Identification and Characterization of Butyrate-Producing Species in the Human Gut Microbiome

Grace Maline
gmaline@unomaha.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/university_honors_program

Part of the Alternative and Complementary Medicine Commons, Bacteria Commons, Carbohydrates Commons, Community Health and Preventive Medicine Commons, Digestive System Diseases Commons, Enzymes and Coenzymes Commons, Immune System Diseases Commons, Lipids Commons, and the Therapeutics Commons

Recommended Citation
https://digitalcommons.unomaha.edu/university_honors_program/148

This Dissertation/Thesis is brought to you for free and open access by the University Honors Program at DigitalCommons@UNO. It has been accepted for inclusion in Theses/Capstones/Creative Projects by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.
Identification and Characterization of Butyrate-Producing Species in the Human Gut Microbiome

Grace Maline

Abstract—Inflammatory Bowel Diseases (IBD) including ulcerative colitis, Crohn’s disease, and indeterminate colitis are increasingly common conditions that places a high physical and financial burden on individuals and global healthcare systems. Though treatments exist for these conditions, their unpredictable nature and causation make them difficult to manage consistently across the variety of IBD patients. Additionally, many of these treatments come with undesirable side effects or modes of delivery. Therefore, we consider the use of Short Chain Fatty Acids (SCFAs) such as butyrate, whose affects in the human gut include decreased inflammation and decreased risk of colorectal cancer. As butyrate is produced primarily by microbial species within the human gut microbiome, this project’s aim was to identify butyrate producing bacteria via metagenomic analysis based on the presence of genes that are essential to pathways of butyrate production. In total, 73 different potentially butyrate-producing species were identified and characterized. Species identified include both gram-positive and gram-negative anaerobic bacteria belonging to the classes Clostridia, Fusobacteriia, Negativicutes, Bacilli, and Bacteroidia. Of these, Clostridia was the highest scoring class of bacteria.

Index Terms—IBD - Inflammatory Bowel Disease, SCFA - Short Chain Fatty Acid

I. INTRODUCTION

DISEASES involving inflammation of the digestive tract are a highly prevalent and debilitating reality for many individuals. Typically, these diseases are referred to as a collection of conditions known as Inflammatory Bowel Disease (IBD). Conditions that fall under the umbrella term of IBD include Crohn’s disease, ulcerative colitis, and indeterminate colitis [1]. In 2017, around 6.8 million cases of IBD were reported worldwide, which is almost twice the number of cases reported just three decades prior [2]. As the prevalence of this disease increases, so does its physical and financial burden both on those who suffer from these conditions and to global healthcare systems. Therefore, it is important that special attention is paid to IBD to establish non-expensive and effective treatments as well as preventative measures.

A. Motivation

IBD conditions are characterized by chronic, persistent relapses of inflammation in the lower digestive tract [3]. They are commonly associated with abdominal pain, rectal bleeding, weight loss, and perianal disease. As a result, patients with IBD report a large amount of discomfort and a substantial reduction in quality of life [2]. Unfortunately, the causes of IBD are often unknown and are overall poorly understood. However, it is suspected that individuals contain genetic predispositions to immune triggers that result in the inflammatory response [3]. It is often speculated that these triggers may come in the form of environmental factors or as a result of the interaction of host and microbial cells in the gut microbiome [4].

Because of their high prevalence and mysterious origins, much work in recent years has been dedicated to studying these diseases, their causes, diagnosis, and treatments. It has been found that induction of inflammation in the gut is often related to two well-known pathways, namely the Nuclear Factor-kappa B (NF-kB) pathway and the Mitogen-activated protein kinase (MAPK) pathway [5]. Though targeting pathway intermediates such as TNF-alpha with anti-inflammatory drugs and monoclonal antibodies can reduce gut inflammation, drugs that are able to do so often come with significant challenges [6]. Because they are non-specific to cell type, they produce undesirable side effects. Additionally, administration of the treatments may be invasive or uncomfortable [6]. Therefore, alternative prevention and treatments as they relate to intervention in these pathways are actively sought and studied.

Potential alternatives include Short Chain Fatty Acids (SCFA), which have shown increasing popularity in recent years. SCFAs such as butyrate, propionate, and acetate can be shown to correlate with decreased gut inflammation and decreased risk of colorectal cancer [7]. This is because SCFAs are the preferred energy source for colonicocytes [8]. Butyrate in particular is a widely studied SCFA and accounts for about 70 percent of the energy supply for colonicocytes, and inadequate supplies of butyrate in the gut are closely tied to inflammation and increased risk for cancer [7]. Butyrate acts by inhibition of NF-kB related inflammatory pathways in epithelial cells as well as by decreasing release of inflammatory cytokines [8]. Because of these anti-inflammatory effects, butyrate, among other SCFAs can be utilized as dietary supplements for gut health. However, there may be limitations to the effectiveness and delivery of direct oral supplementation of butyrate. Orally administrated butyrate products are often difficult to palate due to intense odor and poor taste [9]. Especially for pediatric patients, this can make it difficult or impossible to facilitate regular daily doses [9]. Therefore, more work is needed to investigate the potential for alternative methods of butyrate delivery to cultivate its anti-inflammatory benefits.

Though less explored than direct oral supplementation and prebiotic supplementation for existing flora, the possibility for...
a probiotic solution to this problem is very real. For this to be cultivated, deeper exploration of the individual species in the gut is needed both to identify the butyrate-producing bacteria in the gut and to characterize them by their potential to serve as probiotics [7]. Thus, this project will aim to make steps toward that possibility by identifying and characterizing microbial species in the human gut that could be butyrate-producers based on the presence of genes necessary for its production.

B. Hypothesis

The objective of this project is to identify gut microbial species that have butyrate production capabilities. Their ability to produce butyrate will be based on the presence of pathway genes within their genomes. The expected results will be a list of human gut microbial species that possess such criteria.

In doing so, this project can provide a broad view for further exploration of species that may be probiotic candidates. Future in vitro and in vivo studies to further explore the resulting microbial species could be established based on the findings of this project. After identification, the resulting species will be further evaluated by examining classes of bacteria by phylogenetic analysis and ANOVA analysis.

It is currently speculated that butyrate-producing bacteria are typically anaerobic, gram-positive firmicutes – mostly belonging to clostridial clusters IV and XIVa [10]. It is hypothesized that the results of these analyses will also fall within these classes of bacteria.

C. Background

Progress has already been made to begin identifying butyrate-producing species. Early studies were limited to in vitro observations of butyrate production and acetate utilization by species collected from fecal samples [11]. Enzymatic assays and biochemical assays using primers and qPCR to target important late pathway enzymes were popular methods of studying butyrate production and identifying species [11], [12]. However, these studies are limited by resources and scope. These limitations can be supplemented by sequence analysis and in silico approaches as proposed for this project.

An example of a successful in silico approach couples qPCR enzyme analysis with metagenomic searches for the the butyryl-CoA:acetate CoA transferase gene sequence in a handful of bacterial species suspected of butyrate production. Both methods arrived at primarily the same conclusions and same identified species, showing that both are adequate for prediction of butyrate producing species [12], [13]. Therefore, a widescale metagenomic analysis using pathway genes could provide valuable insight into more microbes with these capabilities.

To identify bacteria capable of producing butyrate, it is important to understand the pathways that bring about its production. Butyrate is typically produced in a pathway that branches from glucose fermentation during glycolysis [14]. Starting with acetyl-coA, a product of glucose fermentation, there are two ways to obtain butyrate. These two pathways shown in Figure 1 stem from a single pathway diverging at its final steps. These diverging branches are known as the Butyrate Kinase pathway and the Butyryl-CoA:acetate CoA transferase pathway [15]. Of these, the butyryl-CoA:acetate CoA-transferase pathway is suspected to be more prevalent in microbial species. Therefore, the butyryl-CoA:acetate CoA-transferase gene, encoding one of the pathway’s key enzymes, has been shown to be one of importance in identifying species that utilize this method of butyrate production [15]. Additionally, butyrate producing pathways with their essential enzymes can also be observed in their overall context in KEGG as a part of the Butanoate Metabolism pathway, (pathway id: ko00650).

II. MATERIALS AND METHODS

The activities for this project can be broken down into multiple steps for data collection, data generation, and analysis. These activities and the flow of data associated with them are summarized in Figure 2. All code for these activities can be found in a public github repository (https://github.com/gmaline/SeniorProject.git).

A. Data Sources

Sources of data for analysis will include whole genome sequences of microbial species of the human gut microbiome that have been made publicly available by the Human Microbiome Project (HMP) [16]. There are 457 microbial species from the gastrointestinal tract available at https://www.hmpdacc.org/hmp/HMRGD/. Much work has been put into sequencing and annotating genome data from HMP since its development [17]. Therefore, the quality of the reference genomes will be sufficient for this project.
Fig. 2. *Flow Diagram for Project Execution.* This diagram illustrates step by step the process for completing this project as broken into 6 main tasks: 2 rounds of Data Collection, Sequence Alignments, Scoring, ANOVA Analysis and Phylogenetic Characterization.

In addition, sequences of the genes from the butyrate production pathways will be pulled and investigated against the genomes from HMP. The genes identified as important in the butyryl-CoA:acetate CoA-transferase pathway and the butyrate kinase pathway are summarized in Table 1.

Finally, for the phylogenetic analysis of the resulting species, 16s rRNA sequences will be obtained from Silva (https://www.arb-silva.de/). Silva is a high-quality ribosomal RNA database supported by de.NBI [18].

### C. Collecting Pathway Enzyme Sequences

Additionally, the translated gene sequences of each enzyme were obtained from NCBI and processed such that each protein sequence was placed in its own fasta file with an identifiable header for subsequent analyses. Protein sequence fasta headers were as follows: thiolase, beta hydroxybutyryl-CoA dehydrogenase, crotonase, butyryl-CoA dehydrogenase, electron transfer flavoprotein alpha-subunit, electron transfer flavoprotein beta-subunit, butyryl-CoA-transferase, phosphahe butyryltransferase, butyrate kinase.

### D. Sequence Alignments

For each of the species CDS fasta files, a temporary nucleotide blastdb was built. Commands from the BLAST Command Line Application from NCBI were run on command line from within a python script using the os python library. For each species database built, a tblastn alignment of each protein sequence was performed using the python library NcbitblastnCommandline from Bio.Blast.Applications [19]. The 4,000+ alignment results files were stored in their own folder in an xml format.

#### Table 1. Butyrate Production Pathway Enzymes

<table>
<thead>
<tr>
<th>Enzyme Name</th>
<th>Pathway(s)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiolase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>beta-hydroxybutyryl-CoA dehydrogenase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Crotonase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>beta-hydroxybutyryl-CoA dehydrogenase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Electron transfer protein alpha</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Electron transfer protein beta</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Butyryl-CoA-CoA-transferase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Butyryl-CoA-dehydrogenase</td>
<td>Both</td>
<td>NCBI: DQ987697.1</td>
</tr>
<tr>
<td>Butyryl-CoA-kinase</td>
<td></td>
<td>NCBI: WP_003240701.1</td>
</tr>
<tr>
<td>Butyryl-CoA-dehydrogenase</td>
<td></td>
<td>NCBI: WP_003240701.1</td>
</tr>
</tbody>
</table>

#### Pseudocode for Alignments

1. FUNCTION findButyrateSpecies(Proteins, HMP):
2. butyrate_species $\rightarrow$ empty list
3. FOR microbe in HMP:
4. temporary_blastdb $\leftarrow$ makeblastdb(microbe)
5. FOR protein in Proteins:
6. alignment $\leftarrow$ tblastn(query=protein, database=temporary_blastdb)
7. IF alignment.score $\geq$ 60%:
8. butyrate_species.add(microbe)
9. RETURN butyrate_species

Fig. 3. *Pseudocode for Alignments.* The pseudocode shown describes a high level overview of how the sequence alignments were performed for the purpose of this analysis.
E. Scoring

To parse the xml results, the NCBI XML python library was used from Bio.Blast [19]. Significant hits were stored as csv output based on high scoring pairs with an e-value less than .01. The results of parsing the xml files took note of species, protein, percent identity, percent query coverage, and the aligned query and subject sequences. After these results were obtained, the resulting csv files were easily parsed and scored such that only those species with one or more hits at or above a 60 percent identity level were kept.

F. Phylogenetic and ANOVA Analyses

For comparison of the resulting potential butyrate-producing species, 16s ribosomal RNA sequences were obtained from Silva and compiled with identifiable headers [18]. A multiple sequence alignment and newick tree format were obtained from Clustal Omega [20]. The phylogenetic tree was visualized using Iroki, an automatic customization and visualization of phylogenetic trees tool [21].

In addition to creating the phylogenetic tree, the resulting scored species were labeled with taxon information. Breaking them down into classes of bacteria an ANOVA analysis of the scores was performed to look for differences in the strength of the representation of butyrate producing species by class of bacteria.

III. RESULTS

At a 60 percent identity threshold, 73 species were found to have at least one of the pathway enzymes in their genomes. Though some species had partial hits for the butyrate kinase pathway, only the Butyryl-CoA:acetate CoA Transferase pathway was represented in the results. Only 9 species had all of the pathway enzymes present at the specified identity threshold. However, many of the species that did not make the cut do have some identity with the pathway enzymes below and/or around the threshold. This is best seen by considering Figures 4A-D in which the identities for each protein are plotted with the threshold shown as a red horizontal line at 60 percent identity.

The scored results were categorized into three distinct categories, Definite Capability, Probable Capability, and Requires Further Validation based on having scores of 7, 5-6, and 1-4 respectively. A breakdown of how many species fell into each category is shown in Table 2.

Table 2. Breakdown of Scored Results. Describes the distribution of categorical data for the scored results of protein pathway alignment for the 457 HMP species.

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite Capability</td>
<td>All</td>
<td>9</td>
</tr>
<tr>
<td>Probable Capability</td>
<td>5-6</td>
<td>21</td>
</tr>
<tr>
<td>Requires Further Validation</td>
<td>1-4</td>
<td>43</td>
</tr>
</tbody>
</table>

Though it was hypothesized that the results would be exclusively firmicutes, two other phyla of bacteria were also found among the results of these analyses. These results are
observed in Figure 5A. Firmicutes represented a majority of the results (74 percent). However, 23 percent of the results were Fusobacteria and 3 percent were Bacteroidota which could lend itself to further evaluation of these phyla and their metabolic activities. Similarly, it was suspected that the results would consist of bacteria of the clostridial class exclusively. The analysis results returned the representation of four additional classes of bacteria as summarized in Figure 5B. The classes found include Clostridia, Fusobacteria, Negativicutes, Bacilli, and Bacteroidia.

Fig. 5. Breakdown of Phyla and Class from Potential Butyrate-Producing Bacterial Results Three phyla and Five total classes were observed among the results of the scored alignment analyses.

Using the scored data, an ANOVA analysis was performed between the 5 classes of bacteria represented in Figure 5B. Its purpose was to identify if there was a significant difference in means among the scores for each class. The results of the ANOVA analysis were significant at a 95 percent confidence level with a p-value of 0.0176. Thus, a Fischer’s LSD was performed to identify the significant differences. As summarized in Figure 6 and Tables 3-4, the significant differences indicate the Clostridia class is the highest scoring class among Bacteroidia, Bacilli, and Fusobacteria. Though it does appear that this class is also higher scoring than Negativicutes, there were not enough results in that class to confer statistical significance.

Table 3. Distribution of Score Means and Descriptive Statistics for 5 Classes of Potential Butyrate-Producing Bacteria.

<table>
<thead>
<tr>
<th>Class</th>
<th>mean</th>
<th>std</th>
<th>Min</th>
<th>Max</th>
<th>Q25</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacilli</td>
<td>2</td>
<td>1.7230508</td>
<td>0</td>
<td>3</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>Bacteroidia</td>
<td>1</td>
<td>2.9302378</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Clostridia</td>
<td>4.306122</td>
<td>2.3022378</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>3.117647</td>
<td>0.8574929</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Negativicutes</td>
<td>2</td>
<td>0.0261476</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4. Significant Results of Fischer’s LSD to Identify Differences in Scores of 5 Classes of Potential Butyrate Producing Bacteria.

<table>
<thead>
<tr>
<th>Class</th>
<th>difference</th>
<th>pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacilli</td>
<td>-2.306122</td>
<td>0.0585</td>
</tr>
<tr>
<td>Bacteroidia</td>
<td>-3.306122</td>
<td>0.0261</td>
</tr>
<tr>
<td>Clostridia</td>
<td>1.188475</td>
<td>0.0399</td>
</tr>
</tbody>
</table>

Finally, a phylogenetic analysis of the resulting species revealed the following phylogenetic tree (Figure 7). The tree ultimately divides into two major groups: Group I and Group II. Group I contains the Clostridia, Bacilli, Negativicutes, and Bacteroidia, suggesting that these groups share more similarities with the class of bacteria that were proposed in this project’s original hypothesis. On the other hand, Group II is composed entirely of the Fusobacteria class.

IV. DISCUSSION

The results of this project found phyla and classes of potential butyrate producing bacteria outside of those expected from the literature. It was hypothesized that potential butyrate producers are within the realm of the firmicutes phyla and are majorly found in clostridial class clusters. However, the results of these analyses indicated that Fusobacteria and Bacterioidota phyla as well as the classes Bacilli, Negativicutes, Fusobacteria, and Bacteroidia were also present.

Interestingly, though it was suspected that butyrate producing species include solely gram-positive anaerobes, the classes identified in this analysis also indicate the help of anaerobic gram-negative bacteria. Like Clostridium, the Bacillus class is an aerobac gram-positive bacteria [22]. However, Negativicutes, Bacteroides, and Fusobacteria are all negative straining anaerobic bacteria [23], [24], [25]. Therefore, more exploration is needed to better understand the type of bacteria with these capabilities.

This project was limited due to the scope and time constraints of completion this Spring. Ways that this project could have been improved with an extended timeline include the inclusion of additional microbial genome datasets, such as the integrated Human Microbiome Project (iHMP) or Culturalable Reference Genome (CGR) for a wider breadth of species to analyze [26], [27]. Additionally, the pathway proteins used for the alignments could have provided a better standard for comparison if a consensus sequence of multiple difference versions of the pathway enzyme genes could have been used.
This could provide guided directions for further analysis by targeting only those species that have the properties to fare well as a probiotic supplement. For instance, the class of Bacteroidia would likely not be worth pursuing as a probiotic as these types of bacteria are typically opportunistic pathogens [24]. Their potential anti-inflammatory benefits would be overshadowed by the risk of infection that they may pose if they become unbalanced in the microbiome or if they invade undesired tissues.

Some additional characteristics of a potential probiotic could include ensuring the bacteria is capable of adhesion and colonization in the colonic environment, having tolerance to acid and bile salts, and maintaining stability and vitality in the digestive tract [28]. If there are identifiable genes associated with these characteristics, preliminary analyses could be performed in silico before testing in vivo models.

V. CONCLUSION

In conclusion, butyrate production via microbial species in the human gut is a natural alternative to treating and preventing inflammatory bowel disease, avoiding the various undesirable side effects of current treatments. This project identified 73 potential butyrate producing species. Though the majority of the high scoring results fell within the scope of the hypothesized phyla and class of bacteria (firmicutes; clostridia), a handful of others were identified that are phylogenetically related to clostridia and might be worth further exploration. With further analysis and validation, these bacteria could be candidates for probiotic supplements for the treatment and prevention of IBD.

REFERENCES


