

Short-Chain Fatty Acid Production: Predictive Functional and Genomic Analysis of the
Microbiota in Crohn's Disease for Novel Plasmid-Based Therapy

Alice Khomski

Boston University Academy
Boston, MA

TABLE OF CONTENTS

Introduction.....	3
Purpose	4
Methods	4
Results & Discussion.....	6
Conclusion	17
Acknowledgments.....	18
References.....	19

INTRODUCTION

Crohn's disease and microbial imbalance

Inflammatory bowel disease, which includes Crohn's disease (CD) and ulcerative colitis, afflicts more than 3 million adults in the United States [1]. The prevalence of CD continues to rise, with more than 33,000 new U.S. cases per year [2]. Characterized by incurable, chronic inflammation, CD can affect any part of the gastrointestinal (GI) tract and most commonly manifests in the ileum. Its severity is unarguable: debilitating abdominal pain, rectal bleeding, persistent diarrhea, abscesses, and intestinal blockage comprise the most common symptoms [2-5]. Apart from these indicators, gut dysbiosis is also linked to CD onset [3-10].

Numerous studies have investigated the significant microbial alterations occurring in the GI tract and fecal matter of CD patients [6-10]. At the phylum level, the abundance of pathogenic *Proteobacteria* has been shown to increase almost four-fold in afflicted individuals, as compared to healthy patients. Commensal *Bacteroidetes* decreased more than two-fold in CD patients [9]. Several other reports demonstrate a substantial reduction in the abundance of commensal microbes spanning not only the *Bacteroidetes* but also the *Firmicutes* and *Actinobacteria* phyla. The reduced species, many of which produce short-chain fatty acids (SCFAs), include *Blautia obeum*, *Bacteroides uniformis*, *Odoribacter splanchnicus*, *Akkermansia muciniphila*, *Roseburia inulinivorans*, *Eubacterium rectale*, *Enterocloster lavalensis*, and *Faecalibacterium prausnitzii* [11]. Consistent with the decreases in these bacteria, SCFA levels become reduced as a whole during CD [9, 11].

Short-chain fatty acids (SCFAs)

SCFAs, namely propionate, butyrate, and acetate, are normally synthesized through various pathways encoded by specific genes. In the succinate pathways, for instance, methylmalonyl-CoA carboxyltransferase catalyzes the decarboxylation of methylmalonyl-CoA to propionyl-CoA, which is then reduced to propionate [12-14]. In the butyrate kinase pathway, moreover, butyryl-CoA is phosphorylated to produce butyryl-phosphate, a step catalyzed by phosphate butyryltransferase (encoded by the *ptb* gene) [15-16]. This butyryl-phosphate product can then be transformed to butyrate via the kinase (encoded by *buk*). Acetate is formed through a similar two-step pathway, where acetyl-CoA is converted to acetyl phosphate (with action from phosphate acetyltransferase/phosphotransacetylase, encoded by *pta*) and subsequently reduced to acetate via acetate kinase (encoded by the *ackA* gene). Besides the well-known *ptb*, *buk*, *pta*, and *ackA* genes, thioesterase genes like *fadM*, *menI*, *tesA*, *tesB*, *ybgC*, *yciA*, *entH*, *menI*, *tesA*, *tesB*, *ybgC*, *ybhC*, and *yigI* have also been shown to contribute to acetate and butyrate formation [17]. Regardless of the encoding gene or biosynthesis pathway taken, the presence of SCFAs in the gut microbiome is vital for intestinal health.

Microbiota-derived butyrate, in particular, serves as the primary energy source for intestinal epithelial cells (IECs), and its production is necessary for normal colonic metabolism [16]. Butyrate and propionate have demonstrated immunomodulatory effects as well. Both molecules engender the induction and proliferation of colonic regulatory T cells (cT_{regs}), especially cT_{regs} producing Foxp3 and interleukin-10 (*IL-10*); these are essential for cT_{reg} anti-inflammatory function and consequent limiting of inflammation [18-19]. Butyrate and other SCFAs inhibit the activity of histone deacetylase (HDAC) enzymes, which typically repress the transcription of crucial genes in inflamed colonocytes [20-21]. Through the activation of G protein-coupled receptors (GPCRs) within IECs, SCFAs also inhibit NF- κ B activation, effectively regulating, and in some cases hindering, pro-inflammatory cytokines (Figure 1). On the whole, SCFAs take on many functions in the gut, simultaneously promoting intestinal homeostasis and regulating inflammation [11]. Hence, the decline in microbial SCFA production during CD is a subject of much concern. The value of SCFAs to human health cannot be understated, and their drastic change in abundance during CD calls for advancements in microbiome-targeting treatments.

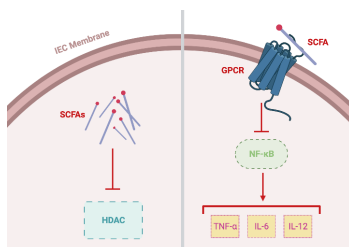


Figure 1. Short-chain fatty acid (SCFA) functions within the intestinal epithelial cell (IEC). SCFAs demonstrate anti-inflammatory effects through several mechanisms, including histone deacetylase (HDAC) inhibition. SCFAs also suppress pro-inflammatory cytokines (TNF- α , IL-6, and IL-12) through inhibition of NF- κ B via G-protein coupled receptors (GPCRs).

Methods of microbial intervention

With increasing evidence linking gut dysbiosis to CD onset, microbiota-based therapies are of considerable importance. Yet, the impacts of current therapeutics on SCFA production are unknown. Their beneficial effects, moreover, prove impermanent. The use of common antibiotics poses risks of emerging communal resistance, reduction in microbial diversity, and infection by *C. difficile* [3-4]. In addition, fecal microbiota transplantation (FMT; a newly introduced option previously used to treat recurrent *C. difficile* infection and currently being tested in clinical trials) has induced remission, but this effect has not been sustained. Side effects such as vomiting, high fever, and inflammatory marker elevation have also been reported; the introduction of pathogenic microbes and the potential for bacterial translocation remain concerns as well [21]. Ultimately, the insertion of foreign bacteria has shown to be problematic. Instead of integrating whole bacteria into the gut flora to produce desired changes, this study proposes another solution: the integration of engineered mobile genetic elements [21].

Mid-biotics have the potential to revolutionize CD treatment. Allowing for the genetic engineering of native gut bacteria and entire bacterial systems, engineered plasmids are the optimal choice for promoting various microbial functionalities. Specifically, plasmids can be used to introduce or remove genes from entire microbial communities. Their broad host range and autonomous transfer capabilities provide a clear advantage in this sense [21]. In contrast to FMT, mid-biotics offer a long-term remedy, as conjugative plasmids can undergo genomic amelioration to ensure permanent maintenance [21-22]. Such a treatment can mimic the beneficial effects of other therapeutics, including FMT, without the associated risks and disadvantages. This study aims to determine the optimal target bacteria and vectors for the induction of community-wide functional changes and promotion of SCFA synthesis.

PURPOSE

The most common treatments for CD do not directly address the diminished production of SCFAs, which are indispensable for healthy intestinal function. It is thus plausible that the nonpermanent or harmful nature of these therapeutics results from disregard for, or even exacerbation of, SCFA decreases. On the other hand, the beneficial results achieved using therapeutics may result from induced increases in SCFA production. In order to fully understand clinical results, these therapies must be evaluated from the perspective of microbial metabolic changes. Given that FMT has effectively induced remission (though mostly temporarily) in past studies, and that SCFAs are vital to intestinal health, I predict that FMT promotes the growth of SCFA-synthesizers and returns SCFA production pathway abundances to healthy levels.

Overall, this study aims to evaluate the effect of FMT on bacterial taxonomy and metabolic processes, specifically in terms of SCFA production, via Quantitative Insights Into Microbial Ecology (QIIME 2) and Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) analyses.

Most importantly, this project aims to provide the basis for a novel treatment option that directly influences SCFA levels in the gut, using the mid-biotics approach. The main objectives in this regard are to 1) establish target gut bacteria through analysis of genomic similarity to SCFA-producing taxa, via the Standard Nucleotide Basic Local Alignment Search Tool (BLAST), and 2) identify the most favorable plasmids through analysis of genomic signature distance between plasmids and select target bacterial chromosomes, via R-based Mahalanobis analysis.

METHODS

16S rRNA sequence data

Illumina MiSeq data was downloaded in the FASTQ file format from the Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>). To investigate the impact of FMT on patient microbiota, raw sequence reads (SRA accession number: SRP272340; 16S rRNA V4-V5 paired reads extracted from the stool) were acquired for CD patients at baseline (n = 42; untreated control), the same patients 3 months after undergoing transplantation (n = 42), and healthy patients (n = 42; overall control). All CD patients were over 10 years old and had been diagnosed for at least 6 months prior to the study [23]. The healthy patient population, moreover, consisted of both transplant donors and general volunteers.

Quantitative Insights Into Microbial Ecology (QIIME 2) analysis

The Quantitative Insights Into Microbial Ecology (QIIME 2) microbiome analysis package was used to assign taxonomy to the downloaded patient sequences. First, the raw sequence reads corresponding to each sample were imported into the QIIME 2 terminal environment in the manifest text file format. The paired sequence reads were then filtered based on quality score [24]. Specifically, the data was denoised with the Divisive Amplicon Denoising Algorithm (DADA2), which was used to truncate all forward reads at position 293 and all reverse reads at position 268. The first 20 base pairs were also trimmed off all sequences, and the forward and reverse reads were

then joined. This allowed for the generation of a representative feature table, along with high-quality, denoised representative sequences. Thereafter, taxonomy was assigned to these sequences using the Greengenes 13_8 Naive Bayes classifier (pre-trained on full-length sequences) within the q2-feature-classifier plugin [25-26].

To visualize the taxonomic differences among comparison groups, genus-level and species-level heatmaps were generated with the qiime2R and tidyverse R plugins. The ANCOM package in QIIME 2 was used to identify features exhibiting differential abundance among the three patient groups [27]. Taxa were considered significantly differentially abundant when the “reject null hypothesis” output was TRUE. Higher W values also indicated more significant differences in abundance between the three patient groups.

In addition to multivariate analysis, pairwise differential abundance was also determined, using ALDEx2 (q2-aldex2 plugin) in QIIME. Before analysis via ALDEx2, the previously generated feature table was filtered to produce three separate tables with only two patient groups each: FMT-treated CD and healthy, healthy and untreated CD, and untreated and FMT-treated CD patients. The three feature tables were then separately analyzed using the full ALDEx2 pipeline, yielding values for the difference between conditions, dispersion, effect size, and Welch’s t/Wilcoxon rank test statistics. Differential abundance was considered statistically significant when $p < 0.05$.

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) analysis

The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) program allowed for predictive analysis of the abundances of metabolic pathways and functional KEGG orthologs per sample. First, the high-quality representative sequences obtained as QIIME 2 outputs were imported into the PICRUSt2 terminal environment [28]. Using the full PICRUSt2 pipeline, the sample amplicon sequence variants (ASVs) were first placed in a phylogenetic reference tree upon their alignment with reference 16S sequences. As part of the hidden state prediction process, the nearest-sequenced taxon index values and the copy number of all Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog (KO) values were predicted for each ASV. Metagenome predictions were generated using the inferred copy numbers, as well as the ASV read depths. Finally, KO number quantities were regrouped to MetaCyc database reactions, and MetaCyc pathway abundances were thereby inferred.

Data profiling via MicrobiomeAnalyst

The abundance of these pathways and corresponding KOs were further investigated using MicrobiomeAnalyst, a web-based tool for metagenomic analysis [29-30]. Features with low counts and low variance were first filtered out of the data using default parameters (minimum count = 4; prevalence filter = 20%; cutoff variance percentage = 10%, based on inter-quartile range). The data was also normalized through centered log-ratio transformation.

Once the filtered and normalized data was obtained, the globaltest algorithm was utilized to test for associations between functional categories and the experimental factor (patient type), simultaneously producing a network visualization of the KO abundance table [31]. Besides pathway enrichment analysis, statistical testing was also done through the metagenomeSeq algorithm [32], which performed a zero-inflated Gaussian fit on the data, as well as through analysis of variance (ANOVA). KO features were identified as exhibiting significant differential abundance when $p < 0.05$. Accordingly, significant features were extracted, and their filtered and log-transformed abundance counts were visualized in box and whisker plots.

Determination of target host bacteria

The untreated CD patient representative sequences acquired through the QIIME 2 pipeline were also imported into Standard Nucleotide Basic Local Alignment Search Tool (BLAST, an online bioinformatics tool that compares nucleotide sequences and calculates the statistical significance of sequence matches) and searched against the 16S rRNA sequences of the previously mentioned SCFA-synthesizing species [33]. The BLAST algorithm was used to align the experimental sequences to those of the SCFA-producing taxa and generate an expected number of chance alignments value (E-value) for each match. Alignment was considered significant when $E < 0.001$, and when percent identity was between 85% and 99%. (Taxonomic features with percent identity greater than 99% were excluded to prevent the selection of naturally SCFA-synthesizing bacteria as targets for vector insertion). CD patient features with significant matches to SCFA-producers were selected as the desired bacterial targets. BLAST was again used to match the features to taxa in the RefSeq Representative genomes (refseq_representative_genomes) database to obtain complete genomes for these mid-biotic targets.

To confirm the BLAST alignment results, a phylogenetic tree was constructed using Molecular Evolutionary Genetics Analysis (MEGA X) software [34-35]. First, 16S rRNA sequences for both the SCFA-synthesizing species and the selected mid-biotic target bacteria were imported into the MEGA graphical user

interface from the NCBI GenBank database. All sequences were aligned using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm with default parameters [36]. Pairwise evolutionary distances between sequences were estimated using the p-distance model, which defines distance as the proportion of nucleotide sites at which two sequences differ [37]. To compute each p-distance, the number of nucleotide site differences was divided by the total number of nucleotide sites compared (Equation 1):

$$p = n_d/n$$

where n_d represents the number of nucleotide site differences and n represents the total number of nucleotides compared. Phylogeny was inferred based on the p-value calculations, via the neighbor-joining method.

Identification of optimal plasmids via Mahalanobis distance calculations

As successfully demonstrated by Tokuda *et al.* [38], the dinucleotide (2-mer) compositions of the target host bacteria were compared with those of select plasmids ($n = 92$; plasmids were included from 13 incompatibility groups, and sequences were downloaded from the GenBank database within NCBI; accession numbers were provided by Suzuki *et al.* [39]) in R. The dinucleotide composition was given by a vector consisting of 16 dinucleotide relative abundance values (x_{ij}). To compute each dinucleotide relative abundance value, the observed dinucleotide frequency was divided by the expected dinucleotide frequency (Equation 2):

$$x_{ij} = \frac{f_{ij}}{(f_i)(f_j)}$$

where f_{ij} represents the frequency of the dinucleotide ij , f_i represents the frequency of the mononucleotide i ($i \in A, C, G, T$), f_j represents the frequency of the mononucleotide j ($j \in A, C, G, T$), and the product of f_i and f_j gives the expected dinucleotide frequency.

The dissimilarity in 2-mer compositions was then measured through Mahalanobis distance (D^2) calculations [39]. Specifically, each plasmid sequence ($n = 92$) was compared with a set of 5-kb segments from each bacterial chromosomal sequence ($n = 13$). The following formula was used to compute the distance between each plasmid and target host bacterium (Equation 3):

$$D^2 = (X - \bar{Y})^T S^{-1} (X - \bar{Y})$$

where X represents the vector of the plasmid's dinucleotide relative abundance values, $\bar{Y}$ represents the mean vector of dinucleotide relative abundance values for the 5-kb bacterial chromosomal segments, S^{-1} represents the inverse variance-covariance matrix of the bacterial dinucleotide relative abundance values, and T represents the transposition operator. Lower D^2 values indicated less difference in genomic signature composition and thus higher genetic similarity.

For each Mahalanobis distance value, a P value (ranging from 0 to 1) was also provided. P values closest to 1 indicated significantly small Mahalanobis distances and thus significantly similar 2-mer compositions. In this study, plasmids with $P \geq 0.50$ were considered significantly close in genomic signature to target host bacteria; these plasmids were selected as the most advantageous vectors.

RESULTS & DISCUSSION

Taxonomic analysis

The abundance of taxa present within patient samples was determined in QIIME 2 and analyzed with the ANCOM test in the q2-composition plugin. Multiple known SCFA-producing species, namely *Butyricicoccus pullicaecorum* (ANCOM $W = 235$), *Faecalibacterium prausnitzii* ($W = 297$), and *Bacteroides uniformis* ($W = 229$), significantly varied among patient groups (FMT-treated, healthy, and untreated). The genera *Roseburia* ($W = 199$), *Faecalibacterium* ($W = 197$), *Ruminococcus* ($W = 180$), *Bifidobacterium* ($W = 154$), *Anaerostipes* ($W = 175$), *Coproccoccus* ($W = 180$), and *Streptococcus* ($W = 265$), which all consist of SCFA-producing species, were also significantly differentially abundant. That is to say, the ANCOM “reject null hypothesis” output was TRUE for all aforementioned taxa.

These results were consistent with the heatmap plots of both genus (Fig. 2A) and species (Fig. 2B) abundances.

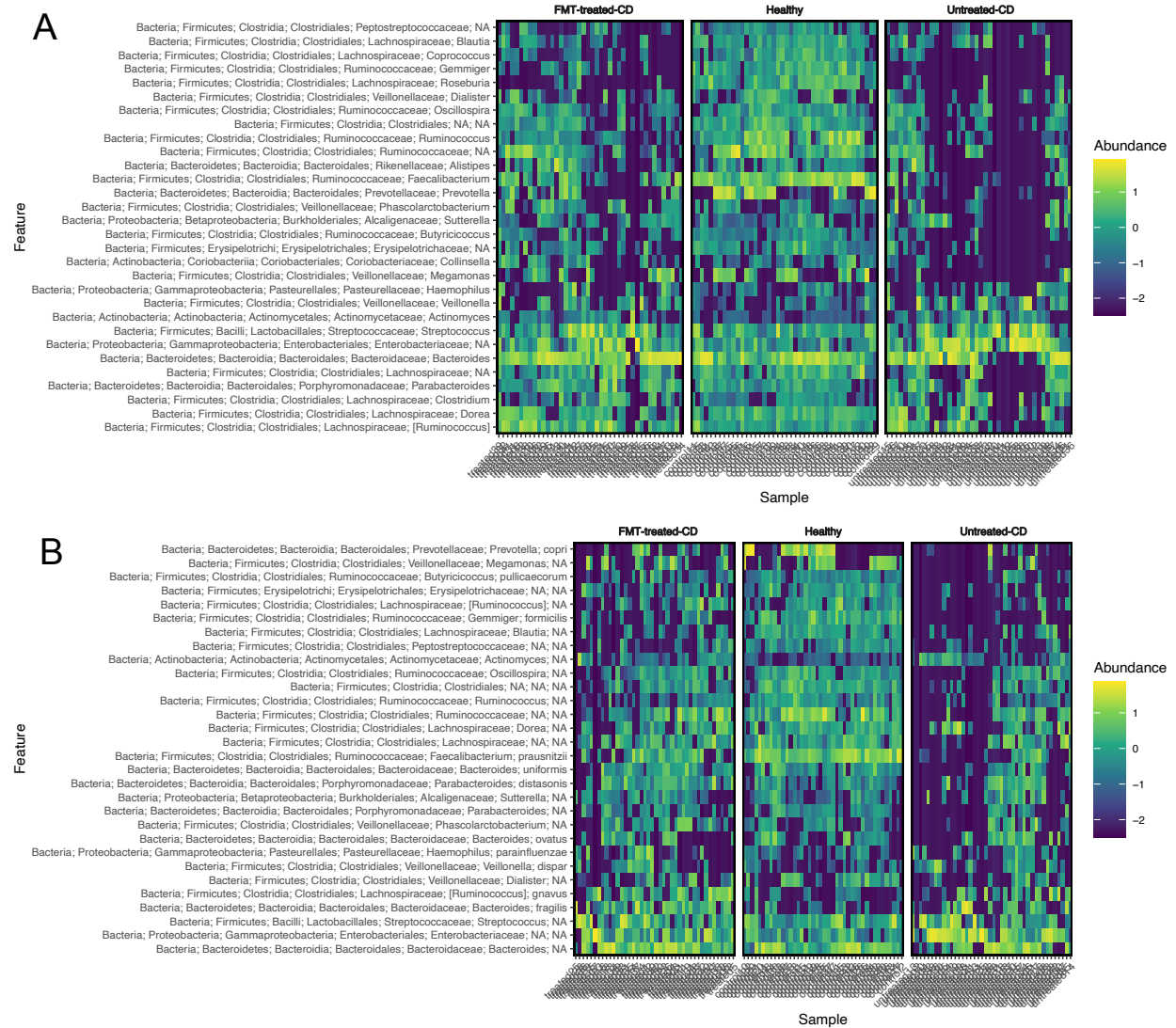


Figure 2. Heatmaps of taxa observed within the patient groups. Relative abundance of **A)** genera and **B)** species is shown for each sample, with yellow indicating highest abundance and dark blue indicating the lowest.

For instance, *Coprococcus*, *Ruminococcus*, *Faecalibacterium*, and *Butyrivibrio* were substantially reduced in the untreated CD patient group as compared to both the FMT-treated and healthy groups (Fig. 2A). FMT-treated patients, meanwhile, achieved abundance levels of these genera similar to those seen in healthy patients. Furthermore, *Bacteroides* was highly abundant in the FMT-treated group, with an overall abundance relatively equal to, and even greater than, that in the healthy group (and clearly greater than the untreated CD group). At the species level, *Butyrivibrio pullicaecorum*, *Gemmiger formicilis*, *Faecalibacterium prausnitzii*, and *Bacteroides uniformis* were considerably more abundant in healthy and FMT-treated patients than in untreated patients (Fig. 2B). Again, the abundances of these SCFA-producers in the FMT-treated group were comparable to those in the healthy group, demonstrating the efficacy of FMT in promoting the growth of SCFA-synthesizing taxa in CD patients.

In addition to multivariate analysis, pairwise comparisons of relative taxonomic abundance were made between all patient groups. Specifically, the ALDEx2 QIIME 2 plugin was used to conduct the Welch's t and Wilcoxon rank tests on the following paired groups: FMT-treated vs. healthy, healthy vs. untreated, and FMT-treated vs. untreated.

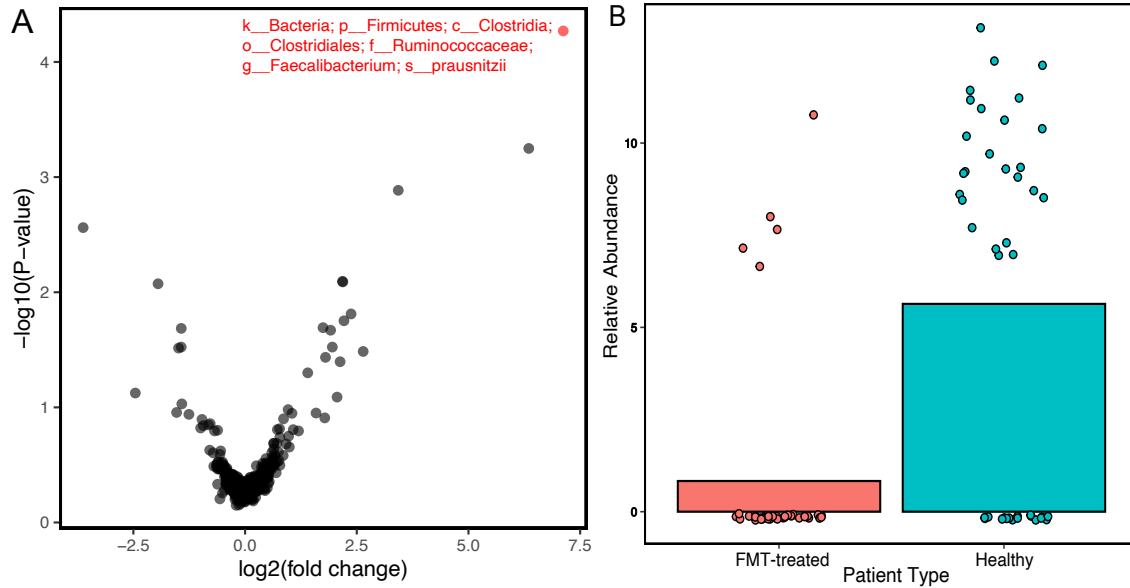


Figure 3. Pairwise comparison of the relative abundance of taxa in FMT-treated and healthy patients. A) A volcano plot was constructed using the Welch's t test and differential abundance results generated through ALDEx analysis. The x-axis shows the binary logarithm of the ALDEx difference output for each taxon, while the y-axis demonstrates the negative common logarithm of the Welch's t test p-value. Features were considered significantly differentially abundant when $p < 0.05$. Significant taxonomic features are highlighted in red. **B)** The centered log-ratio transformed abundance of the significantly differentially abundant taxonomic feature, *Faecalibacterium prausnitzii*, was calculated and plotted for each patient group.

As shown in Fig. 3A and Fig. 3B, *Faecalibacterium prausnitzii* was the only observed taxonomic feature (out of 3,279 total) that significantly differed in relative abundance between the FMT-treated and healthy patient groups ($p < 0.001$). Although the relative abundance of this butyrate-producer was reduced in FMT-treated patients in comparison to healthy controls (Fig. 3B), all other species (including SCFA-synthesizing ones) were at similar abundances in the two groups (Fig. 3A).

In contrast to the FMT-treated vs. healthy pairing, comparison of the healthy and untreated CD groups showed a larger number of significantly differentially abundant taxa, namely *Faecalibacterium prausnitzii* (here corresponding to two different Greengenes-classified features), *Gemmiger formicilis*, and features within the *Enterobacteriaceae* family (Fig. 4A).

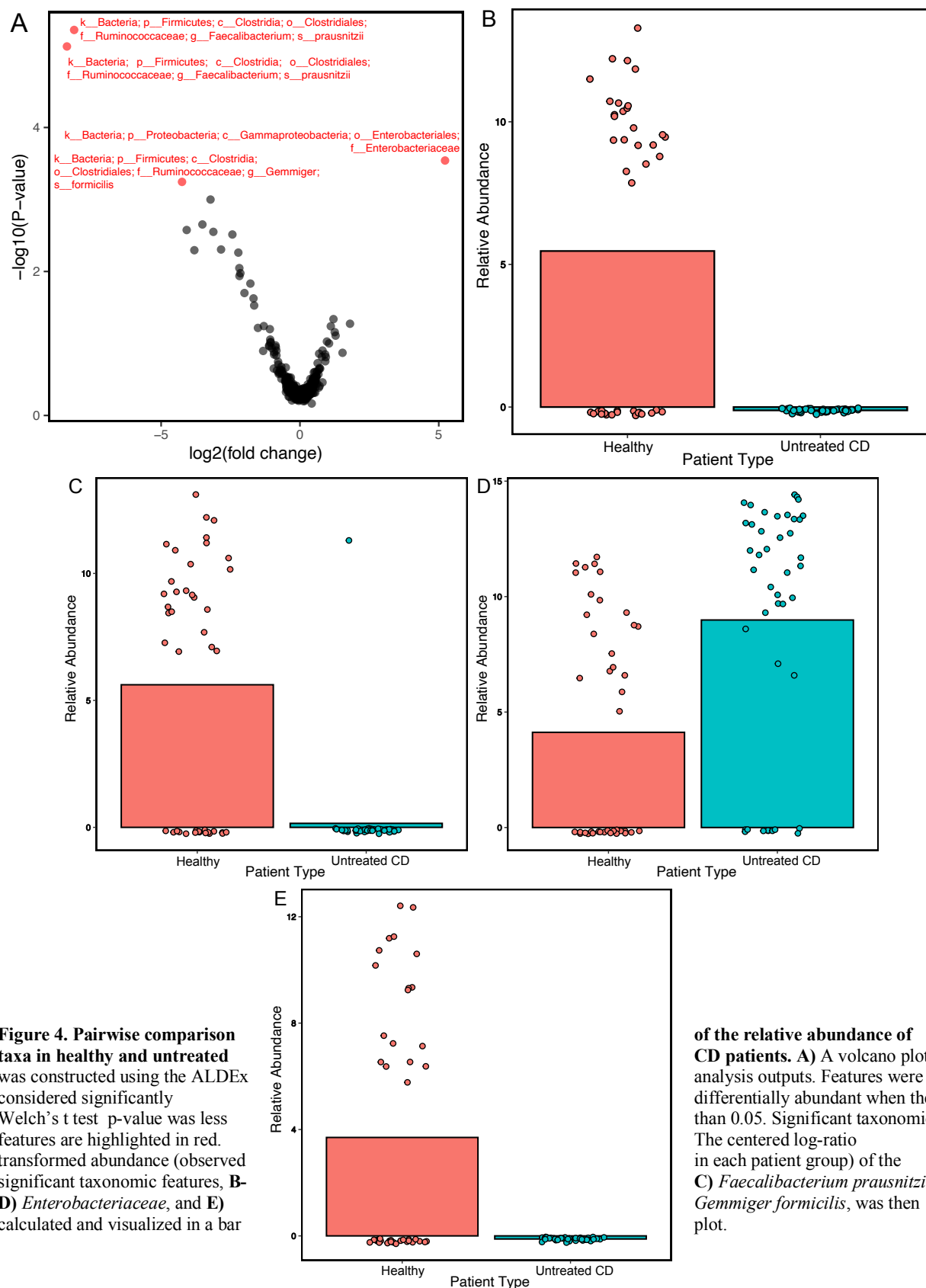


Figure 4. Pairwise comparison taxa in healthy and untreated was constructed using the ALDEx considered significantly Welch's t test p-value was less features are highlighted in red. transformed abundance (observed significant taxonomic features, **B-D**) *Enterobacteriaceae*, and **E**) calculated and visualized in a bar

of the relative abundance of CD patients. A) A volcano plot analysis outputs. Features were differentially abundant when the than 0.05. Significant taxonomic The centered log-ratio in each patient group) of the **C)** *Faecalibacterium prausnitzii*, *Gemmiger formicilis*, was then plot.

Similar to the FMT-treated patient group, levels of *Faecalibacterium prausnitzii* were significantly lower in the untreated group than in the healthy group (Fig. 4B, Fig. 4C; $p < 0.001$). However, while in the FMT-treated group the relative abundance of *Faecalibacterium prausnitzii* was greater than 0, with several outlier values between 5 and 10 (Fig. 4B), the relative abundance in the untreated group was merely 0 (Fig. 4B, Fig. 4C). The relative abundance of *Enterobacteriaceae*, on the other hand, was substantially greater in the untreated CD group than in the healthy group (Fig. 4D; $p < 0.001$). Conversely, *Gemmiger formicilis* was highly diminished in untreated CD patients (Fig. 4E; $p < 0.001$).

Gemmiger formicilis also exhibited differential abundance upon comparison of the FMT-treated and untreated patient groups (Fig. 5A). As shown in Fig. 5B, the abundance of this SCFA-fermenter significantly increased after FMT treatment ($p < 0.05$).

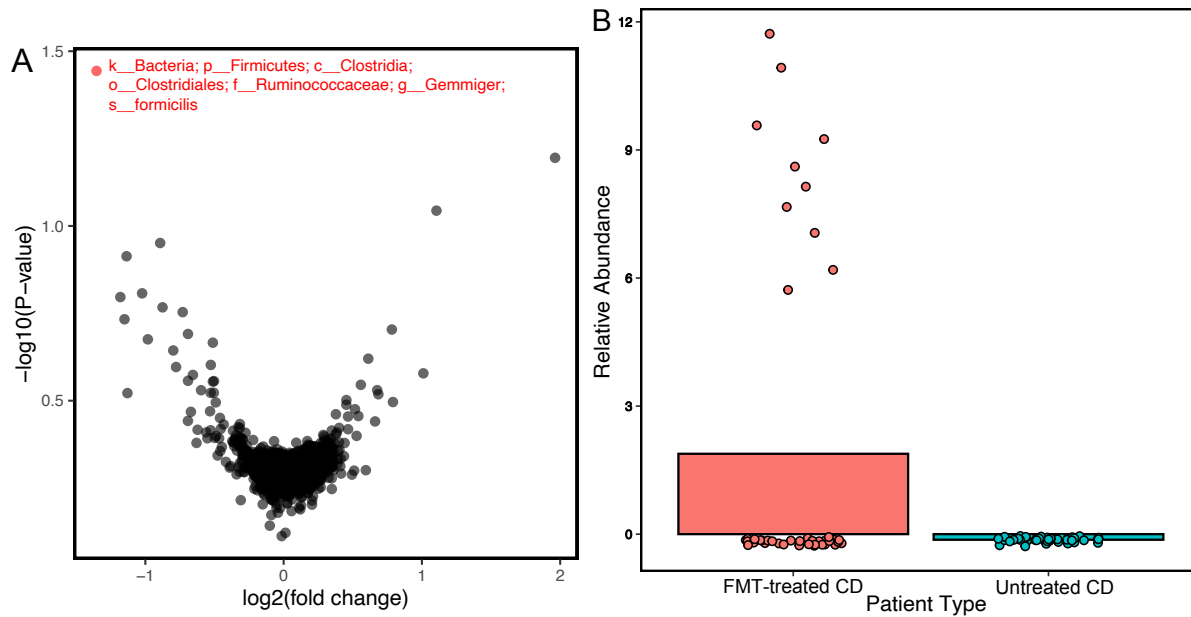


Figure 5. Pairwise comparison of the relative abundance of taxa in FMT-treated and untreated CD patients. A) A volcano plot was constructed using the ALDEx analysis outputs. Features were considered significantly differentially abundant when the Welch’s t test p-value was less than 0.05. Significant taxonomic features are highlighted in red. The centered log-ratio transformed abundance of the significant taxonomic feature, **B)** *Gemmiger formicilis*, was then calculated for the two patient groups and visualized in a bar plot.

Interestingly, while significant differential abundance was not directly observed for many features (indeed only one) between the untreated and FMT-treated groups, there was still more taxonomic similarity between the FMT-treated group and healthy group than between the untreated CD group and healthy group (with only one differentially abundant taxon in the former comparison, and four differentially abundant taxonomic features in the latter). This indicates that the microbiota in CD patients post-FMT treatment was significantly altered to the point of more closely resembling the microbiota of healthy patients.

Functional analysis

Metabolic pathway and KO abundances were predicted using the full PICRUSt2 pipeline and statistically analyzed through Shotgun Data Profiling in MicrobiomeAnalyst. Functional association analysis was conducted with the globaltest algorithm, revealing that the “fatty acid biosynthesis” functional category was significantly associated with patient type ($p < 0.001$), with 19 total KO hits (Fig. 6).

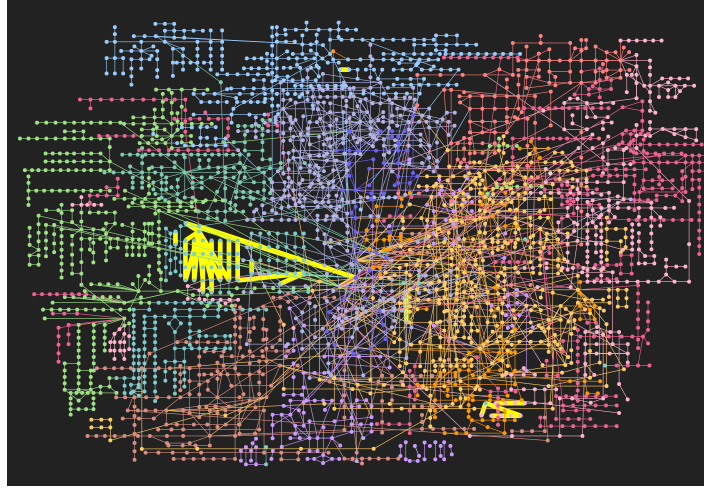


Figure 6. Visualization of the metabolic network observed in the experimental samples. The pathways (catalyzed by the aforementioned 19 KEGG ortholog hits) corresponding to the fatty acid biosynthesis functional category are highlighted in yellow.

The metagenomeSeq algorithm, moreover, was used to analyze orthologs for differential abundance across patient groups. Over 50 significant features were identified, including phosphate butyryltransferase (K00634), methylmalonyl-CoA carboxyltransferase (K17489), phosphate acetyltransferase (K00625), putative phosphotransacetylase (K15024), and acetate kinase (K00925).

The median filtered count of phosphate butyryltransferase in particular was significantly greater in FMT-treated and healthy patients than in their untreated CD counterparts (Fig. 7; $p < 0.01$).

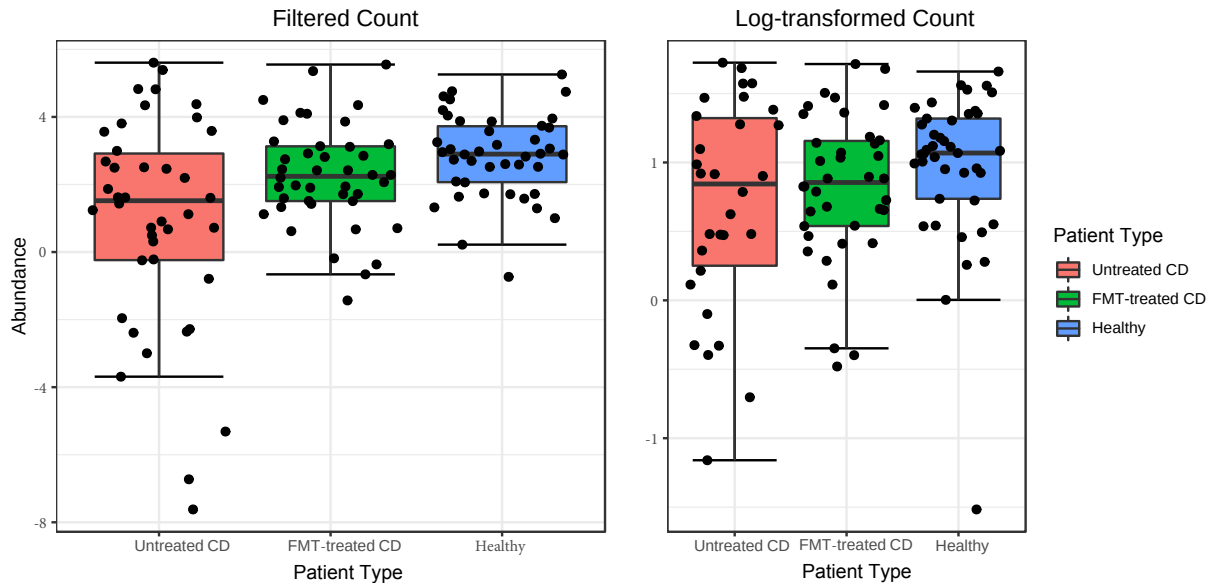


Figure 7. Phosphate butyryltransferase (K00634) abundance in the untreated CD, FMT-treated CD, and healthy patient groups. The first box and whisker plot illustrates the filtered abundance counts of phosphate butyryltransferase in the patient groups, while the second plot shows centered log-ratio transformed abundance counts of this KEGG ortholog.

Although the maximum filtered and log-transformed phosphate butyryltransferase abundance values were relatively similar across all patient groups, the FMT-treated and healthy groups exhibited the highest minimum values. Overall, the FMT-treated group displayed great similarity to the healthy group in terms of phosphate butyryltransferase abundance. FMT treatment induced a relative increase in phosphate butyryltransferase that paralleled healthy levels. Given that phosphate butyryltransferase is a major catalyst of the two-step butyrate production process, the discussed data evidences FMT's efficacy in promoting biosynthesis of SCFAs, particularly butyrate.

The effects of FMT extend past just butyrate biosynthesis: methylmalonyl-CoA carboxyltransferase, an important catalyst of propionate production via the succinate pathway, was also differentially abundant across the patient groups (Fig. 8; $p < 0.001$).

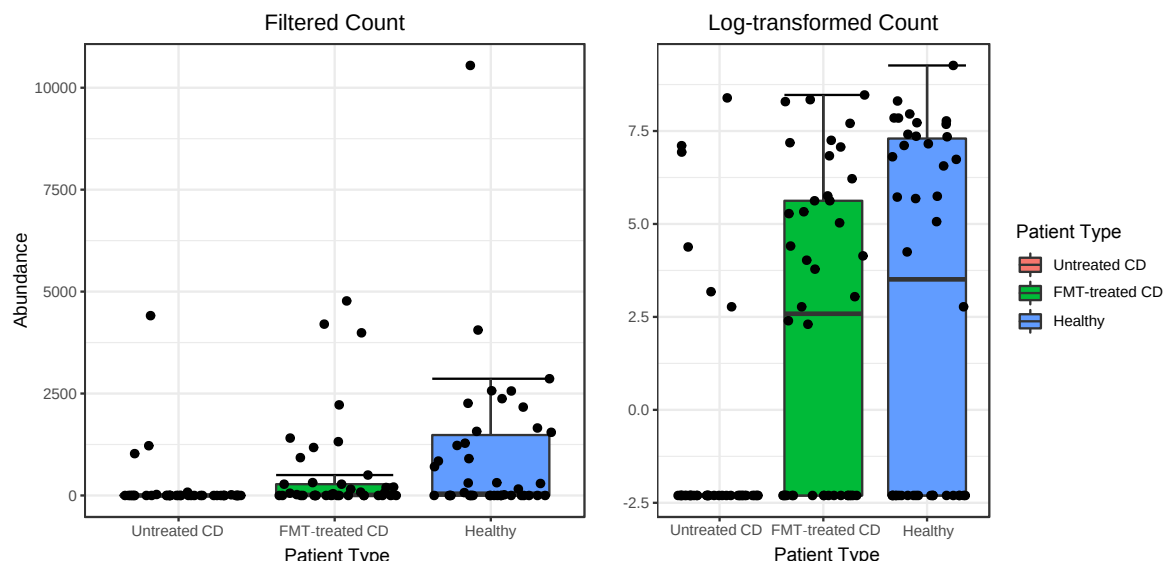


Figure 8. Methylmalonyl-CoA carboxyltransferase (K17489) abundance in the untreated CD, FMT-treated CD, and healthy patient groups. The first box and whisker plot illustrates the filtered abundance counts of methylmalonyl-CoA carboxyltransferase in the patient groups, while the second plot shows the centered log-ratio transformed abundance counts of this KEGG ortholog.

The filtered and log-transformed abundance counts of methylmalonyl-CoA carboxyltransferase in the untreated patient group were so small as to be practically negligible. Compared to those that were untreated, patients post-FMT possessed a markedly greater median abundance (as well as upper quartile) that was also close to the abundance observed in healthy patients. The spread of the log-transformed data for the FMT-treated and healthy groups overlapped almost completely, indicating high similarity. On the whole, these comparisons highlight the ability of FMT to return methylmalonyl-CoA carboxyltransferase abundance in patients to healthy levels, thus actively fostering propionate biosynthesis.

Besides butyrate and propionate biosynthesis, moreover, enzymes involved in acetate production also exhibited differential abundance. The median log-transformed abundance count of phosphate acetyltransferase, for instance, was greater in CD patients after FMT treatment than at baseline (Fig. 9A; $p < 0.05$). The same can be said for the minimum abundance counts, which were substantially higher in healthy and FMT-treated patients than in untreated ones. Unsurprisingly, the abundance of putative phosphotransacetylase was also highest in the healthy group, followed by the FMT-treated and untreated CD patient groups (Fig. 9B; $p < 0.01$). At the same time, however, the mean filtered and log-transformed—as well as minimum and lower quartile filtered—abundance counts of putative phosphotransacetylase were markedly higher in the FMT-treated patient group than the untreated one, signifying that FMT treatment raised the abundance of this enzyme to healthier levels. This was also the case for acetate kinase, which displayed a significantly higher median abundance in healthy and FMT-treated CD patients, compared to CD patients before treatment (Fig. 9C; $p < 0.05$).

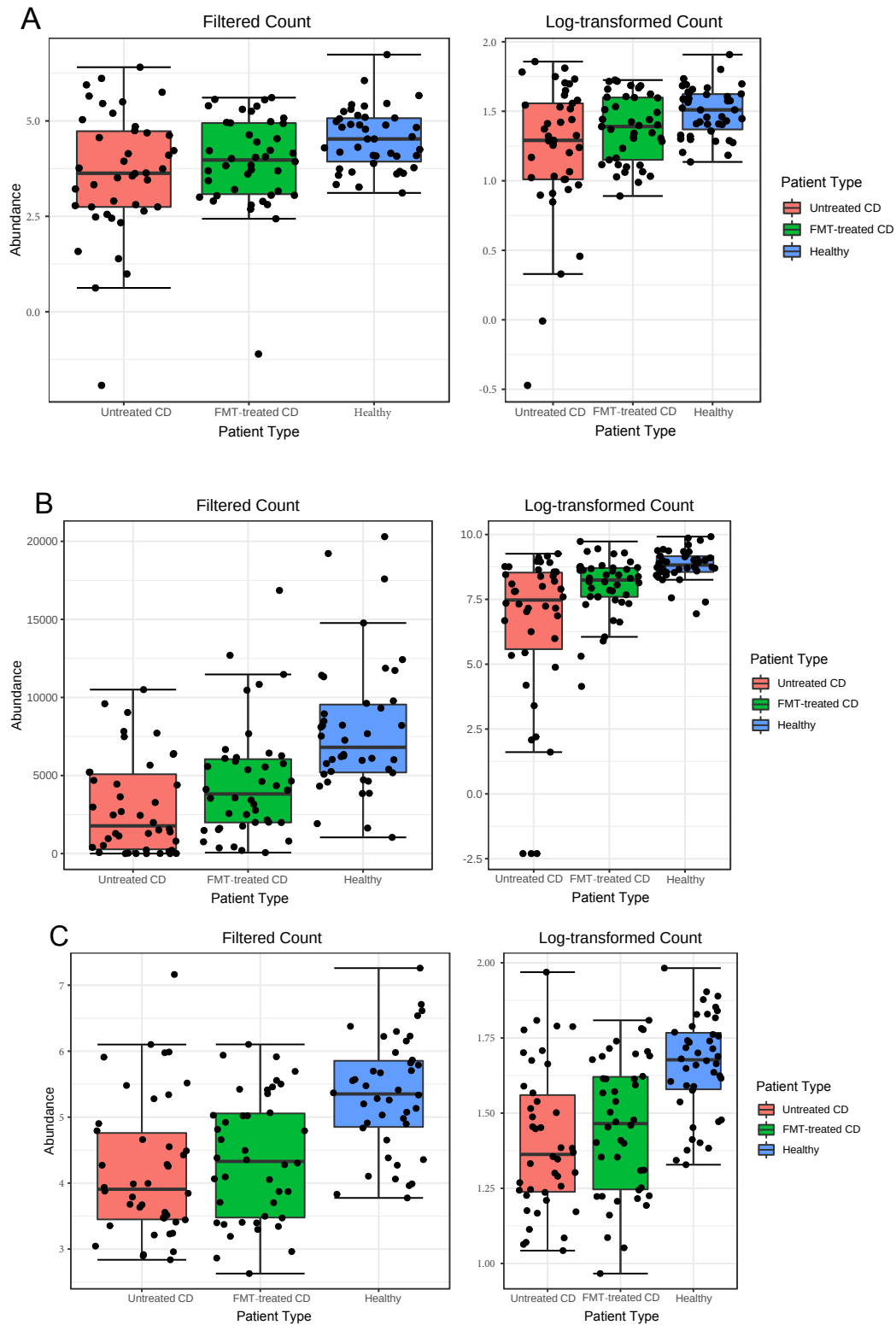


Figure 9. Relative abundance of enzymes involved in acetate production. Box and whisker plots were generated for the filtered and the centered log-ratio transformed abundance counts of **A**) phosphate acetyltransferase (K00625), **B**) putative phosphotransacetylase (K15024), and **C**) acetate kinase (K00925).

Again, it is important to note that phosphate acetyltransferase, putative phosphotransacetylase, and acetate kinase are key players in acetate generation. The former two enzymes catalyze the conversion of acetyl-CoA to acetyl-phosphate, while the latter catalyzes the conversion of acetyl-phosphate to acetate, so these molecules are crucial to acetate biosynthesis via this two-step pathway. Because FMT treatment has increased the abundance of these enzymes in CD patients to levels similar to those in healthy patients, it can be concluded that FMT effectively contributes to the stimulation of acetate synthesis.

Identification of target bacteria

To determine the optimal mid-biotics target bacteria, BLAST was used to match CD patient taxa sequences with those of common SCFA-producing bacteria based on percent identity. 15 taxonomic features significantly overlapped with SCFA-synthesizing taxa (Table 1). All percent identities for features matched with *R. inulinivorans*, *F. prausnitzii*, *E. lavalensis*, *B. obeum*, *B. uniformis*, and *A. halii* were greater than 90% ($E < 0.001$), indicating extremely high similarity. Meanwhile, the feature matched with *A. muciniphila* reached percent identities of nearly 80% ($E < 0.001$) and was also considered significantly similar, along with the features corresponding to *O. splanchnicus* (Percent identity $> 85\%$; $E < 0.001$).

Table 1. Percent identity calculated for each experimental taxonomic feature and SCFA-producing bacterium match. All matches were considered statistically significant, with Expect (E) values less than 0.001. NCBI accession numbers are included for the 16S rRNA sequences of the SCFA-producing bacteria.

SCFA-producing Bacterium	CD Patient Taxonomic Feature	Percent Identity (%)
<i>Roseburia inulinivorans</i> NR_042007.1	8301430ab3908fb88d8c7ca1a01f75ca	95.81
	1ba65f8aa97b650d4885b4c43f81b30a	94.62
<i>Odoribacter splanchnicus</i> NR_044636.1	e8d3b6dde1ebff8221e3615535a44d03	85.33
	70c3042f27b12e5bdd7ed72868e31d75	85.15
<i>Faecalibacterium prausnitzii</i> NR_028961.1	4508db76db9b760c7f337017e3958fb7	92.93
	5ed705d508cfb5eb472f0e5c78099188	92.88
<i>Enterocloster lavalensis</i> NR_044289.1	18b4cbb0c4ce87e9b963dd2060319c2e	96.56
	47b420e563a419876f59612971b751fc	96.54
<i>Blautia obeum</i> NR_118692.1	19bc8104e65eb659c4197672fc95c354	97.04
	e4862a567e6cfe59c80675ab7e1b4c90	96.02
<i>Bacteroides uniformis</i> NR_040866.1	9fd6e830887ac7ea2be5884f89d5cb17	96.8
	a10ef0cfff2d7341ea1be3d78a4607d5	97.29
<i>Anaerobutyricum halii</i> NR_118673.1	8ff919bd64fda87d6184ada9fb6ea175	92.59
	a1b294e731e90116be49e2206608d36b	91.22
<i>Akkermansia muciniphila</i> NR_042817.1	b37b720cd8d125fa1599885da27d735a	76.94

Because of their high genomic similarity to naturally SCFA-producing bacteria, the outlined taxonomic features were deemed optimal mid-biotic targets for the promotion of SCFA production. Accordingly, these features were aligned with taxa in the RefSeq Representative genomes database to acquire whole genome sequences. All

features were matched with one or more taxa at percent identities of at least 85%, and all matches were statistically significant ($E < 0.001$; Table 2).

Table 2. Percent identity calculated for each target feature and representative taxon. All matches were considered statistically significant, with Expect (E) values less than 0.001. The RefSeq database accession number is included for each representative taxon.

Mid-biotic Target Feature	Representative Taxon	Percent Identity (%) [*]	RefSeq Acc. No.
8301430ab3908fb88d8c7ca1a01f75ca	<i>Coprococcus phoceensis</i>	95.01	NZ_FNWC01000007.1
1ba65f8aa97b650d4885b4c43f81b30a	<i>Lachnoclostridium edouardi</i>	92.74	NZ_OESQ01000001.1
e8d3b6dde1ebff8221e3615535a44d03	<i>Culturomica massiliensis</i>	85.15	NZ_LT594621.1
70c3042f27b12e5bdd7ed72868e31d75	<i>Culturomica massiliensis</i>	85.75	NZ_LT594621.1
4508db76db9b760c7f337017e3958fb7	<i>Pygmaibacter massiliensis</i>	94.24	NZ_OAML01000003.1
5ed705d508cfb5eb472f0e5c78099188	<i>Massilioclostridium coli</i>	90.77	NZ_FMIZ01000007.1
18b4cbb0c4ce87e9b963dd2060319c2e	<i>Enterocloster bolteae</i>	100	NZ_CP022464.2
47b420e563a419876f59612971b751fc	<i>Massilistercora timonensis</i>	94.68	NZ_LT990039.1
19bc8104e65eb659c4197672fc95c354	<i>Enterocloster aldenensis</i>	93.30	NZ_QSGP01000008.1
e4862a567e6cfe59c80675ab7e1b4c90	<i>Blautia producta</i>	95.24	NZ_CP039126.1
9fd6e830887ac7ea2be5884f89d5cb17	<i>Phocaeicola dorei</i>	97.07	NZ_LR699004.1
a10ef0cfff2d7341ea1be3d78a4607d5	<i>Bacteroides oleiciplenus</i>	98.10	NZ_CABKRN010000006.1
8ff919bd64fda87d6184ada9fb6ea175	<i>Faecalimonas umbilicata</i>	91.05	NZ_BHEO01000002.1
a1b294e731e90116be49e2206608d36b	<i>Lachnoclostridium edouardi</i>	92.02	NZ_OESQ01000001.1
b37b720cd8d125fa1599885da27d735a	<i>Desnuesiella massiliensis</i>	92.76	NZ_LN879455.1

The representative species ($n = 13$) were not natural SCFA-producers; however, they were commonly from the same orders or classes as the SCFA-producers in Table 1, which may explain their high percent identity to members of this group.

This high genetic similarity between the representative species and the SCFA-producing species was also observed through phylogenetic analysis in MEGA X. For instance, as shown in Figure 10, *E. lavalensis*, an SCFA-producing species, was significantly phylogenetically close to one of its representative BLAST matches (as mentioned above), *E. bolteae* (p-distance = 0.030). Although not a match through BLAST, *E. aldenensis* was also significantly close to *E. lavalensis*, most likely due to their shared genus (p-distance = 0.033; Figure 10).

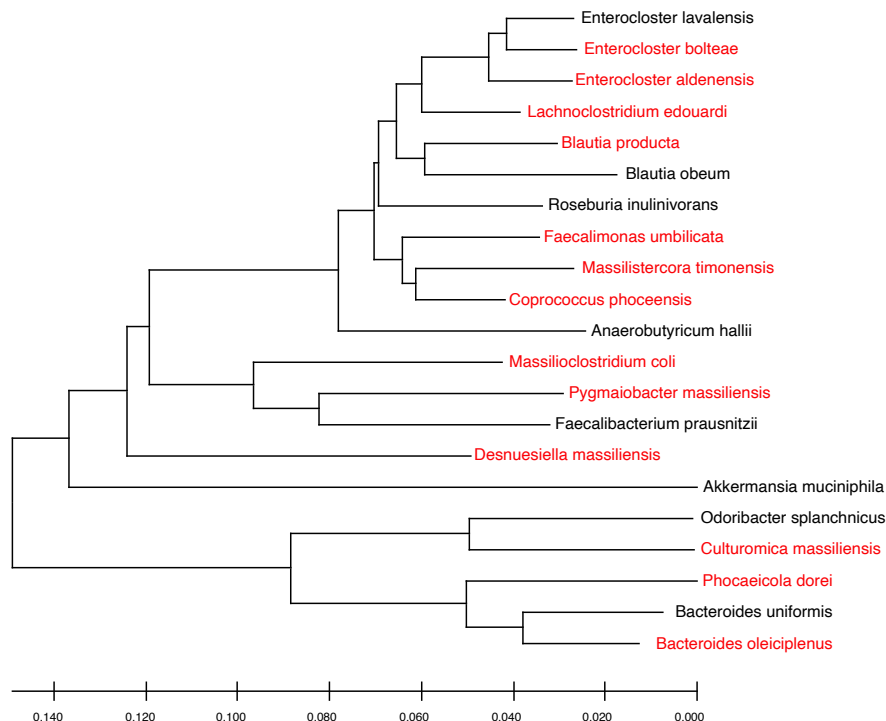


Figure 10. Evolutionary relationships of the representative species and the SCFA-producing species. Evolutionary distances between the bacteria ($n = 21$) were calculated using the p-distance method in MEGA X. A neighbor-joining phylogenetic tree was thereafter constructed, with branch lengths drawn to scale in the p-distance units (number of base differences per site; indicated by the scale bar). The representative species (mid-biotic target host bacteria) are highlighted in red.

Other representative species were highly close to SCFA-synthesizers as well, though to a slightly lesser extent than that observed with *E. lavalensis*: *B. oleiciplenus* and *B. uniformis* (p-distance = 0.056), *C. phoceensis* and *R. inulinivorans* (p-distance = 0.061), *L. edouardi* and *R. inulinivorans* (p-distance = 0.062), *B. producta* and *B. obeum* (p-distance = 0.071), and *E. aldenensis* and *B. obeum* (p-distance = 0.080), *P. dorei* and *B. uniformis* (p-distance = 0.089), and *F. umbilicata* and *A. hallii* (p-distance = 0.091) constituted the most genetically similar pairs. The p-distance values of all remaining pairs were less than 0.140 (with the exception of *D. massiliensis* and *A. muciniphila* [p-distance = 0.216]), indicating relatively high phylogenetic closeness.

Overall, the representative species ($n = 13$; displayed in Table 2) were established as the target hosts for plasmid insertion, due to their current lack of SCFA production capabilities and, most importantly, highly significant genetic similarity to SCFA-producing bacteria, which is crucial for heterologous gene transfer. Indeed, bacterial hosts tend to homogenize the genomic signature of genetic elements replicating within them, largely due to host-specific mutational biases. This inevitable genome amelioration, combined with the exchange of DNA via recombination and transposition, results in similar nucleotide compositions between plasmids and their evolutionary hosts over time. These genetic changes in the plasmid and adaptation with the host reduce the genetic element's fitness costs [40]. Accordingly, the concept of genome amelioration can be applied to the aforementioned results. It is more likely that the chosen plasmids (which, in future testing, will be engineered with genes required for SCFA synthesis; mentioned in the next section) will replicate within their host and exhibit maintenance within the bacterial community if their host has a similar genomic composition. Of course, SCFA-producing bacteria contain the genes encoding SCFA synthesis with which the plasmid will be engineered, and are thus inherently similar. The non-SCFA-producing bacteria, then, (which must be genetically engineered for SCFA synthesis functionality and are thus the targets of mid-biotics) that are most similar to SCFA-synthesizers will be the optimal targets for plasmid insertion. In other words, because these selected target bacteria are genetically similar to SCFA-synthesizers, they will also likely share genomic similarities with the chosen vectors. On the whole, this will minimize fitness costs and ensure compatibility.

Identification of optimal plasmids

The most favorable plasmids for communal introduction were determined through Mahalanobis distance analysis in R. Those with the smallest genomic distance from target bacterial DNA were considered the most advantageous. For context, Mahalanobis distances ranged from approximately 2.51 (this lowest value shown in Table 3) to over 200,000, and P values ranged from 0 to 0.96 (this highest value shown in Table 3 as well). Out of the 92 plasmids tested, 8 showed high genomic similarity to target bacteria, namely R27, R478, pAPEC-O1-R, pMAK1, pK29, pHCM1, pRA1, and pEC-IMP (Table 3). The Mahalanobis distances calculated for all pairs outlined below were less than 10 (and the P value for all matches was greater than or equal to 0.50), indicating high similarity. From these 8 plasmids, R27 was most similar to a target bacterium (*Bacteroides oleiciplenus*), exhibiting the smallest Mahalanobis distance of 2.51 and P value closest to 1 ($P = 0.96$). The next smallest distance ($D^2 = 3.77$) and highest P value ($P = 0.85$) was exhibited by R478 in its match with *Bacteroides oleiciplenus*.

Table 3. Mahalanobis distance calculations for each genetically similar plasmid-target bacterium pair.

Plasmid	Genetically Similar Target Bacteria	Mahalanobis Distance	P value
R27	<i>Bacteroides oleiciplenus</i>	2.51	0.96
	<i>Phocaeicola dorei</i>	3.78	0.79
R478	<i>Bacteroides oleiciplenus</i>	3.77	0.85
	<i>Phocaeicola dorei</i>	6.28	0.50
pAPEC-O1-R	<i>Bacteroides oleiciplenus</i>	4.37	0.76
pMAK1	<i>Bacteroides oleiciplenus</i>	4.54	0.75
pK29	<i>Bacteroides oleiciplenus</i>	4.75	0.75
pHCM1	<i>Bacteroides oleiciplenus</i>	5.60	0.70
pRA1	<i>Enterocloster aldenensis</i>	7.52	0.51
pEC-IMP	<i>Bacteroides oleiciplenus</i>	7.94	0.50

R27 and R478 were thus regarded as the most favorable plasmids for promotion of SCFA production, especially in *Bacteroides oleiciplenus* and *Phocaeicola dorei*, their closest matching target bacteria.

Interestingly, in a previous study, the fatty acid biosynthesis pathway in *E. coli* (which does not natively produce butyrate) was engineered for effective butyric acid production via thioesterases (TesAT, TesBF, and TesBT) [41]. The authors successfully cloned the genes for these thioesterases in pUC57 and pZS21MCS plasmids, commonly used vectors in *E. coli* and usually isolated from *E. coli* strain DH5 α . It must be noted that all plasmids in Table 3 have evolutionary hosts within the *Enterobacteriales* family; the known host of the pAPEC-O1-R plasmid, moreover, is *E. coli* itself [39]. Thus, the previous efficacy (in terms of promoting SCFA formation) of plasmid vectors specific to *E. coli* may point to the future effectiveness of the chosen plasmids in Table 3.

CONCLUSION

Overall, the findings of this study have several implications. First, FMT was shown to promote the growth of SCFA-synthesizing bacteria in CD patients and prompt increases in abundance of enzymes such as phosphate butyryltransferase (K00634), methylmalonyl-CoA carboxyltransferase (K17489), phosphate acetyltransferase (K00625), putative phosphotransacetylase (K15024), and acetate kinase (K00925), which are essential for the microbial production of butyrate, propionate, and acetate. This is in keeping with the results of past studies, where FMT therapy has increased acetate, butyrate, and propionate levels in *C. difficile*-infected patients to a point of SCFA recovery [42-44].

As for the mid-biotics approach, *Coprococcus phoceensis*, *Lachnoclostridium edouardi*, *Culturomica massiliensis*, *Pygmaibacter massiliensis*, *Massilioclostridium coli*, *Enterocloster bolteae*, *Massilistercora timonensis*, *Enterocloster aldenensis*, *Blautia producta*, *Phocaeicola dorei*, *Bacteroides oleiciplenus*, *Faecalimonas umblicate*, *Lachnoclostridium edouardi*, and *Desnuesiella massiliensis* were deemed the target bacteria as a result of their high genomic similarity to SCFA-producing taxa. Meanwhile, the plasmids R27, R478, pAPEC-O1-R, pMAK1, pK29, pHCM1, pRA1, and pEC-IMP were closest in genomic signature to the target bacteria, making these plasmids the most suitable vectors for SCFA synthesis genes.

Future investigations include wet-lab testing to ascertain the effectiveness of the chosen plasmids in terms of compatibility, maintenance within the gut microbiota, and plasmid-host evolution. This can be achieved through *in vitro* experimentation with the target hosts and other gut bacteria in minimal media. Plasmid-host interactions can also be studied with *ex vivo* models using stool samples, intestinal tissue, or intestinal organoids/enteroids. Other potential *ex vivo* tools include the Ussing chamber, which, consisting of apical and basolateral sides covered with electrodes, has been used to examine microbial metabolite production, as well as gut-on-a-chip microfluidic devices, which have been used to evaluate anti-inflammatory probiotics [45]. In the context of mid-biotics research, the technologies can potentially be utilized to study the effect of plasmid introduction on community-wide SCFA biosynthesis.

Ultimately, the results of this study support the effectiveness of FMT as a non-permanent therapy for CD. The data also provide new insight into the effects of FMT on metabolic pathway and KEGG ortholog abundances and overall host functional profile, specifically in terms of SCFA synthesis pathways. Indeed, FMT-induced remission can potentially be monitored by tracking phosphate butyryltransferase, methylmalonyl-CoA carboxyltransferase, phosphate acetyltransferase, putative phosphotransacetylase, and acetate kinase abundance over time, allowing for relapse prediction and, as a result, improved patient outcome. With regards to procedure, this project introduced novel methods for the analysis of FMT's efficacy in terms of SCFA synthesis. While most studies have determined changes in SCFA production through wet-lab measurements of concentration via gas chromatography [43, 46-47], this study shifted the focus to predictive analysis of taxonomy (to quantify SCFA-producing taxa) and functional orthologs (to quantify catalysts of SCFA production). By using predictive programs, rather than wet-lab techniques, in the manner outlined above, scientists can study SCFA production in a more convenient, cost- and time-efficient manner.

Finally, besides the novel approach to studying SCFA changes, this project provides a novel model for CD treatment via mid-biotics. This mid-biotics system can potentially be developed into an ingestible capsule, much like that used for probiotics, to deliver the aforementioned plasmids (to be engineered with genes—such as *ptb*, *buk*, *pta*, *ackA*, *yciA*, *entH*, *menI*, etc.—conferring SCFA production capabilities to target bacterial hosts) directly to the patient's intestinal microflora. Such a therapy has the potential to induce permanent remission and allow for CD patients to reap myriad benefits without having to undergo surgery or other invasive procedures. In other words, mid-biotics treatment could induce increases in SCFA synthesis in CD patients, like FMT has here been shown to accomplish, without the adverse effects and impermanence associated with FMT. On the whole, this line of study has the potential to aid millions in combating CD.

ACKNOWLEDGMENTS

I thank Dr. Serena McCalla and the iResearch Institute for giving me the opportunity to start this project. I thank Emily Gan for her mentorship and for her instruction of essential data analysis techniques.

REFERENCES

- [1] F. Xu, J. M. Dahlhamer, E. P. Zammitti, A. G. Wheaton, and J. B. Croft, "Health-Risk Behaviors and Chronic Conditions Among Adults with Inflammatory Bowel Disease — United States, 2015 and 2016," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 67, no. 6, pp. 190-195, Feb. 2018, doi: 10.15585/mmwr.mm6706a4.
- [2] "The facts about inflammatory bowel diseases," Crohn's & Colitis Foundation, New York City, NY, USA, Nov. 2014. [Online]. Available: <https://www.crohnscolitisfoundation.org/sites/default/files/legacy/assets/pdfs/ibdfactbook.pdf>
- [3] A. D. Kostic, R. J. Xavier, and D. Gevers, "The microbiome in inflammatory bowel disease: Current status and the future ahead," *Gastroenterology*, vol. 146, no. 6, pp. 1489-1499, May 2014, doi: 10.1053/j.gastro.2014.02.009.
- [4] S. M. Vindigni, T. L. Zisman, D. L. Suskind, and C. J. Damman, "The intestinal microbiome, barrier function, and immune system in inflammatory bowel disease: a tripartite pathophysiological circuit with implications for new therapeutic directions," *Therap. Adv. Gastroenterol.*, vol. 9, no. 4, pp. 606-625, Jul. 2016, doi: 10.1177/1756283x16644242.
- [5] X. C. Morgan *et al.*, "Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment," *Genome Biol.*, vol. 13, no. 9, Apr. 2012, doi: 10.1186/gb-2012-13-9-r79.
- [6] S. F. Øyri, G. Müzes, and F. Sipos, "Dysbiotic gut microbiome: A key element of Crohn's disease," *Comp. Immunol. Microbiol. Infect. Dis.*, vol. 43, pp. 36-49, Dec. 2015, doi: 10.1016/j.cimid.2015.10.005.
- [7] H. J. Freeman, "Natural history and long-term clinical course of Crohn's disease," *World J. Gastroenterol.*, vol. 20, no. 1, pp. 31-36, Jan. 2014. doi: 10.3748/wjg.v20.i1.31. [Online]. Available: <https://www.wjgnet.com/1007-9327/full/v20/i1/31.htm>
- [8] B. Yilmaz *et al.*, "Microbial network disturbances in relapsing refractory Crohn's disease," *Nat. Med.*, vol. 25, no. 2, pp. 323-336, Jan. 2019, doi: 10.1038/s41591-018-0308-z.
- [9] M. T. Alam, G. C. A. Amos, A. R. J. Murphy, S. Murch, E. M. H. Wellington, and R. P. Arasaradnam, "Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels," *Gut Pathog.*, vol. 12, Jan. 2020, doi: 10.1186/s13099-019-0341-6.
- [10] H. C. Mirsepasi-Lauridsen *et al.*, "Disease-specific enteric microbiome dysbiosis in inflammatory bowel disease," *Front. Med.*, vol. 5, Nov. 2018, doi: 10.3389/fmed.2018.00304.
- [11] D. Parada Venegas *et al.*, "Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases," *Front. Immunol.*, vol. 10, p. 277, Mar. 2019, doi: 10.3389/fimmu.2019.00277.
- [12] N. Y. Kedishvili, G. W. Goodwin, K. M. Popov, and R. A. Harris, "Mammalian methylmalonate-semialdehyde dehydrogenase," *Methods Enzymol.*, vol. 324, pp. 207-218, 2000, doi: 10.1016/s0076-6879(00)24233-x.
- [13] P. Louis and H. J. Flint, "Formation of propionate and butyrate by the human colonic microbiota," *Environ. Microbiol.*, vol. 19, no. 1, pp. 29-41, Nov. 2016, doi: 10.1111/1462-2920.13589.
- [14] R. A. Gonzalez-Garcia, T. McCubbin, L. Navone, C. Stowers, L. K. Nielsen, and E. Marcellin, "Microbial propionic acid production," *Fermentation*, vol. 3, no. 2, May 2017, doi: 10.3390/fermentation3020021.
- [15] M. Vital, A. C. Howe, and J. M. Tiedje, "Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data," *MBio*, vol. 5, no. 2, Apr. 2014, doi: 10.1128/mbio.00889-14.
- [16] Y. Sun and M. X. O'Riordan, "Regulation of bacterial pathogenesis by intestinal short-chain fatty acids," *Adv. Appl. Microbiol.*, vol. 85, pp. 93-118, 2013, doi: 10.1016/b978-0-12-407672-3.00003-4.

- [17] C. Zhao, H. Dong, Y. Zhang, and Y. Li, "Discovery of potential genes contributing to the biosynthesis of short-chain fatty acids and lactate in gut microbiota from systematic investigation in *E. coli*," *NPJ Biofilms Microbiomes*, vol. 5, no. 1, Jul. 2019, doi: 10.1038/s41522-019-0092-7.
- [18] P. M. Smith *et al.*, "The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis," *Science*, vol. 341, no. 6145, pp. 569-573, Jul. 2013, doi: 10.1126/science.1241165.
- [19] N. Powell, M. M. Walker, and N. J. Talley, "The mucosal immune system: master regulator of bidirectional gut-brain communications," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 14, no. 3, pp. 143-159, Mar. 2017, doi: 10.1038/nrgastro.2016.191.
- [20] M. Li, B. C. A. M. van Esch, P. A. J. Henricks, G. Folkerts, and J. Garssen, "The anti-inflammatory effects of short chain fatty acids on lipopolysaccharide- or tumor necrosis factor α -stimulated endothelial cells via activation of GPR41/43 and inhibition of HDACs," *Front. Pharmacol.*, vol. 9, May 2018, doi: 10.3389/fphar.2018.00533.
- [21] P. Ruotsalainen, R. Penttinen, S. Mattila, and M. Jalasvuori, "Midbiotics: conjugative plasmids for genetic engineering of natural gut flora," *Gut Microbes*, vol. 10, no. 6, pp. 643-653, Apr. 2019, doi: 10.1080/19490976.2019.1591136.
- [22] L. A. Ogilvie, S. Firouzmand, and B. V. Jones, "Evolutionary, ecological and biotechnological perspectives on plasmids resident in the human gut mobile metagenome," *Bioengineered*, vol. 3, no. 1, pp. 13-31, Jan. 2012, doi: 10.4161/bbug.3.1.17883.
- [23] T. Zhang *et al.*, "Alterations of *Akkermansia muciniphila* in the inflammatory bowel disease patients with washed microbiota transplantation," *Appl. Microbiol. Biotechnol.*, vol. 104, no. 23, pp. 10203-10215, Dec. 2020, doi: 10.1007/s00253-020-10948-7.
- [24] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes, "DADA2: High-resolution sample inference from Illumina amplicon data," *Nat. Methods*, vol. 13, no. 7, pp. 581-583, Jul. 2016, doi: 10.1038/nmeth.3869.
- [25] E. Bolyen *et al.*, "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," *Nat. Biotechnol.*, vol. 37, no. 8, pp. 852-857, Aug. 2019, doi: 10.1038/s41587-019-0209-9.
- [26] D. McDonald *et al.*, "An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea," *ISME J.*, vol. 6, no. 3, pp. 610-618, Mar. 2012, doi: 10.1038/ismej.2011.139.
- [27] S. Mandal, W. Van Treuren, R. A. White, M. Eggesbø, R. Knight, and S. D. Peddada, "Analysis of composition of microbiomes: A novel method for studying microbial composition," *Microb. Ecol. Health Dis.*, vol. 26, May 2015, doi: 10.3402/mehd.v26.27663.
- [28] G. M. Douglas *et al.*, "PICRUSt2 for prediction of metagenome functions," *Nat. Biotechnol.*, vol. 38, no. 6, pp. 685-688, Jun. 2020, doi: 10.1038/s41587-020-0548-6.
- [29] J. Chong, P. Liu, G. Zhou, and J. Xia, "Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data," *Nat. Protoc.*, vol. 15, no. 3, pp. 799-821, Mar. 2020, doi: 10.1038/s41596-019-0264-1.
- [30] A. Dhariwal, J. Chong, S. Habib, I. L. King, L. B. Agellon, and J. Xia, "MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data," *Nucleic Acids Res.*, vol. 45, no. W1, pp. W180-W188, Jul. 2017, doi: 10.1093/nar/gkx295.
- [31] J. J. Goeman, S. A. van de Geer, F. de Kort, and H. C. van Houwelingen, "A global test for groups of genes: testing association with a clinical outcome," *Bioinformatics*, vol. 20, no. 1, pp. 93-99, Jan. 2004, doi: 10.1093/bioinformatics/btg382.

- [32] J. N. Paulson, O. C. Stine, H. C. Bravo, and M. Pop, "Differential abundance analysis for microbial marker-gene surveys," *Nat. Methods*, vol. 10, no. 12, pp. 1200-1202, Dec. 2013, doi: 10.1038/nmeth.2658.
- [33] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403-410, Oct. 1990, doi: 10.1016/S0022-2836(05)80360-2.
- [34] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, "MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms," *Mol. Biol. Evol.*, vol. 35, no. 6, pp. 1547-1549, Jun. 2018, doi: 10.1093/molbev/msy096.
- [35] G. Stecher, K. Tamura, and S. Kumar, "Molecular Evolutionary Genetics Analysis (MEGA) for macOS," *Mol. Biol. Evol.*, vol. 37, no. 4, pp. 1237-1239, Apr. 2020, doi: 10.1093/molbev/msz312.
- [36] R. C. Edgar, "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res.*, vol. 32, no. 5, pp. 1792-1797, Mar. 2004, doi: 10.1093/nar/gkh340.
- [37] M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*, 1st ed. Oxford, UK: Oxford Univ. Press, 2000, pp. 33-34.
- [38] M. Tokuda *et al.*, "Determination of plasmid pSN1216-29 host range and the similarity in oligonucleotide composition between plasmid and host chromosomes," *Front. Microbiol.*, vol. 11, Jun. 2020, doi: 10.3389/fmicb.2020.01187.
- [39] H. Suzuki, H. Yano, C. J. Brown, and E. M. Top, "Predicting plasmid promiscuity based on genomic signature," *J. Bacteriol.*, vol. 192, no. 22, pp. 6045-6055, Nov. 2010, doi: 10.1128/JB.00277-10.
- [40] M. Sota *et al.*, "Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein," *ISME J.*, vol. 4, no. 12, pp. 1568-1580, Dec. 2010, doi: 10.1038/ismej.2010.72.
- [41] K. Jawed, A. J. Mattam, Z. Fatma, S. Wajid, M. Z. Abdin, and S. S. Yazdani, "Engineered production of short chain fatty acid in *Escherichia coli* using fatty acid synthesis pathway," *PLoS One*, vol. 11, no. 7, Jul. 2016, doi: 10.1371/journal.pone.0160035.
- [42] N. Ohmiya *et al.*, "Changes in fecal microbiota, short chain fatty acids, and bile acids after fecal microbiota transplantation for recurrent *Clostridioides difficile* infection, ulcerative colitis, and Crohn's disease," *Gastroenterology*, vol. 158, no. 6, May 2020, doi: 10.1016/S0016-5085(20)31885-0.
- [43] A. M. Seekatz *et al.*, "Restoration of short chain fatty acid and bile acid metabolism following fecal microbiota transplantation in patients with recurrent *Clostridium difficile* infection," *Anaerobe*, vol. 53, pp. 64-73, Oct. 2018, doi: 10.1016/j.anaerobe.2018.04.001.
- [44] J. Allegretti *et al.*, "Short chain fatty acid profiles are altered by fecal microbiota transplantation for the treatment of inflammatory bowel disease and recurrent *Clostridioides difficile* infection," *Am. J. Gastroenterol.*, vol. 114, pp. S484-S485, Oct. 2019, doi: 10.14309/01.ajg.0000592884.95396.04.
- [45] S. C. Pearce, H. G. Coia, J. P. Karl, I. G. Pantoja-Feliciano, N. C. Zachos, and K. Racicot, "Intestinal *in vitro* and *ex vivo* models to study host-microbiome interactions and acute stressors," *Front. Physiol.*, vol. 9, Nov. 2018, doi: 10.3389/fphys.2018.01584.
- [46] R. Chen *et al.*, "Transplantation of fecal microbiota rich in short chain fatty acids and butyric acid treat cerebral ischemic stroke by regulating gut microbiota," *Pharmacol. Res.*, vol. 148, Oct. 2019, doi: 10.1016/j.phrs.2019.104403.
- [47] M. El-Salhy, J. Valeur, T. Hausken, and J. G. Hatlebakk, "Changes in fecal short-chain fatty acids following fecal microbiota transplantation in patients with irritable bowel syndrome," *Neurogastroenterol. Motil.*, vol. 33, no. 2, Feb. 2021, doi: 10.1111/nmo.13983.