Identification of Metabolites that May Reinforce Colonization Resistance Against Infection in an Antibiotic Augmented Gut

A thesis submitted by

Carol Chan

in partial fulfillment of the requirements

for the degree of

Master of Science

in

Bioengineering

Tufts University

May 2017

Adviser:
Kyongbum Lee, PhD

Abstract

Under homeostatic conditions, the commensal microbes of the microbiota protect themselves and their host against enteric pathogens through colonization resistance. Colonization resistance comprises of three key defense mechanisms: microbial release of antimicrobials, crowding out through competition for nutrients and/or through the microbial release of molecules to stimulate the host immune response. When antibiotics are applied to the microbiota there is a decrease in microbial diversity, weakening the colonization resistance and leaving the host susceptible to pathogenic invasion, infection, and disease. To assess the host and microbial activity under antibiotic stress, preliminary metabolic profiles of differently antibiotic-conditioned mice were generated using untargeted LC-MS/MS detection and a novel annotation pipeline. Analysis of the metabolic profiles identified several metabolites that may reinforce colonization resistance to deter pathogenic establishment in the gut during an antibiotic-stressed state.

Acknowledgements

The success of this thesis project would not have been possible without the support and collaboration from the many great minds at Tufts University. I would first like to acknowledge my advisor, Kyongbum Lee, for his guidance, patience and for imparting his expansive knowledge on this project. Secondly, I would like to thank my lab mates Sara Manteiga, Smitha Krishnan, and Nicholas Alden for their indispensable help with all matters LC-MS related and for their collaboration to make this project feasible. Next, I would like to thank Soha Hassoun and Emmanuel Tzanakakis for taking the time to be a part of my thesis committee. And, I would like to thank the faculty and staff of the Chemical and Biological Engineering Department at Tufts University for making my studies at this institution rigorous yet enjoyable.

I would also like to recognize my family and friends for their unconditional support through my academic career at Tufts as well as through all my career endeavors. Specifically, I would like to thank my sister, Helen, for ensuring I was not malnourished for the past two and a half years. And, thanks Amy Hunter, for keeping me sane during my time at Tufts. Lastly, I would like to dedicate this work in memory of my mother, who inspired me to pursue a career in the sciences.

Table of Contents

List of Equations:

List of Tables:

List of Figures:

1.0 Introduction

1.1 Role of the Microbiota in Host Health

From birth, an individuals' microbiome - the collection of commensal microorganisms that inhabit the human body - begins to form. Over time, environmental factors, diet, genetics, and modulators morph and define the dynamic yet symbiotic ecosystem of microbes.[1] Roughly equaling the number of cells in the human body, these resident microbes are essential to the development and maintenance of human health.[2,3]

Of the estimated 10^{13} microbes inhabiting the average human body, approximately 70 % reside in the gastrointestinal tract.[3,4] In healthy individuals, the gut is predominantly comprised of *Firmicutes* and *Bacteroidetes* with smaller contributions from *Proteobacteria*, *Actinobacteria*, *Verrucomicrobia* and *Cyanobacteria*.[5,6] Contributing to maintenance of the host's health the intestinal microbiota perform key functions of bile acid metabolism, vitamin synthesis, digestion and fermentation of otherwise non-digestible carbohydrates and proteins, and immune response stimulation.[1,4,7]

Similar to any ecological system, the human body is capable of self-regulation. Under small perturbations, such as minor bouts of sickness, the body and its associated microbes are capable of managing the imbalance in their ecosystem, ultimately returning the system back to a stable state. However, more substantial changes to the host's microbiome through major life events or the use of modulators may lead to unrecoverable alterations in the commensal microbe population.[8,9] This unfavorable change to the host's ecology is termed dysbiosis and is commonly associated with prolonged sickness and disease.

1.2 Defense Against Enteric Pathogens

With minor disturbances to the host's microbiota, resident microbes defend their territory using three colonization resistance defense mechanisms to deter enteric pathogens.[10,11] The first line of defense against invading pathogens are antimicrobial effector molecules, such as bacteriocins and metabolic byproducts, that directly inhibit the intruders.[11] If the pathogens are able to evade direct inhibition, the resident microbes of the gut will deter the pathogens by competing for nutrients, essentially crowding out pathogens by leeching all the available nutrients needed to survive.[11] If the pathogens continue to persist in the gut, the microbiota will indirectly inhibit the intruders by releasing microbial products, such as lipopolysaccharides and peptidoglycans.[11] The release of these microbial products will spread and be sensed by the host's epithelial cells, triggering the host's immune response to further combat the pathogenic microbes.

1.3 Dramatic Shifts in the Microbiota

Following a more severe or sustained disturbance, the microbiota may be slow to return to the original stable state or may transition to a state of dysbiosis.[12,13] One common treatment for enteric infections is to administer antibiotics. Most of these compounds achieve their antibiotic effect by inhibiting either cell wall synthesis or protein synthesis, and thus can affect a broad swath of bacterial species, whether they are pathogenic or beneficial to the host. Many studies have demonstrated how antibiotics can alter the microbial community by drastically decreasing the diversity of the microbiota and thus reducing the colonization resistance.[9,15,16] With a reduction in the diversity of the microbiota, there is a subsequent change in the microbial metabolic activity in the gut. Recently, Theriot et al. found decreases in secondary bile acids,

free fatty acids, glucose, and dipeptides with increases in primary bile acids and sugar alcohols in mice treated with antibiotics.[12] The decrease in microbial competition in the gut and increase in available substrates can fuel pathogenic takeover of the intestinal environment and lead to further progression and/or persistence of disease.

1.4 Infectious Diseases – *Clostridium difficile*

As the causative agent in most cases of infectious post-antibiotic colitis, *Clostridium difficile* research has become of great interest. *Clostridium difficile* is an anaerobic gram-positive, spore-forming, toxin-producing bacterium belonging to the phylum *Firmicutes* and class *Clostridia*. It is an opportunistic pathogen that can exist in a dormant state within a healthy intestinal microbiota. But, when presented with the right environmental conditions this pathogen will germinate and proliferate while releasing virulence toxins, TcdA and TcdB, causing a range of clinical pathologies in the host.[17] Infection with *C. difficile* can present as pseudomembranous colitis (inflammation of the colon), or in more severe cases, can lead to toxic megacolon and/or death.[18] To gain a better understanding of the disease development, studies have examined the spore germination and expansion of *C. difficile* within the gut after antibiotic treatment. Findings demonstrated that primary bile acids, such as cholate and taurochenodeoxycholate, promote the germination of *C. difficile* spores into actively growing vegetative cells.[19] Normally, these primary bile acids would be transformed into growth-inhibiting secondary bile acids by the commensal microbes residing in the intestine. However, exposure to high doses of broad-spectrum antibiotics can eliminate a significant fraction of the commensal microbes that are responsible for deconjugating and dehydroxylating primary bile acids into secondary bile acids. This renders the indigenous intestinal microbiota vulnerable to the overgrowth of *C. difficile*.

1.5 Studying Mechanistic Interactions Through Metabolic Profiling

In the last decade, advances in sequencing and mass spectrometry technologies have spurred a paradigm shift in the study of host-microbiota interactions, from the classical culture-based experimentation to culture-independent methods.[10,20] Although isolation and culture of individual strains have been very useful for physiological and molecular characterization of specific bacteria under well-defined conditions, this approach has limitations when studying complex bacterial communities. First, not every commensal bacterium can be cultured individually under laboratory conditions. Second, identifying an appropriate culture condition to mimic the *in vivo* environment is difficult due to the complex and as yet incompletely characterized milieu of host and microbial factor present in the intestine; consequently, studies use simplified conditions that often lead to skewed bacterial communities that do not resemble the intestinal community.[20]

In contrast, culture-independent techniques involving animal models afford observations on intact bacterial communities. In conjunction with various '-omic' approaches, these culture-independent methods have been used to study bacterial functions and interactions with the host. These methods have featured prominence in two large efforts, the Human Microbiome Project and the European MetaHIT initiative, which are aimed at defining bacterial communities in different individual populations. A key scientific goal of these efforts is to link the detected composition of a bacterial community with the community's functions, often using metagenomics.[21,22] Metagenomics has been very useful in addressing the question of 'who is there' and estimating the gene functions encoded in a bacterial community. However, this provides a picture of possible bacterial activity rather than the functions that are actually engaged.

Building upon the metagenomic analyses, transcriptomics and proteomics allow for the identification of active genes as well as their degree of activity.[13] Studying RNA and protein abundances provides valuable insight into the active biochemical pathways. However, this does not necessarily reveal the molecules that mediate the interactions with the host. There is increasing evidence that the products of microbial metabolism, e.g., secondary bile acids, play important roles as signaling molecules in the interactions between the intestinal microbiota and the host. In this light, metabolomics, defined as systematic identification and quantification of the small molecule metabolic products of a biological system, represents a valuable '-omics' approach that complements sequencing based approaches for characterizing the functional outputs of a microbiota. The metabolic profile of a biological system represents the integral of all enzymatic activities that have occurred in the system, and thus can be used to characterize different phenotypical states at the molecular level. By comparing the metabolite profiles of a microbiota under different physiological conditions, e.g. pre- and post-infection, insights can be obtained into the molecular mediators involved in establishing the different conditions. These insights can in turn identify potential mechanisms that underlie the transition from one physiological state to another.

In this thesis study, metabolic profiles of intestines from mice subjected to infection and treatment are measured and analyzed to identify microbiota-derived metabolites that could explain and/or modulate resistance against infection by *C. difficile* (Figure 1). Previous (unpublished) observations of deterred *C. difficile* growth due to *Candida albicans* overgrowth has led to the hypothesis that aspects of the metabolic activity of *C. albicans* may enforce resistance to *C. difficile* outgrowth. Based on this hypothesis, the commensal fungus *C. albicans* was used as a conditional treatment to elucidate metabolites that may confer resistance against

colonization by *C. difficile*. To develop the metabolic profiles, murine cecum metabolomes were surveyed using liquid chromatography followed by tandem mass spectrometry (LC-MS/MS). Accurate mass and mass fragmentation data were utilized together to annotate the detected metabolites. These annotated metabolites as well as other, as yet unannotated detected compounds were further evaluated across conditions to determine the similarities and differences between conditions, the active pathways within each condition, and to identify potential regulatory metabolites that may confer resistance to opportunistic pathogens such as *C. difficile*.

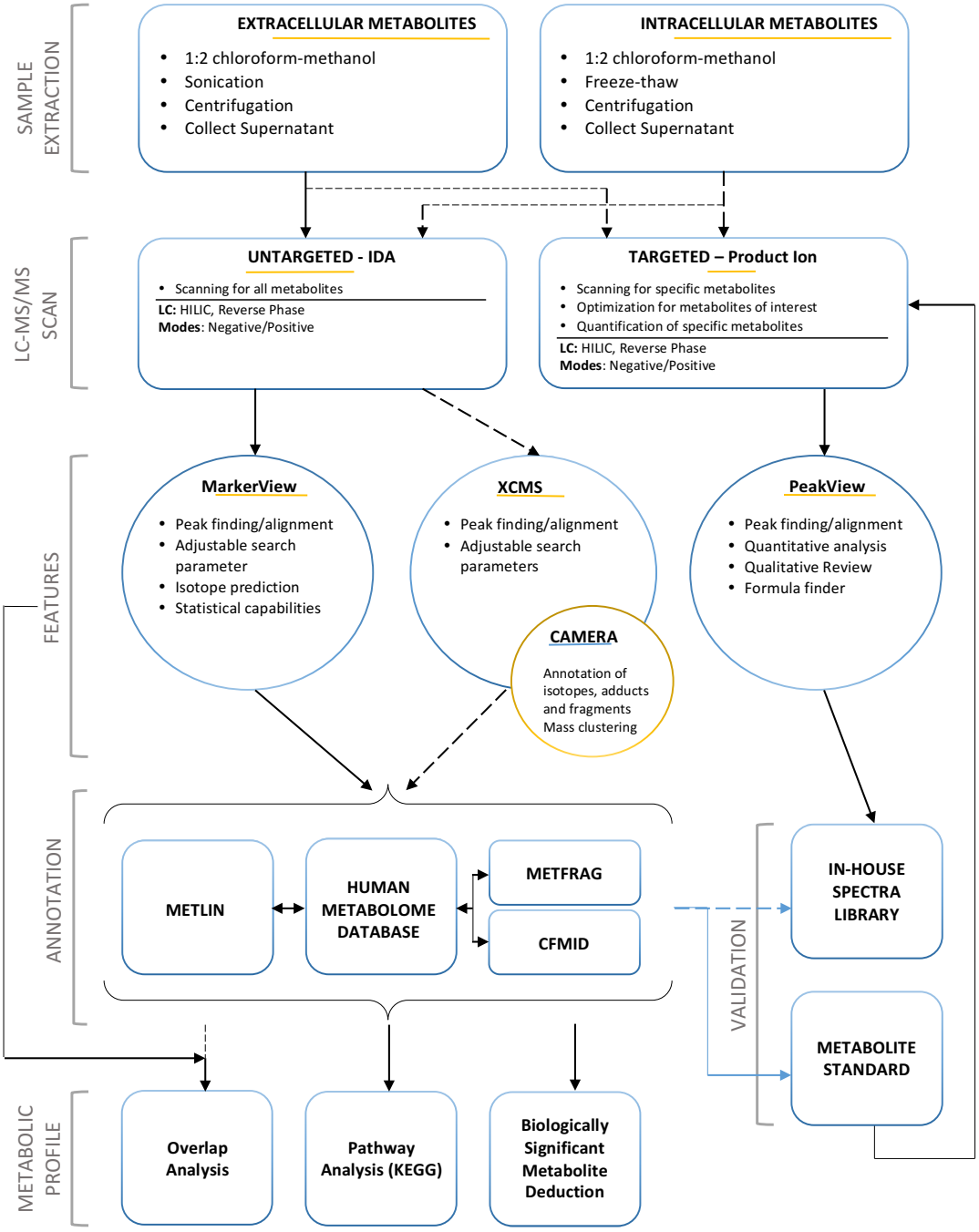


Figure 1 - Untargeted Metabolomics Pipeline. Framework for the process used for this project from sample extraction to metabolic profile development. Solid arrows indicate the steps followed during this particular study while dashed arrows represent alternative steps in the pipeline.

2.0 Surveying Metabolic Profiles

To obtain the metabolic profiles of the murine cecum samples, the metabolites were extracted from the cecum samples using a modified solvent-based extraction method adapted from Sellick et al.[23] The metabolite samples were first analyzed by untargeted analysis to broadly characterize the metabolome of the samples. Two different LC methods were used for the untargeted analysis, hydrophilic interaction chromatography (HILIC) and reverse-phase chromatography, to separate a large range of both polar and nonpolar metabolites.[24] Following the untargeted analysis of the cecum samples, features of the chromatographic spectrums were identified. The identified chromatographic features were normalized and preprocessed based on their significance and interest to this study. After narrowing down the features of interest, the selected features and associated ionization fragmentation spectra were analyzed using several databases for annotation. Annotated features were further scored based on the confidence of annotation to arrive a set of confidently identified metabolites that are present at significantly different levels between the sample groups. Literature searches on these metabolites yielded a panel of metabolites that could potentially play a biologically significant role in intestinal colonization resistance. The chemical identities of these putatively identified metabolites were then validated through targeted analysis using high-purity standards.

2.1 Chemicals

All of the reagents and solvents used for sample extraction and metabolite analysis were of HPLC grade or of high purity. Unless otherwise noted, all reagents and solvents used throughout this project were purchased from Sigma-Aldrich or Fisher Scientific.

2.2 Murine Cecum Samples: Procurement and Metabolite Extraction

Murine experimentation and cecum sample extractions used in this project were performed and provided by the Kumamoto lab at Tufts University School of Medicine. The treatment regimen applied during the murine study was adapted from Theriot et al.[12] This study examined the metabolic shifts in the gut based on an analysis of the cecum samples from three sets of female C57BL/6 mice (n = 4 per group) subjected to different 30-day treatment regimens (Figure 2). Mice in the control group were not treated with antibiotics nor challenged with *C. albicans* throughout the 30-day treatment period. The other two groups of mice were treated with the antibiotic cefoperazone, which was administered through their drinking water at a dose of 0.5mg/mL, for the first 10 days of treatment. One set of the antibiotic treated mice were subsequently inoculated with *C. albicans* while the other set was not inoculated with the yeast. After another 20 days, the second and third set of mice were administered clindamycin by injection at the dose of 10mg/kg. On day 31, all three sets of mice were euthanized by CO_2 and their cecum samples were harvested. The extracted cecum samples were stored in liquid nitrogen until metabolite extraction was performed.

To extract the metabolites of the pre-weighed cecum samples, 1.5mL of chloroform/methanol (1:2 v/v) was added to each sample. Each sample was homogenized for 2 minutes under the application of a tip sonicator and centrifuged at 22,065 RCF for 10 minutes under refrigeration at 4°C. The supernatant was collected and 0.6mL of HPLC grade water was added to the collected supernatant. Each sample was homogenized for approximately 15 seconds using the tip sonicator. The samples were then centrifuged at 22,065 RCF for 2 minutes under refrigeration at 4°C to allow phase separation. Then, the top phase was carefully collected as to not disturb the biphasic interface. Throughout the extraction process, the samples were kept on

ice between extraction steps. The extracted metabolite solutions were stored at -20°C until analysis.

From here on in, the untreated control mice will be referred to as the "healthy" condition. The mice treated only with antibiotics will be referred to as the “antibiotic” condition. And, the set of mice treated with antibiotics and challenged with *C. albicans* will be referred to as the “Candida” condition.

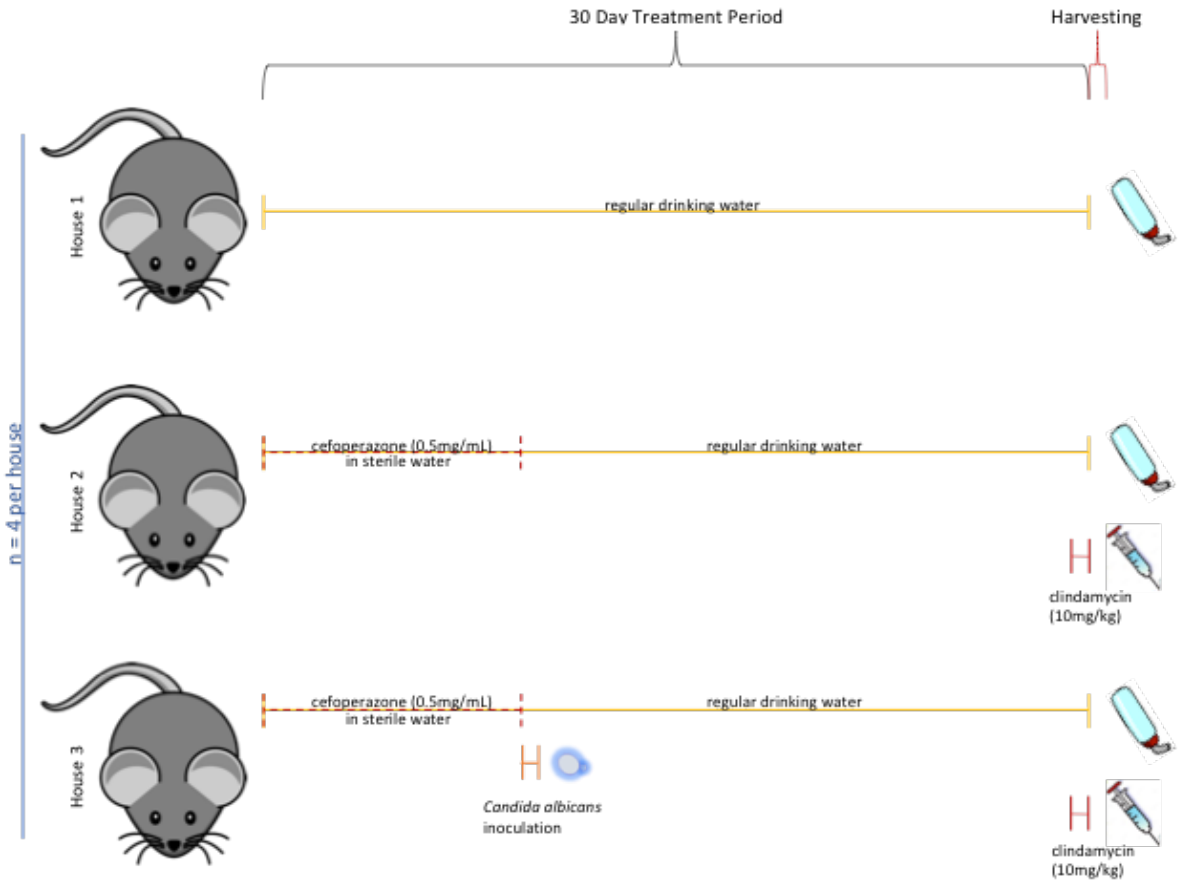


Figure 2 – Mouse experimentation. Three houses of mice (n=4 per house) were subjected to different antibiotic treatment conditions over a 30-day period. The first set of mice, the control, were fed regular drinking water for all 30 days with no further treatments. The second and third set of mice were treated with the antibiotic cefoperazone administered through their drinking water at a dose of 0.5mg/mL for the first 10 days followed by regular drinking water for the remaining 20 days. On day 10, the third set of mice were challenged with *Candida albicans*. Injections of antibiotic, clindamycin at 10mg/kg, were administered to the second and third set of mice on day 30. All mice were euthanized by CO_2 the following day for cecum sample harvesting.

2.3 Untargeted LC-MS/MS Analysis

2.3.1 Instrumentation

Untargeted and targeted metabolite analyses were performed using a triple quadrupole time-of-flight mass spectrometer (5600+ TripleTOF, AB SCIEX, Foster City, CA) coupled with a binary pump HPLC (Agilent Technologies 1200 series, Foster City, CA). The mass spectrometer was controlled by Analyst TF 1.6 software (AB SCIEX, Foster City, CA). Gas parameters of the mass spectrometer were optimized to obtain chromatographic reads with the best separation and reduced noise for each method. Operating parameters for each method and ionization mode are listed in the appendix.

2.3.2 HILIC Method

The hydrophilic interaction chromatographic separation method used was adapted from Bajad et al.[25] A HILIC column (Luna 5 µm NH2 100 Å 250 x 4.6 mm, Phenomenex, Torrance, CA) was used. Solvent A consisted of 95:5 (v/v) water/acetonitrile with 20mM of ammonium acetate. The pH of Solvent A was adjusted to 9.45 with the addition of ammonium hydroxide. Solvent B was pure acetonitrile. The gradient applied for both negative and positive ionization modes were as follows: t = 0 min, 85 % B; t = 15 min, 0 % B; t = 28 min, 0 % B; t = 30 min, 85 % B; t = 50 min, 85 % B. A 300µL/minute flow rate was applied to the column. The sample injection volume was 10µL.

2.3.3 Reverse-Phase Method

The reverse-phase (RP) chromatographic separation technique used was adapted from Lu et al.[26] A C18 column (Synergi 4 µm Fusion-RP 80 Å 100 x 2 mm, Phenomenex, Torrance, CA) was used. Solvent A was 0.1 % (v/v) formic acid in HPLC-grade water and Solvent B was 0.1 % (v/v) formic acid in methanol. The gradient applied for both negative and positive ionization

modes are as follows: t = 0 min, 3 % B; t = 8 min, 3 % B; t = 38 min, 95 % B; t = 45 min, 95 % B; t = 47 min, 3 % B; t = 55 min, 3 % B; t = 65 min, 3 % B. A 100 µL/minute flow rate was applied to the column with a 10µL sample injection volume.

2.4 Data Processing and Annotation

MarkerView software (Version 1.2.1, AB SCIEX, Foster City, CA) was used to find and align peak features of each sample using default settings. Also, peak features appearing in less than 3 of the samples were filtered out. The remaining peaks were normalized, removed of contaminants based on features detected in the blank samples, and used as inputs for the annotation pipeline. A more detailed explanation of the MarkerView peak processing can be found in the appendix.

Peaks of interest and their associated MS/MS spectra, when available, were analyzed using the following databases and *in silico* fragmentation tools for preliminary identification: Metlin (Scripps Center for Metabolomics, La Jolla, CA), HMDB: The Human Metabolome Database (Wishhart et al, Edmonton, AB, Canada), CFM-ID: Competitive Fragmentation Modeling for Metabolite Identification (Allen et al, Edmonton, Canada) and MetFrag (Wolf et al, Hallee, Germany). Metlin is a repository of metabolite information with experimentally derived, high-resolution MS/MS metabolite data and *in silico* predicted MS/MS metabolite data. Encompassing data for a broad range of molecules, Metlin includes metabolites ranging from lipids to plant-derived and bacterially-derived metabolites. HMDB is a database of human metabolite information with high-quality, experimentally derived and predicted MS/MS metabolite data.

While CFM-ID and MetFrag are both *in silico* fragmentation tools for predicting the identity of a given mass based on computational fragmentation, each software approaches the mass fragmentation in a different manner. CFM-ID predicts fragmentation patterns based on competitive fragmentation modeling. Using probability and machine learning techniques, CFM-ID was trained to determine fragmentation parameters using datasets drawn from Metlin.[27] In contrast, MetFrag uses a combinatorial fragmentation approach based on bond dissociation.[28] Provided a neutral mass, ionization mode, and experimental fragmentation data, both *in silico* tools draws mass candidates from compound databases, such as KEGG, fragments the mass candidates, and then ranks the candidates based on the similarity between the computed fragmentation pattern and the queried fragmentation spectra observed in the sample. Each annotated peak feature was assigned three scores for the purposes of peak feature consolidation across LC-MS methods and modes, and for identity selection for features with multiple annotations. The three scoring criteria included a feature intensity score, a confidence level for the accuracy of the metabolite identification, and an annotation score based on the scores derived from the annotation databases.

Feature intensity scores were calculated to capture the ratios of metabolite detected at significantly different levels between experimental groups. Statistical comparisons were performed and feature intensity score ratios were computed for the following pairs: healthy vs. antibiotic and antibiotic vs. Candida. A threshold of five-fold change between the healthy and antibiotic conditions was applied to identify metabolites that are significantly altered by the antibiotic treatment. A smaller threshold of a two-fold change between the antibiotic and Candida conditions was applied to identify metabolites that are significantly altered by the infection with *C. albicans*.

Confidence in annotation was categorized based on which database(s) were the sources of the annotation (Table 1). The highest level of confidence was assigned to annotations that represent a consensus identification across all four sources. The lowest level of confidence was assigned to annotations that represent matches with entries in two databases. Preference was given to databases with MS/MS spectra generated by experimentation, allowing for high-scoring single database matches from Metlin to received a low level of confidence as well.

Table 1 – Annotation Confidence Scoring. Confidence scores of high, medium and low were assigned to the annotated features. High confidence was given to an annotation matched across all four databases. Medium confidence was assigned to features that matched across three databases. And, low confidence was given to annotations with two database matches or for Metlin annotations with a score greater than 0.9. Original Metlin scores are out of 100 but were scaled to 1 for this project to conform with scores generated by the other databases.

Annotation Confidence Strength	# Database Matches	Databases
High	4	Metlin, HMDB, CFM-ID, MetFrag
Medium	3	Metlin, HMDB, CFM-ID
		Metlin, HMDB, MetFrag
		Metlin, CFM-ID, MetFrag
		HMDB, CFM-ID, MetFrag
Low	1-2	Metlin, HMDB
		Metlin, CFM-ID
		Metlin, MetFrag
		HMDB, CFM-ID
		HMDB, MetFrag
		CFM-ID, MetFrag
		Metlin > 0.9

Annotation scores were calculated using the identification scores provided by each database or *in silico* fragmentation tool. The scores were weighted such that annotations from experimentally derived databases would receive a greater weight. Due to the limited coverage of

HMDB, the weighting was adjusted such that a match with an entry generated by both *in silico* tools would be weighted equally as a match with an entry found in HMDB.

2.5 Metabolic Profile Analyses

2.5.1 Similarity Analysis

Similarity analysis of the metabolomes across different sample groups was performed using a technique adapted from Bashan et al.[29] This analysis used the full, unannotated peak features list obtained following peak alignment and identification in MarkerView. Pairwise similarity comparisons of the metabolic profiles between each sample was calculated with the following equations:

$$\boldsymbol{s}(x_i, y_i) = \begin{cases} \frac{2 \cdot \min(x_i, y_i)}{x_i + y_i}, & x_i + y_i \neq 0 \\ [removed], & x_i + y_i = 0 \end{cases} \quad \text{Eq. 1}$$

$$\boldsymbol{O}(X, Y) = \frac{\sum_{i=1}^{n} \boldsymbol{s}(x_i, y_i)}{n} \quad \text{Eq. 2}$$

The overlap (***s***) of each feature i, x_i and y_i, between two samples was determined by calculating the percentage of maximum shared feature overlap; in other words, by dividing twice the minimum feature intensity value between the two samples by the combined feature intensity values of both samples. If the feature was absent in both murine samples, the feature was removed from ***s*** feature set. If a feature was present in one sample, but not the other, the shared overlap score was assigned a value of 0, indicating the greatest dissimilarity. If a feature was present in both murine samples with equivalent feature intensities, the shared overlap score was assigned a value of 1, indicating the greatest similarity. The shared metabolic profile overlap (***O***)

between two samples was calculated by averaging the entire set of metabolic feature overlap values.

To determine the dissimilarity of the shared overlap $S \equiv X \cap Y$ between two sample metabolic profiles *X* and *Y*, the feature data set was first filtered by eliminating features that were not shared between two samples. Then, the dissimilarity of the pared profiles $\tilde{X}$ and $\tilde{Y}$ were calculated using the classical Yue-Clayton equation:

$$\boldsymbol{D}\left(\tilde{X},\tilde{Y}\right) = 1 - \left[\frac{\sum_{i=1}^{S}\tilde{x}_i\tilde{y}_i}{\sum_{i=1}^{S}(\tilde{x}_i-\tilde{y}_i)^2 + \sum_{i=1}^{S}\tilde{x}_i\tilde{y}_i}\right] \quad \text{Eq. 3}$$

Filtering the data prior to calculating the dissimilarity of the shared overlap set *S* was a necessary step because this allowed for the dissimilarity of the overlap to be solely defined by the differences in the shared overlap rather than being skewed by features that were not shared (***s*** = 0).

2.5.2 Pathway Analysis

Using the identification results from the above annotation workflow, a pathway analysis was performed on the annotation features to further refine the annotation. When annotating features to entries in Metlin, HMDB, CFM-ID and MetFrag, the databases provides the user with the KEGG pathway associated with each potential compound match for a given feature, provided that the potential compound match is cataloged in KEGG and has been assigned a KEGG compound number. Using this information, the frequency with which a pathway was associated with the dataset was tabulated to estimate the most actively engaged pathways. The results from this frequency analysis were used to reduce the ambiguity associated with annotating a feature that was matched to multiple compounds. For a given feature of interest, it was assumed that an annotation corresponding to a metabolite in a more active pathway (as determined by the

frequency analysis) was more likely to be correct compared to an annotation corresponding to a metabolite in a less active pathway.

2.6 Metabolite Identity Confirmation

2.6.1 Selection of Biologically Relevant Metabolites

One of the central questions of this project is whether the presence of *C. albicans* could confer resistance against *C. difficile* infections by altering the intestinal metabolite profile. We hypothesized that colonization by *C. albicans* would generate regulatory metabolites that are also present in a healthy state and confer resistance to colonization by *C. difficile*.

To select metabolites that may confer resistance against *C. difficile*, the panel of putatively annotated features was pared down through significance testing. For simplicity, student's t-tests was used for all pairwise statistical comparisons. The null hypothesis of identical sample means was rejected when the p-value was less than 0.05. Both the healthy and Candida conditions were compared to the antibiotic condition, and the significantly different features were retained for further analysis.

These significantly different annotated peaks were divided into two broad categories: up- or down-regulation. A five-fold change difference was applied as a threshold for the healthy vs. antibiotic comparison while a two-fold change difference was used for the Candida vs. antibiotic comparison. Putatively annotated features that were elevated in the healthy state were considered to be potentially biologically relevant because these metabolites would also be depleted in the antibiotic state, possibly weakening the defenses of the microbiota and causing an increased susceptibility to *C. difficile*. These biologically relevant features were further analyzed in the Candida condition for renormalization to determine features that may explain the resistance to *C.*

difficile under homeostatic conditions and the resistance to *C. difficile* seen in patients with Candida overgrowth. Literature searches were performed on the list of potentially biologically relevant metabolites to determine if they may be involved in colonization resistance. Metabolites for which there was literature evidence were selected to validate the annotations and confirm their presence in the samples. Additionally, several other metabolites were selected for validation to evaluate the performance of the annotation workflow in terms of accuracy.

2.6.2 Targeted Metabolite Validation Using Targeted Analysis

To confirm the identity of an annotated feature, the retention time, accurate mass, and MS/MS spectrum of the feature were compared against high-purity standards. Product ion scans were performed for the following pure standards: uracil, n-acetyl-dl-leucine, indole-3-carboxylic acid, enterodiol, pantetheine, 4-pyridoxic acid, inosine, orotic acid, l-(+)-arabinose, alpha-d-fucose and cyclic AMP. The instrument (QTOF) operating parameters used for these validation experiments were identical to the method and MS mode with which the samples were originally detected. Fragmentation patterns (MS/MS spectra) of the standard and metabolite of interest detected in the sample were compared using the Jaccard similarity coefficient and Spearman's rank correlation. The Jaccard similarity coefficient,

$$J(X,Y) = \frac{X \cap Y}{X+Y-X \cap Y} \quad \text{Eq. 4}$$

determines the similarity between shared peaks based on their relative intensities. The Spearman's rank correlation,

$$r_s(X,Y) = 1 - \frac{6\,\Sigma\,(rank(X_i) - rank(Y_i))^2}{n(n^2-1)} \quad \text{Eq. 5}$$

determines the similarity of the shared peaks between the mass spectra, but does not take into account the peaks' intensities. To confirm the identity of a metabolite and thus validate the annotation, we required a high Jaccard similarity score and a high Spearman's correlation score.

3.0 Metabolic Profiling Results

Due to the prolonged lag time between the execution of each method, features derived from the Synergi experimentation were expected to be significantly degraded compared to the HILIC analysis. Despite this fact, the Synergi results were used in conjunction with the HILIC features to provide a more comprehensive view of the similarity between the conditioned mice in this study. While both separation techniques were used in the pathway analysis, biologically relevant metabolites for validation were strictly derived from the HILIC dataset.

3.1 Similarity Analysis

Evaluation of the unannotated HILIC and Synergi features shows a high metabolic overlap between co-housed mice (Figure 3). The healthy mice had approximately a 58 % overlap in their metabolomes amongst each other, while the antibiotic treated mice and Candida treated mice had an average of 69 % and 64 % overlap in their metabolomes respectively. The similarity in the metabolic profiles of the antibiotic and Candida treated mice can also be seen during the cross comparison of the two conditions, with a 65 % average overlap in their metabolomes. When cross compared against the healthy condition, the antibiotic and Candida profiles comprised a significantly lower number of shared metabolites with a 20 % average overlap in their metabolic profiles. While the similarity analysis depicted a decrease in overlap for both

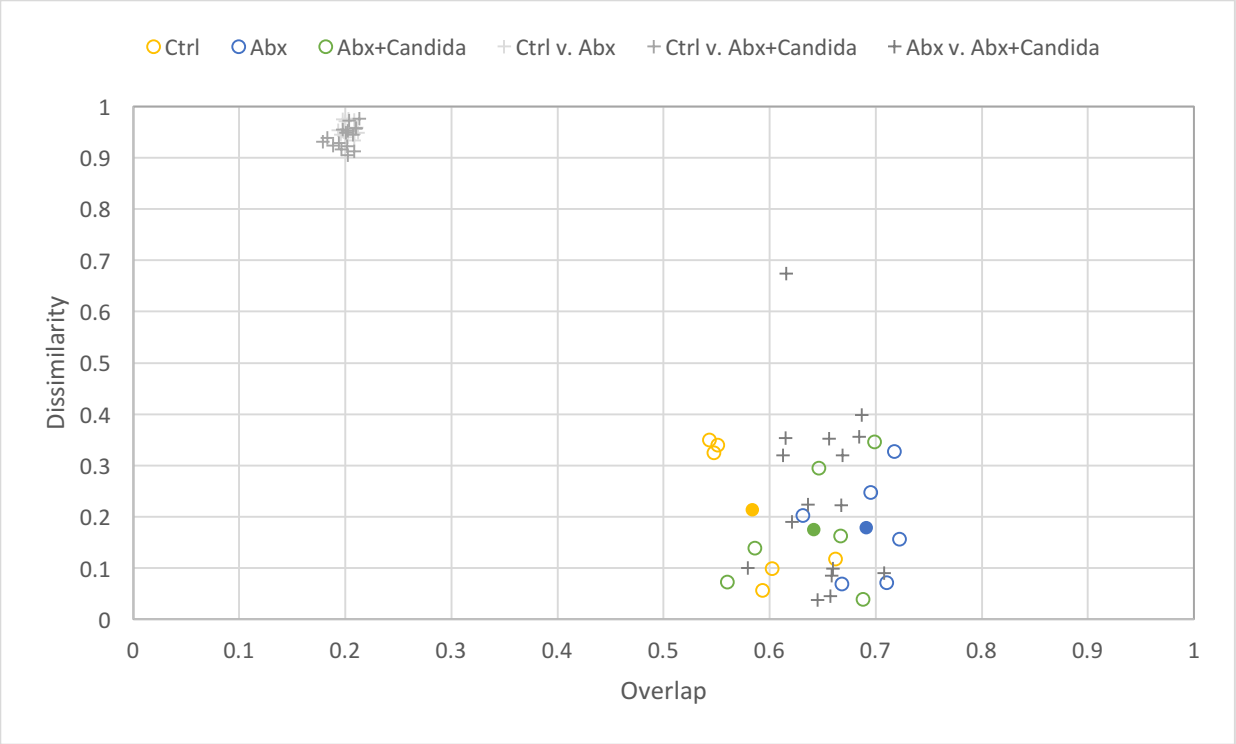


Figure 3 – Metabolome Overlap Comparisons. Similarity analysis of the differently conditioned mice shows high similarity between co-housed mice. The healthy (Ctrl) mice exhibited the lowest degree of metabolic overlap with an average overlap (solid yellow circle) of 58 % with low dissimilarity in abundance production of similar metabolites. No significant difference was found between the antibiotic treated (Abx) and Candida treated (Abx+Candida) conditions with an average overlap of 69 % (solid blue circle) and 64 % (solid green circle) respectively. Both the antibiotic and Candida treated mice possessed low dissimilarity in metabolite production of similar metabolites. The lack of significant difference between the antibiotic and Candida conditions indicates the addition of one strain of yeast, *Candida albicans*, does not drastically alter the overall metabolic activity. In contrast, there was a significant difference in the metabolome with the use of antibiotics, reflected by a 66 % decrease in similar metabolites from the healthy condition. When compared to the healthy metabolome, the antibiotic and Candida treated metabolomes both possess roughly a 20 % metabolic overlap with high dissimilarity in metabolite production. However, there was no significant difference found between the healthy and Candida conditions, suggesting the increased presence of *C. albicans* in an antibiotic augmented gut shifts the metabolic activity towards the healthy baseline.

---

pairwise comparisons to the healthy condition, statistical testing denoted a significant difference between the healthy vs. the antibiotic metabolic overlap ($p = 0.002$) but a lack of significant difference between the healthy vs. the Candida metabolic overlap ($p = 0.156$).

A closer examination of the similar metabolites within the overlapping sections of each pair of metabolomes shows that with lower dissimilarity there is greater variance in the abundance of metabolites. The healthy, antibiotic treated, Candida treated and antibiotic vs. Candida treated mice showed this high span of low abundance dissimilarly among their shared

metabolites, with a standard deviation ranging between 10 and 17 %. In contrast to the other dissimilarity comparisons, the healthy vs. antibiotic and healthy vs. Candida mice assessments produced a low variance of high dissimilarity results, with a standard deviation of only 1 to 2 %.

3.2 Metabolic Profile Annotation Statistics

After preprocessing the features list generated by MarkerView, a total of 16,505 features from the dual modes of the HILIC and Synergi methods were analyzed using the annotation workflow. After merging the features based on similar m/z values (within a 20 ppm tolerance window), the annotated features were assigned scores for their feature intensities, confidence in annotation, and annotation score. Using these three scoring criteria, the list of annotated features was narrowed to 2,241 unique compound identifications. A comprehensive breakdown of the annotation statistics can be found in the appendix.

3.3 Pathway Analysis

Pathway information for the 2,241 putative annotations was gathered and categorized according to the KEGG Brite Hierarchy to determine the most frequently occurring pathways. Pathway analysis of the putative identifications indicated amino acid metabolism, lipid metabolism, nucleotide metabolism, biosynthesis of other secondary metabolites, and metabolism of cofactors and vitamins to be the top five most active pathway categories based on high confidence identifications across all experimental conditions.

The putative annotations show a broader spectrum of differentially active pathways between the healthy vs. antibiotic treated mice when compared to the Candida vs. antibiotic

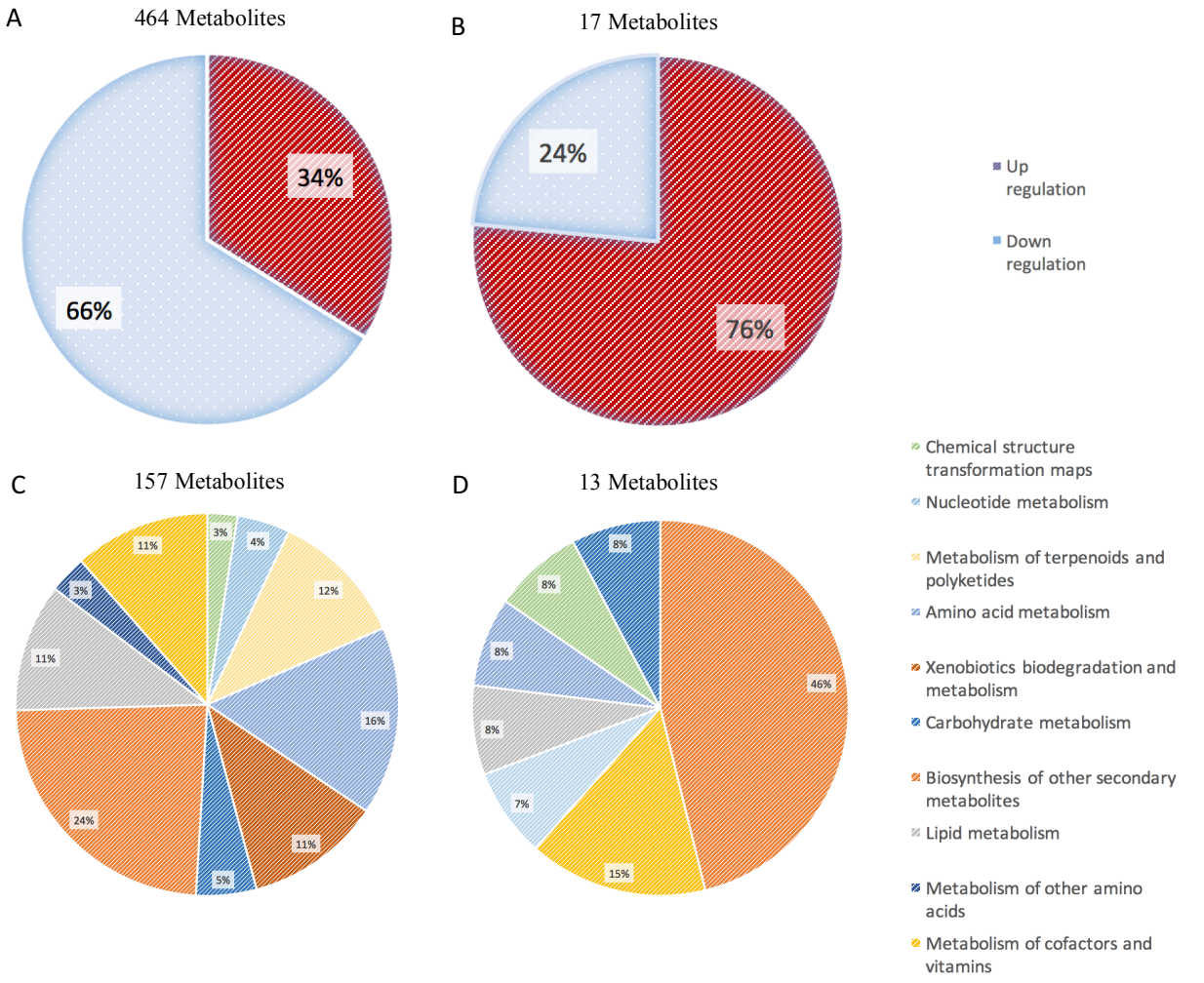


Figure 4 – Significant Pathway Comparisons. (**A**) Significantly different pathways between the healthy and antibiotic treated mice. (**B**) Significantly different pathways between the Candida and antibiotic treated mice. (**C**) Up-regulated pathways between the healthy and antibiotic treated mice. (**D**) Up-regulated pathways between the Candida and antibiotic treated mice. When comparing the healthy mice to the antibiotic treated mice, there is a large portion of down-regulated metabolites. The large percentage of down-regulated metabolites in the healthy mice is likely due to the regular metabolism of the down-regulated compounds by microbes that no longer exist after the use of antibiotics, causing a buildup in the antibiotic condition. The number of significantly different metabolites markedly reduced when comparing the Candida to antibiotic treated mice, supporting the similarity in their metabolomes. The high up-regulation seen in the Candida condition is likely due to the production of metabolic products by the yeast strain. The inclusion of the yeast strain enhances the biosynthesis of secondary metabolites as well as the metabolism of terpenoids and polyketides. Similar up-regulated pathways between the healthy and Candida conditions can provide insight into potential metabolites that may confer colonization resistance during an antibiotic-stressed state.

---

treated mice (Figure 4). For the comparison between healthy and antibiotic mice analysis, 464 putatively annotated metabolites were identified as significantly different between the two conditions. Of the 464 metabolites, 34 % were elevated in the healthy condition while 66 % were

elevated in the antibiotic condition. Far fewer metabolites, only 17, were identified as significantly different between the Candida and antibiotic treated mice groups. Of these metabolites, 76 % were elevated in the Candida treated mice while 24 % were elevated in the antibiotic treated mice.

Compared to the antibiotic treated condition, both the healthy mice and Candida treated mice displayed a higher percentage of elevated metabolites in the following pathway categories: biosynthesis of other secondary metabolites and metabolism of cofactors and vitamins. Less active pathway categories for these conditions include amino acid metabolism, lipid metabolism, carbohydrate metabolism, nucleotide metabolism, and chemical structure transformation maps. Also, when compared to the antibiotic treated mice, both the healthy and Candida treated mice were depleted of metabolites in the following pathways: biosynthesis of secondary metabolites and metabolism of cofactors and vitamins (results available in the appendix).

3.4 Validation of Selected Metabolites

Targeted scans of the selected metabolite standards were used to validate the presence of uracil, acetyl-dl-leucine, enterodiol, pantetheine, 4-pyridoxic acid, and inosine in the cecum samples (Table 2). The tandem mass fragmentation patterns for orotic acid, l-(+)-arabinose, alpha-d-fucose and cyclic AMP did not match the patterns detected in the samples, thus disqualifying these metabolite annotations. Quantitatively, comparisons of MS/MS spectra for the verified metabolites against their respective chemical standards yielded a Jaccard coefficient of at least 0.7 and a Spearman's rank of at least 0.9.

Table 2 – Validation of Metabolite Annotation. Features and their predicted identities are presented in this table along with their relative abundance when compared to the antibiotic condition. How each feature was annotated is indicated under the annotation workflow section. Features with Jaccard similarity coefficients ≥ 0.7 and a Spearman's rank ≥ 0.9 were used as criteria for a positive identification. Also, for an annotation that was identified by more than one database, if two or more of the retrieved database scores (scaled) are ≥ 0.8, the putative identification was designated a confident annotation. A dash in the annotation workflow section indicates no feature matches for the given database. A dash for the Spearman's rank indicates an insufficient amount of fragments to calculate the correlation while N.O. signifies no overlap between the standard fragmentation and sample fragmentation patterns. *Up-regulated in the healthy condition with renormalization in the Candida condition. **Up-regulated in the healthy condition with a two-fold up-regulation in the Candida condition (but not significantly different).

Neutral Mass	Presence in Healthy Mouse	Metabolite	Annotation Workflow				Jaccard Similarity Coefficient	Spearman's Rank	Identity Confirmed
			Metlin	HMDB	CFMID	MetFrag			
173.0974	High*	Acetyl-DL-Leucine	1.00	-	-	-	0.99	1	✓
112.0216	High*	Uracil	-	0.81	0.43	0.95	1	-	✓
161.0406	High**	Indole-3-Carboxylic Acid	-	-	0.28	0.80	N.O.	N.O.	✗
302.1469	High	Enterodiol	0.40	0.99	0.80	1.00	0.74	0.92	✓
278.1246	High	Pantetheine	-	0.93	0.52	1.00	0.73	1	✓
183.0463	High	4-Pyridoxic Acid	0.74	-	0.93	0.94	0.72	1	✓
268.0753	High	Inosine	0.63	-	0.80	1.00	0.84	1	✓
150.0455	Low	L-(+)-Arabinose	1.00	-	0.27	0.78	0.22	-	✗
156.0099	Low	Orotic Acid	0.16	0.74	0.67	1.00	0.36	-0.50	✗
329.0578	Low	cAMP	-	0.87	-	0.70	N.O.	N.O.	✗
164.0611	Low	Alpha-D-Fucose	0.47	-	-	-	N.O.	N.O.	✗

4.0 Discussion

The use of antibiotics is known to alter the gut microbiota and subsequently the metabolome, leaving the host susceptible to infection by enteric pathogens. Presented in this study is a first glimpse into the structure of the metabolome and how the use of antibiotics can perturb the milieu of metabolites that contribute to the functions of microbes residing in the gut. In this thesis, to better understand the functions of the microbiota, the changes that were seen in the metabolome were redefined in terms of active pathways according to the frequency of observed metabolites in the pathways. Metabolites that may offer resistance against infection by opportunistic pathogens in an antibiotic-stressed state were elucidated from the preliminary metabolic profiles based on a study of the published literature and their presence in the samples validated using authentic chemical standards.

4.1 Untargeted Analysis

Due to the difficulty of annotating untargeted LC-MS/MS data and the importance of accurate identifications, much of the work in this thesis projected involved the feature annotation process. In metabolomics studies of the intestinal microbiota, the metabolic profiling is often outsourced to a third-party vendor for analysis, who annotate sample data against their own in-house reference library using proprietary software.[12,30] Instead of relying on a third-party for data processing, this study presents a detailed approach, from sample extraction to data annotation, for the generation of metabolic profiles. By using this metabolic profiling framework, potential regulatory and/or resistance enforcing metabolites were identified and confirmed for further functional testing.

4.1.1 Annotation Workflow

As previously stated, one of the biggest challenges in this project was feature annotation of the untargeted data. Without a standard method for annotating unknown features, data annotation proved to be a bottleneck in the discovery potential of this untargeted analysis. Prior to filtering annotations by feature intensity, confidence level, and annotation score, there were 1.5 times as many annotations as there were queried features. As noted by this over-annotation of the data, the feature preprocessing with MarkerView and the feature annotation through the use of multiple databases in conjunction with *in silico* fragmentation tools can lead to false positives due to a number of different factors.

False positives, or misidentifications, can be partially attributed to misinterpreting isotopic isomers and adducts produced during the LC-MS runs. Although MarkerView does take note of potential isotopes, the presence and misidentification of adducts within a sample can skew the metabolic profile. When a molecule is ionized during an LC-MS run, the expected

result is to form a distinct ion, such as [M+H]. However, in practice, the ionization can form other ion products such as [M+K] or [M+H-H2O]+.[31] These additional product ions may be identified as unique features instead of belonging to a single parent molecule, confounding the analysis and thus contributing to the misidentifications. To resolve these issues, isotopes and adducts should be identified and taken into consideration for future metabolic analyses as well as for metabolite quantification purposes.

One method to identify isotopes and adducts is by preprocessing untargeted data using the XCMS (The Scripps Research Institute, La Jolla, CA) software with the Bioconductor package CAMERA: Collection of Algorithms for Metabolite Profile Annotation. Employing XCMS to align peak features and CAMERA for the recognition of isotopes and adducts, with clustering of production ions that may be derived from a single parent ion, could reduce false positives during the annotation process to produce more accurate metabolic profiles.[31]

4.1.2 Annotation Databases

Of the 11 metabolites selected for validation, 54 % were shown to be identified correctly using the workflow developed in this project. As shown by the annotation validation results, the strength of the annotations should not be dictated by a single database. Instead, leveraging data from multiple annotation sources increases the accuracy of annotation.

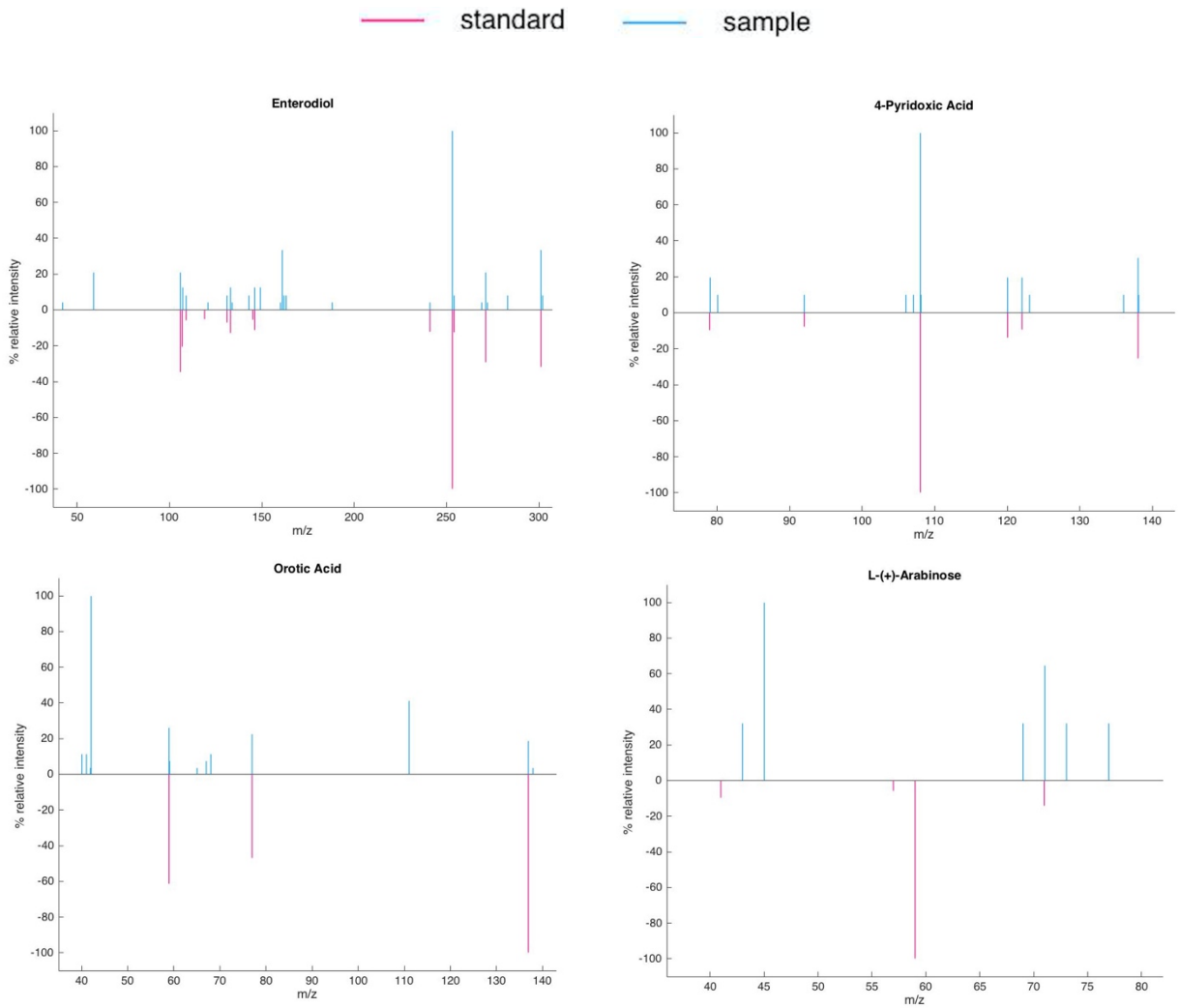


Figure 5 – Select Metabolite Validation Fragmentation Spectra. Mirror plots show the metabolite standard (red) and sample metabolite (blue) fragmentation patterns with their relative fragment intensities. When compared to their metabolite standard the putatively identified enterodiol and 4-pyridoxic acid samples presented similar fragmentation patterns and relative peak intensities indicating a positive match. The fragmentation pattern for the orotic acid sample possesses matching fragments to the metabolite standard but the relative intensities of the matching peaks are quite dissimilar, making orotic acid a misidentification. However, based on the matching fragment peaks, the identity of the sample feature may be a compound closely related to orotic acid. L-(+)-Arabinose is a negative match, with poorly matching fragmentation patterns between the metabolite standard and the sample.

Currently, as one of the more comprehensive metabolite databases, Metlin possesses information on 961,829 metabolites from various sources but only 1.45 % of those metabolites have experimentally derived high-resolution MS/MS information.[32] HMDB is another database containing information for 42,003 metabolites found in the human body but only 13.74 % of

those metabolites have experimentally derived MS/MS information.[33] Due to the low quantity of publically available quality metabolite tandem mass spectral data, such databases can be further supplemented by *in silico* fragmentation tools, including CFM-ID and MetFrag, that are designed to predict fragmentation patterns for a specified metabolite. Because of limited publically available experimental metabolite spectral data and the innate limitation of *in silico* tools by design assumptions, combining the power of these tools can further improve the certainty of putative annotations.

Despite the possibility of obtaining correct annotations through the use of a single database, such as the identification of acetyl-dl-leucine by Metlin, contributions from multiple databases can bridge data gaps presented in any given database (Table 2). For instance, the MS/MS spectra for the 112.0216 Da compound was undetermined by Metlin but was positively identified by HMDB, CFM-ID and MetFrag. Beyond the increase in confidence of an annotation prediction that comes with its identification across multiple databases, the spectral comparison scores provided by each database should be taken into consideration to understand the likelihood of the overall prediction.

As seen with the putative identification of orotic acid, the feature was identified across all four databases. But, the identification of orotic acid received low comparison scores for three of the four databases. Such low scores for a majority of the databases lowers the confidence in the prediction which proved to be an accurate indication of a negative match between the sample spectra to the orotic acid standard spectra (Figure 5). In contrast, enterodiol was identified across all queried databases with high comparison scores for three of the four databases. With a majority of high comparison scores, there is higher confidence in the annotation of enterodiol.

This indication proved accurate with a positive verification when the sample spectra were compared against the enterodiol standard spectra.

The putative identification of l-(+)-arabinose is a prime example of why using multiple databases can be informative in the prediction process. With a perfect match to the Metlin spectra, one may assume l-(+)-arabinose to be an accurate identification because the Metlin database contains experimentally derived MS/MS information. However, l-(+)-arabinose was also predicted by both *in silico* tools but the prediction scores produced suggests the l-(+)-arabinose identification was unlikely. The identification was later verified to be an incorrect annotation.

Upon further inspection as to why Metlin gave a perfect score for the l-(+)-arabinose annotation lead to the finding that their modified X-rank scoring method may be misleading based on their available spectral information.[34] Examination of the spectral comparisons of the Metlin l-(+)-arabinose fragmentation to the sample and in-house standard fragmentations presented good matches between the patterns, with matches to 2/3rds of the sample peaks and a match to all standard peaks (spectral comparisons are available in the appendix). Yet, when the sample was cross compared to the in-house standard there was only one fragmentation peak in common.

With a closer review of the Metlin spectral data, it appears the tandem mass spectral data for different collision energies can comprise of an amalgamation of predicted fragmentations as well as experimentally derived fragments. Traditionally, the X-rank algorithm compares two lists of sorted peak intensities to find correlations and is relatively efficient at discriminating and identifying compounds.[35] But in cases resembling the l-(+)-arabinose prediction, the X-rank algorithm does not appear to account for the predicted fragments in the Metlin spectra. This lack

of differentiation of predicted to experimentally derived fragments likely caused an inflation in the comparison score provided by Metlin that ultimately led to a false identification. By querying multiple databases misidentifications, such as with the case of l-(+)-arabinose, can be circumvented.

4.2 Metabolic Profiles

According to the similarity analysis, the healthy mice have a lower similarity in metabolomes compared to the antibiotic and Candida conditions. The presence of an overlap in the metabolomes of the healthy mice suggests a baseline of microbes present in each mouse, which agrees with previous findings by Antonopoulos et al. that healthy mice have a microbial community with an average Bray-Curtis similarity of 0.765 ± 0.014.[36] And, when antibiotics were introduced into the mice, the microbial communities of the intestines became more similar with an average Bray-Curtis similarity of 0.845 ± 0.083.[36] As reflected by the increased metabolic overlap of the antibiotic and Candida conditioned mice, with increased similarity of the microbial communities, the metabolomes of the mice also become more alike. While there was increased similarity in the metabolomes of mice treated with antibiotics, the antibiotic regimen employed during the experiment may have suppressed the similarity potential of mice treated with antibiotics. Administration of broad-spectrum antibiotics via drinking water affects the dosage obtained by each mouse depending on how much water is consumed by each mouse. The varying dosages of antibiotic each mouse consumed can distort the types of bacteria and the degree of their presence in the antibiotic treated mice, influencing their metabolic diversity while stunting their metabolic similarity. Mice that were inoculation with *C. albicans* following antibiotic treatment showed insignificant change from the antibiotic treated mice; indicating a

minute impact a single species can have on the metabolome. However, the insignificance between the metabolic overlap of the healthy and Candida conditions suggests the addition of *C. albicans* introduces more diversity to the metabolome and the yeast could be producing metabolites that may reinforce colonization resistance.

When cross compared to the healthy condition, the antibiotic and Candida treated mice showed an immense decrease in metabolic overlap, with a 66 % drop in the baseline of similar metabolites. The narrow range of metabolic overlap in the antibiotic and Candida conditions depicts the drastic impact the use of antibiotics can have on a healthy metabolome as well as how antibiotic use can alter the production levels of the remaining similar metabolites. This dramatic change in the metabolome caused by the use of antibiotics was also reflected in the microbiota, with a 72 % decrease in the average Bray-Curtis similarity between the microbial communities of the healthy and antibiotic treated mice.[36] To expand upon changes seen in the similarity analysis, pathway analysis was used to further examine the alterations between the metabolomes of the differently-conditioned mice in greater detail.

Corresponding to the findings of the similarity analysis, the smaller number of differentially identified metabolites and active pathways indicates that the Candida condition is more closely related to the antibiotic than the healthy condition. When compared to the antibiotic condition, the healthy metabolome up-regulates 34 % and down-regulates 66 % of the 464 significantly different metabolites. The higher degree of up-regulated metabolites in the antibiotic condition is likely caused by a decrease in the microbial population that typically consumes these metabolites, causing a buildup of diet-derived metabolites in the antibiotic condition.[37] On the other hand, incorporation of *C. albicans* post-antibiotic treatment increased metabolic activity with up-regulation in 76 % of the 17 significantly different metabolites

between the Candida and antibiotic conditions. This increase in up-regulation seen in the Candida condition is likely a reflection of the added production of metabolic products by the yeast strain.

4.3 Identified Biologically Relevant Metabolites

Searching the intersection of the elevated metabolites between the healthy and Candida treated mice revealed two compounds, uracil and acetyl-dl-leucine, as potential metabolites that may enforce resistance to *C. difficile* infection in the gut in relation to *C. albicans* overgrowth. Other metabolites that were strictly elevated in the healthy mice that may confer resistance to infection include: enterodiol, pantetheine, 4-pyridoxic acid and inosine.

4.3.1 Uracil

Commonly known as a building block of RNA, uracil has been shown to regulate immune health through the microbe-host relationship in *Drosophila melanogaster*. A study conducted by Lee et al. demonstrated bacterially-secreted uracil activates the dual oxidase (DUOX) pathway which induces production of the antimicrobial effector molecule, microbicidal reactive oxygen species (ROS).[38] In addition to fending off pathogenic species, the DUOX-dependent production of ROS has been associated with the regeneration of epithelial cells by activation of intestinal stem cells during infection in the gut.[38–40] Uracil production under normal, homeostatic conditions was only observed in exogenous microbes even though certain symbiotic bacteria, bacteria with pathogenic-like characteristics, are capable of releasing uracil. Under certain conditions, such as dysbiosis, symbiotic bacteria can release uracil causing chronic activation of the DUOX immune response leading to a pathology similar to inflammatory bowel disease.[38]

Although the physiological relevance of uracil release by bacteria remains to be fully elucidated, it has been hypothesized that the epithelial cells in the gut (in the fruit fly) have evolved to possess G-protein coupled receptors (GPCRs) that recognizes the variations in exogenous uracil.[41,42] Leveraging the secretion of uracil by pathogens, the host can employ a first line of defense against the invading microbes by detecting the exogenous uracil and producing antimicrobials to antagonize the foreign microbes before the host's immune response can acquire resistance against the pathogen.[38] While murine physiologically is obviously different from that of the fruit fly, there is a high degree of conservation between the two organisms in terms of signaling pathways involved with intestinal development, regeneration, and disease between mammals and insects. This opens up the intriguing possibility that mice and perhaps humans may respond to bacterially-derived uracil in a similar manner as the *D. melanogaster* to promote homeostatic conditions in the host.[43]

4.3.2 Acetyl-DL-Leucine

Commonly recognized as building blocks for proteins and polypeptides, amino acids also regulate metabolic activities associated with maintenance, growth, reproduction and immunity.[44] In particular, leucine is known to regulate intracellular protein turnover by stimulating the phosphorylation of the serine/threonine protein kinase, mTOR1, in a cell-specific manner. [45–49] Intracellular protein turnover, or the synthesis and degradation of proteins within a cell, is an important function for cell turnover, tissue repair, production of immunological proteins, and immune responses among other tasks.[44] Other than regulating protein turnover, mTOR activity impacts oxygen consumption, allocation of substrates for energy production, and defense against reactive oxidants.[50] Under stressed conditions, such as the occurrence of inflammation or trauma, overproduction of ROS by dysfunctional mitochondrion has been linked to organ failure.[51] In

order to prevent organ failure and promote cell and tissue health, mitochondrial function must be preserved. Capable of mitochondrial biogenesis modulation and reactive oxidant defense, leucine and other branched chain amino acids are necessary to maintain cell and tissue viability under stressed conditions.[52]

As an acetylated derivative of leucine, acetyl-dl-leucine has been examined by Strupp et al. to determine the efficacy of this amino acid as a treatment of cerebellar ataxia. Characterized by gait and muscle coordination problems, cerebellar ataxia is a disorder that occurs when the cerebellum becomes inflamed or damaged. In a case study involving a group of 13 patients with cerebellar ataxia, each patient was treated with acetyl-dl-leucine at a dose 5g/day for a week.[53] At the end of the case study there was overall improvement in the ataxic symptoms. While acetyl-dl-leucine has been used for over 50 years to treat vertigo and dizziness, the pharmacological mode of action of this molecule has not been fully elucidated.[53] However, based on the known functions of leucine, it is plausible that acetyl-dl-leucine reduced inflammation in the cerebellum and/or repaired the cerebellar neuron damage to restore gait and muscle coordination. Such improvements seen in patients with cerebellar ataxia after the supplementation of acetyl-dl-leucine suggests that this amino acid may be able to improve the environmental conditions of an inflamed gut.

4.3.3 Enterodiol

Due to the potential health benefits of lignans, there has been growing interest in studying these plant-derived polyphenolic compounds. Studies have shown lignan-rich foods can promote anti-tumor, anti-inflammatory, anti-viral and antihepatotoxicity properties.[54–58] Enterodiol is an enterolignan product of the metabolism of lignans by the intestinal flora. In a study conducted by Corsini et al. the immunomodulatory effects of enterodiol and its oxidized counterpart,

enterolactone, were explored. Their study found that enterodiol and enterolactone are capable of inhibiting nuclear factor-κB (NF-κB) signaling. The antioxidant activity of enterodiol and enterolactone likely prevented the ROS activation of NF-κB through inhibitory-κB degradation that subsequently suppressed TNF-α release and T cell proliferation.[58,59] Also, Corsini demonstrated enterodiol and enterolactone were able to by-pass the intestinal barrier without disrupting the integrity of the membrane to suppress cytokine production *in vitro*.[58] With the ability to modulate the immune response without harming the intestinal barrier, enterodiol may be able to shift the microbial environment of the gut from an antibiotic-stressed state back to a homeostatic state.

4.3.4 Pantetheine

A naturally occurring thiol dimer, pantethine consists of two pantetheine residues joined by a disulfide bond. When pantethine is consumed, the dimer reduces to pantetheine for the production of coenzyme A (CoA).[60,61] Pantethine is well associated with antioxidant, hypolipidemic and hypocholesterolemic properties.[62,63] Based on pantethine's effect on lipid metabolism, Van Gijsel-Bonnello et al. explored the modularity effects pantethine and pantetheine may have on the functions of T cells. Migratory and signal transduction of T cells are dependent on the lipid rafts, or the microdomains, of the cells' membrane.[64–67] Composed of high concentrations of saturated phospholipids chains interspersed with cholesterol, the rafts of a T cell membrane could be a good target for pantethine given its propensity for inhibiting fatty acid and cholesterol synthesis. Findings from Van Gijsel-Bonnello's study suggests pantethine can alter T cell membrane composition by decreasing saturated fatty acids and cholesterol while increasing polyunsaturated fatty acids.[60] Such modifications to the T cell membrane by pantethine down-regulates: cell adhesion, CXCL12-driven chemotaxis, transendothelial

migration and circulating effector T cells.[60] As for the pantetheine treatment, subtle changes to the T cell membrane were seen with no notable changes to CXCL12 binding nor T cell migration.[60] While pantetheine does not possess the desired immune altering properties that may stabilize an inflamed gut, the pantetheine dimer may be a suitable compound for immunomodulation.

4.3.5 4-Pyridoxic Acid

The catabolite 4-pyridoxic acid is a downstream product of vitamin B6 metabolism. Levels of 4-pyridoxic acid (PA) are used in conjunction with pyridoxal (PL) and pyridoxal-5'-phosphate (PLP) in the ratio, PA: (PL + PLP), to measure vitamin B6 status and systemic inflammation.[68–71] Regarded as the primary catabolic product of vitamin B6 metabolism, not much research has been done on PA beyond the its importance as a biomarker. On the contrary, vitamin B6 and its active form, PLP, have been studied for potential health benefits and identified as a cofactor for over 100 enzymatic reactions.[72] These enzymatic reactions are responsible for the metabolism of amino acids as well as carbohydrates and lipids to maintain cellular homeostasis.[72] A previous mice study by Doke et al. indicated vitamin B6 can affect the cytokine production profile and lymphocyte proliferation that characterizes an inflammatory response.[73] The cytokine profile for vitamin B6 deficient mice trended towards type 2 immunity, with enhanced T-helper 2 cell cytokine production and reduced lymphocyte proliferation.[73] Although type 2 immunity possesses host-protective properties, the type 2 cytokine production can suppress type 1 immunity against a wide range of pathogens, allowing for the development of an infection.[74] While the exact mechanisms underlying the association between plasma PLP and inflammation biomarkers have yet to be elucidated, the broad functionality of vitamin B6

and its association with inflammation warrants an examination of the immune modulating potential of vitamin B6 and its vitamers in relieving a stressed intestinal state.

4.3.6 Inosine

The purine nucleoside adenosine can modulate the immune response system and inflammation by repressing cytokine production and neutrophil function.[75,76] Formed by the degradation of adenosine, inosine is able to modulate immune responses in a similar manner to its parent compound.[77] Found in extracellular excess during times of metabolic stress, Hasko et al. demonstrated that inosine is capable of reducing the production of pro-inflammatory cytokines, such as TNF-α, IL-1, IL-12 and IFN-γ while promoting the anti-inflammatory cytokine production of IL-10.[76,78–80] In an endotoxemic mouse model, inosine was able to suppress pro-inflammatory cytokine production and improved the mortality rate, supporting the ability of inosine swaying a stressed system back into a healthy state.[76] Because inosine is a degradation product of adenosine, it is conceivable inosine plays a role in the immunomodulatory effects of adenosine. Both adenosine and inosine should be further investigated for their ability to enforce resistance against infection during an antibiotic-stressed state.

Spectral validation has verified the presence of the uracil, acetyl-dl-leucine, enterodiol, pantetheine, 4-pyridoxic acid and inosine in the healthy and/or the Candida treated mice. Despite the presence of these metabolites in the samples and literature evidence supporting their immunomodulatory effects, their functionality as candidates for resistance against infections has yet to be seen. Further experimentation is needed to determine if one or a combination of these metabolites, or their parent metabolites, are capable of shifting the gut microbiome away from

dysbiosis by reinforcing the colonization resistance to prevent infection by opportunistic pathogens.

4.4 Limitations, Considerations, and Improvements

With the methods and techniques employed as well as the assumptions made during this project, there are limitations and considerations to be taken into account when drawing conclusions. Firstly, the extraction method used for this study was developed with the preference for amino acids, which may distort the detectable metabolites in the cecum samples. Also, while untargeted LC-MS/MS analysis provides the least amount of bias for metabolite discovery, such a method does possess drawbacks. For instance, the cursory nature of untargeted analysis limits the ability of the user to optimize the ionization and fragmentation mass spectrometry parameters. The lack of mass spectrometry parameter optimization allows for a broader search while causing a loss in sensitivity by overlooking metabolites with low abundances.[81] This trade-off not only constrains the scope of the metabolic profile to high abundance metabolites but also allows for low abundance biomarkers to go undetected.

A prime example of undetected metabolites in this study would be the absence of bile acids. Synthesized from cholesterol by hepatic enzymes, bile acids are transported to the lumen of the intestines to promote digestion, transportation and absorption of nutrients, fats, and vitamins.[82] High concentrations of conjugated bile acids can be found throughout the small intestines but only roughly 5 % of the bile salts evade reabsorption back to the liver and proceeds to the large intestines.[83] Previous studies have shown the presence of bile acids in the gut and how bile acids can influence colonization resistance as well as promote situational spore germination.[19,84,85] Against expectations, primary and secondary bile acids eluded detection in

this study. The lack of detection of these bile acids is likely due to the extraction method as well as low concentrations of these molecules in the cecum samples.

It is important to note that the aim of this project was not for the development of an untargeted workflow but to identify potential colonization resistance metabolites using untargeted analysis. With that said, the framework presented in this paper should be treated as a preliminary metabolic profiling assessment. Using the presented annotation workflow, 2,241 features, or 13 %, of the queried features were annotated with some degree of confidence. This number of annotated features is comparable to the results derived from an *in silico* annotation of available genome sequences and exceeds the annotations by third-party vendors, which were typically in the hundreds.[12,30,86] Results drawn from this annotation pipeline is equal to if not greater in magnitude than alternative methods of annotation but faults in the confidence of the identifications. With this annotation workflow, it is important to note the vast number of annotations are strictly putative until further verified against a pure standard while the limited annotations provided by the third-party vendor are confirmed identifications.

Also, as previously mentioned, there are improvements that can be made to this pipeline to increase the annotation predictability power for future studies. By changing the preprocessing software to XCMS isotopes and adducts can be identified and accounted for, reducing the number of misidentifications during the annotation phase. And, the trend from the validation study indicates annotations that were predicted by at least two databases with high scores were accurate identifications. Because the validation study was comprised of a small selection of metabolites, a larger study should be conducted to determine if the trend is statistically significant. If the trend is significant, the scoring criteria for acceptable annotations can be

adjusted to include this new development to increase the predictability power of this annotation pipeline.

5.0 Future Work

5.1 Functional Testing of Metabolites

To further this work, the confirmed metabolites of uracil and acetyl-dl-leucine can be functionally tested to verify the biological significance that *C. albicans* overgrowth inhibits the development of *C. difficile* infection. Other metabolites, and/or their parent compounds, that were found to be strictly elevated in the healthy murine should be functionally tested as well, for they may be possible infection prevention interventions.

Simple *in vitro* studies can be used as a preliminary functionally assessments. To mimic the intestinal barrier and the immunological response *in vitro*, human epithelial colorectal adenocarcinoma (CaCo-2) cells and helper T cells can be co-cultured.[58,87] The co-cultured system may be subjected to an antibiotic regimen with metabolite supplementation prior to being challenged with *Clostridium difficile*. Cytokine production should be monitored throughout the study to determine the impact of the supplementation on the immune response to antibiotic use as well as to monitor the progression of the *C. difficile* infection. At the end of the study, *C. difficile* enumeration should be performed to determine if the metabolite supplementation has indeed reduced/prevented *C. difficile* outgrowth. Other aspects to examine during the *in vitro* stage would be the metabolite supplementation regime as well as the prevention potential of combination supplementation of these reinforcing metabolites.

For metabolites that exhibit the desired resistance functionality *in vitro*, *in vivo* murine studies can be used to verify the resistance functionality in the presence of a functioning

microbiota. Adapting the method used by Theriot et al., the mice will be subjected to a 10-day antibiotic treatment.[12] Prior to being challenged with *Clostridium difficile*, metabolite supplementation would begin. Time of supplementation intervention and supplementation dosage should be guided by results obtained from the *in vitro* studies. After the 30 days of treatment, the cecum samples should be harvested for microbial and metabolic analyses to determine the efficacy of the metabolite supplementation on the reduction/prevention of *C. difficile* infection. If the metabolites show promise in preventing *C. difficile* infection *in vivo*, the questions of when is the most beneficial time for supplementation intervention and the duration of supplementation should be further refined.

5.2 Incorporating other –Omic Studies

While changes in the metabolome provide insight into how the bacterial and host metabolic activities affect the host phenotype, the microbes associated with each phenotype are still unknown along with how these microbes interact with each other and the host. To further unravel the mechanisms that lead to vulnerability and the onset of infection within the gut, other '-omic' studies should be performed. Linking the findings from this metabolic analysis with associated metagenomic, transcriptomic and proteomic studies can further isolate the essential microbes and their role in host immunity. A detailed mechanistic understanding of the vulnerability to infection in the gut would be a useful tool for the development of biotherapeutic interventions.

6.0 Conclusion

Presented in this study is an investigation of antibiotic-altered intestinal metabolomes that increases susceptibility to disease. Using untargeted LC-MS/MS analysis and a novel annotation pipeline, preliminary metabolic profiles were generated. Findings from the metabolic profile analysis indicate a 66 % decrease in the baseline metabolites with the use of antibiotics, corresponding to previous findings of decreased activity due to decreased microbial diversity. Based on the hypothesis that *Candida albicans* overgrowth can prevent *Clostridium difficile* infection, the metabolomes of healthy mice and *C. albicans* treated mice were examined for similarly elevated metabolites that may confer resistance to infection. The elevated presence of uracil and acetyl-dl-leucine within both the healthy and *C. albicans*-treated metabolomes supports the possibility that *C. albicans* overgrowth may prevent *C. difficile* infection. Other metabolites, namely enterodiol, pantetheine, 4-pyridoxic acid, and inosine, which were strictly elevated in the healthy mice, or their parent metabolites, may confer resistance to infection as well. Once these metabolites are functionally tested to confirm their ability to prevent infection, the metabolite(s) can be used as a prebiotic treatment to deter pathogenic invasion in an antibiotic-stressed gut.

Appendices

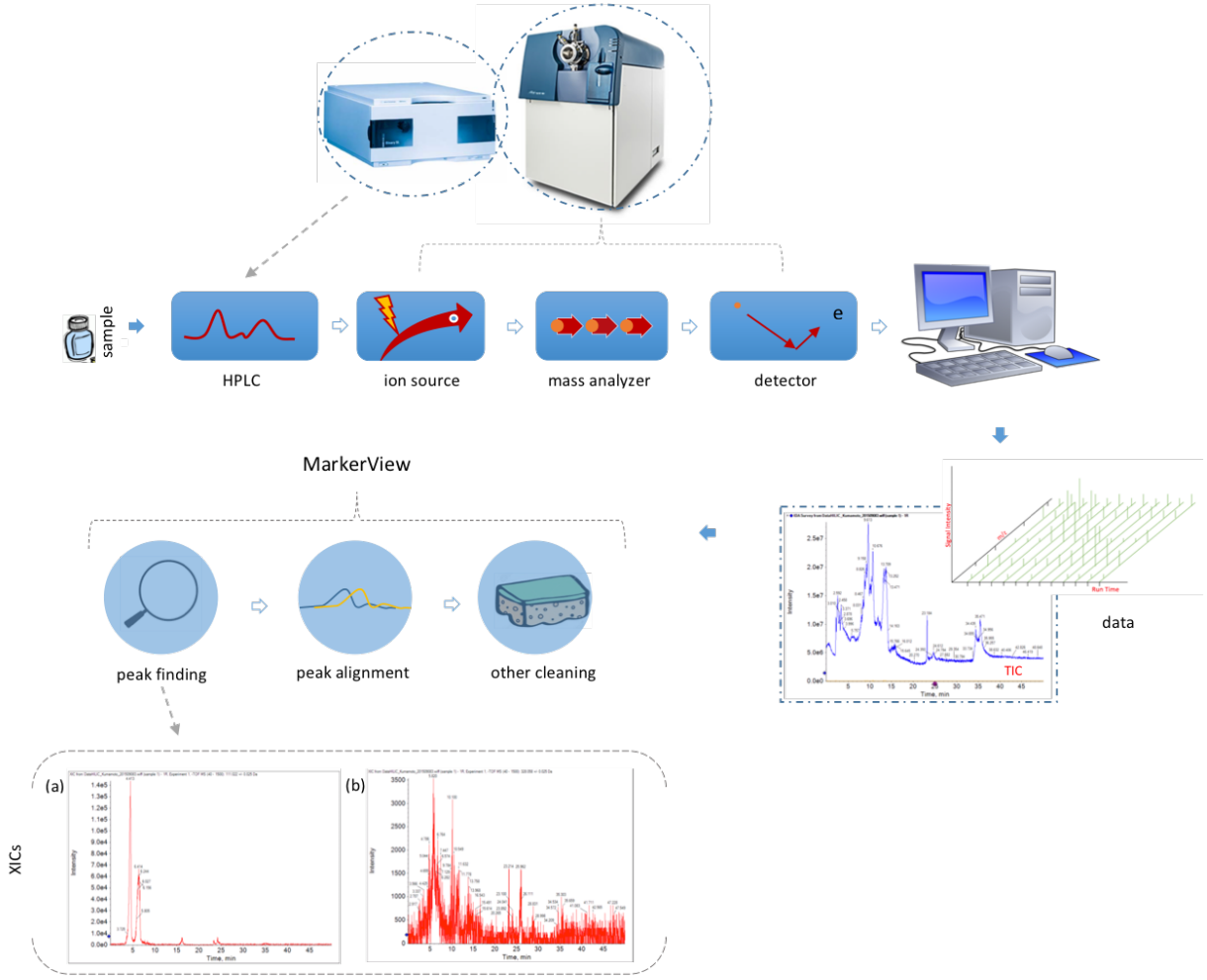
Appendix I: LC-MS/MS Operational Parameters

	HILIC		Synergi	
Mode	**-**	**+**	**-**	**+**
Instrument Parameters				
GS1 (psi)	35	35	40	40
GS2 (psi)	45	45	50	50
CUR (psi)	25	25	30	30
ISVF (V)	4500	5500	4500	5500
TEM (C)	450	450	450	450
Autosampler Temperature (C)	4	4	4	4
Column Temperature (C)	25	25	15	15
Experimental Parameters				
CE	-45	45	-45	45
CES	15	15	15	15
DP	-80	80	-80	80
Ions to Fragment	8	4*	8	8
Flowrate (uL/min)	300	300	100	100
Injection Volume	15*	10	10	10

*deviations in experimental parameters

Appendix II: Data Procurement

A sample is injected in to the HPLC for initial physical separation before entering the mass spectrometer. The separated compounds that enter the mass spectrometer are subjected to electrospray ionization to produce either protonated or deprotonated ions. After ionization, the quadrupole mass filter sorts the ions by mass-to-charge ratios (m/z) for ion fragmentation by collision with inert gas molecules to form product ions. Product ions enter the time-of-flight (TOF) region for further mass sorting before detection. The information from the detector is transferred and recorded by Analyst TF software. Raw data retrieved from the LC-MS/MS analysis comprises of m/z values for the duration of the experiment and their associated signal intensities. The total ion chromatograph (TIC) is the summation plot of all m/z signal intensities for the duration of the experiment. Programs such as MarkerView assist in sorting through the raw data to find peak features, aligning peak features across samples as well as performing additional data cleaning including the removal of sparse sample features. The extracted ion chromatographs (XICs) show the signal intensity across time for a specified m/z. As seen in the XIC(a), there are two distinct peak features that are easily identified as peaks. However, in many cases like the XIC(b), the peak features are less distinct due to the increased presence of noise. Figure adapted from May et. al.[88]



Appendix III: Untargeted Metabolomics Pipeline Statistics

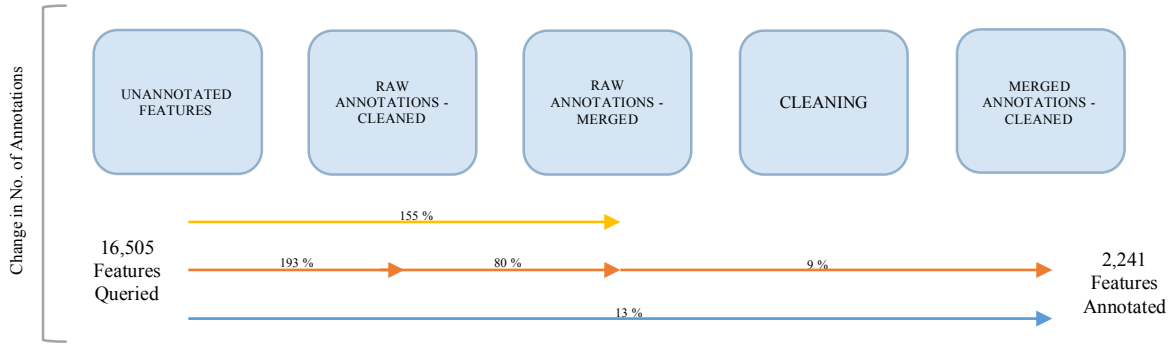
Presented here is a breakdown of the feature identifications during the metabolic profiling pipeline. 16,505 features were queried in the annotation workflow and resulted in a 1.93-fold increase in identifications across the four databases. After common features were merged, common identifications were matched, peptides were removed, duplicate compounds were removed, duplicate m/z (features with multiple identities) were removed, and feature annotations with no confidence were removed, 2,241 annotated features remained. Of the 16,505 features queried, 13 % were annotated with a degree of confidence.

		HILIC (-)	HILIC (+)	SYN (-)	SYN (+)
UNANNOTATED FEATURES	RAW	12199	1500	1500	1500
	MV FILTERED	3686	2337	5215	6290
	(BLANKS)	3667	2324	4876	5638
RAW ANNOTATIONS - CLEANED	METLIN	341	1195	89	239
	HMDB	1808	895	1781	2745
	CMF-ID	2221	1366	2798	4257
	METFRAG	3159	1582	3127	4276

		total				unique				total		unique			
		HILIC (-)	HILIC (+)	SYN (-)	SYN (+)	HILIC (-)	HILIC (+)	SYN (-)	SYN (+)	HILIC	SYN	HILIC	SYN	HILIC/SYN	total
RAW ANNOTATIONS - MERGED	METLIN, HMDB, CFM-ID, METFRAG	61	73	7	78	6	10	0	27	47	3	16	0	55	114
	METLIN, HMDB, CFM-ID	65	85	7	87	11	17	2	39	46	1	24	0	52	145
	METLIN, HMDB, METFRAG	5	6	4	4	0	2	3	1	4	0	1	0	4	11
	METLIN, CFM-ID, METFRAG	77	56	25	56	27	16	6	17	31	11	15	3	44	128
	HMDB, CFM-ID, METFRAG	326	202	241	434	74	42	119	188	133	90	55	54	224	756
	METLIN, HMDB	60	99	17	98	25	59	13	62	23	0	12	0	40	211
	METLIN, CFM-ID	90	59	21	66	56	23	12	39	24	4	17	3	29	179
	METLIN, METFRAG	11	7	3	5	7	6	0	3	1	1	1	1	3	21
	HMDB, CFM-ID	312	184	270	509	125	55	179	310	91	61	47	51	178	945
	HMDB, METFRAG	100	36	107	95	47	15	63	39	19	28	7	24	48	243
	CFM-ID, METFRAG	1905	1038	2142	3008	757	388	1189	1624	466	728	205	482	1127	5772
	METLIN	261	1073	74	153	220	1033	67	131	29	1	25	1	27	1504
	HMDB	1665	752	1763	2674	1430	621	1568	2387	86	135	42	109	238	6395
	CFM-ID	1820	1018	2388	3567	1164	554	1774	2602	280	484	160	415	680	7349
	METFRAG	710	225	712	588	451	118	506	318	94	139	56	121	216	1786

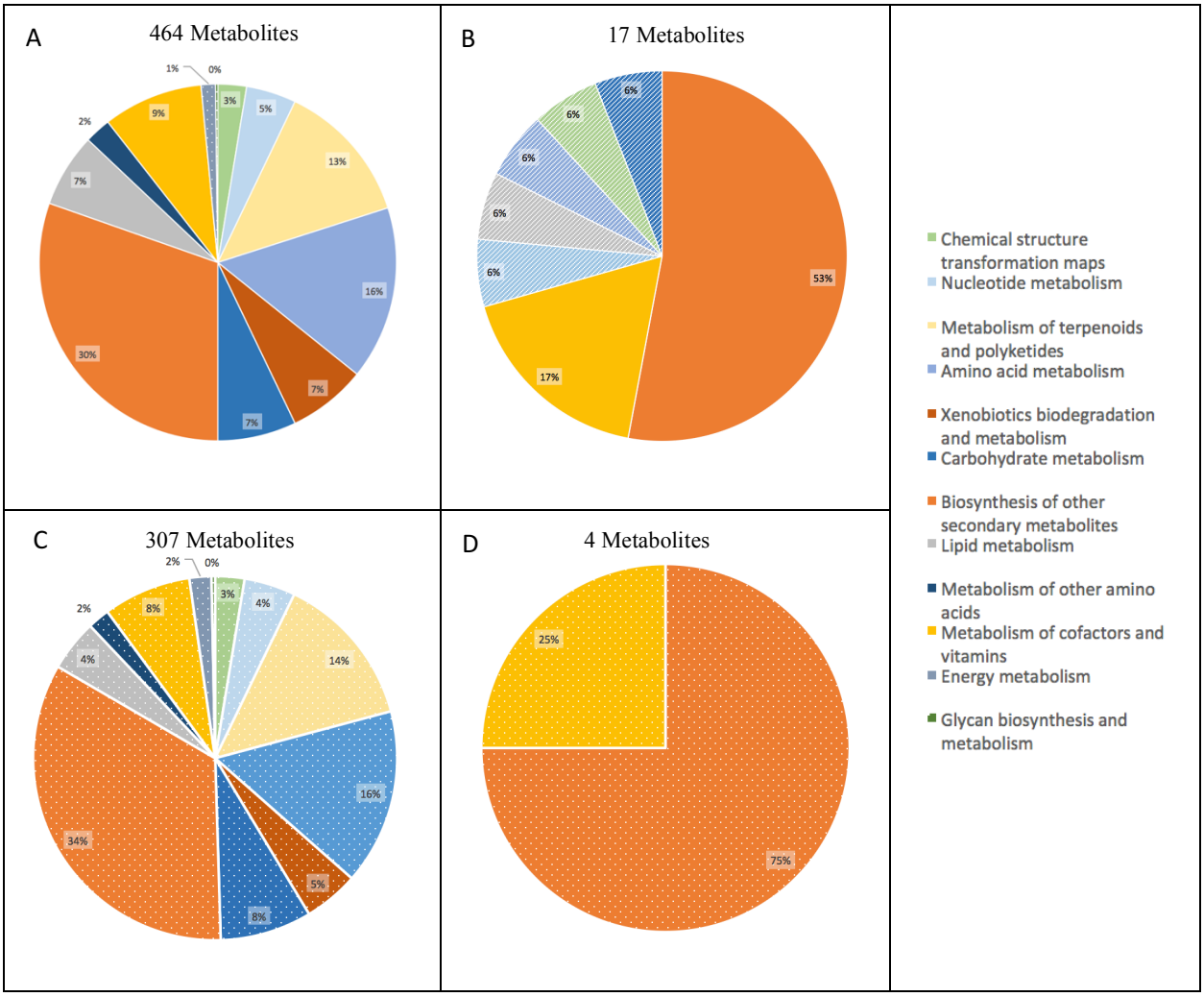
CLEANING	NO. PEPTIDES REMOVED	789
	NO. DUPLICATE COMPOUNDS REMOVED	15066
	NO. DUPLICATE M/Zs	5770
	NO. NO CONFIDENCE REMOVED	1693

		total				unique				total		unique			
		HILIC (-)	HILIC (+)	SYN (-)	SYN (+)	HILIC (-)	HILIC (+)	SYN (-)	SYN (+)	HILIC	SYN	HILIC	SYN	HILIC/SYN	total
MERGED ANNOTATIONS - CLEANED	METLIN, HMDB, CFM-ID, METFRAG	30	36	6	42	4	8	0	17	21	2	4	0	29	62
	METLIN, HMDB, CFM-ID	8	4	3	4	1	0	0	2	4	0	2	0	5	10
	METLIN, HMDB, METFRAG	3	3	1	3	0	1	1	1	2	0	1	0	2	6
	METLIN, CFM-ID, METFRAG	26	18	10	22	8	5	3	8	12	5	6	3	13	46
	HMDB, CFM-ID, METFRAG	123	68	140	178	28	20	67	79	38	54	19	32	86	331
	METLIN, HMDB	8	13	4	3	6	10	3	2	2	0	1	0	2	24
	METLIN, CFM-ID	1	0	1	1	0	0	0	1	0	0	0	0	1	2
	METLIN, METFRAG	3	1	1	1	2	1	0	1	0	0	0	0	1	5
	HMDB, CFM-ID	5	1	13	18	4	1	6	10	0	7	0	7	1	29
	HMDB, METFRAG	47	14	45	28	21	5	27	5	7	9	4	8	24	94
	CFM-ID, METFRAG	474	310	677	806	169	129	381	405	112	223	61	161	313	1619
	METLIN > 90	8	2	3	0	8	2	3	0	0	0	0	0	0	13
	HMDB														
	CFM-ID														
	METFRAG														



Appendix IV: Pathway Analysis: Significant Differences

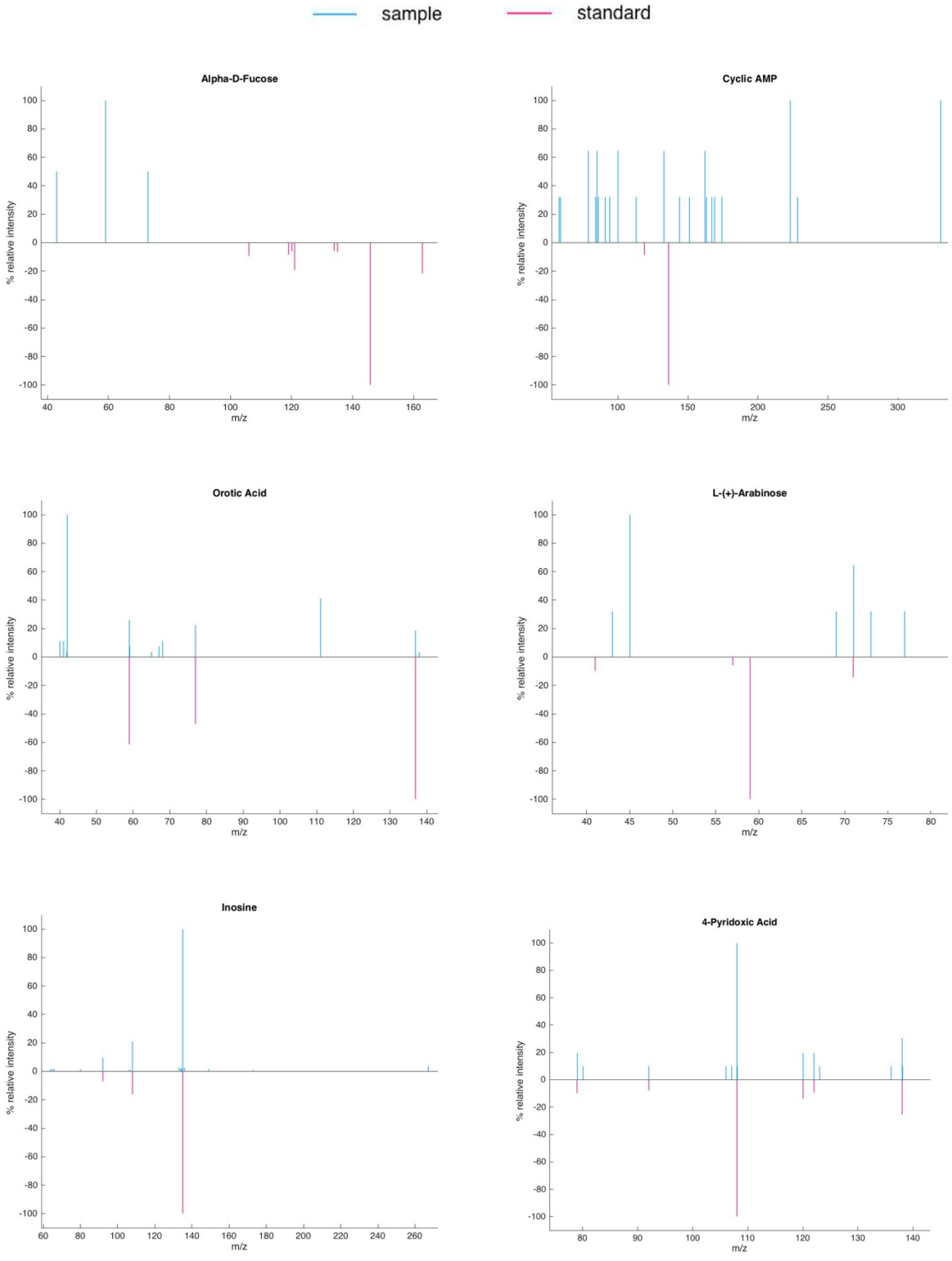
(A) All significantly different pathways between healthy and antibiotic treated mice. (B) All significantly different pathways between Candida and antibiotic treated mice. (C) Down-regulated pathways between healthy and antibiotic treated mice. (D) Down-regulated pathways between Candida and antibiotic treated mice. Solid colors represent pathways with up regulated and down regulated metabolites. Stripped pathways indicate up-regulated metabolites and dotted pathways indicate down-regulated metabolites.

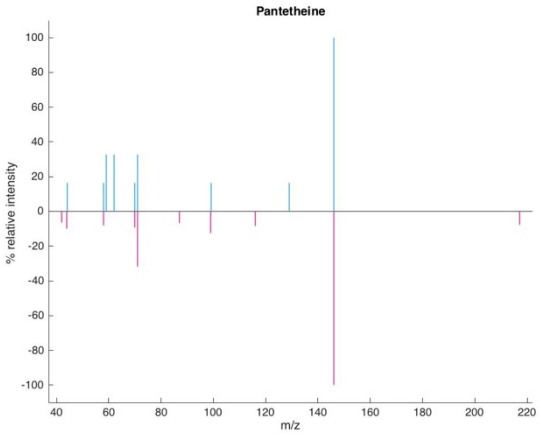


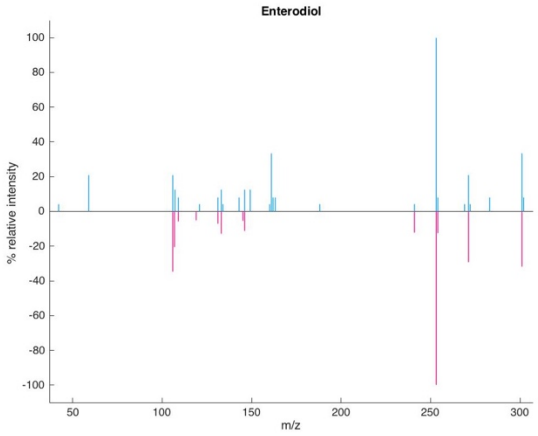
Appendix V: Metabolite Spectral Comparisons

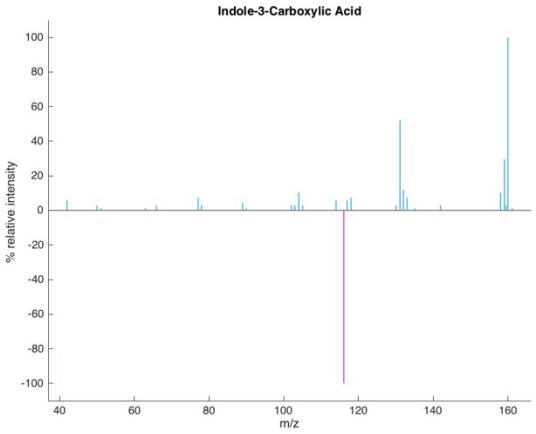
Appendix Va: Metabolite Standard vs. Sample

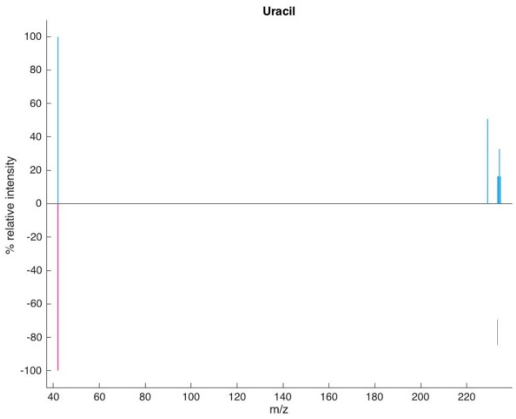
Mirror plots, in-house standards vs. sample, of all metabolites selected for validation.

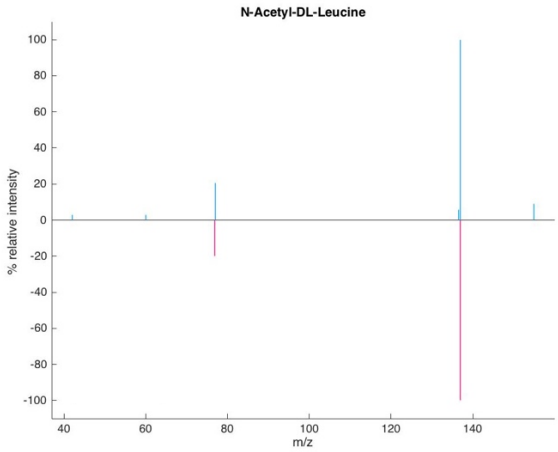



Pantetheine
% relative intensity
m/z


Enterodiol
% relative intensity
m/z

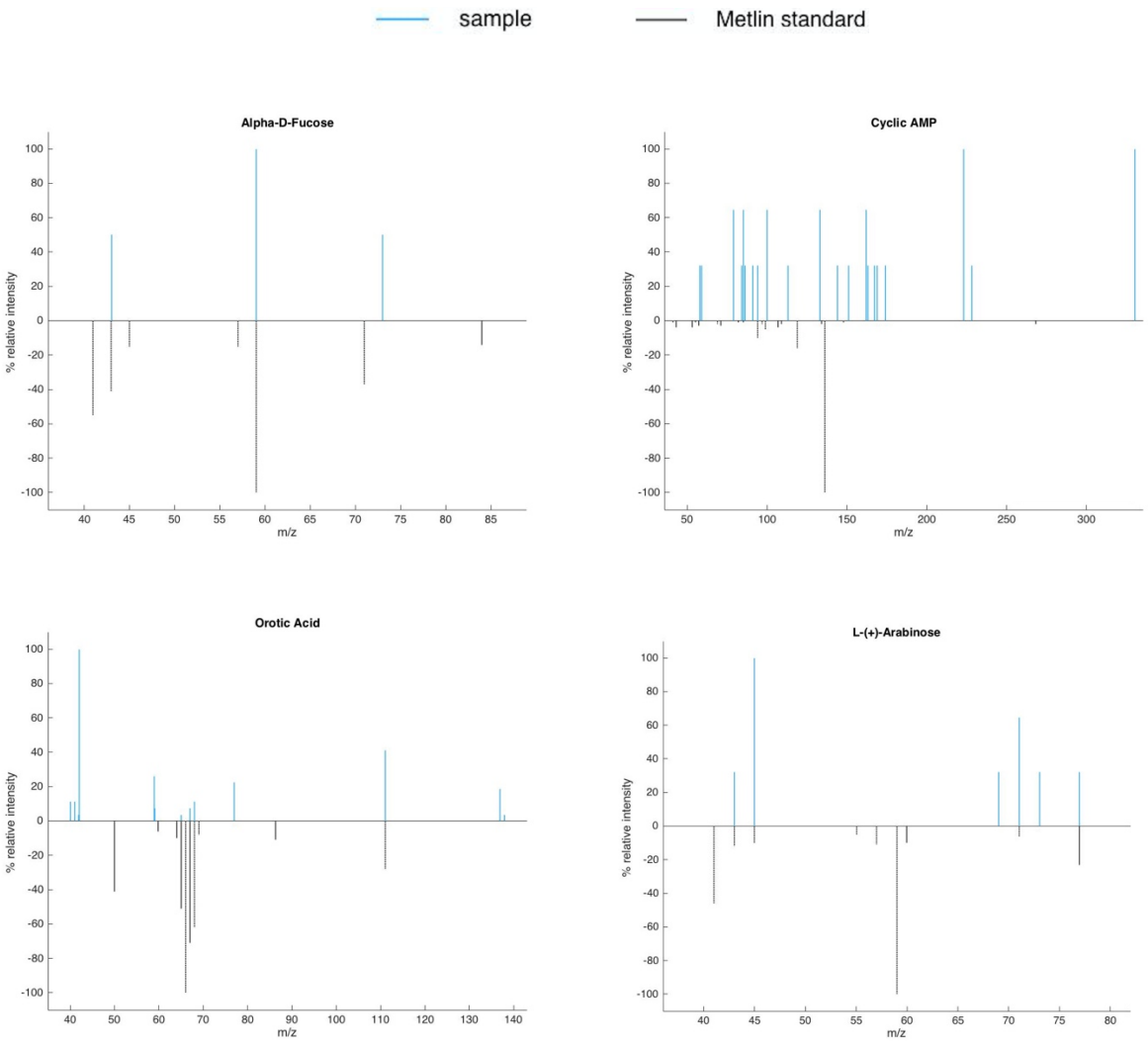

Indole-3-Carboxylic Acid
% relative intensity
m/z

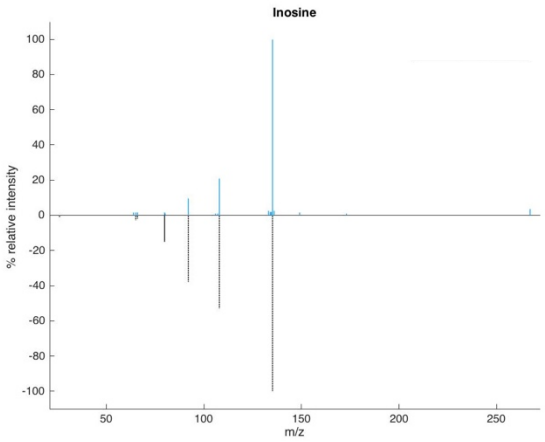

Uracil
% relative intensity
m/z

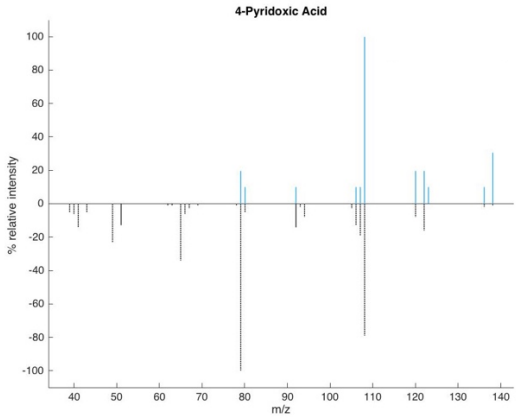

N-Acetyl-DL-Leucine
% relative intensity
m/z

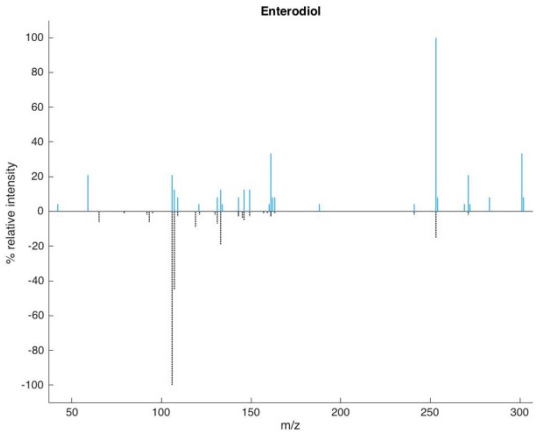
Appendix Vb: Sample vs. Metlin Standard

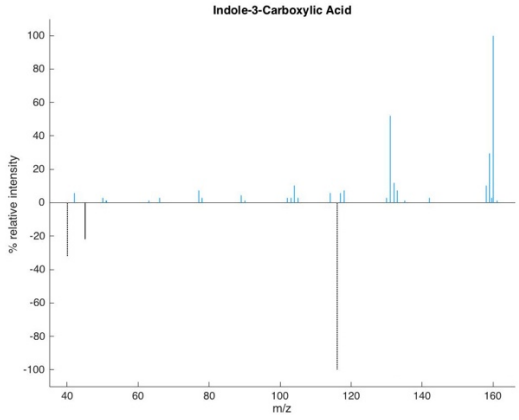
Mirror plots, sample vs. Metlin standard, for all metabolites selected for validation. Pantetheine was excluded due to lack of available Metlin data. Dotted black lines indicate *in silico* predicted fragments for a given Metlin standard. Use of predicted fragments along with experimental data in the Metlin mass fragmentation data can lead to overly-confident false predictions by Metlin, as seen with l-(+)-arabinose. When the putative l-(+)-arabinose spectra was compared to the in-house standard, the two spectra shared one fragment. But when queried, the putative l-(+)-arabinose shared four common peaks with the Metlin fragmentation data, with three of the four fragments being predicted by Metlin.

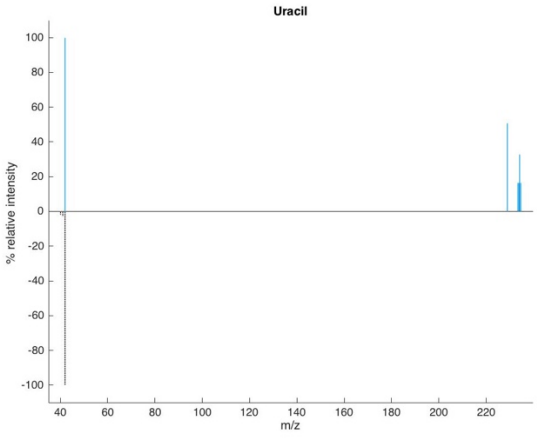


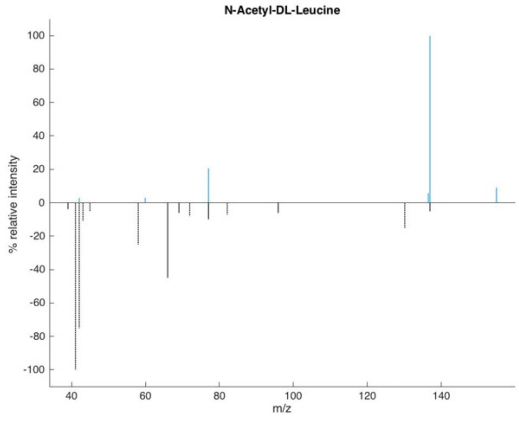

Inosine
% relative intensity
m/z


4-Pyridoxic Acid
% relative intensity
m/z


Enterodiol
% relative intensity
m/z

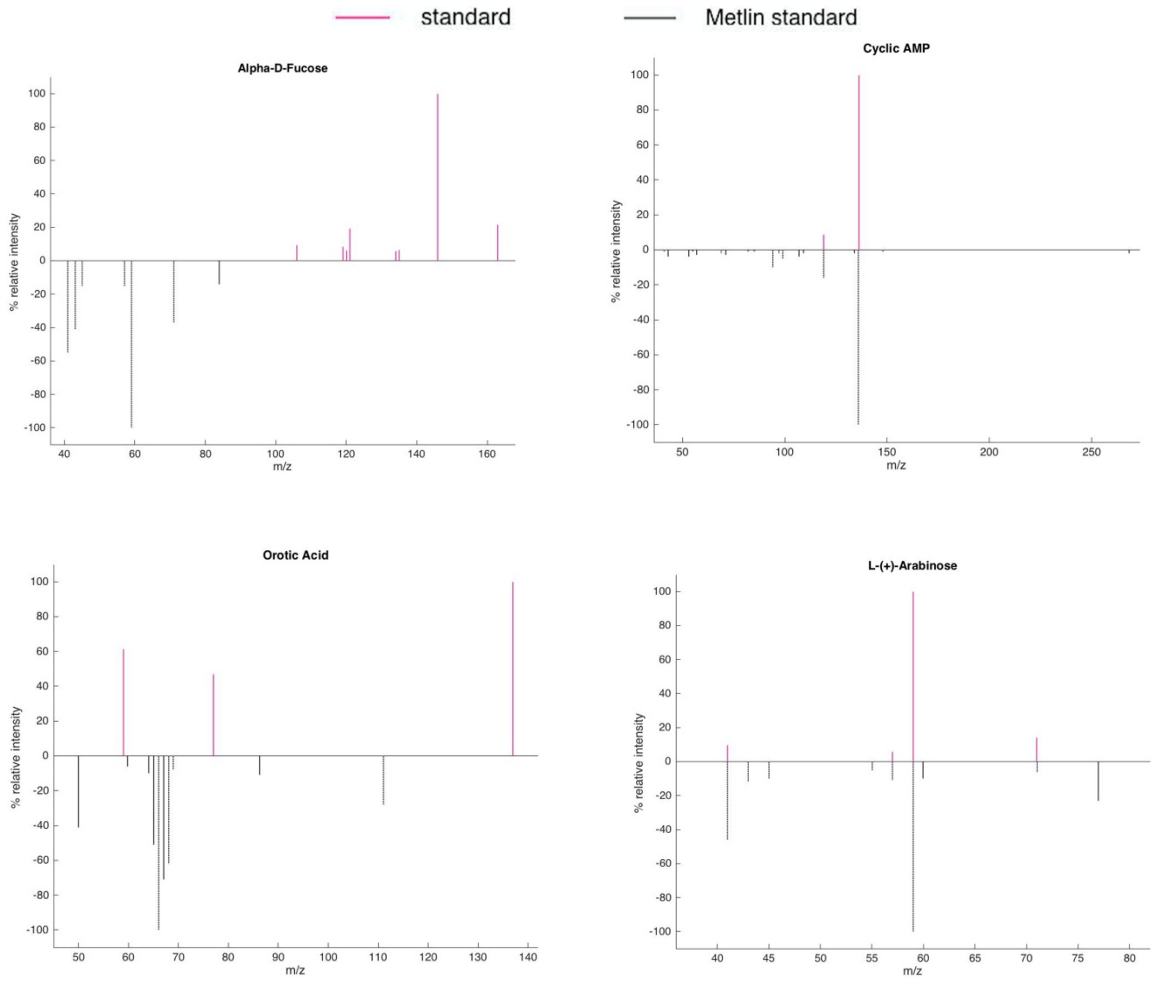

Indole-3-Carboxylic Acid
% relative intensity
m/z

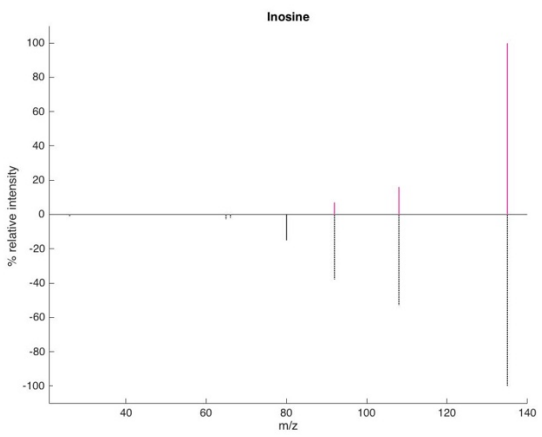

Uracil
% relative intensity
m/z

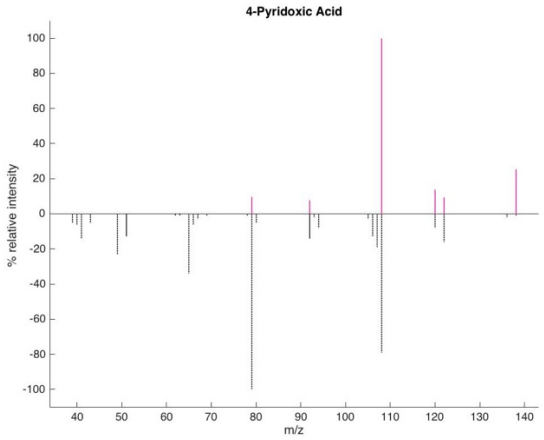

N-Acetyl-DL-Leucine
% relative intensity
m/z

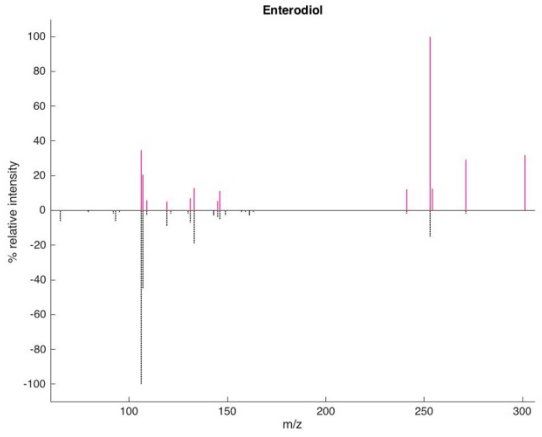
Appendix Vc: Metabolite Standard vs. Metlin Standard

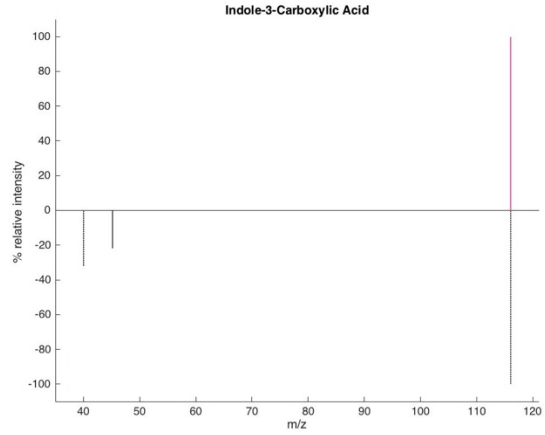
Mirror plots, in-house standard vs. Metlin standard, for all metabolites selected for validation. Pantetheine was excluded due to lack of available Metlin data. Dotted black lines indicate predicted fragments for a given Metlin standard.

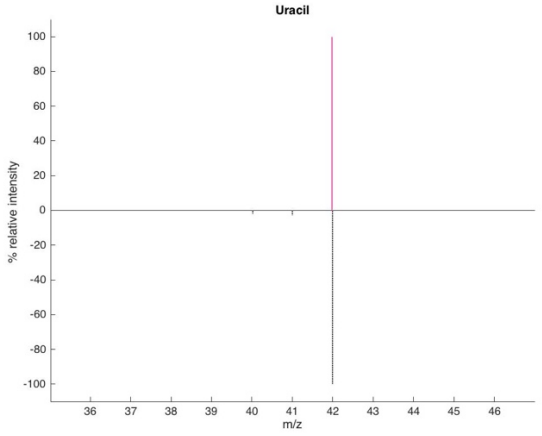


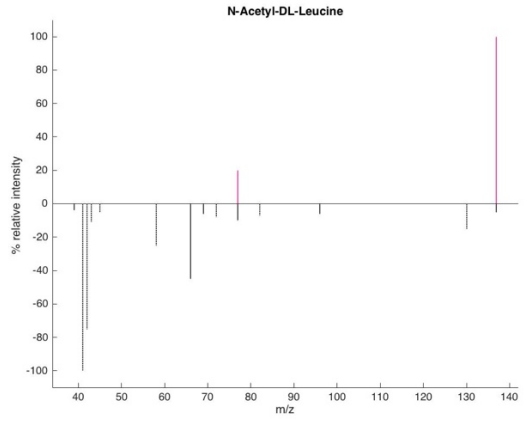

Inosine
% relative intensity
m/z


4-Pyridoxic Acid
% relative intensity
m/z


Enterodiol
% relative intensity
m/z


Indole-3-Carboxylic Acid
% relative intensity
m/z


Uracil
% relative intensity
m/z


N-Acetyl-DL-Leucine
% relative intensity
m/z

Bibliography

1. Lawley, T. D. & Walker, A. W. Intestinal colonization resistance. *Immunology* 138, 1–11 (2013).

2. Abbott, A. *Scientists Bust Myth that Our Bodies Have More Bacteria than Human Cells*. *Nature News* (2016).

3. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *BioRxiv.org* 1–21 (2016).

4. Theriot, C. M. & Young, V. B. Interactions Between the Gastrointestinal Microbiome and Clostridium difficile. 445–464 (2015). doi:10.1146/annurev-micro-091014-104115

5. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science (80-.).* 308, 1635–1638 (2005).

6. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848 (2006).

7. Blaut, M. & Clavel, T. Metabolic diversity of the intestinal microbiota: implications for health and disease. *J. Nutr.* 137, 751S–755S (2007).

8. Cochetière, M. F. D. La *et al.* Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. *J. Clin. Microbiol.* 43, 5588–5592 (2005).

9. Keeney, K. M., Yurist-doutsch, S., Arrieta, M. & Finlay, B. B. Effects of Antibiotics on Human Microbiota and Subsequent Disease. *Annu. Rev. Microbiol.* 68, 217–235 (2014).

10. Eloe-Fadrosh, E. a & Rasko, D. a. The human microbiome: from symbiosis to pathogenesis. *Annu. Rev. Med.* 64, 145–63 (2013).

11. Stecher, B. & Hardt, W. D. Mechanisms controlling pathogen colonization of the gut. *Curr. Opin. Microbiol.* 14, 82–91 (2011).

12. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nat. Commun.* (2014).

13. Theriot, C. M. & Young, V. B. Interactions Between the Gastrointestinal Microbiome and Clostridium difficile. *Annu. Rev. Microbiol.* 69, 445–461 (2015).

14. American Chemical Society International Historic Chemical Landmarks. Discovery and Development of Penicillin. Available at: http://www.acs.org/content/acs/en/education/whatischemistry/landmarks/flemingpenicillin.html. (Accessed: 1st May 2015)

15. Pultz, N. J. & Donskey, C. J. Effect of Antibiotic Treatment on Growth of and Toxin

Production by Clostridium difficile in the Cecal Contents of Mice. *Antimicrob. Agents Chemother.* 49, 3529–3532 (2005).

16. Buffie, C. G. *et al.* Profound Alterations of Intestinal Microbiota following a Single Dose of Clindamycin Results in Sustained Susceptibility to Clostridium difficile-Induced Colitis. *Infect. Immun.* 80, 62–73 (2012).

17. Leffler, D. A. & Lamont, J. T. Clostridium difficile Infection. *N. Engl. J. Med.* 372, 1539–1548 (2015).

18. McCollum, D. L. & Rodriguez, J. M. Detection, Treatment, and Prevention of Clostridium difficile Infection. *Clin. Gastroenterol. Hepatol.* 10, 581–592 (2012).

19. Koenigsknecht, M. J. *et al.* Dynamics and Establishment of Clostridium difficile Infection in the Murine Gastrointestinal Tract. *Infect. Immun.* 83, 934–941 (2015).

20. Deusch, S., Tilocca, B., Camarinha-Silva, A. & Seifert, J. News in livestock research — use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. *Comput. Struct. Biotechnol. J.* 13, 55–63 (2015).

21. Methé, B. a *et al.* A framework for human microbiome research. *Nature* 486, 215–221 (2012).

22. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* 473, 174–180 (2011).

23. Sellick, C. A. *et al.* Evaluation of extraction processes for intracellular metabolite profiling of mammalian cells: Matching extraction approaches to cell type and metabolite targets. *Metabolomics* 6, 427–438 (2010).

24. Contrepois, K., Jiang, L. & Snyder, M. Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)–Mass Spectrometry. *Mol. Cell. Proteomics* 14, 1684–1695 (2015).

25. Bajad, S. U. *et al.* Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* 1125, 76–88 (2006).

26. Lu, W., Kimball, E. & Rabinowitz, J. D. A high-performance liquid chromatography-tandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites. *J. Am. Soc. Mass Spectrom.* 17, 37–50 (2006).

27. Allen, F., Pon, A., Wilson, M., Greiner, R. & Wishart, D. CFM-ID: A web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* (2014).

28. Wolf, S., Schmidt, S., Müller-Hannemann, M. & Neumann, S. In silico fragmentation for

computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* 11, 148 (2010).

29. Bashan, A. *et al.* Universality of human microbial dynamics. *Nature* 534, 259–262 (2016).

30. Jump, R. L. P. *et al.* Metabolomics Analysis Identifies Intestinal Microbiota-Derived Biomarkers of Colonization Resistance in Clindamycin-Treated Mice. *PLoS One* 9, (2014).

31. Kuhl, C., Tautenhahn, R. & Neumann, S. *LC-MS Peak Annotation and Identification with CAMERA*. (2016).

32. Metabolomics, S. C. for. Scripps Center for Metabolomics. Available at: https://metlin.scripps.edu/index.php. (Accessed: 16th December 2016)

33. Wishart, D. S. *et al.* HMDB: Database Statistics. doi:10.1093/nar/gkl923

34. Tautenhahn, R. *et al.* An accelerated workflow for untargeted metabolomics using METLIN database. *Nat. Biotechnol.* 30, 826–828 (2012).

35. Mylonas, R. *et al.* X-Rank: a robust algorithm for small molecule identification using tandem mass spectrometry. *Anal. Chem.* 81, 7604–7610 (2009).

36. Antonopoulos, D. A. *et al.* Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect. Immun.* 77, 2367–2375 (2009).

37. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502, 96–99 (2013).

38. Lee, K. A. *et al.* Bacterial-derived uracil as a modulator of mucosal immunity and gut-microbe homeostasis in drosophila. *Cell* 153, 797–811 (2013).

39. Buchon, N., Broderick, N. A., Chakrabarti, S. & Lemaitre, B. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in Drosophila. *Genes Dev.* 23, 2333–2344 (2009).

40. Lee, W. J. Bacterial-modulated host immunity and stem cell activation for gut homeostasis. *Genes Dev.* 23, 2260–2265 (2009).

41. Wong, A. C. N., Vanhove, A. S. & Watnick, P. I. The interplay between intestinal bacteria and host metabolism in health and disease: lessons from Drosophila melanogaster. *Dis. Model. Mech.* 9, 271–281 (2016).

42. Kim, S.-J. *et al.* Dynamic Response of Mycobacterium vanbaalenii PYR-1 to BP Deepwater Horizon Crude Oil. *Appl. Environ. Microbiol.* 81, 4263–4276 (2015).

43. Apidianakis, Y. & Rahme, L. G. Drosophila melanogaster as a model for human intestinal infection and pathology. *Dis. Model. Mech.* 4, 21–30 (2011).

44. Wu, G. Amino acids: Metabolism, functions, and nutrition. *Amino Acids* 37, 1–17 (2009).

45. Escobar, J. *et al.* Physiological rise in plasma leucine stimulates muscle protein synthesis in neonatal pigs by enhancing translation initiation factor activation. *Am. J. Physiol. Endocrinol. Metab.* 288, E914-921 (2005).

46. Escobar, J. *et al.* Regulation of cardiac and skeletal muscle protein synthesis by individual branched-chain amino acids in neonatal pigs. *Am. J. Physiol. Endocrinol. Metab.* 290, E612-621 (2006).

47. Meijer, A. J. & Dubbelhuis, P. F. Amino acid signalling and the integration of metabolism. *Biochem. Biophys. Res. Commun.* 313, 397–403 (2004).

48. Yao, K. *et al.* Dietary Arginine Supplementation Increases mTOR Signaling Activity in Skeletal Muscle of Neonatal Pigs. *J. Nutr.* 138, 867–872 (2008).

49. Mattick, J. S. A., Kamisoglu, K., Ierapetritou, M. G., Androulakis, I. P. & Berthiaume, F. Branched-chain amino acid supplementation: impact on signaling and relevance to critical illness. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 5, 449–460 (2013).

50. Schieke, S. M. *et al.* The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity. *J. Biol. Chem.* 281, 27643–27652 (2006).

51. Kozlov, A. V *et al.* Mitochondrial dysfunction and biogenesis: do ICU patients die from mitochondrial failure? *Ann. Intensive Care* 1, 41 (2011).

52. Valerio, A., D'Antona, G. & Nisoli, E. Branched-chain amino acids, mitochondrial biogenesis, and healthspan: An evolutionary perspective. *Aging (Albany. NY).* 3, 464–478 (2011).

53. Strupp, M. *et al.* Effects of acetyl-DL-leucine in patients with cerebellar ataxia: a case series. *J. Neurol.* 260, 2556–2561 (2013).

54. Aldercreutz, H. & Mazur, W. Phyto-oestrogens and Western disease. *Ann. Med.* 29, 95–120 (1997).

55. Raffaelli, B., Hoikkala, A., Leppälä, E. & Wähälä, K. Enterolignans. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 777, 29–43 (2002).

56. Duncan, A. M., Phipps, W. R. & Kurzer, M. S. Phyto-oestrogens. *Best Pract. Res. Clin. Endocrinol. Metab.* 17, 253–271 (2003).

57. Arts, I. C. W. & Hollman, P. C. H. Polyphenols and disease risk in epidemiologic studies. *Am. J. Clin. Nutr.* 81, 317S–325S (2005).

58. Corsini, E. *et al.* Enterodiol and Enterolactone Modulate the Immune Response by Acting on Nuclear Factor-κB (NF-κB) Signaling. *J. Agric. Food Chem.* 58, 6678–6684 (2010).

59. Hu, C., Yuan, Y. V. & Kitts, D. D. Antioxidant activities of the flaxseed lignan secoisolariciresinol diglucoside, its aglycone secoisolariciresinol and the mammalian lignans enterodiol and enterolactone in vitro. *Food Chem. Toxicol.* 45, 2219–2227 (2007).

60. van Gijsel-Bonnello, M. *et al.* Pantethine Alters Lipid Composition and Cholesterol Content of Membrane Rafts, With Down-Regulation of CXCL12-Induced T Cell Migration. *J. Cell. Physiol.* 230, 2415–2425 (2015).

61. Rock, C. O., Calder, R. B., Karim, M. a & Jackowski, S. Pantothenate kinase regulation of the intracellular concentration of coenzyme A. *J. Biol. Chem.* 275, 1377–1383 (2000).

62. Panchuk, R. *et al.* Specific antioxidant compounds differentially modulate cytotoxic activity of doxorubicin and cisplatin: in vitro and in vivo study. *Croat. Med. J.* 55, 206–217 (2014).

63. Cighetti, G., Puppo, M. Del, Paroni, R. & Fiorica, E. Pantethine inhibits cholesterol and fatty acid syntheses and stimulates carbon dioxide formation in isolated rat hepatocytes. *J. Lipid Res.* 28, 152–161 (1987).

64. Simons, K. & Toomre, D. Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* 1, 31–39 (2000).

65. Nguyen, D. H., Giri, B., Collins, G. & Taub, D. D. Dynamic reorganization of chemokine receptors, cholesterol, lipid rafts, and adhesion molecules to sites of CD4 engagement. *Exp. Cell Res.* 304, 559–569 (2005).

66. Xavier, R., Brennan, T., Li, Q., McCormack, C. & Seed, B. Membrane compartmentation is required for efficient T cell activation. *Immunity* 8, 723–732 (1998).

67. Horejsí, V. *et al.* GPI-microdomains: A role in signalling via immunoreceptors. *Immunol. Today* 20, 356–361 (1999).

68. Ulvik, A. *et al.* Evidence for increased catabolism of vitamin B-6 during systemic inflammation. *Am. J. Clin. Nutr.* 100, 250–255 (2014).

69. Morris, M. S., Sakakeeny, L., Jacques, P. F., Picciano, M. F. & Selhub, J. Vitamin B-6 intake is inversely related to, and the requirement is affected by, inflammation status. *J. Nutr.* 140, 103–110 (2010).

70. Sakakeeny, L. *et al.* Plasma Pyridoxal-5-Phosphate Is Inversely Associated with Systemic Markers of Inflammation in a Population of U.S. Adults. *J. Nutr.* 142, 1280–1285 (2012).

71. Paul, L., Ueland, P. M. & Selhub, J. Mechanistic perspective on the relationship between pyridoxal 5'-phosphate and inflammation. *Nutr. Rev.* 71, 239–244 (2013).

72. da Silva, V. R., Russell, K. A. & Gregory, Jesse F, I. in *Present Knowledge in Nutrition* 307–320 (2012).

73. Doke, S., Inagaki, N., Hayakawa, T. & Tsuge, H. Effects of vitamin B6 deficiency on cytokine levels and lymphocytes in mice. *Biosci. Biotechnol. Biochem.* 62, 1008–1010 (1998).

74. Wynn, T. a. Type 2 cytokines: mechanisms and therapeutic strategies. *Nat. Rev. Immunol.* 15, 271–282 (2015).

75. Antonioli, L. *et al.* Inhibition of adenosine deaminase attenuates inflammation in experimental colitis. *J Pharmacol Exp Ther* 322, 435–442 (2007).

76. Haskó, G. *et al.* Inosine inhibits inflammatory cytokine production by a posttranscriptional mechanism and protects against endotoxin-induced shock. *J. Immunol.* 164, 1013–1019 (2000).

77. Barankiewicz, J. & Cohen, A. Purine nucleotide metabolism in resident and activated rat macrophages in vitro. *Eur. J. Immunol.* 15, 627–631 (1985).

78. Rego, A. C., Santos, M. S. & Oliveira, C. R. Adenosine Triphosphate Degradation Products After Oxidative Stress and Metabolic Dysfunction in Cultured Retinal Cells. *J. Neurochem.* 69, 1228–1235 (1997).

79. Wang, T., Sodhi, J., Mentzer Jr., R. M. & Van Wylen, D. G. L. Changes in interstitial adenosine during hypoxia: relationship to oxygen supply: demand imbalance, and effects of adenosine deaminase. *Cardiovasc. Res.* 28, 1320–1325 (1994).

80. Bell, M. J. *et al.* Interstitial Adenosine, Inosine, and Hypoxanthine Are Increased after Experimental Traumatic Brain Injury in the Rat. *J. Neurotrauma* 15, 163–170 (2009).

81. Want, E. J., Cravatt, B. F. & Siuzdak, G. The expanding role of mass spectrometry in metabolite profiling and characterization. *ChemBioChem* 6, 1941–1951 (2005).

82. Chiang, J. Y. L. Bile acids: regulation of synthesis. *J. Lipid Res.* 50, 1955–1966 (2009).

83. Ridlon, J. M., Kang, D. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* 47, 241–259 (2006).

84. Paredes-Sabja, D., Shen, A. & Sorg, J. a. Clostridium difficile spore biology: Sporulation, germination, and spore structural proteins. *Trends Microbiol.* 22, 406–416 (2014).

85. Bibbò, S. *et al.* Role of microbiota and innate immunity in recurrent clostridium difficile infection. *J. Immunol. Res.* 2014, (2014).

86. Vogt, S. L., Pena-Diaz, J. & Finlay, B. B. Chemical communication in the gut: Effects of microbiota-generated metabolites on gastrointestinal bacterial pathogens. *Anaerobe* 34, 106–115 (2015).

87. Kleiveland, C. R. in *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models* 293–304 (2015).

88. May, C. *et al.* in *Data Mining in Proteomics* 696, 8 (2011).