity of the representative sequence to the database match. Cell shade signifies the number of sequence reads matching to the specified OTUs. Labels in parentheses were obtained from the RDP classifier algorithm and used when megablast results did not provide taxonomic information. sfb = swine fecal bacterium, rb = rumen bacterium, bpb = butyrate-producing bacterium.
Figure 4-21. Heat map based on prevalent OTUs in either case or control groups at the 92% similarity level (A). Samples 1 to 30 = control group, samples 31 to 60 = polyp group. Scale represents abundance of OTUs within a sample from 1 (sample with the highest number of stated OTU) to 0 (sample with stated OTU absent). The AUC demonstrates the accuracy of the model (B).
Figure 4-22. Number of shotgun sequence reads with closest matches to gamma-Proteobacteria in the case and the control pools. n = 18; 9 cases and 9 controls.
Figure 4-23. Number of shotgun sequence reads with closest matches within bacterial metabolic pathways in the case and the control pools. n = 18; 9 cases and 9 controls.
CHAPTER 5
EFFECTS OF SPECIFIC COMPLEX CARBOHYDRATES ON FECAL MICROBIOTA
COMMUNITY COMPOSITION

Study #1: Resistant Maltodextrin Increases Bifidobacteria Counts in Healthy Males

Introduction

Resistant maltodextrin (RM), a type 3 resistant starch, is a soluble dietary fiber used as a food ingredient to improve qualities such as texture, taste, and health benefits, including reduced blood sugar and improved GI function. The most common form of RM is marketed as Fibersol-2, which consists of 90% RM. The process by which RM is made begins with starch that is hydrolyzed in the presence of heat and acid. The hydrolysis reaction breaks α1-4 and α1-6 glycosidic linkages. Because the amount of water present is limited, transglucosidation occurs between remaining starch polymers randomly forming β1-2, β1-3, β1-4, β1-6, α1-2, and α1-3 linkages. A-amylase is then added to remove α linkages and resulting glucose monomers are removed, leaving behind the remaining polymer defined as RM (35). RM is considered a dietary fiber by the following definition (35):

“Fiber is composed of carbohydrate polymers not hydrolyzed by endemic digestive enzymes, they are derived from food raw materials, and there must be a validated analytical procedure established that measures only the fiber fraction on the ingredient. Also, the material must demonstrate a physiological effect beneficial to humans.”

RM has been shown to induce smaller increases in blood sugar compared with a placebo, allowing it to be considered beneficial to health(34). RM is of interest to food producers in that it exhibits low levels of GI discomfort and provides many of the other benefits commonly observed with other dietary fibers (35). RM was investigated to determine its effects on energy gain and gut microbiota composition, the latter of which
is presented below, under the hypothesis that RM would increase the numbers of bacteria that could selectively ferment RM and induce changes in gut microbiota composition such that similarities arise between RM treated subjects and polyp-free subjects. Data were obtained using DGGE, qPCR, and pyrosequencing as described previously. The data were also compared to those obtained from subjects with polyps and subjects without polyps presented in Chapter 4. Microbiota profiles of RM-treated subjects that are consistent with those from subjects without polyps suggest that supplementation with RM may offer one avenue for increased protection against CRC.

**Study Design**

Fifteen male volunteers were recruited to participate in a randomized, double-blind, placebo controlled, crossover study with three 28 day feeding periods separated by a wash-out period of at least two weeks. 14 of the 15 subjects who began the study completed the entire study protocol. One subject failed to complete the study due to scheduling conflicts. Only data from those 14 subjects who completed the study are included in the data analysis. Fibersol-2® (Matsutani, Inc., Japan) was used as the RM in this study. Subjects were randomly assigned to treatments consisting of 25 g/day of resistant maltodextrin + 25 g/day maltodextrin (RM25), 50 g/day of resistant maltodextrin (RM50) and 50g/day maltodextrin (placebo). Prior to the first fecal collection, volunteers were adapted to a study diet for two weeks. During the three treatment periods, volunteers consumed the same base generic American diet. All foods and beverages were prepared and supplied by the Human Studies Facility at the Beltsville Human Nutrition Research Center (Beltsville, MD). Dinner and breakfast were consumed at the center during the week, and carryout lunches and snacks were provided. Weekend meals and treatments were packaged with instructions for home
consumption. The study protocol was reviewed and approved by the Medstar Research Institute Institutional Review Board. All subjects provided written informed consent and were compensated for their participation.

For microbiota analysis, fecal samples were collected on days 1, 13, and 24 of each intervention for a total of nine fecal samples per subject. Subjects obtained a cooler filled with ice for storage of the sample until delivery to the lab. For the microbiota analysis, all samples were delivered on ice usually within 4 hours of defecation.

**Results**

**Fecal microbiota community diversity**

DGGE, a simple but efficient method for initial profiling of fecal microbiota, suggested that each individual harbored a unique microbiota. A distinct band was consistently and dose-dependently increased after 24 days of RM supplementation in 12 out of 14 subjects (Figure 5-1). This increase was also observed, although to a lesser extent, after 13 days of RM supplementation. DNA isolated from two of these bands was cloned and sequenced. Both resulting sequences matched closest to sequences in the family Lachnospiraceae within the phylum Firmicutes.

**Bifidobacteria qPCR**

Quantitative PCR data was analyzed as described previously using genome equivalents/ng of DNA. Changes in amounts of bacterial groups of interest were determined by comparing the change of the genome equivalents/ng DNA between Day 1 to Day 24 during the RM period with the change between Day 1 to Day 24 during the placebo period. This approach allowed for consideration of the RM independent effect of the study diet.
A statistically significant increase in bifidobacteria from Day 0 to Day 24 was observed during the RM50 period, but not during the placebo period (Figure 5-2). 13 of the 14 subjects showed an increase in bifidobacteria from Day 0 to Day 24 during the RM50 period (50 g/day supplementation) while only 5 of 14 subjects showed such an increase during the placebo period (RM0). Furthermore, the mean numbers of bifidobacteria were higher on Day 24 during RM50 period ($2.7 \times 10^5$ genome equivalents/ng) compared to the placebo period ($1.5 \times 10^5$), although these data did not reach significance.

**454 pyrosequencing analysis of rDNA**

To expand the microbiota analysis from the limited number of targeted groups of gut bacteria, high-throughput microbial community 16S rRNA sequencing using the 454 titanium technology was performed. After removal of low quality and short reads, a total of 466,622 sequences were retained with an average of 5622 sequences per subject and an average length of 296 nucleotides per sequence read. Chao1 based rarefaction curves suggest that overall numbers of estimated OTUs did not differ between the treatments (Fig. 5-3). UniFrac-based PCoA analysis also did not detect differences in overall microbiota composition between treatments (Fig. 5-4). When the distribution of OTUs across treatment was evaluated, many OTUs were increased or decreased by RM (Fig. 5-5). Microbiota was initially compared on days 1 (top block in Figure 5-5) and 24 (middle block in Figure 5-5) during the RM50 period. However, because study subjects not only added RM to their diet but switched from their normal diet to the study diet, the sample collected on day 24 of the placebo treatment (bottom block in Figure 5-5) was included in the analysis. The OTUs that were increased during RM50 matched
closest to clostridia, ruminococci, bacteroidetes and parabacteroides as well as matches to unknown bacteria.

**Conclusion**

Changes within bacterial groups were detected with each experimental method; however, each method revealed changes in different bacterial groups. DGGE revealed in most subjects an increase in a bacterial group within the family Lachnospiraceae, a group of known butyrate producers, while qPCR showed an increase in bifidobacteria as a result of RM supplementation. An increase in bifidobacteria and butyrate production are expected based on previous work (29, 79). OTU-based pyrosequencing results showed increases in many OTUs matching to various taxonomic levels within the class Clostridia and decreases in other OTUs within the same class. These results indicate that RM supplementation has some relatively non-specific effects on higher level taxonomic bacterial groups while simultaneously exhibiting OTU-specific effects within other groups; however, there were no observed differences in overall diversity or community structure. There were no clear associations between OTUs enriched by RM and OTUs that were found to be significantly more prevalent in polyp-free subjects.

**Study #2: The Addition of Whole Grains to the Diets of Middle-school Children: Effects on Fecal Microbiota Community Structure**

**Introduction**

The Food and Drug Administration defines whole grains to be cereal grains that consist of the intact, ground, cracked or flaked fruit of the grains whose principal components -- the starchy endosperm, germ and bran -- are present in the same relative proportions as they exist in the intact grain (http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108598.ht
The bran consists primarily of insoluble fibers including cellulose and hemicelluloses, while the germ consists primarily of phytochemicals and soluble fibers, including oligosaccharides and resistant starch (RS). The endosperm, however, is rich in starch and provides energy for the developing plant (122). The refining process by which whole grains are processed into refined grains involves the removal of the bran and the germ. Although often fortified with B vitamins, this process removes most of the fibers and phytochemicals present in whole grains (3). Phytochemicals, defined as bioactive, non-nutrient plant compounds that are associated with reduced risk of chronic diseases, include many plant phenolic compounds that often serve as antioxidants (144). The demonstrated esterase activity of fecal homogenates suggests that gut microbiota activity may increase bioavailability of phytochemicals, specifically ferulic acids, and thus contribute to the protective effect of whole grains (6, 137, 205).

Other protective effects of whole grains may rely on the ability of whole grains to act as prebiotics. Whole grains may be considered prebiotics due to the presence of oligosaccharides and RS. Substrates including RS that enrich bacterial groups, particularly butyrate producers, offer another mechanism by which whole grain consumption promotes GI health. Butyrate production has been shown to be proportional to RS present in grains (105).

The associations of whole grain intake with beneficial health outcomes and reduced cancer risk, as well as the ability of their components to modulate gut microbiota composition, led to the hypothesis that whole grain consumption shapes gut microbiota community structures such that they display some similarities with gut microbiota communities found in polyp-free patients; however such similarities would
likely be subtle due to dietary differences other than grains and factors other than diet that contribute to polyp formation. The goal of this study was to determine the changes in gut microbiota composition as a result of whole-grain intervention. Stool pH was measured and microbiota analysis will be done using DGGE, qPCR, and pyrosequencing as described previously. These data show that gut microbiota composition may be influenced via dietary fiber intake.

**Study Design**

**Subject recruitment**

Adolescents were recruited from a local middle school (ages 11 to 15 years). A total of 196 adolescents were screened and 91 assented to participate with the consent of at least one parent. Eight participants were excluded because they did not meet inclusion/exclusion criteria \( n=4 \) or were no longer interested \( n=4 \). Eighty-three participants began the 1 week pre-baseline screen and were randomized to either the refined-grain group (RG, \( n=42 \)) or whole-grain group (WG, \( n=41 \)). All study procedures were approved by the University of Florida Institutional Review Board and the School Board of Alachua County. Inclusion criteria included the following:

- Parental/guardian consent
- Student assent
- Willingness to eat three different study foods each day for six weeks
- Willingness to provide two blood samples, 2 saliva samples, and 2 stool samples over the course of the study

Exclusion criteria included the following:

- Currently taking medication for constipation or diarrhea
- Antibiotics within the four weeks before randomization
• Consumption of probiotics of greater than three servings of yogurt per week
• Diseases or illnesses such as gastrointestinal disease (gastric ulcers, Crohn's, ulcerative colitis, etc.), other chronic diseases (diabetes, kidney disease, etc.) or immune-modulating diseases (HIV, AIDS, autoimmune, hepatitis, cancer, etc.)
• Food allergies (wheat, soy, egg, milk, gluten, nuts, or any other food or food ingredient)

Experimental protocol

Participants were enrolled in this six week, randomized, controlled, parallel arm study over a three week period in February, 2010. In the week prior to randomization, one stool sample and two targeted-24-hour recalls were collected, and height (portable stadiometer), weight (digital scale), and birth date were obtained to determine BMI percentile for age. Participants were stratified based on pre-baseline weight category (<95th BMI percentile for age and BMI<30; ≥95th percentile for age or BMI ≥30) and randomized to the RG or WG via sealed envelope. The randomization scheme was generated by a statistician who did not have contact with study participants. A grocery bag of a variety of grain-based foods was picked up at the school or delivered to the participants’ homes each week. During the final week of the study a stool sample was obtained. 83 subjects completed the study protocol, and 59 subjects provided both a baseline and a post-treatment stool sample. Data from 57 subjects were used for 454 pyrosequencing analysis due to barcoding errors in two samples.

Grain-based foods and administration protocol

Four rotating weekly food packages containing approximately ten different grain-based foods (e.g., cereal, pasta, rice, bread, pancake mix, snack foods) were provided to participants and their families. Two single-serving snack packs (i.e., whole-grain cereal [provided by Cereal Partners Worldwide] or 100-calorie cookie packs) were
distributed to study participants on school days. The study foods were primarily wheat-based but also included oats, rice, and corn. One serving of the study foods provided 0 g (RG) or 16±1 g (WG) of whole grains. Participants were told to eat the study foods in place of grains consumed as part of their typical diet, and they were encouraged to eat three different kinds of study foods each day with the goal of 5 oz-equivalents of grains per day.

Results

Fecal pH

Fecal pH was measured in all samples and revealed no differences within diet groups over time or between diet groups post treatment (Figure 5-6).

Fecal microbiota community diversity

The success of DNA extraction in all samples was confirmed by DGGE, but no distinct profiles or individual bands for either diet group or time period were revealed. Using the Shannon-Weiner and Simpson diversity indices, no significant differences in community diversity were detected (Figure 5-7). Profiles within participants remained stable across time periods, suggesting that the intervention did not considerably perturb microbiota profiles (data not shown).

Quantification of targeted bacteria

LAB and bifidobacteria were measured via qPCR. There was a significant diet group by time period interaction; LAB increased (P = 0.0002) between baseline and final time periods in stools from participants in the WG group (Figure 5-8). There was no difference in bifidobacteria between diet groups; however, there was a significant increase (P < 0.0001) in bifidobacteria in both diet groups. There was no effect of baseline weight category on LAB or bifidobacteria.
454 pyrosequencing of 16S rRNA analysis

After preprocessing, 387,933 16S rRNA sequences from 48 subjects were retained, with a mean of 3043 sequences/sample and a mean length of 498 nucleotides. Sequence readings were binned by using ESPRIT-Tree into 11722 OTUs at the 98% similarity level. Chao1-based rarefaction curves did not significantly differ between diet groups or across time periods, suggesting that overall diversity did not change during the study period (Figure 5-9). To evaluate changes in overall microbiota composition, PCoAs based on the unweighted UniFrac metric were generated. No distinct clustering by time period or diet group was detected (Figure 5-10). The proportions of dominant bacterial phyla did not significantly differ between diet groups over time (Figure 5-11). When the effects of WG consumption on individual OTUs were examined, 6 OTUs showed a strong positive association whereas 3 OTUs were negatively associated with WG consumption (Figure 5-12).

Conclusion

qPCR revealed an increase in lactic acid bacteria as a result of WG supplementation, which may explain the claimed beneficial effects in the literature regarding WGs. Enrichment of both lactic acid bacteria and bifidobacteria after WG supplementation has been demonstrated previously; however, no significant changes in bifidobacteria were observed in this study (41, 56, 56). Variation in the source of whole grains may account for this difference. Sequencing analysis revealed no significant differences in the proportions of bacterial phyla as a result of WG supplementation; however, at the individual OTU level, both increases and decreases in the prevalence of specific OTUs were detected. The prevalence of some OTUs matching to butyrate-producing bacteria was decreased while others were increased as a result of WG
supplementation. While these results confirm that gut microbiota composition is affected by WG supplementation, they suggest that the beneficial effect of WG consumption may not occur at the individual OTU level. There are no clear associations between the gut microbiota communities resulting from WG intervention and those found within polyp-free subjects.

Study #3: The Effects of Galactooligosaccharides on Fecal Microbiota Composition in 3.1 Undergraduate Students and 3.2 Aged Adults

Introduction

GOS, described previously, are of interest due to their ability to enrich butyrate-producing bacteria and other bacterial groups that negatively associate with CRC. The demonstrated ability of GOS to modulate gut microbiota community structure led to the hypothesis that subjects who consume GOS will develop some gut microbiota community characteristics that are similar those found in subjects without polyps. Because the diets of study subjects will not be controlled beyond GOS supplementation, dramatic similarities are not expected. Furthermore, polyp formation is a multi factorial process that involves parameters other than diet.

Gut microbiota composition has been shown to differ between age groups (19). Therefore, two studies in two different age groups were conducted to determine the effects of GOS on gut microbiota composition. The first study examined GOS effects in academically stressed healthy undergraduate university students during cold and flu season. The stress model required that all subjects be full time students with at least one final exam during the post-treatment sample collection period. Because stress is associated with weakened immunity and GI discomfort, this model was used to increase the likelihood of observing the beneficial effects of GOS.
The second study examined the effects of GOS on gut microbiota composition in aged adults (60 years or more) over cold and flu season. Studying this population is important because CRC risk increases with age. Data for stool pH and microbiota analyses using DGGE, qPCR, and pyrosequencing were obtained as described previously. Results from both studies were compared to determine both age-dependent and age-independent effects of GOS. Because both CRC risk and gut microbiota composition are dependent on age, GOS may affect these populations differently.

**Protocol 1: Galactooligosaccharides Supplementation in Healthy University Students**

**Study Design**

**Subject recruitment**

Participants from the University of Florida were recruited via list servs, flyers, posters, and announcements in early fall of 2009. Each subject gave written informed consent, and all study procedures were in accordance with the ethical standards of the University of Florida Institutional Review Board. Inclusion criteria included the following:

- Full time undergraduate student taking at least 12 credit hours
- Age of 18 years or older
- Willingness to complete daily assessment forms via computer
- Willingness to discontinue any immune-enhancing dietary supplements (e.g., prebiotics and fiber supplements, probiotics, Echinacea, fish oil, vitamin E > 100% of the RDA or > 15 mg/day
- Willingness to take the fiber daily for 8 weeks
- Willingness to provide a social security number to receive study payment
- A cold in the last 12 months
- At least one final exam during the Fall 2009 exam week, between Saturday,
December 12 and Friday, December 18, 2009

- Daily access to a computer with Internet access for the entire 8-week study

Exclusion criteria included the following:

- Current smoker
- Chronic allergies involving the upper respiratory tract
- Allergy to milk
- Known illnesses or conditions that may impact perceived health such as HIV/AIDS, diabetes, renal or gastrointestinal diseases
- Chemotherapy or other immune-suppressing therapy within the last year
- Antibiotic therapy in the past two months

**Experimental protocol**

Subjects (n = 427) were randomly assigned to a supplement group during the first week in November of 2009 and were followed for eight weeks, including the time of fall final exams. 75 subjects were asked to provide stool samples. Exams were held over a span of one week during the sixth week of the intervention. The number and scheduled time of exams varied with each subject’s academic course schedule. Once subjects completed their last exams, they were on semester break through the remainder of the study. The study was a prospective, randomized, parallel, double-blind, placebo-controlled trial. Subjects were proportionally stratified based on gender (50/50) and randomly assigned over a five-day period via sealed envelopes to receive 0, 2.5, or 5.0 g GOS (Corn Products, Golden, CO). The stratification and randomization schemes were generated by the study statistician who did not have direct contact with any subjects. On the basis of the number of upper respiratory tract infections in GOS-supplemented infants and the number of colds in academically stressed undergraduate
students, which were important for the immunological parameters examined in this study independent of the microbiota work, a sample size of 140 subjects in each group was calculated to observe a 50% reduction in the proportion of colds in the treatment groups with 80% power, $\alpha = 0.05$, and a 10% attrition rate (9, 134). 419 subjects completed the study protocol, of which only 69 provided both a baseline and post-treatment stool sample. Subjects provided baseline stools samples throughout the week before randomization and post-treatment samples throughout the week of final exams.

**GOS administration protocol**

The GOS supplements were provided in coded packets that were similar in size and shape to commercially available single-serving drink mixes. Bakers’ sugar (sucrose) was added to the 0-g and 2.5-g packets so that all packets were the same weight and looked similar. A flow agent (silicon dioxide) was added to all packets to improve emptying of the package contents. The subjects were instructed to pour the contents of the packet into any beverage, mix well, and consume the beverage in its entirety each day for eight weeks. Both the GOS and sucrose had a slightly sweet taste. The subjects were unable to distinguish the GOS packets from the placebo. Regardless of the actual treatment group, the proportions of subjects who thought they were receiving 0, 2.5, and 5.0 g GOS were 44%, 42%, and 14%, respectively. No differences were found between groups, which suggested successful blinding.

**Results**

**Fecal microbiota community diversity**

The success of DNA extraction in all samples was confirmed by DGGE, but no distinct profiles or individual bands for diet groups or time periods were revealed. Using the Shannon-Weiner and Simpson diversity indices, no significant differences in
community diversity were detected (Figures 5-13 and 5-14). Profiles within participants remained stable across time periods, suggesting that the intervention did not considerably perturb microbiota profiles (data not shown).

**Quantification of targeted bacteria**

LAB and bifidobacteria were measured via qPCR. There were no significant diet group by time period interactions for either LAB or bifidobacteria (Figures 5-15 and 5-16).

**454 pyrosequencing analysis of 16S rDNA**

After preprocessing, 553,152 16S rDNA sequences from 69 subjects were retained, with a mean of 5885 sequences/sample and a mean length of 394 nucleotides. Sequence readings were binned by using ESPRIT-Tree into 25595 OTUs at the 98% similarity level. The proportions of dominant bacterial phyla did not significantly differ between diet groups over time (Figure 5-17). Chao1-based rarefaction curves did not significantly differ between diet groups or across time periods, suggesting that overall diversity did not change during the study period (Figure 5-18). To evaluate changes in overall microbiota composition, PCoAs based on the unweighted UniFrac metric were generated. No distinct clustering by time period or diet group (data not shown) was detected. When the effects of GOS consumption on individual OTUs were examined, 50 OTUs showed a strong positive correlation whereas no OTUs were negatively associated with GOS consumption (Figure 5-19).
Protocol 2: Galactooligosaccharides Supplementation in Healthy Aged Adults

Study Design

Subject recruitment

Participants were healthy, older adults aged 60 or older who suffered at least one cold in the previous year. Each subject gave written informed consent, and all study procedures were in accordance with the ethical standards of the University of Florida Institutional Review Board. Inclusion criteria included the following:

- Age of 60 years or older
- Willingness to complete daily and monthly questionnaires
- Willingness to receive the fall influenza vaccination as part of the study protocol
- Willingness to provide three blood samples, three saliva samples, and two stool samples and answer a food frequency questionnaire over the course of the study
- Willingness to discontinue and immune-enhancing dietary supplements (e.g., prebiotics and fiber supplements, probiotics, Echinacea, fish oil, vitamin E > 100% of the RDA or > 15 mg/day)
- Willingness to take the study fiber for 24 weeks
- Willingness to provide a social security number to receive study payment
- A cold within the last 12 months
- Ability to take foods and the study fiber without the aid of another person
- Eligibility and Willingness to receive the influenza vaccine for the current year

Exclusion criteria included the following:

- Current smoker
- Chronic allergies involving the upper respiratory tract
- Currently taking medication for constipation or diarrhea
- Currently taking any anti-inflammatory drugs on a regular basis
Current treatment for Alzheimer's disease
- Milk or serious egg allergies
- Current treatment for any known illnesses or conditions that may impact perceived health such as HIV/AIDS, immune modulating diseases (autoimmune, hepatitis, cancer, etc.), diabetes, kidney disease, gastrointestinal diseases (gastric ulcers, Crohn's, ulcerative colitis, etc.)
- Chemotherapy or other immune-suppressing therapy within the last year
- Antibiotic therapy within the last two months
- Current use of supplemental oxygen

Experimental protocol

Subjects (n = 80) were randomly assigned to a supplement group during the September of 2010 and were followed for six months. The study was a prospective, randomized, parallel, double-blind, placebo-controlled trial. Subjects were proportionally stratified based on gender (50/50) and randomly assigned via sealed envelopes to receive 0 or 5.0 g GOS (Corn Products, Golden, CO). The stratification and randomization schemes were generated by the study statistician who did not have direct contact with any subjects. The GOS supplements were provided in 0 g or 2.5 g coded packets and subjects were instructed to consume the GOS in a beverage.

Based previous work with this population (135), a significant difference in cold and flu symptom intensity score will be observed with 80 subjects (includes 20% attrition rate). The rationale for using a sample size based on cold and flu symptoms is based on other primary outcomes of the study protocol not related to microbiota analysis. 81 subjects completed the study protocol, and 80 provided both a baseline and a post-treatment stool samples. Subjects provided a baseline stool sample during the randomization week and a post-treatment sample after two weeks of supplementation.
**GOS administration protocol**

The GOS supplements were provided in coded packets that were similar in size and shape to commercially available single-serving drink mixes. Bakers’ sugar (sucrose) was added to the 0-g and 2.5-g packets so that all packets were the same weight and looked similar. A flow agent (silicon dioxide) was added to all packets to improve emptying of the package contents. The subjects were instructed to pour the contents of the packet into any beverage, mix well, and consume the beverage in its entirety twice daily for 6 months. Both the GOS and sucrose had a slightly sweet taste. The subjects were unable to distinguish the galactooligosaccharide packets from the placebo.

**Results**

**Fecal pH**

Stool pH was measured in all samples and revealed no differences between diet groups over time or between diet groups post treatment (Figure 5-20).

**Fecal microbiota community composition**

The success of DNA extraction in all samples was confirmed by DGGE, but no distinct profiles or individual bands for either diet group or time period were revealed. Using the Shannon-Weiner and Simpson diversity indices, no significant differences in community diversity were detected (Figure 5-21). Profiles within participants remained stable across time periods, suggesting that the intervention did not considerably perturb microbiota profiles (data not shown).

**Quantification of targeted bacteria**

LAB and bifidobacteria were measured via qPCR. There were no significant diet group by time period interactions for either LAB or bifidobacteria (Figure 5-22). However,
proportions of bifidobacterial genome equivalents were significantly greater in the GOS-treated group than the placebo group. (Figure 5-23). Proportions of genome equivalents detected using qPCR represent the number of target bacterial genome equivalents divided by the number of total genome equivalents determined by the V3 universal primer set.

The data were then fit into a model which predicted higher increases in bifidobacteria genome equivalent proportions for subjects with higher baseline values. Bifidobacteria as a proportion of all bacteria measured in the fecal samples were transformed using arcsin square root and analyzed as a function of the baseline proportion of bifidobacteria (also transformed), intervention group, age group, and the two-way interactions of these main effects using a linear model. Non-significant interactions were removed. GOS-treated subjects displayed significantly greater increases than placebo-treated subjects with the same baseline levels. (Figure 5-24). Lactic acid bacteria displayed similar behavior between treatment groups (data not shown). Furthermore, subjects aged 65 years or more displayed significantly greater increases than those aged less than 65 years. A similar relationship was observed between baseline value and age group for lactic acid bacteria (data not shown).

**454 pyrosequencing analysis of 16S rDNA**

After preprocessing, 719,946 16S rRNA sequences from 80 subjects were retained, with a mean of 8999 sequences/sample and a mean length of 480 nucleotides. The percentages of dominant bacterial phyla did not significantly differ between treatment groups over time (Figure 5-25). Sequence readings were binned by using ESPRIT-Tree into 27597 OTUs at the 98% similarity level. Chao1-based rarefaction curves did not significantly differ between diet groups or across time periods,
suggesting that overall diversity did not change during the study period (Figure 5-26).

To evaluate changes in overall microbiota composition, a PCoA based on the unweighted UniFrac metric were generated. No distinct clustering by time period or diet group was detected (Figure 5-27). When the effects of GOS consumption on individual OTUs were examined, several OTUs were enriched by GOS supplementation (Figure 5-28)

**Conclusion**

Both study populations exhibited no significant change in diversity indexes as result of GOS treatment. Similarly, qPCR results showed no significant changes in genome equivalents for Bifidobacteria or lactic acid bacteria in either study population; however, when genome equivalents were converted to proportions, there was a significantly higher proportion of bifidobacteria in the aged-adult GOS-treated subjects compared to their placebo counterparts post treatment, but this was not observed within the student study population. A dose-dependent increase in GOS was expected based on previous studies (63, 64). qPCR is used for bifidobacteria and lactic acid bacteria quantification because the low abundance of these groups. 16S sequencing is not optimized for quantification of organisms with low abundance because the technologies to date only capture a small percentage of the entire microbial community, and organisms present at low abundances may be missed or underrepresented. In these studies, 16S sequencing data for bifidobacteria did not parallel qPCR results. This is likely due to the lack of sufficient sequencing depth needed to accurately quantify lesser abundant organisms. Because of the incongruence between these two methods, it is difficult to determine which bacterial groups were reduced as proportions of bifidobacteria increased.
The Firmicutes:Bacteroidetes ratios were consistently higher across diet groups and both time points in the aged adult study compared to both diet groups and time points measured in the student study population; however, these ratios were not significantly affected by GOS treatment. Together with the bifidobacteria results, these data suggest that gut microbiota composition differs by age group and that GOS affects microbiota community composition differently depending on host age, which parallels the results presented Biagi, et al, (18).

The results from both studies revealed an increase in the prevalence of several OTUs as a result of GOS intervention. In most cases, GOS altered gut microbiota composition by increasing the prevalence of particular OTUs. In the aged-adult study, only the prevalence of nine out of forty OTUs significantly affected decreased. These data suggest that GOS mainly acts by enriching certain bacterial groups with little activity towards reducing bacterial groups. The OTUs that were increased in both GOS study cohorts matched to various Ruminococcus species, Faecalibacterium prausnitzii, and unidentified human intestinal Firmicutes. The data also show a greater number of OTUs significantly enriched in the student study population compared to the aged-adult study population. These results further suggest that GOS exhibits different effects within different age groups. However, in both groups GOS significantly increased the prevalence of many butyrate producers, which suggests a beneficial effect of GOS independent of age. The gut microbiota profiles of GOS-treated subjects did not display and striking similarities with the gut microbiota profiles found in polyp-free subjects. It is difficult to make meaningful comparisons across studies because the study populations were not consistent. Future studies should examine the affects of GOS in subjects with
polyps previously removed to determine the effects of GOS on subsequent polyp
development. Though direct comparisons to the polyp study are difficult, the increase in
the proportion of bifidobacteria as a result of GOS treatment in the aged adult study,
combined with results in the literature suggesting the protective effects of bifidobacteria,
suggests that GOS may be beneficial in protecting against CRC via this activity.
Figure 5-1. DGGE gel showing the enrichment of a particular band as a result of RM treatment. Numbers below lanes represent treatment doses or RM. Letters below lanes represent individual subjects. M = marker.
Figure 5-2. qPCR results showing abundances in Bifidobacteria genome equivalents between treatment groups over time (N = 14). Error bars represent the standard deviation. * = p < 0.05.
Figure 5-3. Rarefaction curves for RM treatment based on Chao1, which is a measure of estimated diversity if sequenced to completion (N = 14). Top: day 24 placebo; center: day 1 RM50; bottom: day 24 RM50.
Figure 5-4. UniFrac-based PCoA comparing RM baseline, RM post-treatment, and placebo treatment periods (N = 14).
Figure 5.5. Heat map showing the distribution of selected individual OTUs that significantly differed in stools between time points at the 95% similarity level. Each column represents data from a single OTU. Cell shade represents the proportion of a single OTU within a subject, summed with the proportion of that OTU in all subjects (N = 14). Numbers within cells represent the number of sequence reads matching to the specified OTU. Top block = RM50, day 1; center block = RM50, day 24; bottom block = RM0, day 24.
Figure 5-6. Differences in stool pH between WG and RG treatment groups over time (n = 59; 29 WG and 30 RG). Error bars represent the standard deviation.
Figure 5-7. Diversity indices in stool communities between WG and RG treatment groups over time (n = 59; 29 WG and 30 RG). Error bars represent the standard deviation.
Figure 5-8. qPCR results showing abundances of Bifidobacteria and Lactic acid bacteria (LAB) genome equivalents between WG and RG treatment groups over time \((n = 59; 29 \text{ WG and 30 RG})\). Error bars represent the standard deviation. \(* = p < 0.001\).
Figure 5-9. Rarefaction curves for WG treatment based on Chao1, which is a measure of estimated diversity if sequenced to completion (n = 57; 27 WG and 30 RG).
Figure 5-10. UniFrac-based PCoA comparing WG and RG treatment periods (n = 57; 27 WG and 30 RG).
Figure 5-11. Relative abundances of bacterial phyla in stools in WG-treated subjects at baseline (A), WG-treated subjects post-treatment (B), RG-treated subjects at baseline (C), and RG-treated subjects post-treatment (D) (n = 57; 27 WG and 30 RG).
Figure 5-12. Heat map showing the distribution of individual OTUs at the 98% similarity level from subjects within the WG treatment group that significantly differed in stools between time and treatment groups (n = 27). Each column represents data from a single subject. Rows represent the OTU. Cell shade signifies the number of sequence reads matching to the specified OTU.
Figure 5-13. Shannon diversity indices in stool communities between treatment groups over time in the GOS student study population (n = 69; n = 24 (0g), n = 21 (2.5g), n = 24 (5g)). Error bars represent the standard deviation.
Figure 5-14. Inverse Simpson diversity indices in stool communities between GOS treatment groups over time in the student study population (n = 69; n = 24 (0g), n = 21 (2.5g), n = 24 (5g)). Error bars represent the standard deviation.
Figure 5-15. qPCR results showing no significant differences in Bifidobacteria genome equivalents between GOS treatment groups over time in the student study population (n = 69; n = 24 (0g), n = 21 (2.5g), n = 24 (5g)). Error bars represent the standard deviation.
Figure 5-16. qPCR results showing no significant differences in Lactic acid bacteria (LAB) genome equivalents between GOS treatment groups over time in the student study population (n = 69; n = 24 (0g), n = 21 (2.5g), n = 24 (5g)). Error bars represent the standard deviation.
Figure 5-17. Relative abundances of bacterial phyla in stools in GOS-treated subjects (5.0g) at baseline (A), GOS-treated subjects (5.0g) post-treatment (B), placebo-treated subjects at baseline (C), and placebo-treated subjects post-treatment (D).
Figure 5-18. Rarefaction curves for GOS treatment in the student study population based on Chao1, which is a measure of estimated diversity if sequenced to completion ($n = 48$; $n = 24$ (0g), $n = 24$ (5g)).
Figure 5-19. Heat map showing the distribution of individual OTUs that significantly differed in stools between time and GOS treatment groups in the student study population. Each column represents data from a single subject (n = 48; n = 24 (0g), n = 24 (5g)). Rows represent the OTU. Cell shade signifies the number of sequence reads matching to the specified OTU.
Figure 5-20. Measurements of stool pH between GOS treatment groups over time in the aged adult study population (n = 80; n = 37 (0g), n = 43 (5g)). Error bars represent the standard deviation.
Figure 5-21. Diversity indices in stool communities between GOS treatment groups over time in the aged adult study population (n = 80; n = 37 (0g), n = 43 (5g)). Error bars represent the standard deviation.
Figure 5-22. qPCR results showing no significant differences in Bifidobacteria and Lactic acid bacteria (LAB) genome equivalents between GOS treatment groups over time in the aged adult study population (n = 80; n = 37 (0g), n = 43 (5g)). Error bars represent the standard deviation.
Figure 5-23. qPCR results showing differences in Bifidobacteria genome equivalent proportions between GOS treatment groups over time in the aged adult study population (n = 80; n = 37 (0g), n = 43 (5g)). Error bars represent the standard deviation. * = p < 0.05.
Figure 5-24. The predicted proportion of bifidobacteria in fecal samples after two weeks of supplementation by the proportion of bifidobacteria in baseline samples for healthy older adults ≥65 y (A) or <65 y of age (B) receiving 5 g GOS (dashed lines) or placebo (solid lines). Gray lines represent the 95% confidence interval.
Figure 5-25. Relative abundances of bacterial phyla in stools in placebo-treated subjects (5.0g) at baseline (A), placebo-treated subjects (5.0g) post-treatment (B), GOS-treated subjects at baseline (C), and GOS-treated subjects post-treatment in the aged adult study population (n = 80; 37 (0g), 43 (5g)).
Figure 5-26. Rarefaction curves for GOS treatment in the aged-adult study population based on Chao1, which is a measure of estimated diversity if sequenced to completion (n = 80; n = 37 (0g), n = 43 (5g)).
Figure 5-27. UniFrac-based PCoA comparing GOS and placebo treatment periods in the aged adult study population (n = 80; 37 (0g), 43 (5g)).
Figure 5-28. Heat map showing the distribution of individual OTUs that significantly differed in stools between collection periods in the GOS-treated group in the aged adult study population. Each column represents data from a single subject (n = 43). Rows represent the OTU. Cell shade signifies the number of sequence reads matching to the specified OTU.
CHAPTER 6
DISCUSSION AND CONCLUSION

The primary purpose of this work was to test the hypothesis that microbiota contributes to colorectal carcinogenesis. This hypothesis was tested by determining differences in gut microbiota composition between subjects with large intestinal polyps and those without polyps. Previous work has determined gut microbiota associations between CRC patients and healthy counterparts, but has not examined subjects in the pre-cancerous polyp stage (2, 204, 211). The results from this study are important in that they provide information about specific bacterial groups that may potentially be linked to polyp formation, as studies involving diseased subjects cannot determine whether the disease developed first and subsequently contributed to the gut microbiota environment or vice versa.

In this study differences between racial groups in microbiota composition were detected at the OTU level. Observed differences between cases and controls were also expected as data from the literature on CRC subjects suggests an association between the gut microbiota and the CRC microenvironment (26, 207, 215, 226). However, in contrast to previous studies we did not detect differences in bifidobacteria, which has been suggested to be protective against CRC (172, 183). This may be due to differences in methodologies, study designs, or the disease state of the subjects (pre-cancerous vs. developed CRC).

The results of this study may be affected by the limitations of each of the methodologies used. However, a variety of microbiota analysis methods were used to minimize the effects of method-specific biases. DGGE is limited in that it does not provide taxonomic classification of organisms, nor does it determine absolute
abundance. The 454 pyrosequencing generated roughly 250,000 sequence reads; however, gut luminal contents have been reported to contain approximately $10^{12}$ organisms per gram (203). Clearly, only a subset prevalent bacteria present in the gut has been sequenced. Many rare organisms are not detected using this method, though these organisms may not be large contributors to disease development. qPCR is a more specific method, but is limited in that there are not currently designed primer sets for each gut organism, and individual reactions for every organism, if possible, would be extremely laborious. Furthermore, PCR reactions used to generate amplicons introduce a bias as the universal primer sets, when used to calculate proportions of targeted bacterial groups, may have differing affinities for different 16s rDNA sequences. However, bacterial groups that are present at low abundances, such as lactic acid bacteria, are more accurately quantified by qPCR verses 16S sequencing. Fluorescence in-situ hybridization techniques can be used, but similar to qPCR, probes for all gut organisms have not been developed.

With the increase in sequencing capacities, the bioinformatics of analyzing the massive amounts of sequence data generated has become an important concern. Sequence reads must be grouped by similarity level into bins. Bins can be classified using taxonomy-dependent approaches which rely on matching representative sequences to a database. This method introduces a limitation in that only organisms present in the database can be matched. Taxonomy-independent methods can be used, but this creates difficulty in comparing results across studies in that no common label is used. Binning algorithms differ between programs used to process sequencing data, and data have shown that different results can be obtained from a single dataset
based on the binning procedure used. A hierarchical clustering algorithm can be used, which may produce more accurate bin assignments, but may require too much computational power for large data sets. A greedy heuristic algorithm can be executed with less memory but will produce different results depending on the seed sequence used.

Determining the percent similarity between sequence reads within the data set can be done different ways, which can also bias results. A multiple sequence alignment can be used, which may produce more accurate similarity assignments across the entire data set, but it is more computationally costly. A pair-wise sequence alignment is faster but does not consider all sequence reads simultaneously. Furthermore, using different similarity thresholds (e.g., 95% identity vs. 98% identity) can produce inconsistent binning results. Until a common binning procedure, sequence identity procedure, and a comprehensive database are used among studies, specific results at the OTU and phylotype levels may not always be consistent between studies.

Another consideration when interpreting these results is geographical location, which has been shown to associate with gut microbial community composition. These studies were conducted at only two study sites; therefore, generalization from these results applied to the global population would not be appropriate. Combining these results with other studies in meta analyses is also difficult due to differences in methodologies.

To further elucidate the associations between gut microbiota and CRC development, a prospective cohort study such as this one would need to continue until a number of subjects developed CRC such that there was sufficient power to detect
differences between these subjects and cancer-free controls. Because of the low incidence of CRC of roughly 45 per 100,000 (http://www.cdc.gov/cancer/colorectal/statistics/), the sample size needed to achieve the power to detect meaningful associations would be nearly 20,000 to achieve a power of 0.80. A study of this scale would require years of follow-up and an abundance of funding. Further development of next-generation DNA sequencing techniques may provide the tools necessary to detect associations between less abundant members of the gut microbiota and CRC development.

The results from the diet studies demonstrate the ability of fibrous dietary substrates to alter microbiota community composition as expected; however, these alterations did not clearly shape gut microbiota community composition to resemble that of polyp-free subjects. Therefore, the hypothesis that dietary fiber intake contributes to the development of a more protective gut microbiota composition cannot be confirmed.

A large limitation of comparing the results from the diet studies to those of the polyp study is that the controls were not the same. A more appropriate experiment would examine the changes in microbiota composition resulting from dietary fiber intake in subjects with previously polyps and matched cases to determine if dietary fibers help prevent future polyp formation.

Results in the literature often do not report the types of fiber investigated or differentiate between them. This work used specific substrates as opposed to foods that contain various types and amounts of dietary fiber, and showed that the affects of fiber on OTU enrichment and depletion are dependent upon the type of fiber supplemented; however, it is difficult to relate the specific results of this work to the generalized
epidemiological results in the literature. Furthermore, potential confounding variables, including dietary intake other than fiber, are often not addressed in the literature and are difficult to control.

In conclusion, gut microbiota appears to be associated with polyp prevalence. This work is significant in that the observed differences in microbiota diversity or the presence of indicator bacteria in this study might represent a novel target for future CRC screening tests. Diet-mediated changes in gut microbiota composition may be one method to reduce CRC risk, but the data from this work are inconclusive regarding the affective varieties and amounts of dietary fibers. Though it is now clear that there are differences in gut microbiota community composition between subjects with polyps and those without polyps, further studies should be done to determine the effects of dietary fiber on polyp development over time.


78. **Fasano A.** 2012. Leaky Gut and Autoimmune Diseases. Clinical Reviews in Allergy & Immunology **42**:71-78.


Johansson MEV, Larsson JMH, Hansson GC. 2011. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proceedings of the National Academy of Sciences of the United States of America 108:4659-4665.


BIOGRAPHICAL SKETCH

Tyler Culpepper was born in Orlando, FL. He moved to Gainesville, FL in 2004 after graduating from high school. He then attended the University of Florida where he earned dual Bachelor of Science degrees in chemistry and microbiology and cell science. After graduation, he matriculated in graduate school in attempt to earn a Doctor of Philosophy in microbiology and cell science.

As an undergraduate student, Tyler’s research interest expanded while working in Dr. Valerie de Crecy’s laboratory. He then started in the University of Florida Department of Microbiology and Cell Science (MCS) graduate program in August 2008 under the guidance of Dr. Volker Mai. As a graduate student, he has attended research symposiums such as the Florida Genetics Institute Annual Symposium (2010), the Microbiology and Cell Science Graduate Student Symposium (2009, 2010, 2011, 2012), the Moffit Cancer Center Scientific Retreat (2010), the Emerging Pathogens Institute Research Day (2011), the America Society of Microbiology General Meeting (2010), and the Experimental Biology General Meeting (2012). He received the Graduate Student Council Travel Award and the MCS Graduate Student Symposium first-place poster presentation award in 2012.

In addition, he has served as a mentor for three undergraduate students, Madeline Bost, Fleeta Netter, and Adeeb Rohani, and has assisted graduate students Varinder Pannu, Venkateswaran Ganesan, Tyler Caton, and Shu-Jui Hsu in thesis work or laboratory rotations. He served as the faculty/student liaison and assisted in organizing the MCS Graduate Student Symposium (2010-2011). He also assisted as a poster judge for the 2011 Undergraduate Microbiology Research Symposium.
The work presented in this document has generated five manuscripts, some of which are in preparation. Upon completion of his degree, Tyler plans to transition into medical science by first pursuing a Doctor of Medicine degree while continuing to conduct research and subsequently completing a residency and fellowship to prepare for a career in academic medicine.