Wright State University
CORE Scholar

Browse all Theses and Dissertations | Theses and Dissertations

2015

Differences in Fecal Metabolite Profiles from Geographically Distinct Populations of Adolescents

Jessica Moncivaiz
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

 Part of the Molecular Biology Commons

Repository Citation

Moncivaiz, Jessica, "Differences in Fecal Metabolite Profiles from Geographically Distinct Populations of Adolescents" (2015). *Browse all Theses and Dissertations*. 1628.
https://corescholar.libraries.wright.edu/etd_all/1628

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

Differences in fecal metabolite profiles from geographically distinct populations of adolescents

A thesis submitted in partial fulfillment of the requirements

For the degree of Masters of Science

By:

Jessica Danielle Moncivaiz

B.S., Wright State University, 2014

2015

Wright State University

Wright State University

Graduate School

Date: November, 2015

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Jessica Moncivaiz ENTITLED Differences in fecal metabolite profiles from geographically distinct populations of adolescents BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science

_____

Oleg Paliy, Ph.D.

Thesis Director

_____

Madhavi Kadakia, Ph.D.

Chair, BMB Department

College of Science and Mathematics

Committee on Final Examination

_____

Oleg Paliy, Ph.D.

_____

Weiwen Long, Ph.D.

_____

Nicholas Reo, Ph.D.

_____

Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

Abstract

Moncivaiz, Jessica Danielle. M.S., Biochemistry and Molecular Biology M.S. Program, Wright State University, 2015. Differences in fecal metabolite profiles from geographically distinct populations of adolescents.

Microbiota of the gastrointestinal tract have a variety of functions within the human body. They participate in protection of the host from pathogens, aid in immune system development and regulation, and carry out a variety of metabolic functions. This study focuses on the ability of gut microbiota to create metabolites through the degradation of food products. Using ^{1}H NMR on fecal water extracts, I compared the metabolite profiles of two geographically distinct cohorts: healthy adolescents from Egypt (n=28) and healthy adolescents from the United States (n=14). Multivariate statistical analyses of binned NMR data confirmed that samples separated into groups corresponding to sample class. Quantification of metabolites revealed several metabolites that differed between groups. For example, levels of short chain fatty acids were higher in the Egyptian adolescents and most quantified amino acids were higher in the US adolescents. Multivariate statistical analyses of the quantified metabolite data showed separation based on the variability within the samples and placed samples into the correct class. Therefore, I concluded that fecal metabolite profiles differ between Egyptian and United States adolescents, and that these differences in metabolite levels may be linked to dietary differences between these two studied cohorts.

Table of Contents

List of Figures

Page

List of Tables

Page

Acknowledgements

The many months of hard work and research accomplished for my thesis could have not been done without the help of certain people. First, I would like to thank my advisor, Dr. Oleg Paliy, for his guidance and assistance during my time in the lab. I would also like to acknowledge Dr. Nicholas Reo and his laboratory for providing ^{1}H NMR analyses, because our work would not have been possible without their time and hard work. I would also like to thank my lab-mates Vijay Shankar, Richard Agans, and Alex Gordon for their critiques and advice throughout this study. I would like to thank the Biochemistry and Molecular Biology Department for their support. Finally, I would like to thank my family and especially my parents Joanna Moncivaiz, Steve Bilderback, and Dan Moncivaiz for always supporting me and inspiring me to pursue my dreams in the science industry.

1. Introduction

Microbes

Microbes are very diverse organisms that cannot be seen by the naked eye but are abundant in every environment on earth. They have the capability to thrive in air, soil, rock, and even water [1]. They can be classified into six different groups': bacteria, viruses, fungi, protozoa, algae, and archaea [2]. Many communities of microbes are located in these different environments, ranging from hundreds to even thousands of species [3]. These communities are important for a variety of ecosystem processes including energy flow, element cycling, and biomass production. They can also play a role in energy and nutrient processing [3]. Due to the complex nature of these communities and their interactions with the host and within themselves, attempts to analyze them have been challenging.

Identification of microbes

Despite the many challenges of studying microbes, researchers have made significant progress in understanding them through the advancements in molecular and high through-put technologies. Classical identification of microbiota located in areas like the gastrointestinal tract (GIT) have been studied through cultivation on Petri dishes. However, a lot of bacteria located in the GIT are hard to cultivate due their inability to survive aerobic conditions [4]. To obtain a broader view of the communities present in the GIT, molecular techniques have been developed for the sequencing of the small ribosomal subunit RNA (16S rRNA) gene. These approaches use polymerase chain reaction (PCR) primers, complementary to flanking conserved regions of the 16S rRNA

gene, selecting this gene for amplification and enriching it in a sample. The resulting amplicons are further scrutinized by certain variable regions to classify sequences into operational taxonomic units (OTUs) [5]. These OTU profiles allow scientists to differentiate between samples by analyzing the microbial community composition. These types of analyses are critical to better understand microbial functions and their relationships to human health.

Microbes association with human health

Microbiota-human interactions are generally thought to be beneficial to health, but can be harmful under certain conditions [6]. For example, the skin is normally inhabited by phyla of bacteria such as Actinobacteria, Bacteroidetes, Firmicutes, or Proteobacteria. This commensal bacteria of the skin can produce defense molecules or natural antibiotics to protect form pathogens [7]. However, increased colonization of *S. aureus* (a member of the Firmicutes phyla) can cause a chronic, inflammatory infection called atopic dermatitis (AD) [8]. According to a review by Wade, the human oral cavity is dominated by genera such as *Actinomyces, Streptococcus, Neisseria,* and *Veillonella* [9]. These commensal microbiota of the oral cavity defend against pathogens by colonization resistance and can also produce molecules to protect the host from infection. However, when species of microbiota that produce acid shift the microbiome in the oral cavity, dental carries (tooth decay) can occur and dissolve the tooth structure [9]. Finally, another important example of this interaction are the human GIT microbiota and their ability to interact with metabolites, which can contribute to host health. This relationship will be explained in much greater detail in the following sections.

Human gastrointestinal tract microbiota establishment

The human GIT is made up of epithelial cells, lamina propia, and muscularis mucosae and is home to around 10^{14} microbial cells [10]. The GIT microbiota are established at birth. Prior to birth the infants microbiota are undifferentiated across all body habits but based on the birthing method (vaginal or caesarian section) and the type of feeding (breast vs. formula) the developing microbiota can change composition [11]. Vaginally delivered infants' GIT microbiota are dominated by genera such as *Lactobacillus, Prevotella,* and *Atopobium,* bacteria found in the vaginal canal, whereas infants born via C-section are dominated by Staphylococci, bacteria found on the skin [12]. After about a year, the infants microbiota profile starts to become similar to those of adults, having abundance of the phyla's Bacteroidetes and Firmicutes [13, 14]. These varying factors can contribute to a difference in microbial composition found within the GIT.

Human gastrointestinal tract microbiota distribution and differences

Not only do gut microbiota profiles differ by age but they also differ with respect to the region of the GIT they inhabit. The GIT is a large muscular tube that includes the mouth, esophagus, stomach, small intestine (duodenum, jejunum, and ileum), and large intestine (appendix, cecum, colon, and rectum) [15]. The microbiota differ due to physicochemical conditions such as pH, motility, redox potential, nutrients, and secretions by the host. As the GIT progresses through the body the microbial cell density (per gram of intestinal content) increases from 10^{1}-10^{4} cells in the stomach and duodenum, 10^{4}-10^{8} cells in the jejunum and ileum, and 10^{10}-10^{12} cells in the colon [16].

Along with the density increase the type of bacteria present in various sections of the GIT is different. For example, in a study done by Frank et. al., the colon and small intestines were analyzed and found to have different microbiota compositions. The small intestine had a higher abundance of *Bacillus* and a lower abundance of Bacteroidetes and family *Lachnospiraceae* [17]. Overall, compared to the large intestine the small intestine contained less phylogenetic diversity of Bacteroidetes and Firmicutes [17]. Differences within the sections of the GIT can cause a difference in microbiota composition.

Functions of microbiota of the gastrointestinal tract

The microbiota of the GIT have many critical functions including protection against pathogens, development of the immune system, and metabolic functions. For example, commensal bacteria protect the gut environment from pathogenic bacteria through competitive inhibition [10]. The commensal bacteria also produce toxins and metabolites that can inhibit or stunt the growth of potential pathogenic bacteria. If this protective layer of commensal bacteria is disrupted by antibiotics then pathogenic bacteria such as *Clostridium difficile* can colonize and flourish. This bacteria can cause pathophysiological conditions such as *Clostridium difficile* infection (CDI) which presents symptoms like diarrhea, nausea, and abdominal cramping for the host [18]. Therefore, the competitive environment against pathogenic bacteria established by the commensal bacteria is very important for the health of the host.

Another function of the GIT microbiota is its role in the development of the immune system and response. The gut immune system consist of peyer's patches, lamina propia lymphocytes, intra-epithelial lymphocytes, and mesenteric lymph nodes that have

gut-associated lymphoid tissue (GALT) [10]. IgA, an important immunoglobulin secreted into the GIT can coat and protect commensal bacteria and also respond to pathogens on mucosal surfaces. It has been shown that IgA is important for maintaining GIT microbiota homeostasis and composition [19]. A study by Suzuki et. al., showed that mice with a deficiency of activation-induced cytidine deaminase (AID) produced more IgM than IgA. This dysregulated the GIT microbiota community and caused increase activation of GALT [19]. Dysregulation of the immune system can result in further host related problems, including a variety of autoimmune diseases.

Microbiota are also important for a variety of metabolic functions due to their ability to create metabolites and vitamins that can be used by the host through the degradation of various food products. For example, a review by Tremaroli et. al. showed GIT microbiota can degrade non-digestible carbohydrates, that cannot be degraded earlier in the GIT by the host, including cellulose, xylans, resistant starch, and inulin [20]. The end products of this metabolism creates short chain fatty acids (SCFAs) acetate, propionate, and butyrate [20]. Peptides and proteins can also be metabolized by GIT microbiota to produce amino acids, SCFAs, and even some toxic substances such as ammonia and indols [21]. Gut microbiota can also metabolize dietary products, such as choline, into methylamines [11]. Understanding the ability of microbiota in the GIT to regulate metabolic functions can help researchers to better understand their impact on overall human health.

Metabolite functions and importance in human health

Short chain fatty acids

A review by Macfarlane et. al. highlights SCFAs as a major metabolic end products produced in the human large intestine by bacteria [22]. They are primarily produced from digestion of polysaccharides, oligosaccharides, and proteins [20]. Enzymes degrade these large macromolecules and allow the microbiota to ferment the sugars. Fermentation substrates can include starches and plant cell-wall polysaccharides (celluloses, pectins, and xylans) that are not degraded or absorbed in the small intestine. The main SCFAs produced by this fermentation are acetate, propionate, and butyrate. However, other SCFAs such as formate and valerate are formed during catabolism of branched-chain amino acids [20]. Even lactate, ethanol, and succinate, can to some extent, be metabolized into SCFAs in the GIT [22].

SCFAs have many effects in the host and have been implicated in several health benefits. SCFAs can be used by the host as energy substrates for colonic epithelium, peripheral tissues inflammation modulators, vasodilators, and promote wound healing [20]. More specifically, butyrate enhances fatty acid oxidation and thermogenesis in muscle and liver tissue [23]. Propionate and butyrate are known to activate intestinal gluconeogenesis which promotes metabolic benefits on weight and glucose control [23]. Acetate can reduce appetite by changing expression of neuropeptides through the activation of the tricarboxylic acid cycle (TCA) [23].

SCFAs have also been shown to have anti-inflammatory and anti-cancer properties by several studies. The anti-inflammatory property of SCFAs could be used as a potential treatment of gastrointestinal inflammatory diseases such as Crohn’s disease

and inflammatory bowel disease (IBD) [24]. In a study done by Tedelind et. al., they found that SCFAs inhibit lipopolysaccharide-stimulated release of tumor necrosis factor alpha (TNFα), a cytokine involved in systemic inflammation in human neutrophils. They also showed that SCFAs inhibited the NF-kB pathway, an inflammatory pathway activated by cytokines [24]. Other studies have shown SCFAs role against colon cancer. For example, studies have shown that SCFAs act as ligands for G-protein coupled receptors like GPR43, which are found in normal human colon tissue. In a study done by Tang et. al., they showed that GPR43 expression was lower in human colon cancer cell lines. However, when they restored GPR43 in colon cancer cells along with the presence of propionate, it retarded growth and increased apoptosis [25]. SCFAs have the capability to promote positive health effects and could be used as a potential treatment for inflammatory GIT diseases.

Amino acids

Proteins can also be degraded in the large intestine by certain proteolytic microbiota species into amino acids. Proteins that aren't broken down earlier in digestion reach the colon, where bacterial proteases and left over pancreatic endopeptidases break them down into amino acids. Around 12 grams of proteinaceous material enters the colon a day, 49-51% coming from protein and 20-30% coming from peptides [26]. Protein can be further broken down by hydrolysis creating amino acids groups such as aromatic, sulfur, and other amino acids. Then amino acids can be broken down into other compounds. Aromatic amino acids like tryptophan go through an elimination reaction and become phenolic and indolic compounds, while other amino acids can go through deamination, decarboxylation, or fermentation which create ammonia, amines, or other

organic acids [27]. The protein that makes it to the colon is believed to come from around 50% of dietary sources and levels can vary based on certain foods [26].

High protein intake from food products such as meat has been linked by several studies to certain diseases. When meat is cooked it creates N-nitroso compounds (NOCs), heterocyclic amines (HCAs), and polycyclic aromatic hydrocarbons (PAHs); all these compounds are cacogenic and can cause DNA damage [28]. Also wide variety of cancers (colorectal cancer, lung cancer, liver) have been associated with red meat and processed meat intake [29]. In a study done by Smolinska et. al., researchers performed meta-analysis on seven cohort studies that took place from 1994-2007 to determine risk of colon cancer. They found that red meat intake of over 50 g/day can cause colonic carcinogenesis and red meat intake more than once per day can cause colonic and rectal cancers [28]. Another study done by Pan et. al., linked red meat to increased risk of type 2 diabetes (T2D). This study followed three different cohorts of over 4 million US citizens', male and female. The researchers accessed red meat consumption (processed and unprocessed) and T2D by questionnaires. Using statistics and meta-analysis they found a positive correlation between T2D and red meat consumption [30]. Amino acids especially from intake of meat products can cause a variety of disease that are harmful to the host.

Bile acids

Bile acids are important regulators for cholesterol balance and for digestion/absorption of lipids within the small intestine. They act as emulsifiers of lipids and fat-soluble vitamins and are the primary mediators of cholesterol catabolism. When bile acid synthesis occurs in the liver, primary bile acids such as cholic acid and

chenodeoxycholic acid (CDCA) are formed. Before bile acids are stored in the gallbladder as micelles with phospholipids and cholesterol, primary bile acids are conjugated with taurine or glycine to enhance hydrophilicity. After a meal is ingested the gallbladder releases the bile acids into the intestinal lumen to help with digestion of food products. When bile acids reach the GIT around 5-10% are deconjugated by the microbiota to create secondary bile acids such as deoxycholic and litocholic acids (DCA, LCA), and can then be reabsorbed or eliminated in the feces [31].

Bile acids have been implicated as possible causative agents for various cancers of the GIT and have been shown to be involved in inflammatory pathways in GIT cells. Individuals on high fat diets have been shown to have high amounts of bile acids and this can cause cells to be exposed to reactive oxygen species (ROS) and reactive nitrogen species (RNS). These agents can cause DNA damage, increased mutations, and apoptosis which could lead to various GIT cancers [32]. In a study by Bernstein et. al., mice were fed a diet that had high physiological levels of DCA, a common bile acid found in feces, to see if it had the potential to induce colon cancer. DCA was given to the mice for 8-10 months, DCA induced colonic tumors in 17 of the 18 wild type-mice [33]. Another study by Schlottman et. al., showed that bile acids had the ability to cause apoptosis in human colon cancer cell lines. Looking at HT-29, Caco-2, and SW480 colon carcinoma cell lines they found that varying concentrations of DCA caused apoptosis. They also showed caspase cleavage and activation, molecules responsible for biochemical change in cells undergoing apoptosis, occurred after the addition of deoxycholate [34]. Another study by Morbraten et. al., showed that triggering TGR5, a bile acid cell surface receptor, enhances lipopolysaccharide inflammatory response in monocytic cell lines. Bile acids

DCA and CDCA, activated MAP kinases ERK 1/2, p38, and JNK, major pathway activated by stress and inflammation. They also activated NF-kB signaling (inflammation pathway) and the release of inflammatory cytokines. The researchers believe that since LPS may contribute to insulin resistance, that TGR5 activation may also play a role [35]. This gives evidence that bile acids can have damaging effects on cells and their potential to lead to a variety of cancers.

Taurine

Taurine is an important dietary nutrient often referred to as a semi-essential amino acid (even though it lacks a carboxyl group) because mammals have limited ability to synthesize taurine and have to get it through their diet [36]. Taurine is acquired through the synthesis of cysteine and methionine, through absorption of food, excretion of taurocholate (bile salt), or unconjugated taurine in the urine. The taurine source in the intestines is mainly from the deconjugation of bile acids (taurocholic acid, taurochenodeoxycholic acid) by GIT microbiota through bile salt hydrolases [11]. Taurine has been shown to be important in many biological functions of the body. Taurine has a zwitterionic capability making it highly soluble in water and low lipophilicity. These properties contribute to taurine's membrane impermeability, high dipole nature, interactions with cations, membrane modulatory, and conjugation with bile acids. It has also been shown in studies to play a role in post-translational modification of mitochondrial tRNAs. High taurine levels can be found in foods such as beef, pork, fish, and shellfish [36]. It has also been shown to be high in energy drinks [36].

Taurine is an important nutrient in the body for normal function but when combined or taken in high amounts can cause adverse health problems. When the taurine

transporter (TauT) is knocked out in mice many health problems can occur such as reduced exercise capability, loss of retinal photoreceptor function, degeneration of the inner ear, cardiomyopathy, reduced T-cell memory generation, and hepatitis/liver fibrosis. Taurine can regulate cell volume, contribute to pool of organic osmolytes, stabilize membranes, serve as an antioxidant, and has anti-inflammatory/anti-apoptotic effects [36]. Taurine has also been shown to have adverse effects especially when combined with caffeine or in the high amounts present in energy drinks. It has been reported to increase systolic blood pressure and increase heart rate. Long term taurine consumption has also been implicated in human and animal studies to cause hypoglycemia [37].

Choline

Choline is a water soluble nutrient that is utilized in many pathways in the body. Choline can be obtained in the diet through foods such as eggs, liver, meats, and wheat germ or produced endogenously [38]. Choline is a key component to make membrane phospholipids, used as a methyl donor in the betaine production, and it is acetylated to make acetylcholine, an important neurotransmitter [39]. Choline can also influence other processes such as lipid metabolism, activation of nuclear receptors, circulation of bile and cholesterol, plasma membrane fluidity, and the bioenergetics of mitochondria. Dietary choline has been shown to be excreted in urine as N, N-dimethylglycine and leads to the production of creatine and creatinine. It can also be converted into methylamines by GIT microbiota after high fat consumption [40].

Choline has been associated with several health issues including liver and cardiovascular disease. In a study done by Sha et. al., they found that metabolomics

profiling of choline and related metabolites were able to predict which participants fed a certain diet (high or low choline) would develop liver dysfunction. They found that of the 53 patients, 9 developed liver dysfunction or muscle dysfunction when given a diet low in choline and 23 participants developed hepatic steatosis (fatty liver) [39]. Another study by Dumas et. al., mice fed a high fat diet had reduced bioavailability of choline. This is because the GIT microbiota converted the choline into methylamines (trimethylamine, dimethylamine, trimethylamine-N-oxide) which mimicked effects of a choline-deficient diet and caused non-alcoholic fatty liver disease (NAFLD) [40]. Choline is needed by our bodies for a variety of functions and can protect from disease like NAFLD.

Nuclear magnetic resonance use in metabolomics

Nuclear magnetic resonance (NMR) spectroscopy is the measurement of absorption or emission of frequency energy by nuclei in a magnetic field. The output of this process is specific resonances (peaks) that correspond to different compounds. NMR spectroscopy is a dynamic tool that is nondestructive, capable of identifying structural and conformational analyses of complex molecules, and can quantitate complex mixtures of molecules [41]. NMR spectroscopy has recently been shown as a useful technique in metabolite analyses, especially in the metabolic profiling of biofluids (such as fecal material). The reasons it is so effective is because it requires little to no sample preparation and is nondestructive/noninvasive to the sample [42]. These qualities make NMR a great technique to use for metabolite profiling.

Metabolite profiling of the human gut

In order to gain a better understanding of the impact of nutrition on the human gut microbiome Jacobs et. al., introduced the idea of ^{1}H NMR metabolite profiling of feces. In this study the researchers' first generated metabolite profiles by water and methanol extraction of lyophilized feces and they obtained a variety of metabolites e.g. amino acids, SCFAs, organic acids (e.g. succinate, pyruvate), uracil, ethanol, cholate, and glucose [43]. Then to validate this approach they did another study where they placed human subjects on either a placebo, grape juice extract, or grape juice wine extract for 4 weeks. The goal of the researchers was to see if metabolite profiles of feces changed based on dietary changes. They found that the grape juice extract did not change the metabolite profiles of individuals compared to the placebo. However, the individuals placed on the grape juice wine extract did cause a reduction in isobutyrate [43]. Another study similar to Jacobs et. al., was performed by Ndagijimana et. al., looking at effects of synbiotic food on the metabolite profiles of the gut using ^{1}H NMR. They feed 16 healthy subjects synbiotic food for a month and looked at the difference prior to consumption and after consumption. They were able to detect 150 metabolites and the biggest changes between start and end of diet was that SCFAs increased and amino acids decreased [44]. These studies show the efficiency of ^{1}H NMR metabolic profiling on fecal material to determine differences based on dietary changes.

Metabolite differences related to geography and regional diet

Differences in metabolites and microbiota can also be determined by geographical location and what foods are available in those different locations. A study done by De Filippo et. al., compared the fecal material between two groups of adolescents: an African group called the Mossi from Burkina Faso (BF) and European adolescents from Florence, Italy (EU). The BF diet is low in fat and animal protein but rich in starch, fiber, and plant polysaccharides. The EU diet is rich in animal protein, sugar, starch, and fat but low in fiber. The researchers found that the groups had enrichment of certain GIT bacteria that are known to degrade polysaccharides and BF adolescents had significantly more SCFAs compared to the EU. They believe these differences are due to their differing diet [45].

Another study, done by Schnorr et. al., compared metabolites and microbiota from the Hadza hunter gatherers from Tanzania and adults from Bologna, Italy. The Hadza diet consists of wild foods such as meat, honey, baobab, berries, and tubers. As where the Italian cohort's diet followed more of a Mediterranean diet consisting of plant foods, fruit, pasta, bread, and olive oil. The Italian cohort dietary source of carbohydrates mostly came from digestible starch and sugar, with little indigestible fiber. The study showed the Hadza group has higher microbial richness and diversity than the Italian group. The Hadza group also had a higher enrichment of bacteria involved with the digestion of fibrous plant foods. The researchers also found that Italian samples had higher levels of butyrate and the Hazda had higher levels of propionate [46].

Differences between U.S. and Egyptian diets

In this study, we profiled metabolite differences of Egyptian and US teenagers. Diets of these two populations are thought to be different which can impact GIT microbiota. Figure 1.1 and 1.2 give an overview of the main diet differences between Egyptians and United States adolescents.

Figure 1.1: Egyptian diet

Figure 1.1 shows foods of high and low intake in the Egyptian diet

Figure 1.2: United States diet

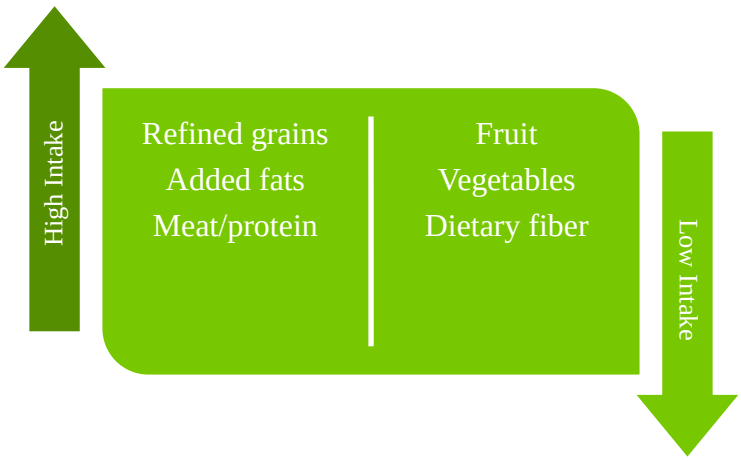

Figure 1.2 shows foods of high and low intake in the United States diet

According to a review published by H. Hassan-Wassef explaining the basic features of the Egyptian diet; bread made from wheat flour is one of the main food staples in the Egyptian diet and eaten with almost every meal. Cereals and legumes are also very popular in the diet and are the main source of calories and protein. Egyptians regularly intake fresh dark green leafy vegetables such as cabbage, spinach, and grape leaves. Also small amounts of cheese and dairy products are consumed regularly. Meat consumption is lower because it tends to be regarded as a luxury item and due to religious and cultural reasons cannot be consumed at certain times. Egyptian diet is traditionally low in saturated fats and oil is more commonly used than butter [47].

The United States diet patterns vary from that of the Egyptians. According to a review published by D. Grotto on the US diet, there are foods eaten in excess and foods under-consumed by the US population. These numbers are calculated based on a 2,000 kilo-calorie per day intake recommended by the Dietary Guidelines for Americans (DGAs) created by the United States Department of Agriculture (USDA) and U.S. Department of Health and Human Services. It is recommended that Americans eat 6ounce-equivalents (oz-eq) of grain per day, however it has been shown that they eat 8.1 oz-eq per day. Most of these grains are refined (7.2 oz-eq) and not the recommended whole grains. It is suggested to ingest 4.5 cups of fruits and vegetables (leafy green, legumes, orange, and starchy vegetables) per day; 2 cup from fruit and 2.5 from vegetables. The US population is currently consuming 0.9 cups of fresh and processed fruit and 1.7 cups of fresh and processed vegetables. Dietary fiber recommended intake is around 28g per day but the US population only consume 15.2 g per day. Meat and beans (not including legumes), however are over consumed. It is recommended 5.5 oz-eq per

day, and actual intake is 6.5 oz-eq per day. The recommended total intake of fat to be between 25%- 35% of total calories a day. American's have been shown to consume around 71.6g of added fats and oils per person per day [48].

Thesis overview

The work in this thesis involved the comparison of metabolite profiles of two geographically distinct cohorts, healthy Egyptian adolescents (n=28) and healthy United States adolescents (n=14). Fecal water extracts to obtain water soluble metabolites were made from collected fecal samples and ^{1}H Nuclear Magnetic Resonance (^{1}H NMR) on the extracts was performed by Dr. Nicholas Reo's Laboratory at Wright State University, Dayton, Ohio. The binned NMR data obtained from the ^{1}H NMR were ran through multivariate statistical tools to determine differences between samples. Next, applying spectrum deconvolution, a method for differentiating peaks of interest, quantification of 32 metabolites was done. Metabolite peak validation procedure (spike-in) was used to confirm peak assignments for metabolites with suspect peaks. Finally, after confirmation and quantification of all the metabolites, the metabolite data was transformed to center-log ratio and multivariate statistical tools were ran.

2. Materials and Methods

Sample Collection

Samples were collected from two different groups; healthy adolescents from Egypt (egkHLT) and healthy adolescents from the United States (uskHLT). egkHLT samples were collected by Dr. Laila Hussein from The National Research Center located in Giza Governorate, Egypt and uskHLT samples were collected by Dr. Oleg Paliy's laboratory and Dr. Sonia Michail's at Wright State University and Dayton Children's Hospital located in Dayton, Ohio. Volunteers defecated into sterile collection containers that were then immediately stored at -80℃. The cohort consisted of 28 male samples from egkHLT and 14 male samples from uskHLT. Table 2.1 and 2.2 details sample name, age, gender, and BMI

2.1: Volunteer information for uskHLT samples

Sample (uskHLT)	Age	BMI
uskHLT05	12	19.9
uskHLT08	15	22.1
uskHLT13	10	22.6
uskHLT14	11	23.8
uskHLT17	10	17.2
uskHLT20	12	21.5
uskHLT22	18	25.6
uskHLT24	16	16.7
uskHLT26	10	17.3
uskHLT27	9	18.4
uskHLT29	12	26.5
uskHLT30	14	18.4
uskHLT31	15	25.8
uskHLT32	16	20.4
Average	**13.0 ± 0.7**	**21.2 ± 0.9**

Table. 2.2: Volunteer information for egkHLT samples

Sample (egkHLT)	Age	BMI
egkHLT01	14	19.5
egkHLT02	15	23.7
egkHLT03	14	22.0
egkHLT04	14	17.6
egkHLT05	15	21.8
egkHLT06	13	17.3
egkHLT07	14	17.4
egkHLT08	13	17.3
egkHLT10	14	16.9
egkHLT11	13	15.5
egkHLT12	13	17.3
egkHLT14	14	18.9
egkHLT15	14	17.3
egkHLT16	14	16.9
egkHLT17	14	18.2
egkHLT19	13	19.6
egkHLT20	15	17.5
egkHLT22	14	19.2
egkHLT23	15	24.5
egkHLT24	14	17.9
egkHLT25	13	17.1
egkHLT26	14	15.6
egkHLT27	14	17.7
egkHLT28	14	18.0
egkHLT29	14	19.9
egkHLT30	14	25.5
egkHLT31	14	20.3
egkHLT32	14	19.2
Average	**14.0 ± 0.1**	**18.9 ± 0.5**

Fecal water extraction

Fecal water extractions were performed to obtain water soluble metabolites for ^{1}H NMR. The phosphate buffer solution used in the protocol contained 2.0 mg KCl, 11.5 mg

Na_2HPO_4-$7H_2O$, 2.0 mg KH_2PO_4, and 10 ml H_2O. After the phosphate buffer was made the fecal water extraction protocol was performed. The protocol is the following:

a) 250 mg of fecal sample was weighed and placed in 2 ml eppendorf tubes
b) 1.25 ml of 1x phosphate buffer was added to the tube and vortexed at maximum speed for 5 minutes to homogenize samples
c) The homogenized samples were centrifuged at maximum speed in the chromatography fridge for 5 minutes at 1600g in 4 ℃
d) Supernatant was collected and filtered through Whatman GDX filter using a syringe and eluted into a 1.5 eppendorf tube
e) Eluate was centrifuged in the chromatography fridge for 15 minutes at 1600g in 4 ℃
f) The supernatant was aliquoted into two clean 1.5 ml eppendorf tubes and placed in the -70℃ freezer

Proton nuclear magnetic resonance ^{1}H NMR analysis of fecal water extracts

Proton NMR was done to determine metabolite composition within the collected fecal samples. This was carried out by Dr. Nicholas Reo's Laboratory at Wright State University in Dayton, Ohio. A 550-600 µl aliquot of the fecal extract sample was transferred into a 5 mm NMR tube with 150 µl of 9 mM trimethylsilypropionic-2.2, 3, 3-d4acid (TSP) in D_2O. ^{1}H NMR spectra were acquired at 25 ℃ using a Varian INOVA operation at 600 MHz (14.1 Tesla). The TSP was a chemical shift standard used for

reference and quantification. The spectra were pre-processed using a standard protocol employed in Dr. Reo's laboratory [49].

To use this data for multivariate analyses, the spectral data are binned. Binning is the process of dividing the spectrum into several hundred non-overlapping regions of the same size. Binning attempts to lessen the effects of variation in peaks position that can be caused by pH, ionic strength, composition, and reduce the dimensionality for multivariate statistical analyses [50]. The method for binning on the data set is a dynamic programming-based adaptive binning technique developed by Anderson et al 2011 and was carried out by Dr. Reo's laboratory [50].

NMR spectrum deconvolution

Deconvolution algorithm was employed in MATLAB R2014a (Mathworks Inc) to determine peak areas of the 32 selected metabolites [50]. The purpose of deconvolution is to identify known peaks from other peaks that can overlap in the same region. The deconvolution software works by fitting a defined spectral region using a combination of baseline shapes (spline, v-shaped, linear, or constant) and a Gauss-Lorentz peak-fitting function. Figure 2.1 shows an example (uracil, ppm ~ 5.80) of the deconvolution process before the line fitting and peak validation and Figure 2.2 is after the line fitting and peak validation by the user. After deconvolution is done a better estimate of the peak area can be determined. Areas for peaks of interest are outputted in text files.

Figure 2.1: Before deconvolution of uracil

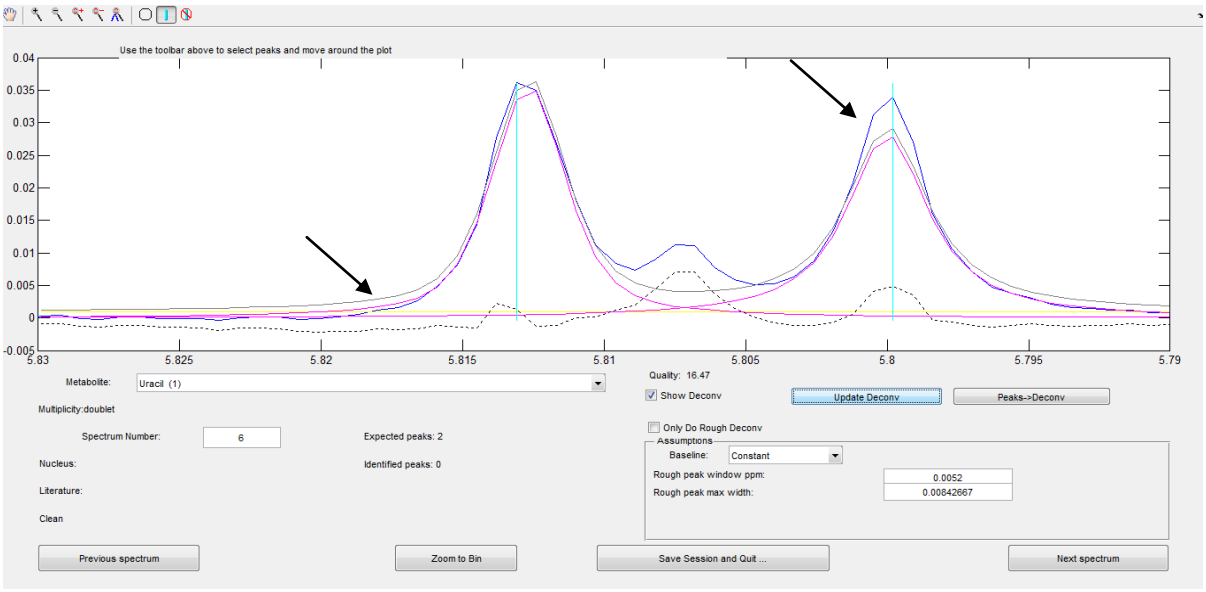

Figure 2.1 shows the spectrum of uracil before using the deconvolution algorithm to correct peak fitting. The two black arrows point to unfitted areas of the spectra.

Figure 2.2: After deconvolution of uracil

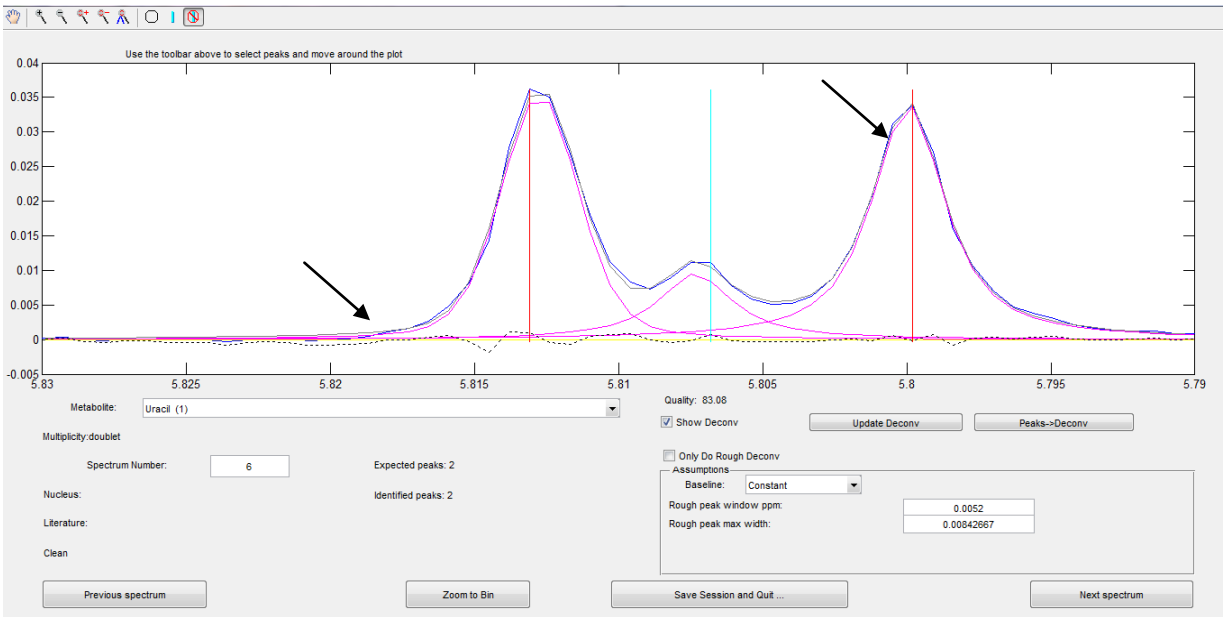

Figure 2.2 shows the spectrum of uracil after the deconvolution algorithm to correct peak fitting. The two black arrows point the areas that were fitted by the algorithm.

Metabolite quantification

After deconvolution all the metabolite peak intensities were corrected for equivalent number of protons then normalized relative to TSP signal intensity through an excel template provided by Dr. Reo's laboratory. The extracts were prepared from identical stool weight so the relative comparisons between the samples were fully valid. Peak assignments are based on literature and databases (such as the Human Metabolome Database) [51].

Validation of metabolite peaks

Due to crowding of metabolite peak signals within some regions of the spectra a peak validation was done. Using sample uskHLT, four fecal water extracts following the same protocol were made. Next, nine 100 mM metabolite solutions at volume of 500 μl were prepared, Table 2.3 shows the nine metabolites molecular weights and mass in the 100 mM solution in milligrams. The final concentration was approximately 1.5 mM.

Table 2.3: Molecular weights and mass of 100 mM solution in mg of each metabolite

Metabolite	Molecular Weight g/mol	Mass to obtain 100 mM Solutions (Volume)
Isoleucine	131.2	6.6
Leucine	131.2	6.6
Tryptophan	204.2	10.2
Glucose	180.2	9.0
Methylamine	31.1	1.6
Galactose	180.2	9.0
Sucrose	342.3	17.1
Uracil	112.1	5.6

The solutions were made with phosphate buffer and H_2O. Uracil and tryptophan solutions were made with H_2O and the other metabolite solution were made in phosphate buffer. 4 μl of NaOH was added to the tryptophan solution because tryptophan is more soluble in alkali hydroxides [52]. Then the fecal water extracts and metabolite solutions were given to Dr. Reo's Laboratory for ^{1}H NMR analysis. Several metabolites were ran in each extract if their peaks did not overlap, Table 2.4 shows which metabolites were ran in the same extract.

Table 2.4: Fecal water extract and contents

Extract	Metabolites
1	Leucine, Isoleucine, Methylamine, Tryptophan
2	Glucose
3	Sucrose
4	Galactose, Uracil

Statistical data analyses

Analyses of the compositional ^{1}H NMR binned and metabolite data was through multivariate analyses. Prior to multivariate analyses several processing steps on the binned and metabolite quantification data had to be done. First, sample-to-sample within-group variance was reduced by normalization of binned data where egkHLT and uskHLT were auto-scaled to both group means [54]. Then center log ratio transformation *(clr*) using a custom script developed in our laboratory was performed on the metabolite data set. Since the data is compositional and is based on a relative scale, this transformation can remove the effect of constant-sum constrain on the co-varying variables [55]. The multivariate statistic tests used on this data set were principle component analysis (PCA), orthogonal projections to latent structures discriminate analysis (OPLS-DA), and

Random Forest (RF). PCA and OPLS-DA were ran in MATLAB R2014a (Mathworks Inc.) using custom scripts (Shankar et al 2013), and RF was ran in R x64 3.2.2 [49].

PCA reduces dimensionality in the data set while retaining most of the variation by creating principle components. The principle components represent the variability within the data set and are re-ranked according to the amount of variation they represent, therefore PC1 would have the most variation, PC2 the second most and so on. This technique is unsupervised meaning that the tool is not told which samples belong to which groups [56]. PCA was developed in 1901 by Karl Pearson [57].

OPLS-DA finds linear combinations of variables that separate groups. It is another technique that combines dimensionality reduction with a form of classification called discriminant analysis. It takes the variability of X that corresponds to Y and puts it in the T axis and takes the rest of the variability and puts it in the T orthogonal axis. The X data is the bins/metabolites and Y data is the class information. OPLS-DA is a supervised model meaning that it told which samples belong to which group. Q^2 values are used to assess the predictive ability of the metric. As Q^2 approaches 1 it means that the model exhibits predictive capability, however if it less than zero it means it has no predictive power [56]. OPLS-DA was developed by Trygg and Wold, 2002 [58].

Random forest is a method of classification that uses a collection of classification trees to provide variable importance. It works by using data that relies on random sampling with replacement (bootstrap) method and random variable selection to build the trees. RF creates a collection of trees that have low bias and low variance. Then RF then determines importance scores based on their mean gini scores. Random forest was developed by Leo Breiman in 2001 [59].

To evaluate statistical significance of metabolite levels, a Mann-Whitney U-Test was ran in SPSS (SPSS Inc.).

Correlations analysis was done to find the relationships among metabolites. To do this, SparCC, method and software established by Freidman and Alm was used [60]. First, correlations were found using the script in Python following these parameters –i 20 inference iterations to average over, -x 10 number of exclusion iterations to remove strongly correlated pairs, and -t 0.1 correlation strength exclusion threshold. Second, to calculate pseudo p-values first the bootstrap procedure using the MakeBootstraps script was done generating 100 shuffled data sets. Next, SparCC was ran again (using the same parameters) on all 100 of these shuffled data sets to give two sided p-values.

3. Results

Ordination analyses of binned data

Principle component analysis

Figure 3.1 shows PCA carried out in MATLAB R2014a demonstrates that the samples could form two distinct clusters based on overall variability. A few samples did overlap into opposite clusters meaning that their variability identified more with their opposite classification. These samples were found to be uskHLT 29 being more related to egkHLT samples and egkHLT 14 being more related to uskHLT samples. The model's p-value ($p < 0.001$) is statistically significant

Figure 3.1 PCA plot of binned data

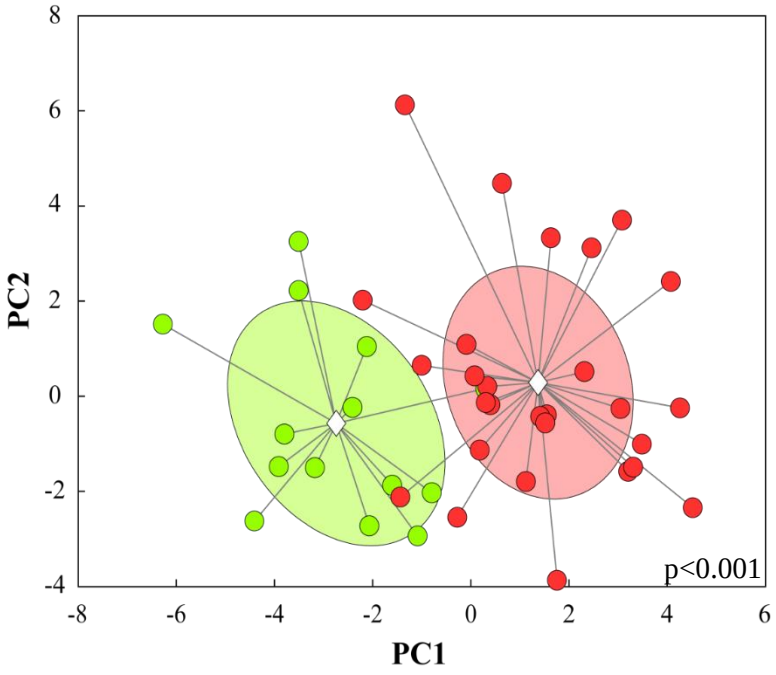

Figure 3.1 is the plot of principle components analysis of egkHLT (red) and uskHLT (green) samples. This data is based on the binned NMR data provided by Dr. Reo's laboratory.

Orthogonal projection to latent structure discriminant analysis

Figure 3.2 demonstrates that OPLS-DA had large group separation of T and little separation of T orthogonal. The model also demonstrates that the binned NMR data is separated into two groups and samples were placed in the correct groups. The models p-value ($p < 0.001$) statistically significant. Q^2 value was 0.81 meaning the model had strong predictive ability.

Figure 3.2: OPLS-DA plot of binned data

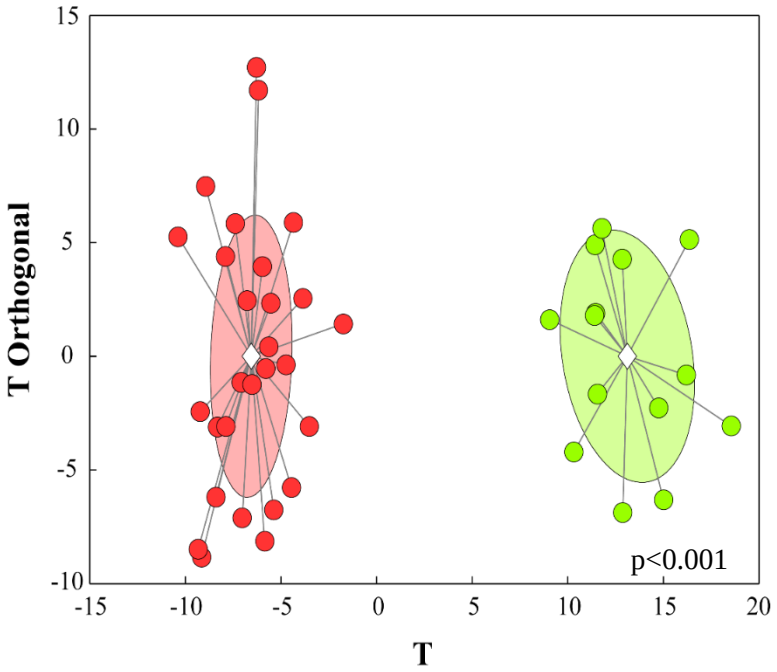

Figure 3.2 is the plot of orthogonal projection to latent structure discriminate analysis of egkHLT (red) and uskHLT (green) samples. This data is based on the binned NMR data provided by Dr. Reo's laboratory.

Metabolite measurements

Using ^{1}H NMR analysis, 32 metabolites were identified and the deconvolution algorithm was used to isolate individual metabolite peaks and allow robust quantification. These values calculated are arbitrary units respective to TSP, an internal chemical shift standard used in NMR. Tables 3.1-3.3 are all the metabolites identified in the egkHLT and uskHLT spectra with their chemical shift values in ppm. Figure 3.3 shows the egkHLT and uskHLT spectra with the peaks identified based on ppm.

Table 3.1: Quantified SCFAs with chemical shifts in ppm

Short Chain Fatty Acids	Chemical Shift ^{1}H ppm
Acetate	1.92 (s)
Butyrate	1.56 (tq)
Formate	8.46 (s)
Propionate	1.06 (t)
Valerate	0.86 (t)

Table 3.1 shows the SCFAs identified and their peak locations along with peak type, (s) singlet, (d) doublet, (t) triplet, (dd) doublet of doublets, and (tq) triplet of quartets.

Table 3.2: Quantified amino acids with chemical shifts in ppm

Amino Acids	Chemical Shift ^{1}H ppm
Alanine	1.48 (d)
Aspartic Acid	2.82 (dd)
Glycine	3.56 (s)
Isoleucine	1.01 (d)
Leucine	0.96 (t)
Lysine	3.03 (t)
Tryptophan	7.54 (d)
Tyrosine	6.90 (m)
Valine	3.61 (d)

Table 3.2 shows amino acids identified and their peak locations along with peak type, (s) singlet, (d) doublet, (t) triplet, and (dd) doublet of doublets

Table 3.3: Quantified metabolites with chemical shifts in ppm

Other Metabolites	Chemical Shift ^{1}H ppm
1-Methylhistamine	3.25 (t)
Cholate Bile Acids	0.73 (m)
Choline	3.20 (s)
Cytosine	7.52 (d)
Ethanol	1.19 (t)
Fumarate	6.52 (s)
Galactose	5.27 (d)
Glucose	5.24 (d)
Hypoxanthine	8.22 (s)
Lactate	1.33 (d)
Malate	4.30 (dd)
Methylamine	2.61 (s)
Pyruvate	2.47 (s)
Succinate	2.41 (s)
Sucrose	4.22 (d)
Taurine	3.45 (t)
Trimethylamine	2.90 (s)
Uracil	5.80 (d)

Table 3.3 shows the other metabolites identified and their peak locations along with peak type, (s) singlet, (d) doublet, (t) triplet, and (dd) doublet of doublets.

Figure 3.3: ^{1}H NMR spectra of egkHLT and uskHLT

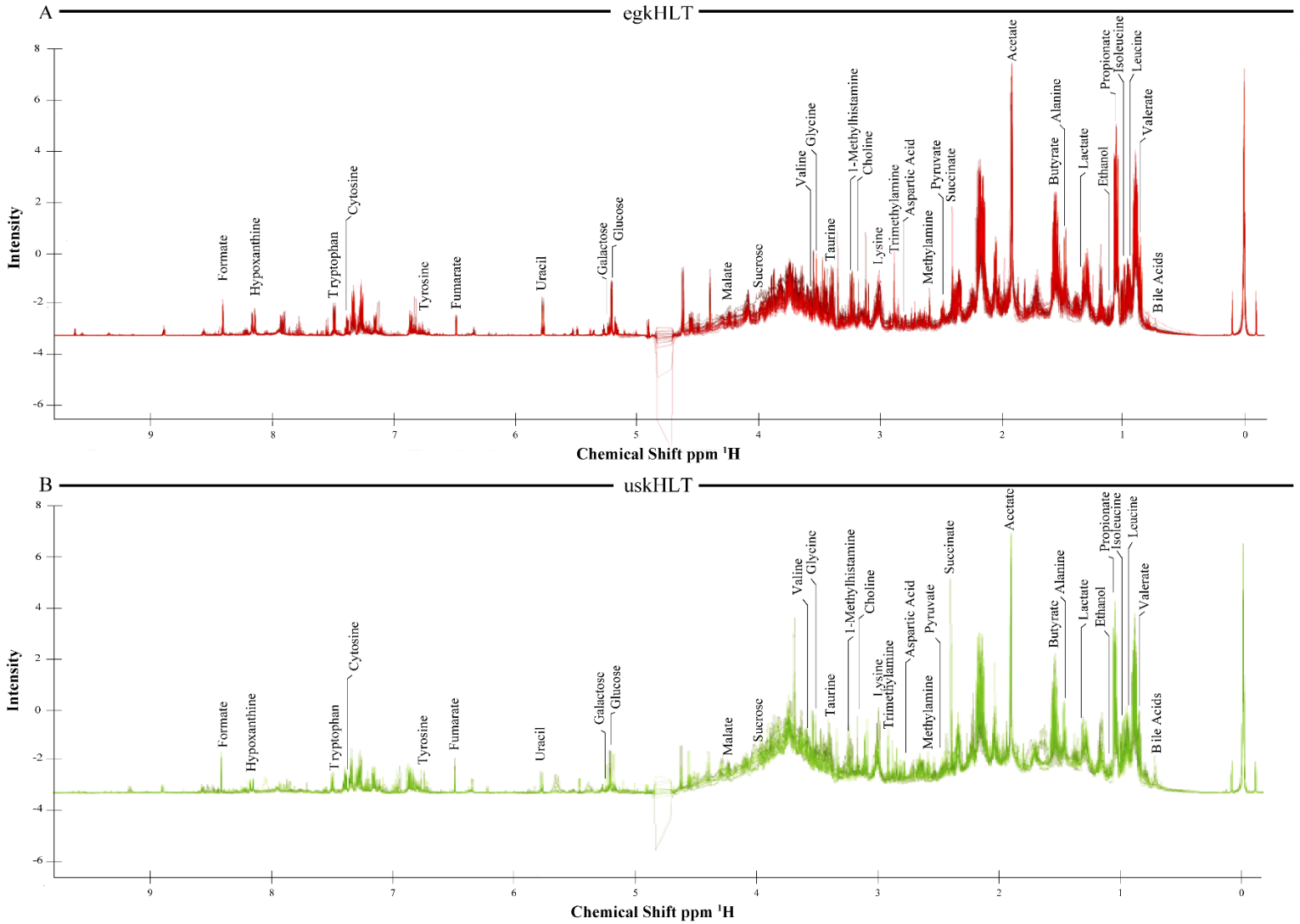

Figure 3.3 shows the spectra acquired by the ^{1}H NMR of egkHLT (red) and uskHLT (green) samples along with the identification of the 32 metabolites peaks

Short chain fatty acids

Table 3.4 shows the quantified (arbitrary units respective to TSP) SCFAs values. Overall, the SCFAs are higher in egkHLT compared to uskHLT. Acetate, butyrate, and propionate are all higher in egkHLT. However, formate and valerate are higher in uskHLT. Mann-Whitney U-Test indicated that differences in acetate, propionate, and SCFAs total were statistically significant ($p \leq 0.05$).

Table 3.4: Short chain fatty acids quantification results

Metabolite	egkHLT	±SE	uskHLT	±SE	Mann-Whitney U Test
Acetate	42.12	0.72	38.36	1.49	**0.040**
Butyrate	16.04	0.65	15.15	1.02	0.471
Formate	0.55	0.03	0.60	0.06	0.594
Propionate	23.25	0.56	19.66	1.27	**0.008**
Valerate	1.17	0.13	1.39	0.22	0.455
Total	**83.29**	**1.04**	**75.10**	**2.38**	**0.001**

Table 3.4 shows SCFAs quantification results in fractional values of egkHLT and usKHLT, standard error values, and results of Mann-Whitney U test (p-values, bolded numbers are statistically significant).

Amino acids

Table 3.5 shows the quantified (arbitrary units respective to TSP) amino acid values. Amino acid totals are higher in uskHLT compared to egkHLT. Alanine, aspartic acid, isoleucine, leucine, lysine, tyrosine, and valine are all higher in uskHLT. Glycine and tryptophan are higher in egkHLT. Mann-Whitney U-Test indicated that differences in aspartic acid, isoleucine, leucine, lysine, tryptophan, tyrosine, valine and the amino acid total were statistically significant ($p \leq 0.05$).

Table 3.5: Amino acids quantification results

Metabolite	egkHLT	±SE	uskHLT	±SE	Mann-Whitney U Test
Alanine	1.97	0.19	2.41	0.18	0.078
Aspartic Acid	0.35	0.02	0.73	0.06	**0.000**
Glycine	0.67	0.08	0.51	0.11	0.311
Isoleucine	0.50	0.05	1.01	0.11	**0.001**
Leucine	0.66	0.07	1.41	0.16	**0.000**
Lysine	2.01	0.27	3.68	0.47	**0.001**
Tryptophan	0.78	0.08	0.56	0.06	**0.048**
Tyrosine	0.28	0.03	0.58	0.05	**0.000**
Valine	0.14	0.01	0.28	0.04	**0.000**
Total	**7.57**	**0.64**	**11.43**	**0.74**	**0.001**

Table 3.5 shows amino acid quantification results in fractional values of egkHLT and usKHLT, standard error values, and results of Mann-Whitney U test (p-values, bolded numbers are statistically significant).

Other metabolites

Table 3.6 shows the quantified (arbitrary units respective to TSP) values of the other metabolites looked at in our data set. Metabolites that are higher in egkHLT include cytosine, glucose, hypoxanthine, and uracil. Metabolites higher in uskHLT are 1-methylhistamine, bile acids, choline, ethanol, fumarate, galactose, lactate, malate, methylamine, pyruvate, succinate, sucrose, taurine, and trimethylamine. Mann-Whitney U-Test indicated that differences bile acids, choline, ethanol, hypoxanthine, lactate, malate, methylamine, pyruvate, succinate, sucrose, taurine, and uracil were statistically significant ($p \leq 0.05$).

Table 3.6: Metabolites quantification results

Metabolite	egkHLT	±SE	uskHLT	±SE	Mann-Whitney U Test
1-Methylhistamine	0.90	0.10	1.28	0.21	0.122
Bile Acids	0.04	0.00	0.13	0.05	**0.006**
Choline	0.04	0.01	0.09	0.02	**0.001**
Cytosine	0.81	0.08	0.58	0.06	0.066
Ethanol	0.60	0.07	1.34	0.34	**0.002**
Fumarate	0.08	0.01	0.12	0.02	0.364
Galactose	0.09	0.01	0.18	0.12	0.575
Glucose	1.34	0.16	1.02	0.22	0.311
Hypoxanthine	0.56	0.06	0.32	0.04	**0.002**
Lactate	0.85	0.05	1.30	0.12	**0.005**
Malate	0.99	0.09	1.59	0.25	**0.010**
Methylamine	0.18	0.01	0.24	0.02	**0.015**
Pyruvate	0.09	0.01	0.21	0.03	**0.000**
Succinate	0.30	0.09	1.27	1.13	**0.013**
Sucrose	0.42	0.06	0.94	0.17	**0.004**
Taurine	0.65	0.06	1.49	0.27	**0.001**
Trimethylamine	0.08	0.01	0.08	0.01	0.790
Uracil	0.75	0.08	0.46	0.05	**0.022**

Table 3.6 shows metabolite quantification results in fractional values of egkHLT and usKHLT, standard error values, and results of Mann-Whitney U test (p-values, bolded numbers are statistically significant).

Figure 3.4 shows the comparison of select metabolites that differed between groups. All the SCFAs are grouped showing a higher percentage in egkHLT (83%) compared to uskHLT (75%). All the amino acids are grouped showing a higher percentage in uskHLT (11%) compared to egkHLT (8%). Bile acids, choline, and taurine are all grouped because of their involvement with bile acids and are higher in uskHLT (0.13%, 0.09%, 1.49%) compared to egkHLT (0.04%, 0.04%, 0.7%), respectively. Galactose, glucose, and sucrose were also grouped because they are all sugars. Galactose and sucrose are higher in uskHLT (0.18%, 0.94%) compared to egkHLT (0.09%, 0.42%). Glucose is higher in egkHLT (1.34%) compared to uskHLT (1.02%).

Figure 3.4: Fractional abundance percentages of metabolites

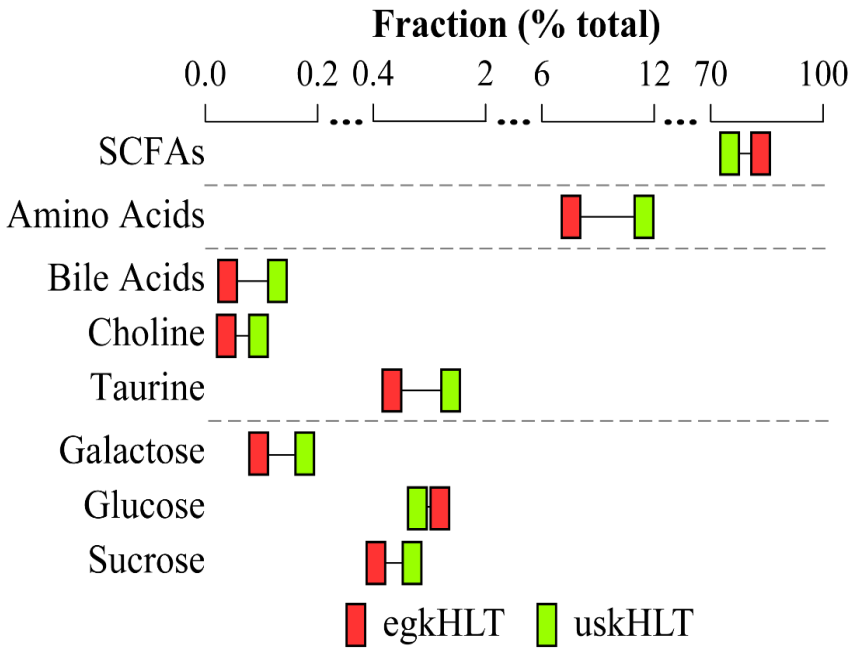

Figure 3.4 shows fractional abundance percentages of selected metabolite quantification results, egkHLT (red) and uskHLT (green). Groups include SCFAs, amino acids, compounds involved with bile acids, and sugars.

Metabolite validation

Figures 3.5-3.11 show uskHLT31 spectrum before any addition of metabolite solutions and the spectra after the metabolite solutions were added. The spike-ins confirmed peak assignments that were located in crowded regions of the ^{1}H NMR egkHLT/uskHLT spectra. Validations were successful and yielded the expected peaks based on a database search (HMBD).

Figure 3.5: Galactose validation

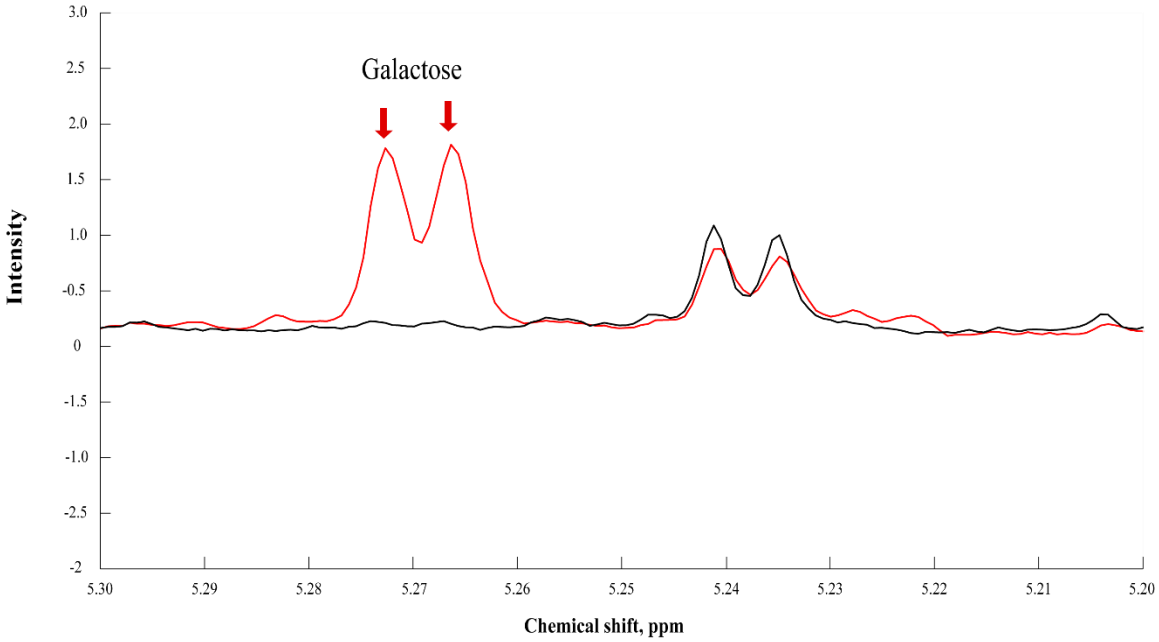

Figure 3.5 shows galactose validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.6: Glucose validation

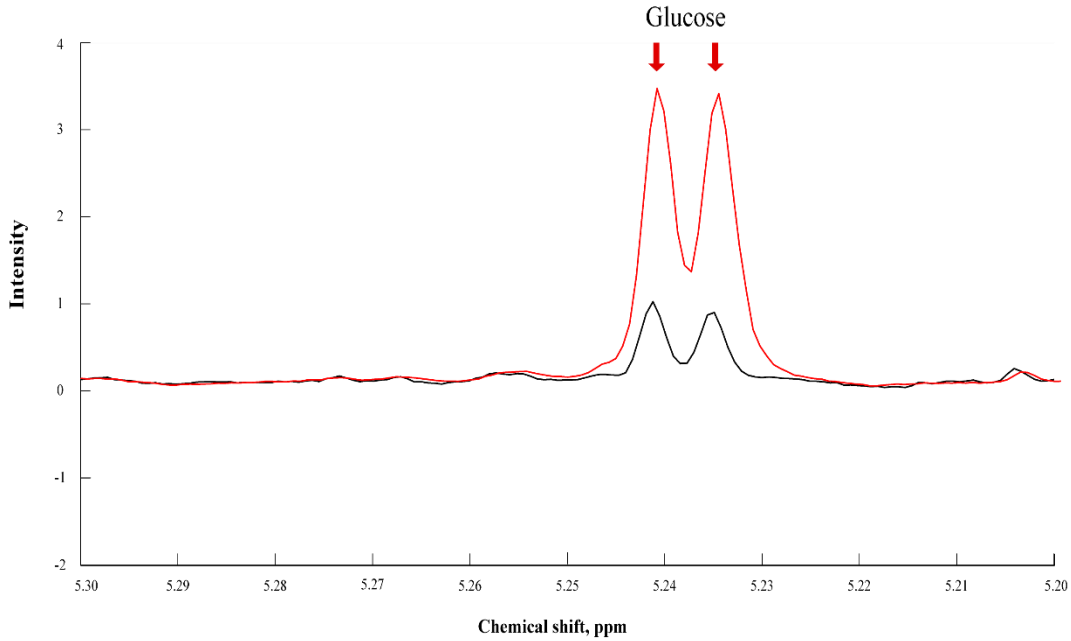

Figure 3.6 shows glucose validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.7: Methylamine validation

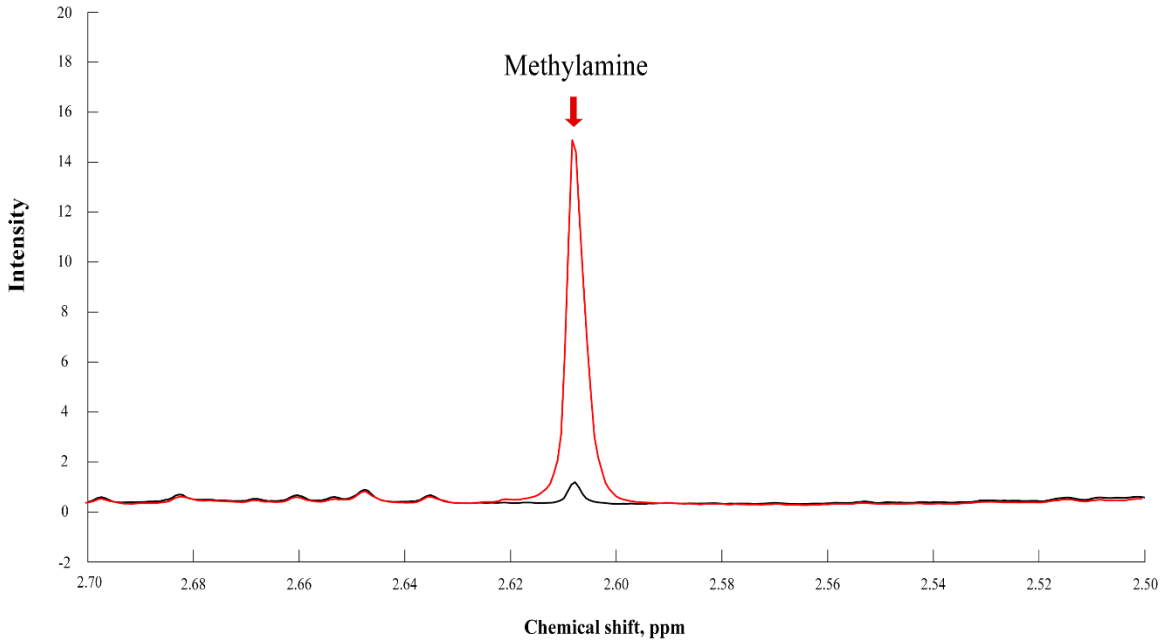

Figure 3.7 shows methylamine validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.8: Sucrose validation

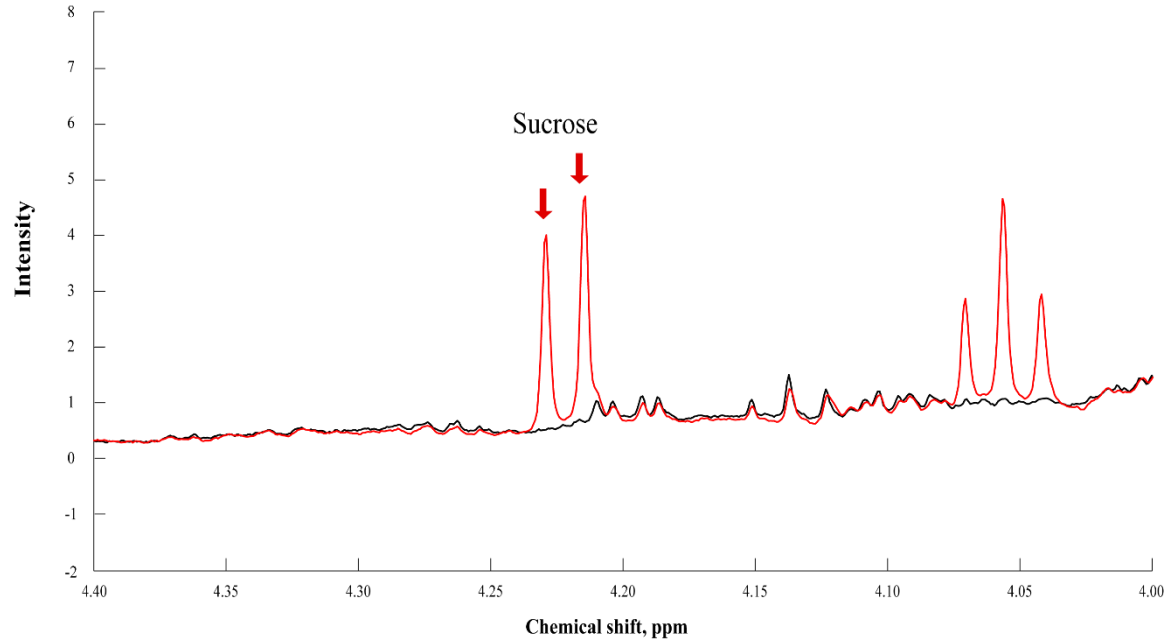

Figure 3.8 shows sucrose validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.9: Isoleucine and leucine validation

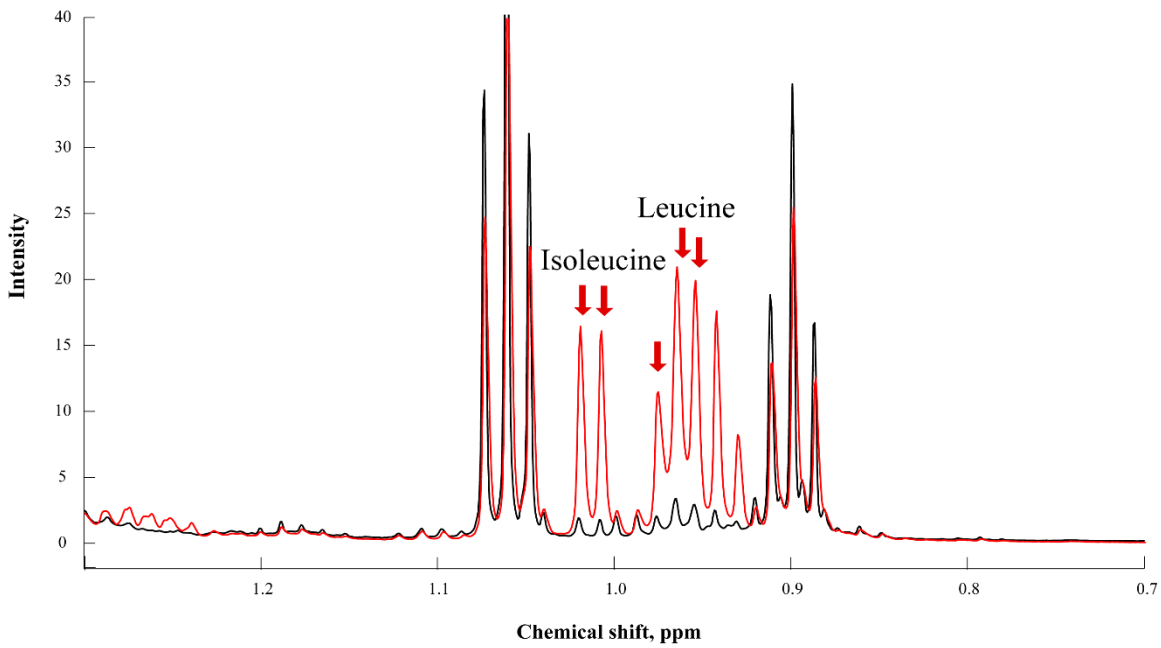

Figure 3.9 shows isoleucine and leucine validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.10: Uracil validation

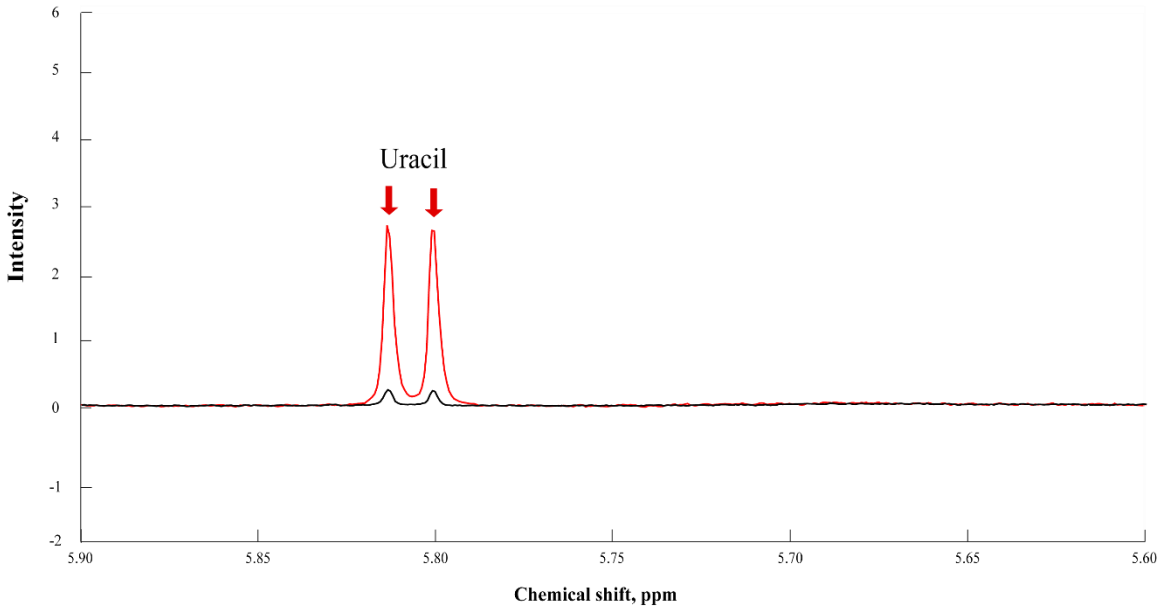

Figure 3.10 shows uracil validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after.

Figure 3.11: Tryptophan validation

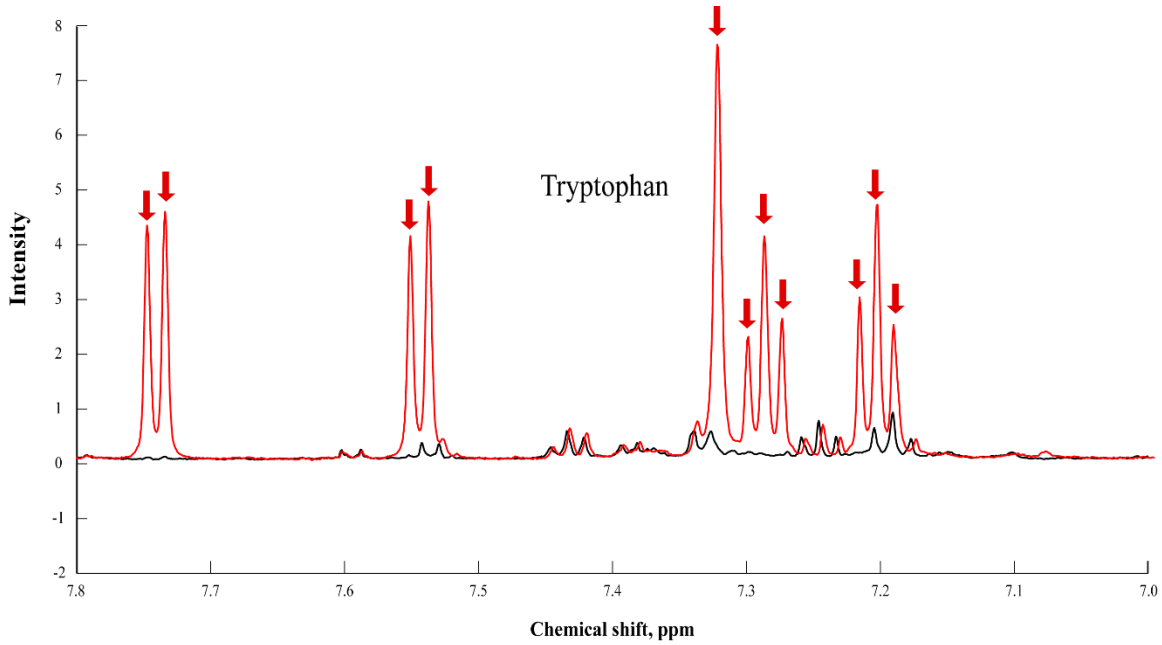

Figure 3.11 shows Tryptophan validation through the spike-in procedure. The black line is the spectrum before the metabolite solution was added and the red line is after. The peak that was validated is at 7.54 ppm

Ordination analyses of metabolite data

Principle component analysis

Figure 3.12 shows PCA carried out in MATLAB R2014a. This model demonstrates that there is enough variability within the measured metabolites levels that can two distinct clusters that correspond to nationality of groups. The model's p-value ($p < 0.001$) is statistically significant.

Figure 3.12: PCA plot of metabolite data

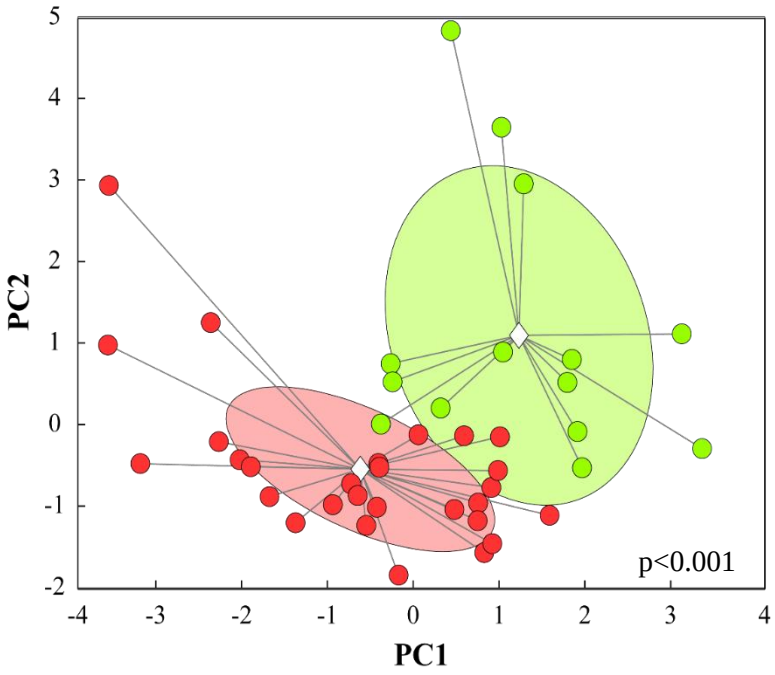

Figure 3.12 is the plot of principle components analysis of egkHLT (red) and uskHLT (green) samples. This data is based on the quantified metabolite data.

Orthogonal projection to latent structure discriminate analysis

Figure 3.13 shows the OPLS-DA model with large group separation of T and little separation of T orthogonal. The model demonstrates that the metabolite variability in samples is separated into two groups and samples were placed in the correct groups. The model's p-value ($p < 0.001$) is statistically significant. The Q^2 value was 0.73 meaning the model had predictive ability.

Figure 3.13: OPLS-DA plot of metabolite data

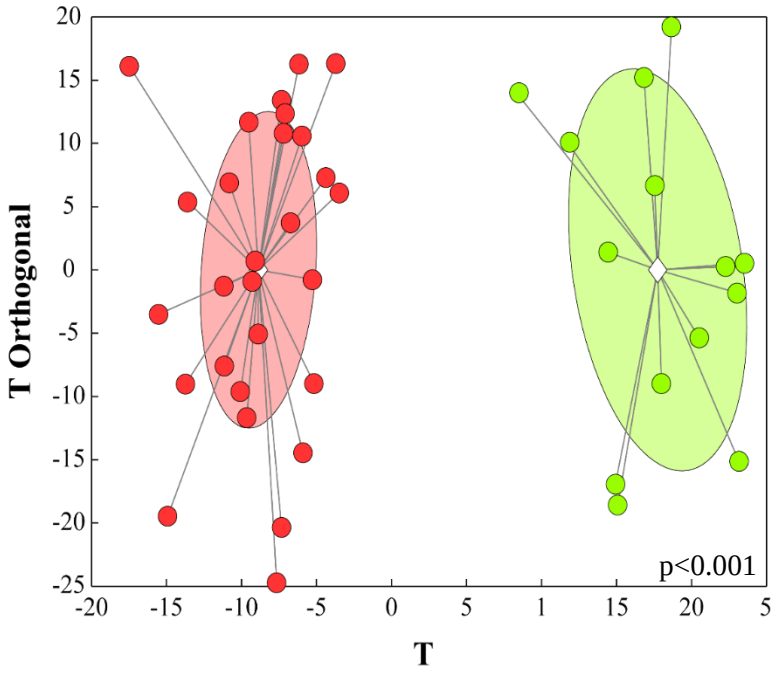

Figure 3.13 is the plot of orthogonal projection to latent structure discriminate analysis of egkHLT (red) and uskHLT (green) samples. This data is based on the quantified metabolite data.

Random forest

Figure 3.14 shows Random forest ability to classify samples based on variability in metabolite levels into nationalities. RF also determined metabolites that are important for separation of the two nationalities. The mean decrease gini is the importance of each variable in classifying the data. The top 12 important metabolites are shown in Table 3.7, uracil has the highest score

Figure 3.14: Random forest plot

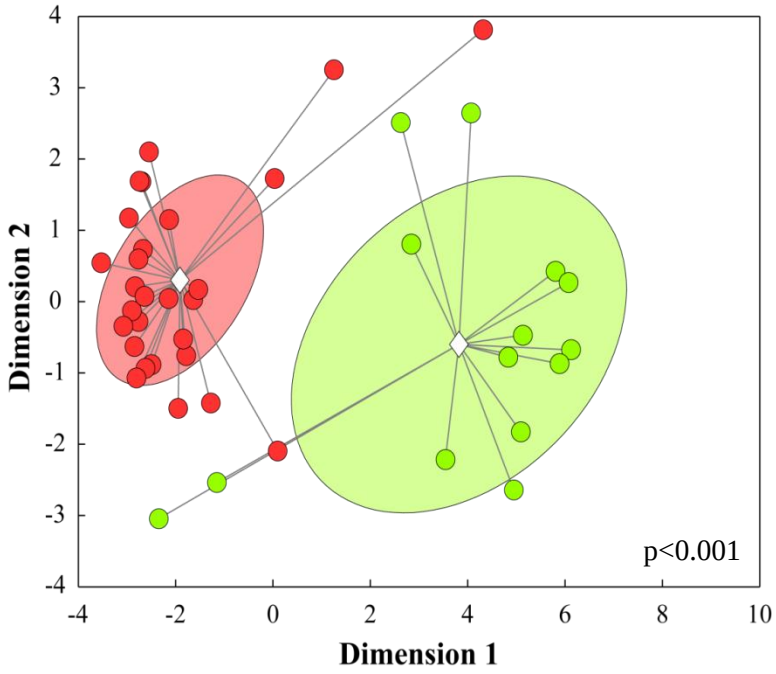

Figure 3.14 is the plot of random forest of egkHLT (red) and uskHLT (green) samples. This data is based on quantified metabolite data.

Table 3.7: Mean decrease gini scores

Metabolites	Mean Decrease Gini
Uracil	2.5
Tryptophan	2.3
Cytosine	2.0
Hypoxanthine	1.7
Propionate	1.0
Acetate	1.0
Aspartic Acid	0.7
Choline	0.6
Leucine	0.6
Butyrate	0.5
Formate	0.4
Isoleucine	0.4

SparCC

Correlation analysis provided by SparCC demonstrated positive and negative correlations among metabolites and provided information about correlations between groups of metabolites. Table 3.8 shows the negative correlations between all metabolites ($\leq$ -0.3 and $p \leq 0.01$). Table 3.9 shows positive correlations between all metabolites ($\geq$ 0.3 and $p \leq 0.01$). Table 3.10 shows the correlations between SCFAs and based on their median (0.41) they are positively correlated among each other. Table 3.11 shows the correlations between the amino acids and based on their median (0.17) they are positively correlated among each other. Table 3.12 shows the correlations between the amino acids, tryptophan and glycine, and based on their median (0.18) they are positively correlated. These two are grouped separately because they are higher in egkHLT. Table 3.13 shows the correlations between nucleotides (uracil, cytosine) and a nucleotide pre-cursor (hypoxanthine) and based on their median (0.40) they are positively correlated among each other. Table 3.14 shows the correlations between metabolites involved in

carbohydrate metabolites and based on their median (-0.4) they are negatively correlated. Table 3.15 shows the correlations between sugars and based on their median (0.08) they are positively correlated. Table 3.16 shows the correlations between other metabolites in the data set and based on their median (0.01) they are positively correlated. Table 3.17 show the correlations between different groups of metabolites. It shows positive correlations between glycine/tryptophan and SCFAs (0.08), nucleotides and SCFAs (0.11), metabolites involved in carbohydrate metabolism and SCFAs (0.01), and metabolites involved in carbohydrate metabolism and amino acids (0.03). It shows negative correlations between SCFAs and amino acids (-0.11), glycine/tryptophan and amino acids (-0.07), nucleotides and amino acids (-0.02), sugars and SCFAs (-0.04), sugars and amino acids (-0.12), other metabolites and SCFAs (-0.07), and other metabolites and amino acids (-0.01).

Table 3.8: Negative SparCC correlations

Negative Correlations (≤ -0.3)	
Metabolites Correlated	SparCC Value
formate → taurine	-0.35
glucose → lactate	-0.36
glycine → lactate	-0.40
glucose → leucine	-0.36
glycine → leucine	-0.35
glucose → valerate	-0.49
glycine → valerate	-0.46

Table 3.9: Positive SparCC correlations

Positive Correlations (≥ 0.3)	
Metabolites Correlated	SparCC Value
1-methylhistamine → glycine	0.39
1-methylhistamine → taurine	0.51
acetate → butyrate	0.73
acetate → propionate	0.84
alanine → leucine	0.45
butyrate → propionate	0.71
formate → propionate	0.33
glucose → glycine	0.81
isoleucine → leucine	0.55
lactate → leucine	0.34
lactate → valerate	0.40
lysine → taurine	0.33
propionate → valerate	0.34
uracil → tryptophan	0.57

Table 3.10: SCFAs SparCC correlations

SCFAs Correlation	
Metabolites Correlated	SparCC Value
acetate → butyrate	0.73
acetate → propionate	0.84
acetate → formate	0.21
acetate → valerate	0.16
butyrate → propionate	0.71
butyrate → formate	0.18
butyrate → valerate	0.34
formate → propionate	0.33
formate → valerate	0.20
propionate → valerate	0.37
Median	**0.41**

Table 3.11: Amino acid SparCC correlations

Amino Acid Correlations	
Metabolites Correlated	SparCC Value
alanine → aspartic acid	-0.05
alanine → isoleucine	0.41
alanine → leucine	0.45
alanine → lysine	0.38
alanine → tyrosine	0.27
alanine → valine	-0.02
aspartic acid → isoleucine	0.16
aspartic acid → leucine	0.15
aspartic acid → lysine	0.15
aspartic acid → tyrosine	0.10
aspartic acid → valine	0.03
isoleucine → leucine	0.55
isoleucine → lysine	0.21
isoleucine → tyrosine	0.38
isoleucine → valine	0.14
leucine → lysine	0.27
leucine → tyrosine	0.46
leucine → valine	0.17
lysine → tyrosine	0.17
lysine → valine	0.03
tyrosine → valine	0.05
Median	**0.17**

Table 3.12: Tryptophan and glycine SparCC

Tryptophan and Glycine Correlations	
Metabolites Correlated	SparCC Value
tryptophan → glycine	0.18
Median	**0.18**

Table 3.13: Nucleotide SparCC correlations

Nucleotide Correlations	
Metabolites Correlated	SparCC Value
cytosine → hypoxanthine	0.40
cytosine → uracil	0.59
hypoxanthine → uracil	0.40
Median	**0.40**

Table. 3.14: Carbohydrate metabolism SparCC correlations

Carbohydrate Metabolism	
Metabolites Correlated	SparCC Value
ethanol → fumarate	0.07
ethanol → lactate	-0.04
ethanol → malate	-0.09
ethanol → pyruvate	0.29
ethanol → succinate	0.03
fumarate → lactate	-0.10
fumarate → malate	-0.07
fumarate → pyruvate	-0.02
fumarate → succinate	-0.10
lactate → malate	-0.21
lactate → pyruvate	0.22
lactate → succinate	-0.25
malate → pyruvate	-0.05
malate → succinate	0.03
pyruvate → succinate	-0.04
Median	**-0.04**

Table. 3.15: Sugars SparCC correlations

Sugars Correlations	
Metabolites Correlated	SparCC Value
galactose → glucose	0.24
galactose → sucrose	0.08
glucose → sucrose	-0.04
Median	**0.08**

Table. 3.16: Other metabolite SparCC correlations

Other Metabolite Correlations	
Metabolites Correlated	SparCC Value
1-methylhistamine → bile acids	0.12
1-methylhistamine → choline	-0.04
1-methylhistamine → methylamine	0.14
1-methylhistamine → taurine	0.51
1-methylhistamine → trimethylamine	-0.05
bile acids → choline	0.03
bile acids → methylamine	0.00
bile acids → taurine	0.06
bile acids → trimethylamine	-0.09
choline → methylamine	-0.04
choline → taurine	-0.01
choline → trimethylamine	0.01
methylamine→ taurine	0.11
methylamine → trimethylamine	-0.02
taurine → trimethylamine	-0.06
Median	**0.01**

Table 3.17: Correlations between different groups of metabolites

Correlations Between Groups	
Metabolites Correlated	Median
SCFAs → amino acids	-0.11
nucleotides → SCFAs	0.11
nucleotides → amino acids	-0.02
glycine/tryptophan → SCFAs	0.08
glycine/tryptophan → amino acids	-0.07
carbohydrate metabolism→ SCFAs	0.01
carbohydrate metabolism → amino acids	0.03
sugars → SCFAs	-0.04
sugars → amino acids	-0.12
other metabolites → SCFAs	-0.07
other metabolites → amino acids	-0.01

Figure 3.15 demonstrates the correlation network of the metabolite data. SCFAs are correlated positively among each other and negatively between amino acids. Nucleotides are correlated positively among each other and positively between SCFAs. Amino acids are positively correlated among each other, except glycine and tryptophan which are negatively correlated to the other amino acids. Although a lot of these correlations are weak it gives us a better understanding as to what metabolites might interact with each other and their possible influence on each other.

Figure 3.15: SparCC correlation network

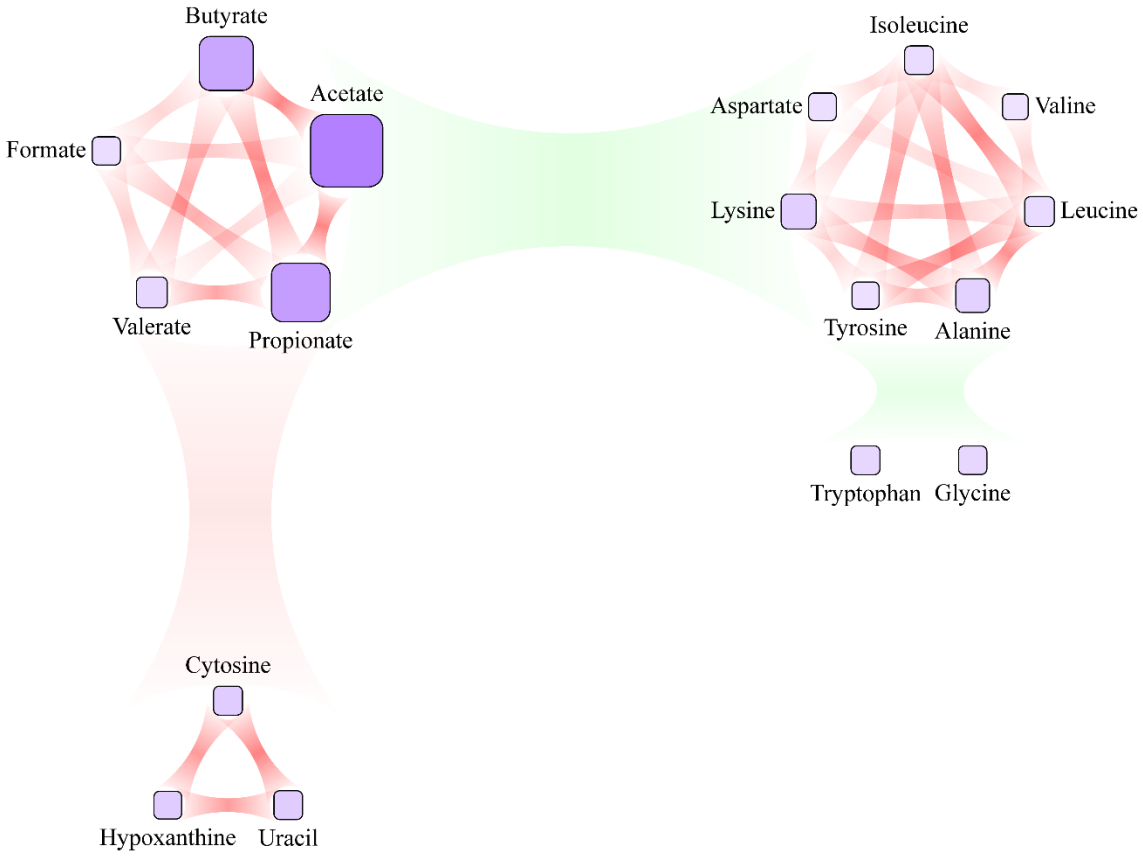

Figure 3.15 demonstrates the correlation network based on statistically significant SparCC values. Red lines are positive correlations and green lines are negative correlations. Saturation of the line corresponds to the strength of the correlation. The boxes represent metabolites and their abundance (lighter purple = less abundant, darker purple = more abundant) within the data set.

*Image was created by V. Shankar

4. Discussion

This work aimed to determine if there is a difference in metabolite profiles of fecal material collected from adolescents in two distinct geographical locations. Using ^{1}H NMR, egkHLT and uskHLT fecal metabolite levels were quantified and compared. Differences are thought to be an effect of diet differences of the Egyptian population and US population. These metabolite differences may cause certain health effects on the host.

Sample ordination analyses

Ordination analyses of the binned NMR data showed that the samples formed two distinct groups and could be classified based on nationality. Ordination analyses of metabolite data demonstrated that variability within the quantified metabolites formed two distinct clusters and samples could be classified based on nationality. Quantified metabolite data analyses also showed that uracil was the most important metabolite for classifying the data. Correlation analysis by SparCC showed positive and negative correlations within the metabolite data set. It also showed inter and intra correlations among the following groups: SCFAs, amino acids, tryptophan and glycine, carbohydrate metabolism metabolites, sugars, and the other metabolites in the data set.

Differences in Metabolite Levels Between egkHLT and uskHLT

The data shows that levels of SCFAs are higher in egkHLT compared to uskHLT. This may be due to the high intake of dietary fiber in the Egyptian diet. Egyptians have been shown to intake around 31g per day of fiber [61]. Dietary fiber is found in foods

such as whole grains, legumes, fruits, and vegetables [62]. The Egyptian's diet main food staples are bread made from high extraction wheat flour, cereals, legumes, and fresh dark leafy green vegetables [47]. The dietary fiber acquired from these foods can be converted into SCFAs in the colon by bacteria. In the United States, the recommended amount of dietary fiber is around 25g per day for women and 38g per day for men. However, Americans are only consuming around 15g per day [62]. This difference in dietary fiber intake could be a major contributor to the difference between the egkHLT/uskHLT fecal metabolite levels. SCFAs have been shown to have important anti-inflammatory effects which could potentially be used as therapies for gastrointestinal inflammatory diseases such as IBD [24].

The levels of almost all amino acid levels were higher in the uskHLT samples. This may also be linked to diet differences and the high amount of protein found in foods such as meats. According to an article about trends in meat consumption in the United States, children ages 12-19 consumed 121.8g per day of meat products [29]. Meat in take in Egypt is only around 72g per day [61]. Meat consumptions may be lower because it is a luxury item and religious and cultural observations prevent some Egyptians from eating meat at certain times [47]. This high meat consumption in the United States could play a key role in the development of several health diseases such as colorectal cancer [28].

Another difference shown in our data is that the amino acid tryptophan is higher in egkHLT compared to uskHLT. Tryptophan is high in foods such as seeds, legumes, soy protein, dairy products, and animal protein [63]. In a study done by Young et. al., they showed that tryptophan is slightly higher in plant protein sources (nuts, seeds, and legumes) than in animal protein [64]. In the US only around 30% of protein intake comes

from plants where as almost 70% comes from animal protein. However, in Africa and Middle East around 14% of protein intake is animal and around 79% is plant [64]. This could account for the levels of tryptophan being higher in egkHLT.

Other metabolites also show a difference between egkHLT and uskHLT. For example, bile acids are higher in uskHLT. Bile acids are increased during high fat intake and high fat foods seem to be consumed more in the US diet [32]. The Dietary Guidelines for American's say that adolescents should only consume 30% of calories from fat and only 10% of that fat from saturated fats [65]. A study by Reedy et. al., determined based on National Health and Nutrition Examination Survey (NHANES) showed that 40% of total energy (798 of the 2,027 kcal per day) consumed by 2-18 year olds, came from solid fat and added sugars. The major sources of these fats were from pizza, grain desserts, whole milk, cheese, and fatty meats [66]. Egyptian diet tends to be lower in saturated fats [47]. In a study done by Galal, dietary energy from fat in Egypt was 28.7% for children, and primarily came from vegetable oil in mixed dishes [67]. This is almost 10% lower than US adolescents and could account for the difference in levels of bile acids. Increased bile acid levels have been linked to a variety of cancers in humans [32].

Another metabolite found higher in uskHLT samples was taurine. Taurine is present in many foods and is high in foods such as meat, fish, and energy drinks. Meat and fish can contain anywhere from 300-6000 μmol per 100g per wet weight of taurine. As where fruits, vegetables, nuts, and legumes contain ranges less than 1 μmol per 100g per wet weight [36]. Meat consumption as mentioned above seems to be more prevalent in the United States diet. Another source of taurine in the United States diet is through energy drinks. Energy drinks are a billion dollar business in the United States and half of

the market is consumed by adolescents (< 12, 12-18 years old) and young adults (19-25 years old) [37]. Energy drinks which usually come in 16 or 24 oz cans, can contain up to 1,000 mg of taurine in an 8.0-8.4 ounce serving [68]. According to one article, thirty-one percent of energy drink consumers are ages 12-17 years old and 28 percent of 12-14 year olds report to regularly consume energy drinks, which is the average age of the uskHLT cohort [69]. The high consumption of meat and energy drinks could account for the higher amount of taurine in the uskHLT samples. Over consumption of taurine can cause adverse effects such as increased heart rate and hypoglycemia [37].

Choline was also found to be higher in uskHLT. Choline is rich in foods such as eggs, liver, and peanuts. In the US diet the major contributors of choline includes meat, poultry, egg-based dishes, and dairy foods [70]. Choline can also be found in wheat and spinach, which are main staples in the Egyptian diet [70]. However, high meat intake as mentioned before in the US diet could be the reason for the slight increase choline levels in uskHLT. Choline deficiency has been linked to liver and muscle dysfunction [39].

Connections to previous studies

Previous studies have shown a link in diet and metabolite differences among different groups of people. However, most of these studies focus on specific groups of metabolites and use techniques other than NMR for metabolite quantification. For example, as mentioned in the introduction an article by De Flippo et. al. compared levels of SCFAs between an African group of adolescents from Burkina Faso (BF) and European adolescents from Florence, Italy (EU) [45]. The researchers believed that diet differences caused changes in the microbiota composition and this resulted in the higher

amount of SCFAs in the BF cohort. The SCFAs quantifications were determined by mass spectrometry [45]. Another study compared individuals on two different short-term diets, one consisting of plant-based foods and one of animal product based foods. Using gas chromatography and mass spectrometry they determined that people on the animal-based diet had higher levels of bile acids (specifically DCA) and people on the plant-based diet had higher levels of SCFAs [71]. These examples prove that diet differences can be linked to differences in metabolites and we wanted to expand on that concept by looking at more metabolites using the nondestructive and noninvasive technique of NMR.

Conclusion

^{1}H NMR analysis on fecal material has been previously shown by researchers as a valuable tool to determine metabolite profiles [39, 40]. This study confirms this approach and concludes that there are differences in metabolite profiles between Egyptian and US adolescents. Major differences include, SCFAs being higher in egkHLT samples and most amino acids being higher in uskHLT samples. These differences in metabolite levels is thought to be due to the varying diets of the two populations. Based on a literature, I was able to link certain metabolites to dietary products and determine their involvement with human health. Although my analysis is limited to fecal material and 32 metabolites, my hope is that other researchers can use the same concepts and methods to look at other parameters (blood, urine) and gain a more comprehensive profile of metabolites between individuals. I also hope this work can help scientist and medical professionals gain a better knowledge as to what foods influence metabolite profiles and how this could promote or be detrimental to human health.

References

1. *Understanding Microbes in Sickness and in Health* 2009, National Institue of Health

2. *Introducing Microbes*. 2015 [cited 2015 October 29]; Available from: http://www.microbiologyonline.org.uk/about-microbiology/introducing-microbes.

3. Paliy, O. and V. Shankar, *Applications of Phylogenetic Microarrays to Profiling of Human Microbiomes*.

4. Hayashi, H., M. Sakamoto, and Y. Benno, *Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods.* Microbiol Immunol, 2002. **46**(8): p. 535-48.

5. Dethlefsen, L., et al., *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing.* PLoS Biol, 2008. **6**(11): p. e280.

6. Neish, A.S., *Microbes in gastrointestinal health and disease.* Gastroenterology, 2009. **136**(1): p. 65-80.

7. Baviera, G., et al., *Microbiota in healthy skin and in atopic eczema.* Biomed Res Int. **2014**: p. 436921.

8. Sanford, J.A. and R.L. Gallo, *Functions of the skin microbiota in health and disease.* Semin Immunol. **25**(5): p. 370-7.

9. Wade, W.G., *The oral microbiome in health and disease.* Pharmacol Res. **69**(1): p. 137-43.

10. Zhang, Y.J., et al., *Impacts of gut bacteria on human health and diseases.* Int J Mol Sci. **16**(4): p. 7493-519.

11. Nicholson, J.K., et al., *Host-gut microbiota metabolic interactions.* Science. **336**(6086): p. 1262-7.

12. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns,* National Acad Sciences. p. 11971-11975.

13. Palmer, C., et al., *Development of the human infant intestinal microbiota.* PLoS Biol, 2007. **5**(7): p. e177.

14. Guinane, C.M. and P.D. Cotter, *Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ.* Therap Adv Gastroenterol. **6**(4): p. 295-308.

15. Diseases, N.I.o.D.a.D.a.K. *Gastrointestinal Tract (GI Tract).* 2015 [cited 2015 November 10]; Available from: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0022855/.

16. Gerritsen, J., et al., *Intestinal microbiota in human health and disease: the impact of probiotics.* Genes Nutr. **6**(3): p. 209-40.

17. Frank, D.N., et al., *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13780-5.

18. Guarner, F. and J.R. Malagelada, *Gut flora in health and disease.* Lancet, 2003. **361**(9356): p. 512-9.

19. Suzuki, K., et al., *Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1981-6.

20. Tremaroli, V. and F. Backhed, *Functional interactions between the gut microbiota and host metabolism.* Nature. **489**(7415): p. 242-9.

21. Macfarlane, G.T., J.H. Cummings, and C. Allison, *Protein degradation by human intestinal bacteria.* J Gen Microbiol, 1986. **132**(6): p. 1647-56.

22. Macfarlane, S. and G.T. Macfarlane, *Regulation of short-chain fatty acid production.* Proc Nutr Soc, 2003. **62**(1): p. 67-72.

23. Kasubuchi, M., et al., *Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation.* Nutrients. **7**(4): p. 2839-49.

24. Tedelind, S., et al., *Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease.* World J Gastroenterol, 2007. **13**(20): p. 2826-32.

25. Tang, Y., et al., *G-protein-coupled receptor for short-chain fatty acids suppresses colon cancer.* Int J Cancer. **128**(4): p. 847-56.

26. Hughes, R., E.A. Magee, and S. Bingham, *Protein degradation in the large intestine: relevance to colorectal cancer.* Curr Issues Intest Microbiol, 2000. **1**(2): p. 51-8.

27. Nyangale, E.P., D.S. Mottram, and G.R. Gibson, *Gut microbial activity, implications for health and disease: the potential role of metabolite analysis.* J Proteome Res. **11**(12): p. 5573-85.

28. Smolinska, K. and P. Paluszkiewicz, *Risk of colorectal cancer in relation to frequency and total amount of red meat consumption. Systematic review and meta-analysis.* Arch Med Sci. **6**(4): p. 605-10.

29. Daniel, C.R., et al., *Trends in meat consumption in the USA.* Public Health Nutr. **14**(4): p. 575-83.

30. Pan, A., et al., *Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis.* Am J Clin Nutr. **94**(4): p. 1088-96.

31. Staels, B. and V.A. Fonseca, *Bile acids and metabolic regulation: mechanisms and clinical responses to bile acid sequestration.* Diabetes Care, 2009. **32 Suppl 2**: p. S237-45.

32. Bernstein, H., et al., *Bile acids as endogenous etiologic agents in gastrointestinal cancer.* World J Gastroenterol, 2009. **15**(27): p. 3329-40.

33. Bernstein, C., et al., *Carcinogenicity of deoxycholate, a secondary bile acid.* Arch Toxicol. **85**(8): p. 863-71.

34. Schlottman, K., et al., *Characterization of bile salt-induced apoptosis in colon cancer cell lines.* Cancer Res, 2000. **60**(15): p. 4270-6.

35. Mobraten, K., et al., *Activation of the bile acid receptor TGR5 enhances LPS-induced inflammatory responses in a human monocytic cell line.* J Recept Signal Transduct Res. **35**(5): p. 402-9.

36. Lambert, I.H., et al., *Physiological role of taurine--from organism to organelle.* Acta Physiol (Oxf). **213**(1): p. 191-212.

37. Seifert, S.M., et al., *Health effects of energy drinks on children, adolescents, and young adults.* Pediatrics. **127**(3): p. 511-28.

38. Hollenbeck, C.B., *An introduction to the nutrition and metabolism of choline.* Cent Nerv Syst Agents Med Chem. **12**(2): p. 100-13.

39. Sha, W., et al., *Metabolomic profiling can predict which humans will develop liver dysfunction when deprived of dietary choline.* Faseb J. **24**(8): p. 2962-75.

40. Dumas, M.E., et al., *Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice.* Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12511-6.

41. Smith, I.C. and D.E. Blandford, *Nuclear magnetic resonance spectroscopy.* Anal Chem, 1995. **67**(12): p. 509R-518R.

42. Reo, N.V., *NMR-based metabolomics.* Drug Chem Toxicol, 2002. **25**(4): p. 375-82.

43. Jacobs, D.M., et al., *(1)H NMR metabolite profiling of feces as a tool to assess the impact of nutrition on the human microbiome.* NMR Biomed, 2008. **21**(6): p. 615-26.

44. Ndagijimana, M., et al., *Effect of a synbiotic food consumption on human gut metabolic profiles evaluated by (1)H Nuclear Magnetic Resonance spectroscopy.* Int J Food Microbiol, 2009. **134**(1-2): p. 147-53.

45. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.* Proc Natl Acad Sci U S A. **107**(33): p. 14691-6.

46. Schnorr, S.L., et al., *Gut microbiome of the Hadza hunter-gatherers.* Nat Commun. **5**: p. 3654.

47. Hassan-Wassef, H., *Food habits of the Egyptians: newly emerging trends.* East Mediterr Health J, 2004. **10**(6): p. 898-915.

48. Grotto, D. and E. Zied, *The Standard American Diet and its relationship to the health status of Americans.* Nutr Clin Pract. **25**(6): p. 603-12.

49. Shankar, V., et al., *The networks of human gut microbe-metabolite associations are different between health and irritable bowel syndrome.* Isme J. **9**(8): p. 1899-903.

50. Anderson, P.E., et al., *Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data*. 2011, Springer. p. 179-190.

51. Wishart, D.S., et al., *HMDB: the human metabolome database*. 2007, Oxford Univ Press. p. D521-D526.

52. Information, N.C.f.B., *L-Tryptophan*. 2015, PubChem Compound Database.

54. van den Berg, R.A., et al., *Centering, scaling, and transformations: improving the biological information content of metabolomics data*. 2006, BioMed Central Ltd. p. 142.

55. Kucera, M. and B.r.A. Malmgren, *Logratio transformation of compositional data: a resolution of the constant sum constraint*. 1998, Elsevier. p. 117-120.

56. Wiklund, S. *Multivariate Data Analysis for Omics*. 2008 [cited 2015; Available from: http://metabolomics.se/Courses/MVA/MVA%20in%20Omics_Handouts_Exercises_Solutions_Thu-Fri.pdf.

57. Pearson, K., *LIII. On lines and planes of closest fit to systems of points in space*. 1901, Taylor & Francis. p. 559-572.

58. Trygg, J. and S. Wold, *Orthogonal projections to latent structures (O-PLS)*. 2002. p. 119-128.

59. Breiman, L., *Random forests*. 2001, Springer. p. 5-32.

60. Friedman, J. and E.J. Alm, *Inferring correlation networks from genomic survey data*.

61. Karamanos, B., et al., *Nutritional habits in the Mediterranean Basin. The macronutrient composition of diet and its relation with the traditional Mediterranean diet. Multi-centre study of the Mediterranean Group for the Study of Diabetes (MGSD).* Eur J Clin Nutr, 2002. **56**(10): p. 983-91.

62. Slavin, J.L., *Position of the American Dietetic Association: health implications of dietary fiber*. J Am Diet Assoc, 2008. **108**(10): p. 1716-31.

63. US Department of Agriculture, A.R.S., Nutrient Data Laboratory, *USDA National Nutrient Database for Standard Reference*. 2015.

64. Young, V.R. and P.L. Pellett, *Plant proteins in relation to human protein and amino acid nutrition.* Am J Clin Nutr, 1994. **59**(5 Suppl): p. 1203S-1212S.

65. Story, M. and J. Stang, *Nutrition needs of adolescents*. 2005. p. 21-34.

66. Reedy, J. and S.M. Krebs-Smith, *Dietary sources of energy, solid fats, and added sugars among children and adolescents in the United States.* J Am Diet Assoc. **110**(10): p. 1477-84.

67. Galal, O.M., *The nutrition transition in Egypt: obesity, undernutrition and the food consumption context.* Public Health Nutr, 2002. **5**(1A): p. 141-8.

68. Clauson, K.A., et al., *Safety issues associated with commercially available energy drinks.* J Am Pharm Assoc (2003), 2008. **48**(3): p. e55-63; quiz e64-7.

69. Simon, M.R. and J.F. Mosher, *Alcohol, energy drinks, and youth: a dangerous mix*. 2007: Marin Institute.

70. Institue, L.P. *Choline*. [cited 2015 October 23]; Available from: http://lpi.oregonstate.edu/mic/other-nutrients/choline#reference27.

71. David, Lawrence A. et al. "Diet Rapidly and Reproducibly Alters the Human Gut Microbiome." *Nature* 505.7484 (2014): 559–563. *PMC*. Web. 7 Jan. 2016.