Taxonomic and Functional Characterization of Human Gut Microbes Involved in Dietary Plant Lignan Metabolism

Isaac M. Elkon

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Washington

2015

Committee:
Johanna W. Lampe
Meredith A.J. Hullar

Program Authorized to Offer Degree:
School of Public Health
Abstract

Taxonomic and Functional Characterization of Human Gut Microbes Involved in Dietary Plant Lignan Metabolism

Isaac M. Elkon

Chair of the Supervisory Committee:
Research Professor Johanna W. Lampe
Department of Epidemiology

Background: Dietary plant lignans, such as secoisolariciresinol diglucoside (SDG), are metabolized to the enterolignans, enterodiol (END) and enterolactone (ENL), by gut microbes. Evidence suggests that enterolignans may reduce risk of cardiovascular disease and several forms of cancer. Our aim was to characterize the microbial community involved in enterolignan production by using an in vitro batch culture system to enrich for lignan-metabolizing organisms.

Methods: Stool samples from eight participants were incubated separately for ~1 week with a mineral salts media containing formate, acetate, glucose, and 6.55 µM SDG. Daily secoisolariciresinol (SECO), END, and ENL concentrations were measured using gas chromatography–mass spectrometry (GCMS). Microbial community in initial stool (Day 1) and in vitro-incubated fecal suspensions (final-day) was assessed via Illumina paired-end 16S rRNA gene amplicon and whole-metagenome shotgun sequencing. 16S rRNA gene sequences were taxonomically annotated using an in-house QIIME pipeline. Metagenomic shotgun sequences
were taxonomically annotated using MetaPhlAn and functionally annotated using DIAMOND and HUMAnN. Annotation-based alpha diversity, organism abundance, and functional gene abundance were used to assess differences between Day 1 and final-day microbial community composition.

**Results:** Stool of all 8 participants converted the majority of SDG to ENL by Day 6 (average ± standard deviation: 88.4% ± 6.9). In most incubations, Bacteroidetes:Firmicutes ratio increased; methanogenic, deglycosylating, and demethylating organisms were enriched; and KEGG Orthology gene families (KOs) corresponding to beta-glucosidases were present at high levels. In some incubations, *Eubacterium limnosum, Akkermansia muciniphila,* and KOs corresponding to demethylases, dehydroxylases, dehydrogenases, methane metabolism, the Wood-Ljungdahl pathway, and acetogenic bacterial O-demethylation were enriched.

**Conclusions:** Organisms and functional genes identified or hypothesized to be involved in lignan metabolism were not unanimously enriched, but sample-specific. These findings may help further elucidate organisms, enzymes, and biochemical pathways involved in lignan metabolism. Future studies with larger sample sizes may help to confirm these results.
TABLE OF CONTENTS

Chapter 1. Background .......................................................................................................................... 1
  1.1 Introduction......................................................................................................................................... 1
  1.2 Plant Lignans in the Diet .................................................................................................................. 1
  1.3 Health Effects of Lignans ................................................................................................................ 2
  1.4 Inter-individual Differences in Enterolignan Production ................................................................. 3
  1.5 Organisms Involved in Lignan Metabolism .................................................................................... 4
  1.6 H₂ Consumers in the Human Gut ..................................................................................................... 5
  1.7 The Potential Role of Acetogens in Lignan Metabolism ................................................................. 6
  1.8 The Potential Role of Methanogens in Lignan Metabolism ........................................................... 7
  1.9 Microbial Enzymes in Lignan Metabolism ..................................................................................... 7
    1.9.1 O-linked Deglycosylation ........................................................................................................... 8
    1.9.2 O-linked Demethylation ........................................................................................................... 8
    1.9.3 Dehydroxylation ...................................................................................................................... 9
    1.9.4 Dehydrogenation ..................................................................................................................... 9
  1.10 Aim, Hypothesis, and Implications ............................................................................................... 10

Chapter 2. Methods ............................................................................................................................... 11
  2.1 Subjects .............................................................................................................................................. 11
  2.2 In Vitro Incubation .......................................................................................................................... 11
    2.2.1 Media ......................................................................................................................................... 11
    2.2.2 Incubation Protocol .................................................................................................................. 12
    2.2.3 Sample Collection .................................................................................................................... 12
  2.3 Bacterial Enumeration via qPCR ...................................................................................................... 13
  2.4 Lignan Metabolite Measurement via GCMS .................................................................................. 13
  2.5 DNA Extraction and Sequencing .................................................................................................... 14
  2.6 Taxonomic and Functional Analyses ............................................................................................... 15
2.6.1 Processing and Taxonomic Annotation of 16S rRNA Gene Amplicon Sequences
with QIIME ................................................................. 15
2.6.2 Processing of Metagenomic Sequences with MG-RAST .................................................. 15
2.6.3 Taxonomic Annotation of Metagenomic Protein Sequences with MG-RAST .... 15
2.6.4 Taxonomic Annotation of Metagenomic Protein Sequences with MetaPhlAn ...... 16
2.6.5 Functional Annotation of Metagenomic Protein Sequences with HUMAnN ........ 16
2.6.6 Analyses of Taxonomic and Functional Output .................................................. 17
2.7 Decision to Use MetaPhlan Instead of MG-RAST .................................................. 17
2.8 Decision to Use Summed KOs Instead of HUMAnN Module Output .................. 18

Chapter 3. Results .................................................................................................................................................. 19

3.1 Bacterial Enumeration via qPCR .................................................................................. 19
3.2 Lignan Metabolite Measurements via GCMS .................................................. 19
3.3 qPCR-normalized GCMS measurements .................................................................. 20
3.4 V4 16S rRNA Gene Taxonomic Analysis (n = 6) ........................................... 20
  3.4.1 Sequence Processing .................................................................................. 20
  3.4.2 Alpha diversity, Phyla and Genera .................................................................. 20
  3.4.3 Organisms Identified or Hypothesized to be Involved in Lignan Metabolism ........ 21
3.5 Metagenomic Taxonomic Analysis (n=3) ................................................................. 22
  3.5.1 Sequence Processing .................................................................................. 22
  3.5.2 Alpha diversity, Top 10 Phyla and Genera .................................................. 22
  3.5.3 Organisms Identified or Hypothesized to be Involved in Lignan Metabolism ........ 22
3.6 Metagenomic Functional Analysis (n=3) .................................................................. 23
  3.6.1 Pathways .................................................................................................. 23
  3.6.2 Modules .................................................................................................. 24
  3.6.3 KOs Corresponding to Enzymes Hypothesized to Act Directly on Lignans ....... 25

Chapter 4. Discussion ........................................................................................................................................... 26

4.1 Metabolite Measurements and Eubacterial Enumeration ........................................ 26
4.2 Taxonomic Analysis .................................................................................................. 26
4.3 Functional Analysis .................................................................................................. 29
ACKNOWLEDGEMENTS

I would like to thank my committee members Johanna Lampe and Meredith Hullar for their expertise and guidance. I would also like to express my gratitude to the whole Lampe Lab, with special thanks to Lisa Levy for recruiting the participants, Wendy Thomas for helping with data generation and providing great general life advice, Seth Yoder for helping collect much of the data and training me in the ways of in vitro incubations, Orsalem Kahsai for imparting her knowledge of molecular biology, Timothy Randolph for statistical and theoretical advice, Keith Curtis for helping analyze the data bioinformatically, and Elizabeth Traylor for helping with data collection when needed most.
Chapter 1. BACKGROUND

1.1 INTRODUCTION

Evidence suggests that lignans, one of the many bioactive phytochemicals in the human diet, may reduce the risk of several chronic diseases.\(^1,2\) Below I briefly review lignan sources, metabolism, and health effects, and discuss the crucial role that gut bacteria play in human exposure to lignan metabolites.

1.2 PLANT LIGNANS IN THE DIET

Lignans are polyphenolic compounds composed of two dimerized cinnamic alcohols.\(^2\) They are found in seeds, grains, leaves, fruits, and woody portions of various plants.\(^1\) Humans ingest many lignan-containing plant products, including: flax, sesame, cloudberry, hemp, and blackberry seeds, cereal grains, dates, apricots, prunes, garlic, asparagus, brassica vegetables, beer, wine, coffee, and tea.\(^3\) Lignans found in the human diet include secoisolariciresinol diglucoside (SDG), secoisolariciresinol (SECO), medioresinol (MED), matairesinol (MAT), lariciresinol (LARI), pinoresinol (PINO), syringaresinol (SYR), and sesamin (SES).\(^4\) The degree to which each of these lignans are found in the diet is subject of debate.\(^5\) SDG is the primary lignan found in flaxseed, the richest known dietary source of lignans.\(^3\)

The gut microbiota are crucial to the metabolism of lignans in humans. When lignans are ingested, metabolism and absorption appears to be quite rapid.\(^6,7\) Because of short transit time in the mouth, oral enzymatic digestion is unlikely.\(^4\) Lignans also appear to be resistant to the low pH of the stomach and are not metabolized by gastric enzymes.\(^8-10\) The majority of lignan bioavailability is due to microbial metabolism in the ileum and the large intestine, where great numbers of human microbes reside.\(^4,10\) This is well supported by the timing of lignan metabolite appearance in the blood in relation to transit time to colon,\(^4\) as well as the disappearance of measured lignan metabolites with antibiotic administration.\(^11\)

Upon entering the colon, lignans are metabolized by microbes into the bioactive compounds enterodiol (END) and enterolactone (ENL), known as enterolignans.\(^1\) I will focus on the
degradation of SDG, but other dietary lignans are metabolized to enterolignans similarly, sharing many of the same reactions and intermediates.\textsuperscript{12}

SDG undergoes several transformations mediated by gut microbes: $O$-linked deglycosylation to SECO, which can be $O$-link demethylated to dihydroxyenterodiol (DHEND), which can be dehydroxylated to END, which can be dehydrogenated to ENL. DHEND can also be dehydrogenated to dihydroxyenterolactone (DHENL), which can be dehydroxylated to ENL.\textsuperscript{4,13,14} These reactions appear to be enantiomer-specific, with chirality of the precursor lignan maintained throughout the reaction series.\textsuperscript{15} Either the (+) or (−)-enantiomer of SDG appears to dominate depending on the flax species.\textsuperscript{16,17}

Evidence suggests that most lignan absorption occurs in the colon.\textsuperscript{18} Enterolignans can be absorbed, conjugated, and excreted by enterocytes into the hepatic portal vein.\textsuperscript{19,20} Unconjugated lignans can also be absorbed from the lumen into the lymph and enter directly into systemic circulation.\textsuperscript{21,22} Enterolignans can be conjugated in the liver, excreted in the bile, acted on again by gut microbes, and reabsorbed – undergoing enterohepatic circulation.\textsuperscript{23} Enterolignans conjugated in the liver can also be released into systemic circulation. Once in the systemic bloodstream, they can be deposited in tissues throughout the body.\textsuperscript{24,25} Enterolignans are excreted from the body via urine – mainly in the mono-glucoronidated form – and feces.\textsuperscript{26}

1.3 \textbf{Health Effects of Lignans}

As reviewed by Clavel and Mapesa in 2013,\textsuperscript{27} evidence suggests that enterolignans may reduce the risk of cardiovascular disease and prostate, breast, and colon cancer. Several human dietary interventions have been conducted showing reduced prostate cancer cell proliferation,\textsuperscript{28} breast cancer tumor growth,\textsuperscript{29,30} and low-density lipoprotein (LDL) concentration\textsuperscript{31,32} with flaxseed supplementation. In a prospective cohort study of women (221 cases, 886 controls), lignan intake was associated with decreased colon cancer risk.\textsuperscript{33} Despite observed results, it is difficult to isolate the effects of lignans in these studies, since dietary sources and supplements contain many bioactive compounds in addition to lignans.\textsuperscript{27}

Few randomized controlled trials (RCT) using pure lignans have been carried out on humans. When 22 post-menopausal women were given a flaxseed-extracted lignan complex containing 500 mg/day SDG for 6 weeks they had significantly lower plasma C-reactive protein levels.\textsuperscript{34} In another study, 78 subjects with benign prostatic hyperplasia that were given a
flaxseed extract containing 300 mg/day SDG for 4 months had improved symptoms and quality of life scores.\textsuperscript{35}

There have also been several interventions conducted in animals. In murine models, SDG administration has lowered mammary tumorigenesis,\textsuperscript{36,37} lung metastasis,\textsuperscript{38} and colon carcinogenesis.\textsuperscript{39} SDG administration was shown to lower rates of atherosclerosis in rabbits\textsuperscript{40} and have cardio-protective effects in rats.\textsuperscript{41}

Several studies using \textit{in vitro} models help explain the bioactive properties of enterolignans. ENL can bind to human pregnane X receptor, affecting xenobiotic detoxification and steroid metabolism.\textsuperscript{42} ENL can weakly stimulate or block estrogen receptors – especially ER-α – potentially modulating estrogen signaling.\textsuperscript{43,44} ENL and END can also bind to sex hormone-binding globulin which may affect circulating steroid hormone levels.\textsuperscript{45} Enterolignans can affect cell proliferation in colon, prostate, and breast \textit{in vitro}-cultured cells.\textsuperscript{27} They have been shown to have stronger antioxidant properties \textit{in vitro} than lignan precursors.\textsuperscript{46,47}

1.4 \textbf{INTER-INDIVIDUAL DIFFERENCES IN ENTEROLIGNAN PRODUCTION}

Average enterolignan blood concentration for an unsupplemented individual is estimated to be 10 - 25 nmol/l,\textsuperscript{4} but varies considerably, and an ENL concentration over 1000 nmol/l has been documented in one individual. Dietary patterns can account for some of this variation and intake of vegetables and grains has been associated with higher serum enterolignan levels.\textsuperscript{48} Furthermore, human dietary intervention studies have shown higher serum enterolignan levels with increased intake of lignan-rich foods.\textsuperscript{18,49} Similarly, human dietary flaxseed supplementation has been shown to increase urinary enterolignan excretion.\textsuperscript{49,50}

Lignan intake levels, however, do not fully account for inter-individual variation. When 12 individuals were administered the same quantity of SDG, there was great variation in plasma enterolignan profiles – both total concentration of enterolignans and ratio of END:ENL.\textsuperscript{18} Inter-individual variation in urinary and plasma END:ENL ratios has also been documented in several dietary intervention studies.\textsuperscript{49,50} When stool from 100 different participants was incubated \textit{in vitro} with a flaxseed extract (final concentration in fecal suspension: 432 – 864 mg SDG per liter), END and ENL were produced in 63% and 39% of the samples, respectively.\textsuperscript{5} Differences in enterolignan status could be explained by many factors in addition to level of dietary intake –
including gut transit time, antibiotic use, and smoking status – but gut microbial composition appears to play a key role. 18,51-53

1.5 Organisms Involved in Lignan Metabolism

The populations of bacteria responsible for SECO to END conversion are subdominant in the human gut, on the order of $10^8$ cells/g stool, compared to $10^{12}$ total bacterial cells/g stool. 5,51 Those responsible for SECO to ENL conversion, however, appear to be on the order of $10^5$ cells/g stool. 18,51 Women appear to possess more enterolignan-producing bacteria than men. 51 One explanation for this is that lignan-metabolizing microbes may also be capable of acting on female sex hormones, such as progesterone. Thus, higher levels of colonic female sex hormones might enrich for lignan-metabolizing microbes. 4

In a cross-sectional study of 115 premenopausal women (ages 40 - 45), the bacterial 16S rRNA gene was amplified and sequenced from stool sample-extracted DNA, and urinary excretion of ENL was measured. 54 ENL production was associated with higher community alpha-diversity. Furthermore, presence of genera Moryella, Acetanaerobacterium, Fastidiosipila, and Streptobacillus in stool samples was significantly greater in the highest tertile of ENL excreters. The genera Fastidiosipila and Streptobacillus are related to bacteria that are capable of cleaving glucose from complex substrates 55-57 and hence may be capable of performing SDG to SECO conversion.

In another cross-sectional study of 84 men, serum concentration of ENL was measured using a time-resolved fluoroimmunoassay, and several different bacterial groups were enumerated in stool samples using fluorescence in situ hybridization. Those with the highest serum ENL concentration had a significantly higher fecal bacterial cell count, particularly of the Lactobacillus-Enterococcus group, when compared to those with the lowest serum ENL concentration. Enterococcus faecalis PDG-1 has been shown to convert pinoresinol to lariciresinol. 15 Lactobacilli and Enterococci both exhibit β-glucosidase activity, and thus may be capable of deglycosylating SDG. 58 Ruminal Prevotella spp. have been shown to deglycosylate SDG as well. 59

Using in vitro incubation of human stool, at least 29 species or strains of bacteria capable of performing lignan metabolism have been isolated and identified, as recently summarized by Yoder et al. 12 Bacteroides fragilis, Bacteroides ovatus, Clostridium cocleatum, Clostridium
saccharogumia, Clostridium ramosum, Bacteroides distasonis, Bifidobacterium bifidum WC 418, Bifidobacterium breve WC 421, Bifidobacterium catenulatum ATCC 27539, Bifidobacterium longum subsp. infantis ATCC 15697, Bifidobacterium longum subsp. longum WC 436, Bifidobacterium longum subsp. longum WC 439, Bifidobacterium pseudocatenulatum WC401, Bifidobacterium pseudocatenulatum WC 402, Bifidobacterium pseudocatenulatum WC 403, and Bifidobacterium pseudocatenulatum WC 407 have been shown capable of lignan deglycosylation.\textsuperscript{8,60-62} Butyribacterium methylotrophicum, Eubacterium callanderi, Eubacterium limosum, Blautia producta, and Clostridiaceae bacterium END-2 are able to demethylate lignans.\textsuperscript{8,61,63}

Several organisms have been shown to dehydroxylate and dehydrogenate lignans, in some cases enantioselectively. Eggerthella lenta, Clostridium scindens, and Lactonifactor longoviformis can dehydroxylate lignans.\textsuperscript{61} Eubacterium sp. ARC-2 is capable of dehydroxylating (-) DHEND and (+) DHENL. Eggerthella sp. SDG-2 can dehydroxylate (-) DHENL.\textsuperscript{64} Lactonifactor longoviformis is capable of dehydrogenating lignans.\textsuperscript{60} Clostridiaceae bacterium END-2 can dehydrogenate (+) END, while Ruminococcus sp. END-1 can dehydrogenate (-) END.\textsuperscript{65} Enterococcus faecalis and Eggerthella lenta can reduce lignans.\textsuperscript{8,15} It is probable that there are more undiscovered microbes that play a role in lignan metabolism.\textsuperscript{4}

1.6 H\textsubscript{2} Consumers in the Human Gut

In the human gut, the main form of microbial energy acquisition is fermentation. Poly- and oligosaccharides are broken down by endogenous and microbial enzymes, and resulting monosaccharides are fermented, generating short chain fatty acids (SCFAs) – including acetate, propionate, and butyrate – CO\textsubscript{2}, and H\textsubscript{2}. H\textsubscript{2} is primarily generated because of the need to reoxidize reduced pyridine and flavin nucleotides. Accumulation of H\textsubscript{2} would lead to a partial pressure that would thermodynamically inhibit this reoxidation. Thus, it is crucial that H\textsubscript{2}-consuming microbes are part of the intestinal community. These organisms fall into three categories based on their form of H\textsubscript{2} metabolism: acetogens, methanogens, and sulfate-reducing bacteria.

Acetogens utilize the Wood-Ljungdahl pathway (WLP) to reduce CO\textsubscript{2} to acetate. This is a multi-reaction process by which a methyl group is transferred from enzyme subunit to subunit. Overall, with H\textsubscript{2} as an electron donor, the total standard free energy change (\(\Delta G^\circ\)) for the
reduction of CO	extsubscript{2} to acetate is \(-95\) kJ per mol. No net ATP is generated, but this process is coupled with chemiosmotic ion gradient-driven phosphorylation and generates 1-2 ATP overall. This yield is estimated to be lower at physiological conditions.\textsuperscript{66}

Methanogens can also utilize the WLP to synthesize acetyl-CoA or methane (CH\textsubscript{4}). The synthesis of CH\textsubscript{4} involves the first half of the WLP, from which the methyl group is transferred to a different subset of enzymes. Overall, the synthesis of CH\textsubscript{4} from CO\textsubscript{2} and H\textsubscript{2} has a $\Delta G^\circ$ of \(-135.6\) kJ per mol\textsuperscript{67} and is estimated to generate 1 ATP per mol of methane.\textsuperscript{68}

Sulfate-reducing bacteria (SRB) reduce sulfate to sulfide for energy conservation. The SRB that live in the human gut generally oxidize H\textsubscript{2}, but can use other organic compounds as well.\textsuperscript{69} Overall, with H\textsubscript{2} as an electron donor, the $\Delta G^\circ$ for the reduction of sulfate to sulfide is \(-152.2\) kJ per mol.\textsuperscript{68} This process is estimated to generate 1 ATP per molecule of sulfate reduced, and even less when energy requirement for uptake of sulfate is taken into account.\textsuperscript{70}

The extent to which each form of metabolism exists in the human intestine is unclear.\textsuperscript{69} Results of some studies suggest that methanogenesis and sulfate reduction dominate because they are thermodynamically more favorable;\textsuperscript{71} however, one metagenomic study of human stool found acetogenesis to be the most prevalent form of microbial H\textsubscript{2} utilization in the gut.\textsuperscript{72}

1.7 The Potential Role of Acetogens in Lignan Metabolism

Acetogens use acetyl-CoA, generated from the WLP, for biomass or ATP generation.\textsuperscript{66,73} Many acetogens belong to the genus Clostridium, but the group is phylogenetically diverse. It is likely that many bacteria have yet to be characterized as acetogens.\textsuperscript{72}

Interestingly, lignan demethylation, dehydroxylation, and dehydrogenation appear to be carried out by bacteria that have been characterized as, or are phylogenetically related to, acetogens. \textit{B. methylotrophicum}, \textit{B. producta}, \textit{E. limnosum} are all documented acetogens.\textsuperscript{73-75} \textit{Ruminococcus sp. END-1} shares a 98% 16S rRNA gene sequence similarity to the lignan-metabolizing \textit{B. producta} strain.\textsuperscript{64} \textit{L. longiformis} and \textit{Clostriadiaceae} bacterium END-2 are highly related to acetogenic species in \textit{Clostridium} cluster XIVa, including \textit{B. producta}.\textsuperscript{60} \textit{Eubacterium sp. ARC-2} shares a 99% 16S rRNA gene sequence similarity to \textit{E. limnosum}, and 98% similarity to \textit{B. methylotrophicum}.\textsuperscript{76} \textit{E. lenta} possesses a formyltetrahydrofolate synthetase (FTHFS) that has a 93% similarity to acetogenic FTHFS. FTHFS PCR amplification is typically used to assess presence of acetogens.\textsuperscript{77} \textit{Eggerthella sp. SDG-2} is closely related to \textit{E. lenta}.\textsuperscript{64} \textit{E.}
callanderi is in *Clostridium* Cluster XV, and most other species in this cluster are acetogens. The genus *Moryella* forms a sub-lineage with *Blautia cocoides*, an acetogen, in *Clostridium* cluster XIVa and produces acetate *in vitro*. As its name suggests, the genus *Acetoanaerobium* contains at least three known acetogens including *A. ruminis*, *A. noterae*, and *A. romashkovii*.

### 1.8 The Potential Role of Methanogens in Lignan Metabolism

To date, no archaea have been identified as capable of lignan metabolism; however, their methods of energy acquisition are highly similar to those of acetogens. Archaea utilize the WLP, many of the same substrates, and highly similar enzymes.

Approximately 50% of the human population possesses intestinal archaea, including *Methanobrevibacter smithii*, *Methanosphaera stadtmannae*, *Candidatus Methanomethylophilus alvus*, *Candidatus Methanomassiliicoccus intestinalis* and *Methanomassiliicoccus luminyensis*. Another order of archaea (Mx-lineage) that contains species closely related to the Rumen Cluster C has also recently been identified. *M. smithii* is both the most abundant and prevalent archaea in the human gut, followed by *M. stadtmannae*. The remaining three identified species, belonging to the order *Methanomassiliicoccales*, appear to be less prevalent.

All archaea found in the human gut thus-far are methanogens, organisms for which methanogenesis is their sole energy-yielding metabolic process. There are two primary forms of methanogenesis that occur in the human gut, both H₂-dependent. Hydrogenotrophic methanogenesis involves using H₂ to reduce CO₂ to methane through the WLP and is utilized by *M. smithii*. The main form of methylotrophic methanogenesis in the human gut involves demethylation of a variety of compounds – including methanol and methylamines – using H₂ to reduce the methyl group to CO₂. This pathway is utilized by *M. stadtmannae* and species of the order *Methanomassiliicoccales*.

### 1.9 Microbial Enzymes in Lignan Metabolism

Deglycosylation is the only step of lignan metabolism that has been connected to a specific bacterial enzyme. Identities of other enzymes involved in lignan degradation can only be hypothesized based on the reactions. Below I use two different number classification schemes: 1) Enzyme Commission (EC) numbers refer to groups of enzymes that carry out the same chemical
reaction, KEGG Orthology gene families (KOs) refer to groups of genes that carry out similar biochemical functions.  

1.9.1  \textit{O}-\textit{linked Deglycosylation}

Microbial $\beta$-glucosidases (BG)[EC 3.2.1.21] can carry out lignan deglycosylation.\textsuperscript{62} Bacteria harboring BG appear to be quite prevalent in the human gut, on the order of $10^{10}$ cells/g stool.\textsuperscript{89} A wide range of genera from human stool, including \textit{Bacteroides}, \textit{Bifidobacteria}, \textit{Butyrivibrio}, \textit{Coprococcus}, \textit{Eubacterium}, \textit{Roseburia}, and \textit{Ruminococcus}, have been shown to possess BG activity.\textsuperscript{90} Many of these genera contain known acetogens.\textsuperscript{73} BuBGL8, coded by the \textit{bgl8} gene from \textit{Bacteroides uniformis} ZL1, was shown to be a BG capable of deglycosylating SDG.\textsuperscript{91}

1.9.2  \textit{O}-\textit{linked Demethylation}

Acetogens, methanogens, and sulfate-reducing bacteria harbor highly homologous O-demethylases that utilize corrinoid proteins to cleave methyl groups from methoxy groups. A common assay used to select for acetogens is to culture on methoxylated aromatics such as vanillate.\textsuperscript{92} Acetogenic O-demethylases cleave off oxygen-bound methyl groups and funnel them into the WLP. \textit{Moorella thermoacetica}, the model acetogenic bacteria, is capable of metabolizing over 20 different types of methoxylated aromatic compounds.\textsuperscript{93} Lignan demethyating \textit{E. limnosum}, \textit{B. methylotrophicum}, and \textit{B. producta} have been shown to O-demethylate several different methoxylated compounds.\textsuperscript{73,94} Because the conversion of SECO to DHEND is a methoxylated aromatic O-demethylation, it is quite plausible that it could be catalyzed by a wide range of acetogens.

Acetogens harbor different types of similarly related O-demethylases. \textit{M. thermoacetica} possesses a demethylase composed of 3 subunits coded by MtvA, -B and -C (K00548).\textsuperscript{93} There have also been several Mtv homologs identified in the genome of \textit{M. thermoacetica}.\textsuperscript{95} The four-subunit O-demethylase possessed by \textit{A. dehalogenans} is composed of an activating enzyme, a corrinoid protein (K16179), and two methyl transfer proteins coded by \textit{odm}A (K14082) and \textit{odm}B (K00548).\textsuperscript{96} \textit{Holophaga foetida} contains several genes belonging to these KOs as well.\textsuperscript{97} Its demethylating activity has been shown to have a broad range of substrate-specificity.\textsuperscript{72}

Methanogens possess demethylases that are highly similar to the acetogenic enzymes coded by mtvABC.\textsuperscript{72,96} All identified human gut methanogens have the gene that encodes methyl-
transferase $\text{mtaABC}$ (K14080, K04480, K14081) which performs demethylation of methanol.$^{82}$ This methyl-transferase may demethylate SECO as well.

Sulfate-reducing bacteria can demethylate methoxylated aromatics. *Desulfitobacterium hafniense*, which has been shown to perform these reactions, appears to possess many putative O-demethylase encoding operons. The proteins of one of these operons, Dhaf_4610, Dhaf_461, and Dhaf_4612, have been characterized and express high similarity to O-demethylase proteins of *A. dehalogens*.$^{98}$

### 1.9.3 Dehydroxylation

Bile acids are metabolized by human gut microbes in the intestinal lumen. Interestingly, microbial bile acid metabolism is highly similar to lignan metabolism, in that bile acids are aromatic compounds that are first deconjugated, and subsequently dehydroxylated and dehydrogenated.

Both lignans and bile acids need to be deconjugated before they can be dehydroxylated.$^{13,99}$ Bile acid dehydroxylation is an energy-yielding process by which bile acids act as electron acceptors.$^{100}$ Most bacteria capable of 7α-dehydroxylating bile acids are highly related and fall within *Clostridium* cluster XIVa, a phylogenetic group that contains many acetogens.$^{101}$ 7α-dehydroxylating species include *C. scindens, C. hiranonis, C. hylemonae, C. sordellii, B. thetaiotaomicron*, and certain species in the genus *Eubacterium*.$^{100,102-104}$ The bile acid inducible (*bai*) genes (K15868, K15869, K15870, K15871, K15872, K15873, K15874) code for enzymes involved in this process.$^{102-104}$ *Eggerthella lenta* possesses corticosteroid 21-dehydroxylase [EC 1.14.99.10], involved in 21-dehydroxylating biliary steroids.$^{105}$ An *E. lenta*-like strain was also identified as having 7α-dehydroxylating ability.$^{106}$ Since *C. scindens, Eubacterium* spp., and *Eggerthella* spp. have been shown to dehydroxylate SECO, it has been suggested that bile acid dehydroxylating enzymes may be involved in the dehydroxylation of lignans.$^{4}$

### 1.9.4 Dehydrogenation

Organisms that catalyze bile acid dehydrogenation may benefit from the reducing equivalents generated from the process.$^{99}$ Bile acid 3-, 7-, 12-, 20- and 21- hydroxyl groups are dehydrogenated by microbial hydroxysteroid dehydrogenases (HSDH) (EC 1.1.1.357, EC 1.1.1.51, EC 1.1.1.159, EC 1.1.1.201, EC 1.1.1.176, EC 1.1.1.238).$^{99,107}$ The taxonomic
distribution of HSDH genes is more widespread than the bai genes, but the acetogen *B. productus* does possess three types of HSDH. Other organisms closely related to known acetogens, including *Eubacterium aerofaciens, Ruminococcus gnavus*N53, *Clostridium absonum, Clostridium sordelii, Clostridium scindens,* and *Clostridium barattii,* can dehydrogenate bile acids.\textsuperscript{108,109,110} Since the conversion of END to ENL requires dehydrogenation of aromatic diols, and all three of the lignan dehydrogenating strains identified – *Clostridium* sp. END-2, *Ruminococcus* sp. END-1, and *L. longiformis*– are highly related to *B. producta,* HSDH may be involved.

Another instance of aromatic hydroxyl dehydrogenation is the conversion of estrone to estradiol. Several intestinal bacteria have been identified that can carry this out including *Akaligenes faecalis, Pseudomonas aeruginosa, Staphylococcus aureus,* and *Streptococcus faecalis.*\textsuperscript{111} These bacteria may harbor enzymes capable of lignan dehydrogenation.

### 1.10 Aim, Hypothesis, and Implications

For the current study we incubated human stool samples anaerobically with a mineral salts media (TCAP2) and SDG in an *in vitro* batch system. We hypothesized that incubation of stool with media containing SDG would enrich for lignan-metabolizing microbes and corresponding genes. We assessed the associated metabolites of SDG (SECO, ENL, and END) and presence of microbial taxa and genes. Through this, we were able to taxonomically and functionally describe a community that metabolizes lignans. Identification of these microbes and enzymes will help characterize an individual’s enterolignan-producing capability. This information has implications for our understanding of cancer and cardiovascular disease susceptibility, and may help in development of dietary interventions and preventative probiotics.
Chapter 2. METHODS

2.1 SUBJECTS

Subjects 6105, 6127, 6128, 6129, 6171, 6178, and 6188 from the FlaxFX study (ClinicalTrials.gov identifier: NCT01619020) provided seven of the eight stool samples used for our experiments. Men and women were eligible to participate if they were 20-45 years, healthy non-smokers with a low vegetable and fiber intake, consented to be in the trial, and consented to provide a fresh stool sample. Individuals were excluded if they had been prescribed antibiotics within the last 3 months, currently taking prescription or over-the-counter medications (including oral contraceptives), consumed more than two alcoholic drinks per day, pregnant or lactating, diagnosed with a chronic medical illness or gastrointestinal disorder, or had experienced a recent weight change of 4.5 kg or greater. The eighth stool sample was derived from the same individual (subject 5100) who provided samples used for methods development of our in vitro platform. Not all samples were used for all analyses (Table 1).

2.2 In Vitro Incubation

2.2.1 Media

TCAP2 media (Table 2) was prepared by adding heat-stable ingredients to de-ionized water, adjusting to a pH of 7.3, and autoclaving for 20 minutes at 121 °C. After cooling, resasurine, vitamin solution, trace element solution, glucose, and cysteine were added by filter-sterilization through a 0.2-micron disc filter. A gas manifold hose was fitted with a 0.2-micron disc filter, followed by a small sterile plastic hose, which was inserted into the bottle of media. The media was sparged with 100% high-purity N₂ (Airgas, Radnor Township, PA) and heated to 70 °C in a water bath until the purple color of resasurine, an anaerobic indicator, disappeared (approximately 20 - 90 minutes). The bottle was capped and immediately placed in a Bactron Anaerobic Chamber that contained a gas mixture of 80% N₂, 15% CO₂, and 5% H₂.
2.2.2 Incubation Protocol

Each participant provided a fresh stool sample for in vitro analysis. The stool was collected in a plastic tub, and within 30 minutes, brought into a Bactron Anaerobic Chamber. Inside the chamber the stool was weighed and combined with anaerobic TCAP2 media to a final concentration of 1.66% weight:volume. The stool was broken apart with a spatula and further mixed with a homogenizer. The fecal suspension was then filtered through sterile cheesecloth and electronically pipetted into 50 mL glass serum bottles (20 mL aliquots). These bottles were inoculated with lignan substrate dissolved in 100% methanol (Thermo Fisher Scientific, Waltham, MA) to a final concentration of 6.55 µM SDG, 6.54 µM ENL, 6.45 µM END, or left un-spiked (a blank “control”).

The bottles were capped with rubber stoppers and brought outside the chamber, where the tops were fitted with aluminum seals and sterilized with 100% ethanol. A gas manifold hose was fitted with a 0.2 micron disc filter followed by a sterile Luer-lok needle, which was inserted into the rubber stopper at the top of the serum bottle. Another Luer-lok needle was inserted into the rubber stopper as a vent and the serum bottle headspace was replaced with 100% high-purity N₂ for 10 minutes at the rate of 0.5-1.0 L/min. Bottles were placed in a rotating incubator (C24 Incubator, New Brunswick Scientific, Enfield, CT) and incubated at 37°C/300 rpm. Samples from 6105, 6127, 6128, 6129, and 5100 were incubated for 7 days. Samples from 6171, 6178, and 6188 were incubated for 6 days.

2.2.3 Sample Collection

Daily aliquots of fecal suspensions from individual incubating serum bottles were taken at approximately the same time every day using anaerobic technique, and stored at -20 °C in aliquots of 500 µL/day for lignan analysis and 100 µL/day for bacterial enumeration. Samples for DNA analysis were taken on Day 1 and the final day of the incubation (Day 6 or 7), and stored at -80 °C. The Day 1 sample was two pea-sized scoops of fresh stool dispersed in 5 mL sterile RNALater (Thermo Fisher Scientific, Waltham, MA) collected in the anaerobic chamber during the previously described procedure. The final-day sample was obtained from pelleted SDG-incubated fecal suspensions. To collect the pellet, fecal suspensions from serum bottles were combined in a sterile 50 mL centrifuge tube and spun down in a Beckman Coulter
Centrifuge. Suspensions from 6105, 6127, 6128, 6129, and 5100 were centrifuged at 6000 rpm/20°C for 10 minutes, supernatant was discarded, and the pellet was re-suspended with 5 mL of sterile RNAlater for storage. Suspensions from 6171, 6178, and 6188 were centrifuged at 14,000 rpm/4°C for 10 minutes. Supernatant was discarded, pellets were re-suspended in TCAP2, and tubes were re-centrifuged using the same conditions. Supernatant was discarded and the pellet was re-suspended with 5 mL of sterile RNAlater for storage.

2.3 Bacterial Enumeration via qPCR

Eubacterial 16S rRNA gene copy number concentration in SDG-inoculated fecal suspensions was estimated via quantitative polymerase chain reaction (qPCR) (7900HT Fast Real-Time PCR System by Applied Biosystems, Carlsbad, CA). Universal primers 330F (5’-ACT CCT ACG GGA GGC AGC AGT-3’) and 530R (5’-GTA TTA CCG CGG CTG CTG GCAC-3’) were used to amplify the 16S rRNA gene\textsuperscript{112,113} with SYBR Green (Invitrogen, Carlsbad, CA), a total reaction volume of 20 µL, and cycling conditions: 50 °C for 2 min; 95 °C for 10 min; [95 °C for 30 sec; 58 °C for 30 sec; 72 °C for 1 min] x40; 95 °C for 15 sec; and 95 °C for 15 sec. We extended the first 95 °C step to 10 minutes in order to ensure cell lysis.

2.4 Lignan Metabolite Measurement via GCMS

For lignan metabolite analysis, SECO, END, and ENL concentrations from fecal suspensions were measured via gas chromatography mass spectrometry (GCMS) using modified methods from Frankenfeld \textit{et al}\textsuperscript{114} as described elsewhere.\textsuperscript{14} Time points from Days 1 through 6 were analyzed for some participants, while time points from only Days 1 and 6 were analyzed for other participants (Table 1). Day 7 was additionally analyzed for participant 6105. It was decided to only use measured metabolite concentrations through Day 6 for analyses because: a) metabolite levels did not seem to change when measured through Day 7 (Figure 1) and b) ENL appeared to be the final metabolite in the pathway.

For each time point, measured metabolite concentration was divided by inoculated substrate concentration to calculate percent conversion. In a separate analysis for samples that had been analyzed by qPCR (Table 1), moles of metabolite was divided by qPCR-estimated copy number at each time point to normalize for bacterial biomass.
2.5 DNA EXTRACTION AND SEQUENCING

DNA was extracted from Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension with the Qiagen Stool DNA Mini Kit (Qiagen, Hilden, Germany)\textsuperscript{115} using a modified protocol of the manufacturer’s instructions. DNA was sent to Research and Testing (Shallowater, TX) for 16S rRNA gene and whole metagenome shotgun sequencing. The 12 samples (from 6 participants) that were V4 16S rRNA gene-sequenced and used for our analyses [6x Day 1 initial stool and 6x final-day pelleted SDG-incubated fecal suspension (Table 1)] were sequenced with 19 samples that were not used for analyses in this paper. There were 6 samples (from 3 participants) [3x Day 1 initial stool and 3x final-day pelleted SDG-incubated fecal suspension (Table1)] that were sent for metagenomic sequencing and used in our analyses.

The V4 region of the 16S rRNA gene was paired-end (2×300) sequenced with the target of 30,000 sequences-per-sample. Primers 515F (5’-GTGCCAGCMGCGCCGCGGTAA-3’ + barcode) and 806R (5’-GGACTACVSGGGTATCTAAT-3’),\textsuperscript{116} were used in a 30 cycle PCR with HotStarTaq Plus Master Mix Kit (Qiagen, USA) to amplify the V4 region of the 16S rRNA gene with the cycling conditions: 94°C for 3 min; [94°C for 30 sec] x28; 53°C for 40 sec; 72°C for 1 min; 72°C for 5 min. PCR products were run in 2% agarose gel to ensure high quality amplification. Samples were pooled in equal ratios based on DNA concentration and molecular weight. Pooled samples were purified with Ampure XP beads (Beckman Coulter, Brea, CA). Libraries were prepared according to the Illumina TruSeq DNA library preparation protocol.\textsuperscript{117} Illumina MiSeq instruments and reagents were used following manufacturer’s instructions.

Samples were sequenced using whole metagenome shotgun sequencing\textsuperscript{118} with the target of 6 million sequences-per-sample. Sample concentrations were measured using the Qubit® dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA). Libraries were prepared using the Nextera DNA Sample preparation kit (Illumina, San Diego, CA) following the manufacturer's instructions. Specifically, samples were diluted to a concentration of 2.5 ng/μL. An aliquot containing 50 ng of DNA was fragmented and tagged using the Nextera Enzyme Mix, which contains a transposase and transposon ends bound to sequencing primer sites. A 5-cycle PCR was carried out, in which unique index was added to the samples. The final concentration and average size of the library was measured using the Qubit® dsDNA HS Assay Kit and the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), respectively. The library (12.5 pM)
was paired-end (2x300) sequenced in MiSeq (Illumina) with the 600 Cycles v3 Reagent Kit (Illumina), and following the manufacturer’s instructions.

2.6 TAXONOMIC AND FUNCTIONAL ANALYSES

2.6.1 Processing and Taxonomic Annotation of 16S rRNA Gene Amplicon Sequences with QIIME

16S rRNA gene sequences were processed through an in-house QIIME\textsuperscript{116} pipeline. Paired-end reads were joined, removed if <200 basepairs (bp) or >700 bp, and truncated with QIIME scripts and a quality score sliding window of 50 bp. They were then clustered into operational taxonomic units (OTUs) at 97% similarity and classified using MOTHUR\textsuperscript{119} and the SILVA database(www.arb-silva.de). For community structure analyses, annotated sequences were grouped into phylum and genus-level phylotypes.\textsuperscript{120} The final output was a data matrix of metagenomic sample by organism abundance.

2.6.2 Processing of Metagenomic Sequences with MG-RAST

Unassembled metagenomic sequences were annotated by the Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline.\textsuperscript{121} Briefly, paired-end reads were joined, and low quality regions were trimmed from FASTQ data using SolexaQA\textsuperscript{122} so that sequences contained at most 5 bases below a Phred quality score of 15. The data were screened, using Bowtie,\textsuperscript{123} to remove reads that mapped to the human genome. Genes were called with FragGeneScan.\textsuperscript{124}

2.6.3 Taxonomic Annotation of Metagenomic Protein Sequences with MG-RAST

Proteins were clustered at 90% similarity with uclust\textsuperscript{125} in QIIME.\textsuperscript{116} sBLAT\textsuperscript{126} was used to map the longest sequence in each cluster to the M5nr database.\textsuperscript{127} Phylum-level taxonomy was assigned to M5nr-annotated reads with an e-value less than 1e-10, an identity greater than 80%, and an alignment length greater than 75 base-pairs\textsuperscript{128} using the Representative Hit tool. Organism abundance tables were collated with in-house R scripts to yield a data matrix of metagenomic sample by organism abundance.
2.6.4  **Taxonomic Annotation of Metagenomic Protein Sequences with MetaPhlAn**

Reads that had been processed through the trimming step of the MG-RAST pipeline were uploaded to the Huttenhower Galaxy server and classified with MetaPhlAn\textsuperscript{129} using the “very sensitive (global)” setting of BowTie\textsuperscript{2} as recommended by MetaPhlAn developers for human-associated microbiomes. The global option utilizes end-to-end alignment, which uses all nucleotides in a query sequence for alignment. The “very sensitive” option maximizes sensitivity and accuracy of the alignment with preset parameters: -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 as defined elsewhere.\textsuperscript{130}

Reads were mapped to a collection of clade-specific marker genes derived from the IMG database,\textsuperscript{131} providing species-level taxonomic assignment. The number of marker genes per clade varies, but the number of hits per clade was normalized by the sum of nucleotides in all marker genes corresponding to that clade. Due to a problem with the server, we were unable to upload the sample from 6128 Day 1, which has been left out of our MetaPhlAn taxonomic analysis. The final output was a data matrix of metagenomic sample by organism abundance.

2.6.5  **Functional Annotation of Metagenomic Protein Sequences with HUMAnN**

Reads that had been processed through the screening step of the MG-RAST pipeline were annotated by DIAMOND\textsuperscript{132} using KEGG database v56. Annotations with an e-value < 1e-5 were analyzed using HUMAnN.\textsuperscript{133} This program normalizes abundance for each KO and then assigns KOs to KEGG modules and pathways. Modules and pathways are groups of KOs that work together in a biochemical pathway.\textsuperscript{88} Modules generally contain fewer KOs and involve fewer reactions than pathways.

During HUMAnN processing, abundance for each KO was weighted by the e-value of its hits, normalized by average sequence length of its gene family, and assigned to KEGG modules and pathways using MinPath.\textsuperscript{134} The mapping of KOs to modules and pathways resulted in the output of (KO, module/pathway) pairs. Modules/pathways that were in clear disagreement with a rough taxonomy of reads in a sample were eliminated. Specifically, a very approximate organism abundance was calculated. Each organism abundance was derived by summing all normalized, weighted hits for that organism. “Expected abundance” for each module/pathway was then calculated by summing the abundances for each organism that reads in that module/pathway had
mapped to. (KO, module/pathway) pairs for KOs that mapped to two or more modules/pathways and corresponded to a module/pathway that possessed an abundance below the “expected abundance” were removed.

The abundance of each organism-specific KO was further normalized by organism-specific gene family copy number. Specifically, KOs with an abundance 1.5 inter-quartile ranges below the median number of hits across all the genes in the module/pathway were replaced with this median. Finally, module/pathway abundance was calculated as the mean of the upper half of its component KO abundances. For individual gene analysis, the “01b-hit-keg-cat” file was used. For module and pathway analysis, the “04b-hit-keg-mpm-cop-nve-nve” and “04b-hit-keg-mpt-cop-nul-nve-nve” files were used, respectively. The final output was a data matrix of metagenomic sample by KO, module, or pathway abundance.

2.6.6 Analyses of Taxonomic and Functional Output

All abundance tables that were not already in the form of relative abundance expressed as a percent, were converted. Relative abundance was calculated by dividing specific organism, KO, module, or pathway abundance by the sum of all organism, KO, module, or pathway (respectively) abundances in the sample and multiplying by 100. Shannon’s index\(^\text{135}\) was used to for assessment of taxonomic alpha diversity. For 16S rRNA gene amplicon sequences, this was carried out as part of our in-house QIIME pipeline with OTUs rarified to 30,766 sequences per sample. For metagenomic sequences, this was carried out using species-level assignments from MetaPhlAn with the VEGAN R package.\(^\text{136}\) Organisms and KOs identified or hypothesized to be involved in lignan metabolism (Table 3 and 4) were searched for using in-house R scripts.

2.7 Decision to Use MetaPhlAn Instead of MG-RAST

Classification of protein sequences into phyla was carried out using both MG-RAST and MetaPhlAn (Figure 2). Notably, MG-RAST classified sequences into many phyla that had never been identified in human stool. Relative abundances of dominant phyla were substantially different. For example abundance of Bacteriodetes was much higher, and Firmicutes and Euryarchaeota much lower, in MG-RAST output.

One sample, 6128 Day 7, was sequenced using both V4 16S rRNA gene amplicon and metagenomic shotgun sequencing. When classification of sequences into dominant phyla was
compared between our in-house V4 16S rRNA gene pipeline, MG-RAST, and MetaPhlAn, MGRAST was the outlier (Figure 3). Because of this, and because MG-RAST output contained spurious phyla, MetaPhlAn was chosen for further taxonomic analysis of metagenomic data.

In an additional analysis (data not shown), for which we used the MG-RAST Representative Hit tool to classify 16S rRNA gene sequences from the 6128 Day 7 shotgun-sequenced metagenomic sample (Database: M5rna, $10^{-12}$ e-value <1e-10, identity >97%, alignment length >150 bp), relative abundances of major phyla were similar to the MG-RAST protein output.

2.8 Decision to Use Summed KOs Instead of HUMAnN Module Output

Several KEGG pathways hypothesized to be involved in lignan metabolism were searched for in the HUMAnN pathway output. Of those that were found (Figure 21a), methane metabolism (ko00680) was present at high levels in all samples. Subsequently, KEGG modules corresponding to ko00680 were searched for in HUMAnN module output (Figure 4). Methane oxidation (M00174) and Coenzyme M biosynthesis (M00358) were the only modules found. This did not make sense according to the V4 16S rRNA gene and metagenomic taxonomic classification. For example, methanogens were highly abundant in Day 7 samples from 6128 and 5100, but the modules for methanogenesis were nonexistent as annotated by HUMAnN.

Furthermore, M00174 was found at very low levels (<0.000025%) in the 6105 samples, and did not seem to explain the relatively high abundance (~0.6%) of ko00680 that appeared in the pathway output. Component KOs of M00174 were then searched for in the HUMAnN KO output and none were found. It was concluded that this must have been a HUMAnN error that occurred during assignment of KOs to modules. It was thus decided to sum component KO relative abundances (from the “01b-hit-keg-cat” file) to yield module and pathway relative abundances. We used this method along with HUMAnN pathway output to assess pathway abundance, but did not use HUMAnN module output for analyses. Modules unique to pathway ko00680 were investigated to further understand the components contributing to its abundance.
Chapter 3. RESULTS

3.1 BACTERIAL ENUMERATION VIA QPCR

SDG incubations from the four participants assessed exhibited a similar pattern of eubacterial cell growth over time, although 16S rRNA gene copy number per µl for 6128 and 5100 changed less over time than for 6105 and 6129. 16S rRNA gene copy number per µl peaked on Day 2 or 3 and ranged between 2,583,751 and 20,342,383 (Figure 5). For most serum bottles, copy number declined after Day 3. For some, especially one bottle from participant 5100, concentrations began to rise from Day 5 to Day 7.

3.2 LIGNAN METABOLITE MEASUREMENTS VIA GCMS

Lignan concentration data from serum bottles inoculated with END, ENL, and left un-inoculated were used as controls to isolate END to ENL conversion, any metabolism of ENL, and presence of any lignans in the original stool sample, respectively. Stool from four participants converted nearly all inoculated END to ENL (86 – 92%) in vitro (Figure 6). 6128 converted roughly 30% END to ENL almost immediately, whereas the rest of the END-incubated samples converted no substrate by the first time point. Most of the inoculated ENL was recovered at Day 6 (minimum 83 %), and no other metabolites were measured in this treatment. Transient low levels of lignans were measured in the “blank” controls. The maximum of these measurements was 0.53 µM ENL (8% of the SDG concentration added to SDG-inoculated bottles).

Stool of all 8 participants converted the majority of SDG to ENL by Day 6 (Figure 7). The average and standard deviation of SDG to ENL percent conversion from all SDG-inoculated serum bottles was 88.4% ± 6.9. For 6105 and 6129, the time course of measured metabolites reflected several steps of the SDG conversion pathway. SDG was metabolized almost immediately to SECO, which was metabolized to END and subsequently ENL. 6105 produced by far the most END. Negligible END was measured for 5100 and 6128, and negligible SECO was measured for 5100.
3.3 QPCR-NORMALIZED GCMS MEASUREMENTS

Normalizing measured metabolites by eubacterial 16S rRNA gene copy number changed the rate of conversion moderately (Figure 8). Most notably, ENL:copy number ratio for 6105 rose steadily and did not plateau, when compared to un-adjusted measurements, and to adjusted measurements of incubations from other participants. The maximum value for this measurement was also substantially higher than adjusted measurements of incubations from other participants. Since all incubations had similarly high levels of percent conversion to ENL, this was largely due to declining copy number during the latter half of the 6105 incubation. ENL:copy number ratio plateaued for incubations from other participants, however for both 6129 and 5100 there was a slight decrease in this value from Day 5 to Day 6. This was mainly due to increase in copy number during this time period for these incubations. Another notable effect of normalization was a high initial SECO:copy number ratio for 6129.

3.4 V4 16S rRNA GENE TAXONOMIC ANALYSIS (N = 6)

3.4.1 Sequence Processing

For the 31 samples sequenced using Illumina MiSeq paired-end sequencing of the V4 region, 4.2 million raw sequences with an average length of 299 bp were processed. Of these sequences, 1.9 million with the average length of 272 bp passed quality control. The twelve samples (out of 31) that were used for analyses, all possessed over 38,000 quality-processed sequences (Table 5).

3.4.2 Alpha diversity, Phyla and Genera

OTU-level alpha diversity, for the most part, did not change consistently with a ~1 week incubation; however, there were moderate decreases in diversity for incubations 6128 and 6129 when comparing Day 1 initial stool to Day 7 pelleted fecal suspension (Figure 9). Taxonomy was first assessed at the level of phylum (Figure 10). Bacteriodetes grew in all incubations, with the exception of 6128. Firmicutes decreased in all incubations. Euryarchaeota grew substantially in 6127, 6128, 6171. Incubations from the rest of the participants demonstrated growth of Euryarchaeota, but at very low levels (Figure 11). None of the samples contained
Verrucomicrobia with a relative abundance over 1.5%, although in incubations from 6127 and 6188, this phylum increased notably.

Taxonomy was then assessed at the level of genus. Growth of Bacteroidetes was primarily accounted for by an expansion of Bacteroides, and to a lesser degree, Parabacteroides and Alistipes (Figure 12). Decrease in Firmicutes was not attributable to one particular genus (Figure 13). Growth of Euryarchaeota was almost entirely due to Methanobrevibacter (Figure 14). Verrucomicrobia was primarily composed of Akkermansia, with the exception of the 6127 Day 7 sample, which possessed an uncultured bacterium of the order vadinHA64 (Figure 15).

### 3.4.3 Organisms Identified or Hypothesized to be Involved in Lignan Metabolism

Subsequently, genera that were identified or hypothesized to be involved in lignan metabolism were investigated for in vitro enrichment (Figure 16, Table 6). The only organism out of these genera that grew substantially in all incubations was Bacteroides. A sulfate-reducing organism of the family Desufovibrionaceae grew markedly in the incubation from 6129, and to a lesser degree in the incubation from 6188. Bifidobacterium was also present at high levels – especially in samples from 6127 – but did not grow consistently in all incubations. An organism of the order Eubacteriaceae grew substantially in the 6128 incubation. Other genera of interest were present at lower abundance. Many genera searched for were not present or detected in our samples.

Most samples contained at least one genus identified or hypothesized to be involved in each step of SDG to ENL metabolism. Lignan deglycosylating, demethylating, and dehydroxylating organisms were found at higher abundance than dehydrogenating organisms. As a group, deglycosylating organisms grew in vitro. As a group, demethylating organisms grew in all incubations except 6178. The 6128 Day 7 sample had particularly high levels of demethylating and dehydroxylating organisms, which was heavily accounted for by the order Eubacteriaceae. Dehydroxylating organisms grew in the 6127, 6128, and 6171 incubations.
3.5 Metagenomic Taxonomic Analysis (n=3)

3.5.1 Sequence Processing

For the 6 samples sequenced using Illumina MiSeq paired-end metagenomic shotgun sequencing, 35.3 million raw sequences were processed, with an average length of 301 bp. After quality control, 28.7 million sequences with the average length of 262 bp were used for analyses.

3.5.2 Alpha diversity, Top 10 Phyla and Genera

The Day 7 pelleted SDG-incubated fecal suspension had a marginally lower alpha diversity than Day 1 initial stool at the species-level for incubations from 6105 and 5100 (Figure 17). Phyla Bacteroidetes and Firmicutes dominated initial stool samples (Figure 18). Firmicutes relative abundance decreased with in vitro incubation. Euryarchaeota relative abundance increased substantially in the 5100 incubation and was dominant in 6128 Day 7. At the genus-level, this was largely due to the presence of Methanobrevibacter in 6128 and Methanobrevibacter and Methanosphaera in 5100. Euryarchaeota was measured at very low levels in 6105 Day 7, but abundance was negligible compared to Day 7 from the other two participants. Verrucomicrobia and Bacteroidetes grew substantially with stool incubated from 6105. The genus Akkermansia accounted for most Verrucomicrobia abundance. At the species-level, these reads were assigned to the species Akkermansia muciniphila (data not shown).

3.5.3 Organisms Identified or Hypothesized to be Involved in Lignan Metabolism

Genera identified or hypothesized to be involved in lignan metabolism were further investigated (Figure 19). Several of these genera were present in at levels above 1.5%. As aforementioned, Methanosphaera and Methanobrevibacter grew substantially in the 5100 incubation, and the latter was present at high levels in 6128 Day 7. Bifidobacterium grew in the 6105 incubation and was present in 5100 Day 1. Ruminococcus was abundant in all samples except 5100 Day 7. Eubacterium generally decreased with incubation, but was highly abundant in 6128 Day 7. Bacteroides was highly abundant in all samples, although less so in 6128 Day 7.

Several genera investigated were present at levels below 1.5%. Clostridium increased substantially in the 6105 incubation, and was also abundant in 6128 Day 7. Lactobacillus was
abundant in both samples from 5100. *Eggerthella* grew substantially in the 6105 incubation. *Streptococcus* abundance decreased in the 5100 incubation and was present in 5100 and 6128 Day 7. *Desulfovibrio* was present in 5100 Day 7.

Species identified or hypothesized to be capable of lignan metabolism were further investigated (Figure 20). Most notably, *Eubacterium limosum* grew substantially with incubation of stool from 5100. It was also present at high levels in the 6128 Day 7 sample. According to the V4 16S rRNA gene data, relative abundance of the family Eubacteriaceae increased by several orders of magnitude (.02 to 23%) from Day 1 to Day 7 for the 6128 incubation. Our V4 16S rRNA gene pipeline was unable to classify this phylotype at a higher resolution than family, but it is likely that it contains the same reads that MetaPhlAn classified as *E. limosum*. *Methanosphaera stadmanae* and *Methanobrevibacter smithii* accounted for the full abundance of their corresponding genera described above. *Ruminococcus bromii* was most abundant in both 6105 samples, but also notably present in 6128 Day 7 and 5100 Day 1. *Ruminococcus lactaris* was present in both 6105 samples and 6128 Day 7.

Several other species identified or hypothesized to be capable of lignan metabolism were present at levels below 1.5%. *Bifidobacterium longum*, *Bacteroides ovatus*, *Eggerthella lenta*, *Bifidobacterium pseudocatenulatum*, *Clostridium leptum*, *Bacteroides thetaiotaomicron*, and *Escherichia coli* grew substantially in the 6105 incubation. *Eggerthella lenta* and *Bifidobacterium pseudocatenulatum* were undetectable or at negligible levels in Day 1 samples, but present in all measured Day 7 *in vitro* samples. *Bifidobacterium catenulatum* and *Bacteroides fragilis* grew in 5100. *Clostridium scindens*, *Clostridium leptum*, *Bacteroides thetaiotaomicron* were present at notable levels in 6128 Day 7. *Desulfovibrio desulfuricans* was found at low levels in 5100 Day 7. Incubated samples from 6128 and 5100 showed increases in highly abundant (>1.5%) species of interest that were not present in the 6105 samples. 6105 showed the largest increase in low abundance (<1.5%) species of interest.

### 3.6 Metagenomic Functional Analysis (n=3)

#### 3.6.1 Pathways

Sequences in our samples mapped to a total of 7276 KOs and 231 pathways using HUMAnN. Of the KEGG pathways hypothesized to be involved in lignan metabolism searched for in
HUMAnN pathway output (Figure 21), methane metabolism (ko00680), starch and sucrose metabolism (ko00500), steroid hormone biosynthesis (ko00140), and primary bile acid biosynthesis (ko00120) were present in the samples analyzed. Methane metabolism and starch and sucrose metabolism were the most abundant of the pathways investigated. Methane metabolism was present in all samples and increased with incubation of the samples from participants 5100 and 6128. The 3 most abundant KOs in the Day 1 sample from these two participants and both samples from 6105 were 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (K01834), 6-phosphofructokinase 1 (K00850), and fructose-bisphosphate aldolase, class II (K01624). Starch and sucrose metabolism was present in all samples and decreased marginally with the 7-day incubation. Steroid hormone biosynthesis was present in all samples and increased slightly with the 7-day incubation. Primary bile acid synthesis was present in all samples except 5100 Day 7. This pathway increased with incubation of samples from 6105 and decreased with incubation of samples from 5100 and 6128.

The results of manually summed KOs for these pathways was almost identical, with two exceptions: 1) abundances for methane metabolism and starch and sucrose metabolism were at lower levels, but maintained roughly the same inter-sample proportions; 2) primary bile acid synthesis was present in all samples at roughly the same abundance. Although several KOs were found in the primary bile acid biosynthesis (ko00120) pathway, choolylglycine hydrolase (K01442) and alpha-methylacyl-CoA racemase (K01796) were the most abundant. The former was the only KO detected in the secondary bile acid biosynthesis pathway (ko00121).

### 3.6.2 Modules

Total abundance of all detected KOs in the WLP did not change substantially (Figure 22). However, acetyl-CoA synthase (K14138) and carbon-monoxide dehydrogenase subunits (K00198, K00197, K00194) increased considerably with the 7-day incubation of stool from 5100 and 6128, but decreased less markedly for participant 6105.

Only two KOs – 11beta-hydroxysteroid dehydrogenase (K00071) and cytochrome P450, family 17, subfamily A (K00512) – belonging to the C21-Steroid hormone biosynthesis module (M00109) were detected in our samples. K00071, which was substantially more abundant than K00512, was present in all samples, and increased with incubation of stool from 6105 (Figure 23).
Summed KOs of each module unique to the methane metabolism pathway all increased after the 7-day incubation of stool from 6128 and 5100 (Figure 24). 3 KOs – acetyl-CoA synthetase (K01895), acetate kinase (K00925), and phosphate acetyltransferase (K00625) – from M00357 were found in high abundance in all Day 1 samples and 6105 Day 7.

3.6.3 **KOs Corresponding to Enzymes Hypothesized to Act Directly on Lignans**

KOs corresponding to enzymes hypothesized to act directly on lignans were further investigated (Figure 25). KOs corresponding to beta-glucosidases were present at high levels in all samples, the most abundant being K05349. All beta-glucosidase KO abundances decreased in samples after the 7-day incubation. K01195 corresponding to beta-glucuronidase, was present in all samples. KOs corresponding to acetogen demethylases were present in all samples, and increased substantially with a 7-day incubation of stool from 6128 and 5100. 5-methyltetrahydrofolate-homocysteine methyltransferase (K00548) was the most abundant of these KOs. KOs corresponding to methanogen demethylases were present at high levels in 5100 and 6128 Day 7 samples. KOs corresponding to bile acid dehydrogenases, as a group, increased in 5100 and 6128 incubations, and decreased slightly in the 6105 incubation. 3alpha(or 20beta)-hydroxysteroid dehydrogenase (K00038) and 7-alpha-hydoxysteroid dehydrogenase (K00076) were the most abundant of this group.
Chapter 4. DISCUSSION

4.1 METABOLITE MEASUREMENTS AND EUBACTERIAL ENUMERATION

The fact that most of the inoculated ENL was recovered at Day 6, and no other metabolites were measured in this treatment, indicates that this is the last metabolite in the process of SDG metabolism. Lack of full recovery could be due to pipetting error or less-than-complete extraction in GCMS preparation, since measured ENL for this treatment remains relatively constant over the 6-day period.

Our incubations performed SDG to ENL conversion to very efficiently compared other studies with similar incubations, however the amount of SDG we inoculated with was substantially lower. The fecal inoculum concentration we used was chosen — based on our methods development — to maximize SDG to ENL conversion as described elsewhere. The negligible levels of SECO and END measured in incubations 6128 and 5100 could indicate more rapid or alternate routes of metabolism. For example DHEND can be metabolized to DHENL and then to ENL, bypassing END. The occurrence of this pathway is unknown in our experiments, since we did not measure DHENL. The close-to-full conversion to ENL in END-inoculated incubations for these participants shows that their stool is at least capable of converting END to ENL, yet their microbes could prefer an alternate route of metabolism when provided different precursors. The high conversion rate to ENL in the 6128 END-inoculated incubation, could indicate a faster metabolism as an explanation for lack of END observed in the 6128 SDG-inoculated incubation. Conversely, the 5100 incubation did not show this difference in rate.

4.2 TAXONOMIC ANALYSIS

MetaPhlAn appeared to outcompete MG-RAST in accurately producing sample phylotype abundance. There are probably several explanations for this. MetaPhlAn uses clade-specific markers instead of all protein reads. This helps account for gene transfer and misclassification of ubiquitous proteins. By using marker genes and normalizing by nucleotide length, MetaPhlAn eliminates the problem of variable genome size. MG-RAST does nothing to account for this,
which may explain the difference in abundance of sequences classified to Bacteroidetes, Firmicutes, and Euryarchaeota.

Abundance of Bacteroidetes, in fact, has been associated with a larger average genome size (AGS), while abundance of Firmicutes has been associated with a smaller AGS. Thus, failing to correct by genome size would probably yield an inflated Bacteroidetes:Firmicutes ratio as seen with MG-RAST output. More metabolically specialized organisms appear to have smaller genomes than generalist organisms. For example, *Methanobrevibacter smithii* has a much smaller genome than *Bacteroides thetaiotaomicron*. This could explain the reason Euryarchaeota had a lower abundance according to MG-RAST. Finally, MetaPhlAn uses nucleotides rather than amino acids for classification, which provides more taxonomic information, since several different nucleotide triplets can code for one amino acid.

OTU and species-level alpha diversity did not decrease markedly for most samples after a ~1 week incubation of stool with a mineral salts media. We had expected to observe a decrease in alpha diversity since the growth substrates available *in vitro* were substantially limited and less diverse compared to the colonic biome. The change in species-level alpha diversity is small when compared to a similar study in which human stool was incubated with crude oil and a dispersant. The fact that MetaPhlAn species-level alpha diversity was lower per sample than QIIME OTU-level alpha diversity may indicate that the 16S rRNA gene methods used provide a more diverse representation of a microbial community that the metagenomic methods used.

An increase in Bacteroidetes:Firmicutes ratio after incubation was observed in taxonomic classification of both the V4 16S rRNA gene and metagenomic data. This is similar to results of *in vitro* incubations with polyphenols from tea and a red wine/grape juice mix. In both studies, Bacteroidetes:Firmicutes ratio increased *in vitro* when media was supplemented with polyphenols. It is possible that the small amount of SDG in our *in vitro* incubations could drive this effect. Regardless, since a large number of SDG-deglycosylating bacteria fall in the phylum Bacteroidetes, and deglycosylation is necessary for the subsequent steps of SDG to ENL conversion, it follows that enriching for Bacteroidetes might promote SDG to ENL conversion.

Euryarchaeota increased in several of the incubations, but it has yet to be determined whether these microbes are involved in lignan metabolism. It is possible that by consuming hydrogen, they are driving fermentation reactions, and in turn promoting the growth of organisms capable of cleaving and utilizing glucose moieties from lignans.
Verrucomicrobia increased notably in several samples after incubation. According to metagenomic analysis, this phylum was almost primarily accounted for by *A. muciniphila*. *A. muciniphila* is a gram-negative, strictly anaerobic, mucin degrading bacteria isolated from human feces. Interestingly, dietary administration of polyphenol extracts from green tea, Concord grape, and cranberry increased cecal levels of *A. muciniphila* in mice. Similarly, *in vitro* incubation of human feces with polyphenol extracts from a red wine/grape juice combination or black tea increased relative abundance of *A. muciniphila*.

The mechanism by which polyphenols promote the growth of *A. muciniphila* is still undetermined. It may be due to an antimicrobial effect, which could reduce the viability of competitors and thus allow *A. muciniphila* to thrive. This is supported by the fact that the microbe has been shown to dominate in human feces following antibiotic administration. Another possibility is that polyphenols promote production of mucin, which is feeding *A. muciniphila*. This conflicts, however, with findings that mucin gene expression was not altered in mice fed Concord grape polyphenols, and would not explain the results observed *in vitro*. A third possibility is that since *A. muciniphila* is strictly anaerobic, and hence sensitive to free oxygen radicals, polyphenols could provide protection by scavenging these radicals. A final option is that polyphenols are a substrate utilized by *A. muciniphila*. *A. muciniphila* does possess glycosidases which could potentially act on glycosylated polyphenols, but this has yet to be investigated.

Organisms hypothesized to be involved in lignan metabolism, as a group, were not enriched after all incubations. Enrichment and presence of these organisms was sample-specific, with each sample containing some of the organisms of interest. Most samples contained at least one genus identified or hypothesized to be involved in each step of SDG to ENL metabolism. This is in agreement with the fact that SDG metabolism is carried out by a consortia of microbes and that each step can be performed by several different types of organisms. The observation that dehydrogenating organisms were found at lower levels than organisms capable of performing other reactions involved in lignan metabolism is in line with studies that have measured abundance of lignan metabolizing organisms in human stool. The fact that *Ruminococcus bromii* and *Ruminococcus lactaris* were found in our samples is interesting in light of a very recent study showing an association between the presence of these species in stool from participants supplemented with flaxseed, and ENL production.
*Bacteroides* prominently increased after most incubations. Though our metagenomic taxonomic classification of incubated samples from three participants did not reveal that this genus was primarily accounted for by species identified in lignan metabolism, it is possible that other species in this genus are involved as well. If this was the case, an increase in their abundance could be promoting SDG deglycosylation and ultimately helping to drive ENL production.

4.3 **FUNCTIONAL ANALYSIS**

There were some differences in HUMAnN pathway vs. KO-summed pathway output. The generally lower abundance of HUMAnN pathway output is likely due to the fact that it averages the most abundant genes in the pathway, instead of summing total abundance. The differences inter-sample ratios observed with primary bile acid synthesis might have been due to taxonomic limitation and/or copy number normalization.

KOs unique to methane metabolism were greatly enriched in 6128 and 5100 Day 7 samples, mirroring the presence of methanogens in these samples detected by our taxonomic analysis. It is possible that methanogens are demethylating SECO in these incubations. The fact that the methane metabolism pathway (ko00680) was present in samples 5100 and 6128 Day 1, and both samples from 6105, in which methanogens were not detected by taxonomic analysis, was surprising. Upon further investigation, the 3 most abundant KOs in these samples are KOs found in many other pathways, in addition to methane metabolism. All KOs corresponding to modules unique to methane metabolism were found at very low levels in the non-methanogen samples, with the exception of three KOs belonging to M00357. These KOs are also part of several other pathways. It can thus be concluded that our detection of ko00680 in non-methanogen containing samples is mainly due to ubiquitous KOs, and that methane metabolism is not occurring at notable levels in these samples.

The purpose of MinPath is essentially to eliminate pathways that are falsely detected due to ubiquitous KOs. Since the HUMAnN pathway output was highly similar to the manually KO-summed pathway output, it seems that MinPath failed to eliminate the correct pathways. Furthermore, HUMAnN eliminated modules that contained non-ubiquitous KOs; yet when these modules were assessed by summing KOs, they were found to be present. Finally, KOs corresponding to a module that was present in HUMAnN module output, were not present in
HUMAnN KO output. Future studies should be conducted with an awareness of these limitations.

Since several acetogens are capable of lignan metabolism, especially the demethylation step, enriching for these organisms may promote the conversion of SDG to ENL. Four KOs that correspond to the last step in the acetyl CoA pathway (Figure 26) were more abundant in the 6128 and 5100 Day 7 samples. K00194, K00197, and K00198 are also utilized in methanogenesis, while K14138 is not. Co-occurrence of these four specific KOs has been shown to be a good indicator of the presence of the WLP in microbes. This tracks well with the fact that \textit{E. limnosum} – an acetogenic bacterium – and methanogens were highly enriched in these samples, since methanogens and the acetogenic bacteria both utilize the WLP. Furthermore acetogens are a phylogenetically diverse group and it is likely that there are many uncharacterized acetogens in the human gut. Thus there may be undetected acetogens in these samples that are additionally accounting for the presence of the WLP.

Furthering this hypothesis, KOs corresponding to acetogenic demethylases increased substantially with incubation of samples from 6128 and 5100. The enzymes at the end of the WLP are those that receive the methyl group from these demethylases, thus the pathway for methoxylated aromatic demethylation appears to be enriched for (Figure 26).

KOs corresponding to beta-glucosidases and beta-glucuronidases were detected at high abundance in our samples and may be responsible for the SDG deglycosylation in our samples. No \textit{bai} genes were found in our investigation. Two abundant KOs involved in bile acid biosynthesis were detected. Choloylglycine hydrolase (K01442) is a hydrolase that cleaves carbon-nitrogen bonds other than peptide bonds. Alpha-methylacyl-CoA racemase (K01796) is a racemase that is involved in converting cholesterol to cholate and chenodeoxycholate. Though these enzymes probably would not act directly on lignans, their presence indicates that bile acid metabolizing microbes are present in our samples.

KOs corresponding to bile acid dehydrogenases were enriched in the 6128 and 5100 incubations, and a KO corresponding to a steroid hormone dehydrogenase was enriched in the 6105 incubation. It is possible that these dehydrogenases are acting to dehydrogenate END in our incubations.
Several different growth media have been used to culture enterolignan-producing microbes. In a study described previously, Brain Heart Infusion broth was used to culture feces from 100 different individuals, yielding variable degrees of enterolignan production. Mt Broth has also been used, yielding an average SECO to enterolignan (END+ENL) percent conversion of 59.9%. Human feces incubated with General Anaerobic Medium converted 90% of SDG to ENL. When MRS broth was used in vitro, microbes were only capable of metabolizing SDG to SECO and did not produce enterolignans. The efficient conversion of SDG to ENL observed in our incubations may be due in part to the ability of our media to promote the survival of lignan-metabolizing organisms.

The reagents in our media provided substrates necessary for acetogens and methanogens to thrive. Our media contained formate. Acetogens do not necessarily possess all enzymes in the WLP, and not all acetogens are capable of growing solely on H₂ and CO₂. Since formate is the intermediate after CO₂ in the WLP, it follows that several acetogens have been shown to grow well on formate, and addition of formate can drive acetogenesis. The acetogen Marvinbryantia formatexigens (isolated from human stool), does not harbor the selenium-containing subunit of formate dehydrogenase, and requires formate for acetogenesis. Addition of sodium formate to the culture media for this species has been shown to greatly increase its growth. Methanogens can also utilize formate, and addition of formate to methanogen culture media has been shown to promote methanogenesis.

Our media also contained acetate. Acetate is standardly used in enrichment media for methanogens, and can be entered into the incomplete reductive tricarboxylic acid (TCA) cycle for energy yield. Methanogens can also utilize ammonium and sulfur, both provided by ammonium sulfate in our media. Our incubations contained a substantial amount of methanol, which can be used as a substrate by acetogens, such as E. limnosum, and methanogens. The standard concentration of methanol used for culturing methanogens is 0.4%, while our media contained 1.5%. Trace metals provided in our media are important for the metalloenzymes that acetogens and methanogens possess. Glucose, contained in our media, can sustain both primary fermenters
and acetogens. The fermentation of glucose would generate CO2, H2, and short chain fatty acids, which could subsequently be utilized by acetogens and methanogens.\textsuperscript{69,151}

Acetogens and methanogens are considered obligate anaerobes.\textsuperscript{67,73} We ensured our media was anaerobic by adding an anaerobic indicator, sparging with nitrogen, and adding the reducing agent Cys-HCl. Both methanogens and acetogens thrive in a neutral pH\textsuperscript{67,73} and our media was brought to a pH of 7.3.

It is unclear why \textit{A. muciniphila} grew well in our incubations. Potentially the small amount of SDG, being a polyphenol, could promote its growth. Additionally, methanol could be exerting antimicrobial effects, and reducing competition.

### 4.5 Limitations

There are several limitations to our methods. For one, the sample size is small, and limits our ability to detect differences between treatments. However, this is part of an ongoing study and future analyses will be conducted with more samples.

Eubacterial 16S rRNA gene copy number estimated by qPCR changed less over time for incubations that contained high levels of methanogens. This is indicative of the fact that we are only measuring the eubacterial, and not the archaeal, 16S rRNA gene. Future studies should use primers that amplify both the archaeal and eubacterial 16S rRNA gene.

We have observed that the community from four different participants converted SDG to ENL \textit{in vitro}. We have also observed presence or enrichment of acetogens, methanogens, bile acid metabolizing microbes, and KOs corresponding to the WLP, methanogenesis, beta-glucosidases, beta-glucuronidases, demethylases, and bile acid and steroid dehydrogenases in these incubations. We may hypothesize that these organisms and pathways are responsible for SDG metabolism, but there is no firm evidence to conclude such. The community in our cultures was very large and it could be argued that the relative abundance of microbes involved in lignan metabolism actually decreases \textit{in vitro}, but remains at just a high enough level to enable conversion of SDG to ENL. Further steps should be taken to solidify understanding.

Isolation of microbes and microbial enzymes, and testing for lignan-metabolizing capability would provide the most concrete answers. Organismal isolation could be carried out as described in Dr. Clavel’s thesis.\textsuperscript{137} If investigation of acetogens or methanogens was desired, substrates used for their enrichment could be added to plates and media. For acetogens one might add
methoxylated aromatic compounds or formate.\textsuperscript{73} Probably the most relevant and useful methoxylated aromatic to add would be SECO. For methanogens one might add methanol, formate, acetate, CO\textsubscript{2}, and H\textsubscript{2}.\textsuperscript{67} Methane, acetate, H\textsubscript{2}, and CO\textsubscript{2} in serum bottles could be measured over time to further characterize hydrogenotrophic metabolism.

Putative lignan-metabolizing enzyme genes could be amplified from identified lignan-converting microbes and subsequently cloned, expressed, and characterized as in Tao et al.\textsuperscript{91} Furthermore, isolates belonging to groups of microbes hypothesized to be involved in lignan metabolism could be ordered from a supplier and subsequently tested for lignan-degrading capabilities.

It may also be advantageous to inoculate media with enough SDG (or other lignans) to potentially promote growth of microbes that are using it as a substrate. The concentration of SDG (6.55 µM) accounts for less than 0.01\% of the total carbon in our media. In addition, hydrogenotrophic metabolism is low-energy yielding.\textsuperscript{67,73} For these reasons it is unlikely that 6.55 µM SDG would be enough to serve as a substantial food source. It may be difficult, however, to add enough lignan substrate to promote growth. For example, in one study, compounds demethylated by acetogenic bacteria had no effect on growth.\textsuperscript{152} This may have been due to several factors, including less ATP generated from metabolism than is required for uptake of the compound or disposal of metabolites. Also, higher levels of SDG might prove toxic to microbes. Flaxseed-cake extract with a concentration of 364 µM SDG was shown to slow growth of \textit{Staphylococcus aureus} in Luria Broth.\textsuperscript{153}

Incubation in media that does not support SDG to ENL conversion might generate useful insights. Microbial populations grown in this media could be compared to populations grown in TCAP2, with the idea that TCAP2-grown communities would be richer in lignan-metabolizers. Differences in phenotype, however, could be gene expression-mediated, in which case taxonomy and functional gene analysis would provide little insight.

This precipitates another limitation and future direction. Changes in functional gene abundance may belie actual changes in function, in that many of the genes measured metagenomically are not expressed. Adding a metatranscriptomic component would give a much clearer picture of actual functions at work. It may be expected that metatranscriptomic changes would be more transient, thus it would be critical to identify time points of interest for functional characterization.
Subsequently, we were only comparing initial microbial community to Day 7, but there were likely changes in community structure throughout the incubation. It also may be the case that DNA sequenced from the pelleted fecal suspension originated from microbes that died early in the incubation. Future studies should pellet fecal suspensions on several days to capture the transformation of the microbial community.

Several taxa in addition to those identified or hypothesized may be involved in lignan metabolism. By focusing our analyses on specific taxa, we may have missed important phylogenetically related organisms. Furthermore there are many microbial genomes that have not been fully sequenced. These microbes would not have been identified by our methods. Similarly we may have missed functional genes that are closely related to the KOs searched for, but have a different KO numbers or are unclassified. Analyzing our data using phylogenetic trees would help to remedy these problems.

Using KOs as a gene identifier also poses problems. A single KO can be found in many different enzymes and pathways. For example K00548, which is part of an acetogenic demethylase, is also used in several pathways including, cysteine and methionine metabolism, selenocompound metabolism, one carbon pool by folate, and biosynthesis of amino acids58 and is found in 67% of gut microbes.154

4.6 Conclusion

In summary, SDG was almost fully converted to ENL in our incubations. In most incubations, Bacteroidetes:Firmicutes ratio increased; methanogens, deglycosylators – especially Bacteroides – and demethylators were enriched; and KOs corresponding to beta-glucosidases were present at high levels. Several phenomena were true of some incubations. Eubacterium limnosum, Akkermansia muciniphila, methane metabolism, and the WLP were highly enriched. KOs corresponding to demethylases, dehydroxylases and dehydrogenases were enriched. The acetogenic demethylation pathway may have been enriched. To conclude, organisms and functional genes identified or hypothesized to be involved in lignan metabolism were not unanimously enriched, but sample-specific. These findings may shed light on organisms, enzymes, and biochemical pathways involved in lignan metabolism. Future studies with larger sample sizes may help to confirm these results.
REFERENCES


Table 1. Participant study number, gender, age, whether qPCR analysis was conducted, time points of aliquots from fecal suspensions that were used for GCMS analysis, time points (Day 1 initial stool and final-day pelleted SDG-inoculated fecal suspension) from which DNA was extracted for sequencing, whether or not DNA was sequenced using 16S rRNA gene sequencing and/or shotgun metagenomic sequencing.

<table>
<thead>
<tr>
<th>Participant #</th>
<th>Gender</th>
<th>Age</th>
<th>qPCR</th>
<th>GCMS timepoints</th>
<th>DNA extraction timepoints</th>
<th>16S rRNA gene sequencing</th>
<th>Shotgun metagenomic sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>5100</td>
<td>Female</td>
<td>50.3</td>
<td>yes</td>
<td>Days 1 - 6</td>
<td>Days 1 &amp; 7</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>6105</td>
<td>Female</td>
<td>41.77</td>
<td>yes</td>
<td>Days 1 - 7</td>
<td>Days 1 &amp; 7</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>6127</td>
<td>Male</td>
<td>40.2</td>
<td>no</td>
<td>Days 1 - 6</td>
<td>Days 1 &amp; 7</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>6128</td>
<td>Female</td>
<td>25.55</td>
<td>yes</td>
<td>Days 1 - 6</td>
<td>Days 1 &amp; 7</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6129</td>
<td>Male</td>
<td>30.89</td>
<td>yes</td>
<td>Days 1 - 6</td>
<td>Days 1 &amp; 7</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>6171</td>
<td>Female</td>
<td>43.63</td>
<td>no</td>
<td>Days 1 &amp; 6</td>
<td>Days 1 &amp; 6</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>6178</td>
<td>Male</td>
<td>25.01</td>
<td>no</td>
<td>Days 1 &amp; 6</td>
<td>Days 1 &amp; 6</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>6188</td>
<td>Female</td>
<td>28.27</td>
<td>no</td>
<td>Days 1 &amp; 6</td>
<td>Days 1 &amp; 6</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>
### Table 2. TCAP2 Media Components

<table>
<thead>
<tr>
<th>TCAP2</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Cysteine-HCl (Aldrich)</td>
<td>0.25</td>
</tr>
<tr>
<td>Glucose (Sigma)</td>
<td>0.9</td>
</tr>
<tr>
<td>Resazurin (1mg/L) (Aldrich)</td>
<td>1</td>
</tr>
<tr>
<td>NaCl (Sigma)</td>
<td>2.78</td>
</tr>
<tr>
<td>K₂HPO₄ (Sigma)</td>
<td>0.42</td>
</tr>
<tr>
<td>Na₂HPO₄·7H₂O (Fisher)</td>
<td>3.3</td>
</tr>
<tr>
<td>NaH₂PO₄·2H₂O (Fluka)</td>
<td>1.2</td>
</tr>
<tr>
<td>(NH₄)₂SO₄ (Fisher)</td>
<td>1.2</td>
</tr>
<tr>
<td>MgCl₂·6H₂O (Sigma)</td>
<td>0.08</td>
</tr>
<tr>
<td>KH₂PO₄ (Fisher)</td>
<td>0.16</td>
</tr>
<tr>
<td>CH₃COONa (Fisher)</td>
<td>0.5</td>
</tr>
<tr>
<td>HCOONa (Fluka)</td>
<td>0.5</td>
</tr>
<tr>
<td>Trace element solution</td>
<td>10 mL</td>
</tr>
<tr>
<td>Vitamin solution</td>
<td>1 mL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trace Element Solution</th>
<th>mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>MnSO₄·H₂O (Fisher)</td>
<td>1159</td>
</tr>
<tr>
<td>FeSO₄·7H₂O (Fisher)</td>
<td>3680</td>
</tr>
<tr>
<td>CoCl₂·6H₂O (MP Biomedicals)</td>
<td>120</td>
</tr>
<tr>
<td>ZnSO₄·7H₂O (Fisher)</td>
<td>440</td>
</tr>
<tr>
<td>CuSO₄·5H₂O (Fisher)</td>
<td>98</td>
</tr>
<tr>
<td>(NH₄)₆Mo₇O₂₄·4H₂O (Fluka)</td>
<td>17.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vitamin Solution</th>
<th>mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menadione (MP Biomedicals)</td>
<td>1</td>
</tr>
<tr>
<td>Biotin (Jackson Immuno Research)</td>
<td>2</td>
</tr>
<tr>
<td>Pantothenate (MP Biomedicals)</td>
<td>10</td>
</tr>
<tr>
<td>Nicotinamide (Acros)</td>
<td>5</td>
</tr>
<tr>
<td>Vitamin B₁₂ (unknown)</td>
<td>0.5</td>
</tr>
<tr>
<td>Thiamine (MP Biomedicals)</td>
<td>4</td>
</tr>
<tr>
<td>p-Aminobenzoic acid (MP Biomedicals)</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 1. SDG to SECO, END, and ENL percent conversion measured by GCMS from daily aliquots of SDG-inoculated fecal suspension over a 7-day incubation of stool from participant 6105.
Table 3.  SDG-metabolizing bacteria, acetogenic bacteria, methanogenic archaea, bile acid dehydroxylators, bile acid dehydrogenators, sulfate-reducing bacteria, and estrone to estradiol converters searched for. Search terms, using custom R-scripts in QIIME and MetaPhlAn output, were genera and full species names when available. Exceptions to these search terms were “desul” for sulfate-reducers and “formaxigenis”, for Marvinbryantia formaxigenis, since its genus had been renamed.

<table>
<thead>
<tr>
<th>SDG-metabolizing bacteria</th>
<th>Acetogenic bacteria</th>
<th>Methanogens</th>
<th>Bile acid dehydroxylators</th>
<th>Bile acid dehydrogenators</th>
<th>Sulfate-reducing bacteria</th>
<th>Estrone to estradiol converters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morvella</td>
<td>Clostridium cocoides</td>
<td>Methanobrevibacter smithii</td>
<td>Bacteroides</td>
<td>Eggerthella lenta</td>
<td>Desulfo bacter</td>
<td>Akliligenes faecalis</td>
</tr>
<tr>
<td>Acetanaerobacterium, Blautia hydrogenotrophica</td>
<td>Methanothermobacter stadiumae</td>
<td>Methanobacterium</td>
<td>Eubacterium</td>
<td>Clostridium perfringens</td>
<td>Desalomonas</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>Fastidiosigella Marvinbryantia formaxigenis</td>
<td>Methanomethylophilus alvus</td>
<td>Candidatus</td>
<td>Eggerthella</td>
<td>Ruminococcus productus</td>
<td>Desulfofotomaculum</td>
<td>Staphylococcus aureus</td>
</tr>
<tr>
<td>Streptococcus Ruminococcus</td>
<td>Methanomasilitllicoccus intestinalis</td>
<td>Candidatus</td>
<td>Clostridium scindens</td>
<td>Bacteroides fragilis</td>
<td>Desulfovibrio</td>
<td>Streptococcus faecalis</td>
</tr>
<tr>
<td>Lactobacillus CS1 Van</td>
<td>Methanomasilitllicoccus luminyensis</td>
<td>Clostridium hiranonis</td>
<td>Bacteroides thetatoaminimon</td>
<td>Clostridium innocuim</td>
<td>Desullobalbus</td>
<td></td>
</tr>
<tr>
<td>Enterococcus CS3Glu</td>
<td>Clostridium hylemonae</td>
<td>Clostridium ahbom</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroides ovatus, HA</td>
<td>Clostridium sorbeli</td>
<td>Clostridium</td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium cocleatum S5a2h</td>
<td>Clostridium scindens</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium saccharogumia VK64i</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium ramoxus IS2</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteroides distasonis</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium bifidum</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium breve</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium catenulatum</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium longum</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium pseudocatenulatum</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butyribacterium methylotrophicum</td>
<td>Eubacterium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eubacterium callanderi</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eubacterium limosum</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blautia producta</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggerthella lenta</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium scindens</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridium scindens</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacnaviril factor longisiformis</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eubacterium sp. ARC-2</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggerthella sp. SDG-2</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clostridiaceae</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruminococcus</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterococcus faecalis</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruminococcus bromii</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruminococcus lactaris</td>
<td>Clostridium</td>
<td></td>
<td>Clostridium</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. KEGG IDs corresponding to KOs, modules, and pathways searched for in HUMAnN output. Modules and pathways that are starred were investigated as summed component KOs (as defined on the KEGG website: [http://www.genome.jp/kegg/](http://www.genome.jp/kegg/))

<table>
<thead>
<tr>
<th>Enzyme, Module or Pathway</th>
<th>KEGG ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-glucosidases</td>
<td>K01188, K05540, K05549</td>
</tr>
<tr>
<td>β-glucoronidases</td>
<td>K01195</td>
</tr>
<tr>
<td>demethylases (acetogenic bacteria)</td>
<td>K00548, K15023, K14084, K14082</td>
</tr>
<tr>
<td>demethylases (methanogen - methanol)</td>
<td>K14080, K04480, K14081</td>
</tr>
<tr>
<td>corticosteroid 21-dehydroxylase</td>
<td>K00513</td>
</tr>
<tr>
<td>bile acid hydroxysteroid dehydrogenases</td>
<td>K05296, K00089, K00037, K04119, K00076, K05295, K00038, K1548, K05883</td>
</tr>
<tr>
<td>Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)</td>
<td>M00377*</td>
</tr>
<tr>
<td>C21-Steroid hormone biosynthesis, progesterone =&gt; cortisol/cortisone</td>
<td>M00108*</td>
</tr>
<tr>
<td>C21-Steroid hormone biosynthesis</td>
<td>M00109*</td>
</tr>
<tr>
<td>Methane oxidation, methanotroph, methane =&gt; formaldehyde</td>
<td>M00174*</td>
</tr>
<tr>
<td>Formaldehyde assimilation, xylulose monophosphate pathway</td>
<td>M00344</td>
</tr>
<tr>
<td>Formaldehyde assimilation, ribulose monophosphate pathway</td>
<td>M00345</td>
</tr>
<tr>
<td>Pyruvate dehydrogenase, serine pathway</td>
<td>M00346</td>
</tr>
<tr>
<td>Methanogenesis, methanol =&gt; methane</td>
<td>M00356*</td>
</tr>
<tr>
<td>Methanogenesis, acetate =&gt; methane</td>
<td>M00357*</td>
</tr>
<tr>
<td>Methanogenesis, methylamine/triethylamine =&gt; methane</td>
<td>M00563*</td>
</tr>
<tr>
<td>Methanogenesis, CO2 =&gt; methane</td>
<td>M00567*</td>
</tr>
<tr>
<td>2-Oxocarboxylic acid chain extension, 2-oxoglutarate =&gt; 2-oxoadipate =&gt; 1-oxopimelate =&gt; 2-oxosuberate</td>
<td>M00608</td>
</tr>
<tr>
<td>Methane metabolism</td>
<td>ko00680*</td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td>ko00500*</td>
</tr>
<tr>
<td>Secondary bile acid biosynthesis</td>
<td>ko00121*</td>
</tr>
<tr>
<td>Steroid hormone biosynthesis</td>
<td>ko00140*</td>
</tr>
<tr>
<td>Primary bile acid biosynthesis</td>
<td>ko00120*</td>
</tr>
</tbody>
</table>
Figure 2. Relative abundance of phyla in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspensions from 3 different participants using MetaPhlAn or MG-RAST for classification.
Figure 3. Comparison of dominant phyla relative abundance for the same sample (6128 Day 7) using 3 different sequencing method/classification tool combinations: V4 16S rRNA gene amplicon/in-house QIIME pipeline, metagenomic shotgun/MetaPhlAn, metagenomic shotgun/MG-RAST.
Figure 4. Relative abundance of modules hypothesized to be involved in lignan metabolism in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants calculated using DIAMOND and HUMAnN module output.
Figure 5. 16S rRNA gene copy number concentration estimated by qPCR from daily aliquots of fecal suspension over a 7-day SDG-inoculated incubation for four different participants.
Figure 6. A) SDG to SECO, END, and ENL percent conversion measured by GCMS from daily aliquots of fecal suspension over 6 days of SDG-inoculated incubations from five participants. B) SDG to ENL percent conversion measured by GCMS from Day 6 fecal suspension of SDG-inoculated incubation from three remaining participants.
Figure 7. END to ENL percent conversion measured by GCMS from daily aliquots of fecal suspension over 6 days of END-inoculated incubations from four participants.
Figure 8. Moles of SECO, END, and ENL (measured by GCMS) divided by 16 rRNA gene copy number (measured by qPCR) from daily aliquots of fecal suspensions over 6 days of SDG-inoculated incubations from four participants.
Table 5. **A)** QC statistics of 16S rRNA gene amplicon and whole metagenome shotgun sequenced DNA extracted from Day 1 initial stool and final-day pelleted SDG-inoculated fecal suspension from 6 different participants. **B)** Post-QC statistics of 16S rRNA gene amplicon DNA from remaining 19 samples that were NOT USED in our analyses.

<table>
<thead>
<tr>
<th>Participant #</th>
<th>DNA extraction timepoint (Day #)</th>
<th>16S rRNA After QC</th>
<th>Metagenomic QC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of sequences</td>
<td>Length in basepairs (avg+/8 stdev)</td>
<td>MG-RAST ID</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Before QC</td>
</tr>
<tr>
<td>5100</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6105</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6127</td>
<td>1</td>
<td>64,146</td>
<td>272 ± 8</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>50,350</td>
<td>273 ± 9</td>
</tr>
<tr>
<td>6128</td>
<td>1</td>
<td>101,936</td>
<td>272 ± 9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>65,188</td>
<td>273 ± 10</td>
</tr>
<tr>
<td>6129</td>
<td>1</td>
<td>68,921</td>
<td>272 ± 11</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>38,966</td>
<td>273 ± 11</td>
</tr>
<tr>
<td>6171</td>
<td>1</td>
<td>55,137</td>
<td>272 ± 9</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>63,333</td>
<td>273 ± 9</td>
</tr>
<tr>
<td>6178</td>
<td>1</td>
<td>43,658</td>
<td>272 ± 8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>60,864</td>
<td>273 ± 9</td>
</tr>
<tr>
<td>6188</td>
<td>1</td>
<td>68,144</td>
<td>272 ± 10</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>68,742</td>
<td>272 ± 8</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>sample name</th>
<th>number of sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFX.6129.END</td>
<td>63182</td>
</tr>
<tr>
<td>FFX.6135.END</td>
<td>58114</td>
</tr>
<tr>
<td>FFX.6135.stool</td>
<td>40963</td>
</tr>
<tr>
<td>FFX.6146.END</td>
<td>53835</td>
</tr>
<tr>
<td>FFX.6146.stool</td>
<td>65898</td>
</tr>
<tr>
<td>FFX.6155.END</td>
<td>87469</td>
</tr>
<tr>
<td>FFX.6155.stool</td>
<td>85840</td>
</tr>
<tr>
<td>FFX.6156.END</td>
<td>88464</td>
</tr>
<tr>
<td>FFX.6156.stool</td>
<td>53580</td>
</tr>
<tr>
<td>FFX.6171.END</td>
<td>55152</td>
</tr>
<tr>
<td>FFX.6171.END.TR</td>
<td>53100</td>
</tr>
<tr>
<td>FFX.6171.SDG.M</td>
<td>67507</td>
</tr>
<tr>
<td>FFX.6178.END</td>
<td>49154</td>
</tr>
<tr>
<td>FFX.6178.SDG.M</td>
<td>49913</td>
</tr>
<tr>
<td>FFX.6188.END</td>
<td>50671</td>
</tr>
<tr>
<td>FFX.6188.SDG.M</td>
<td>52383</td>
</tr>
<tr>
<td>FFX.FHC.A1</td>
<td>51383</td>
</tr>
<tr>
<td>FFX.FHC.A2</td>
<td>57573</td>
</tr>
<tr>
<td>FFX.FHC.A3</td>
<td>79335</td>
</tr>
</tbody>
</table>
Figure 9. Alpha diversity (Shannon Index) of OTUs calculated using an in-house QIIME pipeline and derived from 16S rRNA amplicon sequences of DNA extracted from Day 1 initial stool and Day 6 or 7 pelleted SDG-inoculated fecal suspension from 6 different participants.
Figure 10. Microbial composition (phyla) of Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house QIIME pipeline classified V4 16S rRNA sequences.

Figure 11. Euryarchaeota relative abundance in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from participants 6129, 6178, 6188 calculated using in-house QIIME pipeline classified V4 16S rRNA sequences.
Figure 12. Relative abundance of genera from the phylum Bacteroidetes in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house QIIME pipeline classified V4 16S rRNA sequences.
Figure 13. Relative abundance of genera from the phylum Firmicutes in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house pipeline classified V4 16S rRNA sequences.
**Figure 14.** Relative abundance of genera from the phylum Euryarchaeota in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house QIIME pipeline classified V4 16S rRNA sequences.

**Figure 15.** Relative abundance of genera from the phylum Verrucomicrobia in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house QIIME pipeline classified V4 16S rRNA sequences.
Table 6. Relative abundance of genera that were identified or hypothesized to be involved in lignan metabolism in Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house pipeline classified V4 16S rRNA sequences.

<table>
<thead>
<tr>
<th>Genus</th>
<th>6127 Day 1</th>
<th>6128 Day 7</th>
<th>6129 Day 1</th>
<th>6129 Day 7</th>
<th>6171 Day 1</th>
<th>6178 Day 1</th>
<th>6178 Day 6</th>
<th>6188 Day 1</th>
<th>6188 Day 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteroides</td>
<td>5.45</td>
<td>20.44</td>
<td>13.25</td>
<td>14.01</td>
<td>24.35</td>
<td>39.90</td>
<td>11.73</td>
<td>27.01</td>
<td>5.34</td>
</tr>
<tr>
<td>Clostridium</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
<td>0.00</td>
<td>0.03</td>
<td>0.23</td>
<td>1.03</td>
<td>0.88</td>
</tr>
<tr>
<td>Bifidobacterium</td>
<td>15.99</td>
<td>18.86</td>
<td>1.51</td>
<td>4.22</td>
<td>4.69</td>
<td>8.22</td>
<td>6.39</td>
<td>1.39</td>
<td>1.67</td>
</tr>
<tr>
<td>Blautia</td>
<td>3.62</td>
<td>0.42</td>
<td>4.07</td>
<td>0.39</td>
<td>1.18</td>
<td>0.15</td>
<td>0.99</td>
<td>1.05</td>
<td>0.82</td>
</tr>
<tr>
<td>Eggerthella</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Ruminococcus</td>
<td>0.32</td>
<td>0.07</td>
<td>3.82</td>
<td>0.46</td>
<td>1.36</td>
<td>0.52</td>
<td>1.29</td>
<td>1.37</td>
<td>2.05</td>
</tr>
<tr>
<td>Marvinbryantia</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Thermoanaerobacterales;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methanobrevibacter</td>
<td>0.34</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Methanosphaera</td>
<td>1.17</td>
<td>4.20</td>
<td>0.10</td>
<td>27.74</td>
<td>0.02</td>
<td>0.10</td>
<td>0.67</td>
<td>8.28</td>
<td>0.00</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>0.08</td>
<td>0.33</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Desulfovibrionaceae;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>14.52</td>
<td>0.03</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Bilophila</td>
<td>0.06</td>
<td>0.63</td>
<td>0.18</td>
<td>1.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Desulfovibrio</td>
<td>0.66</td>
<td>0.16</td>
<td>0.04</td>
<td>0.04</td>
<td>0.00</td>
<td>0.18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Figure 16. Relative abundance of genera that were identified or hypothesized to be involved in lignan metabolism in Day 1 initial stool and final-day pelleted SDG-incubated fecal suspension from 6 different participants calculated using in-house pipeline classified V4 16S rRNA sequences: A) grouped by sample B) grouped by sample and identified or hypothesized lignan metabolizing capability.
Figure 17. Species-level alpha diversity (Shannon index) calculated with the VEGAN R package from metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants based on MetaPhlAn output.
Figure 18. Phyla and top 10 genera relative abundance in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using MetaPhlAn for classification.
Figure 19. Relative abundance of dominant (>1.5%) and rare (<1.5%) genera identified or hypothesized to be involved in lignan metabolism in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using MetaPhlAn for classification.
Figure 20. Relative abundance of A) dominant (>1.5%) and B) rare (<1.5%) species identified or hypothesized to be involved in lignan metabolism in metagenomic shotgun-sequenced DNA samples extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using MetaPhlAn for classification.
Figure 21. Relative abundance of pathways hypothesized to be involved in lignan metabolism derived from metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspensions from 3 different participants using DIAMOND and HUMAnN for functional assignment: A) derived from HUMAnN pathway output B) derived from HUMAnN KO output (Note: KOs corresponding to ko00680 and ko00500 not listed because too numerous – over 80 per pathway detected in our samples).
Figure 22. Relative abundance of component KOs in the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) - M00377 - in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using DIAMOND and HUMAnN for functional assignment.
**Figure 23.** Relative abundance of component KOs of C21-Steroid hormone biosynthesis (progesterone => cortisol/cortisone) - M00109 - in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using DIAMOND and HUMAnN for functional assignment.
Figure 24. Relative abundance of KOs corresponding to modules unique to the methane metabolism pathway (ko00680) in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using DIAMOND and HUMAnN for functional assignment.
Figure 25. Relative abundance of KOs hypothesized to be involved in lignan metabolism in metagenomic shotgun-sequenced DNA extracted from Day 1 initial stool and Day 7 pelleted SDG-incubated fecal suspension from 3 different participants using DIAMOND and HUMAnN for functional assignment.
Figure 26. The demethylation of a methoxylated aromatic compound by an O-demethylase (K00548) belonging to *M. thermoacetica*. The methyl group is shunted into the Wood-Ljungdahl pathway. Specifically, it is transferred to the 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase (K15023) and then to CO dehydrogenase/acetyl-CoA synthase, which forms acetyl-CoA. For incubations from 6128 and 5100, KOs highlighted in red had increased relative abundance in Day 7 pelleted SDG-incubated fecal suspension, when compared to Day 1 initial stool.