# Microbiota-metabolites interactions in non-human primate gastrointestinal tract

Ce Yuan[1,2], Melanie Graham[2], Subbaya Subramanian[1,2,3*]

1. Bioinformatics and Computational Biology Program, University of Minnesota, Rochester, MN, 55904.
2. Department of Surgery, University of Minnesota, Minneapolis, MN, 55455.
3. Masonic Cancer Center, University of Minnesota, Minneapolis, MN, 55455

*Contact Information to whom correspondence should be directed:

Dr. Subbaya Subramanian

email: subree@umn.edu

Conflicts of interest: No conflicts of interest

## Abstract

### Background

The microbiota has been recognised as an important part for maintaining human health. Perturbation to its structure has been implicated in many diseases, such as obesity and cancers. The microbiota is highly metabolically active and fills in many niche metabolic pathways absent from the human host. Diseases such as obesity, cardiovascular disease and colorectal cancer has been linked to altered microbiota metabolism. However, there is a gap in the current knowledge on how mucosal-associated microbiota and colon mucosa interact. Here we performed an integrated analysis between the mucosal-associated microbiota and the mucosal tissue metabolites in healthy non-human primates.

### Results

We found that the overall microbiota composition is influenced by both the tissue location as well as the host. We also identified bacteria signatures for different intestinal locations. The distal colon bacterial signature includes *Ruminococcaceae*, *Bacteroidales*, *Christensenellaceae*, *Clostridiales*, *Sphaerochaeta*, *Victivallaceae*, *GMD14H09*, *CF231*, *ML615J-28*, *RF39* and *R4-45B* taxa. In the cecum, the signatures include *Prevotella*, *Anaerovibrio*, *Roseburia*, and *Anaerostipes*. *Desulfovibrionaceae* family is the only taxon that may be a signature for the duodenum. We also found an intricate global relationship between the microbiota and the host tissue metabolome that is mainly driven by the distal colon. Most importantly, we found microbial-centric tissue

metabolites clusters that may have potential implications to studying host-microbiota metabolic interactions.

**Keywords:** Microbiota, metabolome, host-microbiota interactions, non-human primate.

## Background

The human gastrointestinal tract harbors trillions of microorganisms, including thousands of species of bacteria, termed microbiota [1]. It has become evident that the gut microbiota is important in regulating and maintaining the health of the host and is implicated in many diseases, including cancers [1–5]. Despite numerous studies indicating important roles of microbiota in diseases, many of these studies have largely focused on the taxonomic composition of the microbiota. The mechanism of the host-microbiota interaction, however, still remains unclear.

Previous studies suggest, the gut microbiota produces a vast amount of metabolites. Some metabolites - e.g., vitamin K, biotin, vitamin Bs, and short-chain fatty acids (SCFAs) are essential to maintaining homeostasis in the colon microenvironment [6–8]. This metabolic interaction between the host and its microbiota has widespread implications around the body [7]. For example, the obesity associated microbiota has been shown to possess increased metabolic capability to harvest energy from food [9, 10]. The metabolism of L-carnitine by the gut microbiota has been shown to promote atherosclerosis [11]. These studies suggest a potential metabolic adaptations of the microbiota in response to the host metabolic change [10].

The most direct metabolic interaction between the host and its microbiota, however, is at the colon. In fact, the majority (~70%) of energy source required by the normal colon epithelium come from butyrate produced by the microbiota through fermentation of complex carbohydrates [12]. Without a functional microbiota, the colon epithelia will undergo autophagy and fail to maintain its normal structure and function [13]. More

importantly, this metabolic interactions may have important implications in colorectal cancer, the 2nd most deadly cancer in the United States [1, 5–7, 14–16].

Most previous studies in microbiota used fecal samples or biopsy samples due to the ease of sampling. Thus, the mucosal host-microbiota metabolic interactions along healthy gastrointestinal tract is largely unknown. Here, we investigate such metabolic interactions in 10 healthy baboons, a family of Old World monkeys belong to the Papio genus. We collected tissue samples from 6 small and large intestine segments and sequenced the 16S rRNA gene to identify the mucosal-associated microbiota composition. We also performed untargeted metabolomics on the immediate adjacent tissues to profile the tissue metabolites composition. To our knowledge, this is one of the first analyses to comprehensively establish the intestinal host-microbiota metabolic interactions in NHPs.

## Results

### Microbiota along the non-human primate gastrointestinal tract.

We first assessed the NHP GI tissue-associated microbiota composition in 10 baboons using 16S rRNA gene sequencing method (**Supplementary Table 1**). At the phyla level, the NHP GI tissue-associated microbiota is dominated by the bacterial phyla of Firmicutes, Bacteroidetes and Proteobacteria, regardless of the tissue location (**Figure 1A, B**). This microbiota composition is similar to that observed in human GI tissue-associated microbiota, however dissimilar to that observed in mouse fecal samples (**Supplementary Figure 1**). [5] In the NHP samples, most phyla level

composition remain unaltered along the GI tract except for Tenericutes and Lentisphaerae ($p < 0.005$, ANOVA with Tukey post-hoc test). Both of these two phyla are more abundant in the distal colon compared to all other locations. Since many important information were masked at the phyla level, we next analyzed diversity metrics to better understand the microbiota composition difference at the OTU resolution. We first assessed the beta-diversity between tissue locations by performing Principal Coordinate Analysis (PCoA) using both weighted and unweighted UniFrac distance metrics. The unweighted UniFrac distance only consider the presence and absence of a certain OTU, while weighted UniFrac distance will consider the abundance, thus these metrics can give an overview on the microbial structure difference of different tissue locations [17]. The PCoA of unweighted UniFrac distance show clusterings mainly based on the tissue location ($p < 0.01$, PERMANOVA) as well as the sample origin ($p < 0.001$), the weighted UniFrac distance also show a same clustering based on the tissue location ($p < 0.05$) and sample origin ($p < 0.001$) (**Figure 1C, Supplementary Figure 2**). This suggests both host and tissue location may have an impact on the mucosal-microbiota structure in the intestines [1].

We then analyzed the alpha-diversity to discern the microbial diversity within each sample. Consistent with previous reports, we found the small intestinal microbiota has significantly lower phylogenetic diversity ($p < 1 \times 10^{-7}$, two-tailed t-test, **Figure 2A**), lower Shannon index ($p < 0.0005$, **Figure 2B**) and lower chao1 index ($p < 1 \times 10^{-6}$, **Figure 2C**) [18, 19]. This observation is likely due to the microbial concentration gradient along the GI tract, where the small intestine harbor less bacteria due to the

high pH environment. We then assessed the microbiota differences at the genus and the OTU level to discover site specific bacterial community signature. We found 17 bacterial taxa to significantly differ in relative abundance in at least 1 tissue site ($p < 0.1$, ANOVA with Tukey post-hoc test). Twelve of these bacterial taxa were enriched in distal colon, 4 were enriched in cecum and 1 was enriched in the duodenum compared to other locations (**Supplementary Figure 3**). In the distal colon, where there is higher microbiota diversity (**Figure 2**), *Ruminococcaceae*, *Bacteroidales*, *Christensenellaceae*, *Clostridiales*, *Sphaerochaeta*, *Victivallaceae*, *GMD14H09*, *CF231*, *ML615J-28*, *RF39* and *R4-45B* taxa were significantly higher compared to at least 1 other tissue location (**Supplementary Figure 3**). In the cecum, *Prevotella*, *Anaerovibrio*, *Roseburia*, and *Anaerostipes* show higher relative abundance. *Desulfovibrionaceae* family is the only taxon that was more enriched in the duodenum and it is only significantly enriched compared to the jejunum and ileum. It is not surprising that the distal colon and cecum harbors more distinct bacteria taxa compared to other locations. Previous studies have shown that both distal colon and the cecum are where most bacteria fermentation take place.

**Microbiota-Metabolome interactions.**

Using untargeted metabolomics method, we analyzed the tissue metabolome composition in tissue samples immediately adjacent to the tissues used for 16S rRNA gene sequencing. A total of 3,395 compounds were present in at least 2/3 of all the samples analyzed (**Supplementary Table 2**). After searching against the Human Metabolome Database (HMDB) and in house libraries generated by the University of

Minnesota Center for Mass Spectrometry and Proteomics, a total of 292 compounds were assigned putative identity. To analyze the global microbiota-metabolome relatedness, we performed the Procrustes analysis using the *vegan* package in R (**Figure 3**). We first performed Principal Component Analysis (PCA) on the complete metabolome and microbiota data (**Figure 3A, B**), then scaled and superimposed the microbiota PCA onto the metabolome PCA (**Figure 3C**). Globally, the Procrustes analysis shows significant relatedness ($p = 0.039$) between the microbiota and the metabolome. Interestingly, this relatedness is only driven by the distal colon ($p = 0.0088$) (**Supplementary Figure 4**).

We then analyzed the microbiota-metabolites relationships using spearman's ranked correlation on the metabolites with assigned identity. The spearman's ranked correlation is a nonparametric test that can be used to reveal subtle relationship between microbiota and metabolites [20]. We first summarized the microbiota OTU data to the genus level, then performed the correlation with the corresponding metabolites data. Globally, a total of 4,410 significant correlations ($q < 0.1$, False discovery rate adjusted p-value) were found between the microbiota and the known metabolites (**Supplementary Figure 5**). Interestingly, the small intestine had 315 significant interactions (**Supplementary Figure 6A**) versus 139 for the large intestine (**Supplementary Figure 6B**). Additionally, the correlation network for the small intestine is more interconnected as compared to the large intestine. One explanation is that the large intestine harbors more bacterial species compared to the small intestine, thus there could be more functional redundancies in the large intestine which means less

correlations at the lower levels of the OTU hierarchy. Indeed, when we performed a similar analysis using the phyla level data, the large intestine had 78 significant correlations ($q < 0.1$) compared to 39 significant correlations ($q < 0.1$) for the small intestine (**Supplementary Table 3**). Interestingly, both networks appears to have a microbiota-centric correlation, where most metabolites only correlation to few bacteria and such correlations tend to be at the same direction. For example, most metabolites are negatively correlated with CF231, Alphaproteobacteria and Flexispira in the large intestine, and those metabolites correlated with Prevotella and Dorea have mostly positive correlations. The same structure maintains when we performed the global correlation analysis, without regard of the tissue location (**Supplementary Figure 5**). This data warrant further investigation to elucidate the causal-relationship between the mucosal microbiota and metabolites composition change.

Interestingly, among the top correlations, metabolites commonly found in several vegetables have significant correlations with several bacteria. For example, 6-Hydroxypentadecanedioic acid and 1-Isothiocyanato-7-(methylthio)heptane have 19 and 16 significant correlations in the small intestines. The 3H-1,2-Dithiole-3-thione have 7 significant correlations in the large intestine (**Figure 4**). All 3 compounds are commonly found in vegetables, which were included as part of the normal diet enrichment all animals received. All 19 of the bacteria have positive correlations with this compound, including Christensenellaceae, Bifidobacterium and Lactobacillus, all have shown health benefits in humans [21, 22]. This suggests that these compounds may have a prebiotics effect. Additionally, these compounds are present in different

amount in the intestines. Both 6-Hydroxypentadecanedioic acid and 1-Isothiocyanato-7-(methylthio)heptane have lower presence in the small intestines and 3H-1,2-Dithiole-3-thione have lower presence in the large intestine. One explanation is the differences of absorption at different locations of the intestine which can lead to different metabolite concentrations in the intestinal lumen that can affect the bacterial composition. Due to the potential health benefits associated eating brassicas vegetables, this finding warrant additional investigations.

**Discussion**

Currently, there is limited knowledge in the microbiota composition along different sections of the GI tract in either human or NHP samples [23, 24]. Studies in human subjects usually require prior bowel preparation, which have been shown to alter the microbiota [25]. In this study, we collected tissue samples from healthy non-human primates without prior bowel preparation, thus providing an unaltered view of the healthy microbiota. Previous studies have analyzed the GI tract microbiota compositions in mouse, chicken, dog, cow and horse [19, 26–29]. However, due to the anatomical differences, in addition to the dietary and genetic difference, these animals may have different microbiota along the GI tract. To our knowledge, this study is the first report to comprehensively analyze the mucosa-associated microbiota-metabolites interactions along the GI tract in NHPs.

In this study, we performed untargeted metabolomics on the intestinal tissues. Although it is able to identify over 3,395 compounds, we were only able to assign identities to 292 compounds. This lack of positive identification is mainly due to the lack of database

available. It is conceivable that when such database becomes available, we will be able to extract additional information from the data. Another caveat of the current study is that we were not able to identify the metabolites origin. Future studies should aim to separate metabolites originated from the host, microbiota or food source.

Similar to previous reports in humans, we found variations to the microbiota composition between different NHP subjects. In addition, we found that the microbiota composition along the GI tract is also influenced by the host. Previous studies suggest that this variation between individuals can be attributed to factors such as genetics, dietary preferences and other factors. [1, 30, 31] This study also characterized the microbiota signatures along different sections of the GI tract, and supported the previous hypothesis that the small intestines harbor less diverse microbiota compared to the large intestine.

This study also found that the host-microbiota metabolic interactions are microbiota-centric. Most metabolites are correlated with few microbes and in the same direction. It is important to note that this analysis could not identify the directionality of such interactions. Surprisingly, only 4 microbiota-metabolome correlation pairs were found in common between the small and large intestine. As speculated previously, it is possibly due to the higher level of functional redundancy present in the large intestine due to the larger species present. Another possible explanation is that the pH, as well as the nutrient composition are different between the small and large intestines, and these factors together may have additional effects on the microbiota and the tissue metabolome.

Our analysis also found the 6-Hydroxypentadecanedioic acid, 1-Isothiocyanato-7-(methylthio)heptane, and 3H-1,2-Dithiole-3-thione, compounds commonly found in Brassica vegetables, are correlated with higher levels of several beneficial bacteria. This may suggest a potential prebiotic effect of these compounds. Moreover, the location specific correlation may suggest a potential strategy to target beneficial bacteria in different intestinal locations. Notably, 3H-1,2-Dithiole-3-thione has been previously shown as a potent antioxidant and potential chemopreventive agent, by targeting the transcription factor NRF2. [32]

**Conclusions**

In the present study, we report the host-microbiota interactions along the healthy non-human primate lower gastrointestinal tract. Our study provided a global view of the microbiota landscape of healthy NHPs. Our analysis suggests an intricate global relationship between the microbiota and the metabolites along the GI tract. Further functional validation is warranted to establish the directionality of such interactions.

**Methods**

All protocols and procedures were approved by the University of Minnesota Institutional Animal Care and Use Committee, conducted in compliance with the Animal Welfare Act, and animals were housed and cared for according to the standards detailed in the Guide for the Care and Use of Laboratory Animals. The cohort included 10 adult purpose-bred female olive baboons *(Papio anubis)* modeling ACL injury and subsequent repair using regenerative medicine techniques. Animals were between 6.5 and 15.6 years (median, 9.3 years) in age and weighed between 14.4 and 24.9 kg

(median, 20.1 kg). They were pair housed or housed in protected contact with compatible conspecifics. Baboons had free access to water and were fed identical diets that included biscuits (Harlan Primate Diet 2055C, Harlan Teklad) based on body weight and daily enrichment with fresh fruits, vegetables, grains, beans, nuts, and a multivitamin preparation. Semiannual veterinary physical examinations were performed in all animals. Animals participated in an environmental enrichment program designed to encourage sensory engagement, enhance foraging behavior and novelty seeking, promote mental stimulation, increase exploration and play and activity levels, and strengthen social behaviors, together providing opportunities for animals to increase time budget spent on species-typical behaviors. Baboons were trained to cooperate in medical procedures including hand feeding and drinking, shifting into transport cages for sedation and targeting or presentation for examination. Animals were euthanized via barbiturate overdose (Beuthanasia-D ≥86 mg/kg IV) and tissue procurement performed post mortem. No oral medications were used for at least 6 months prior to tissue collection. Approximately 1cm x 1cm tissue sections that included duodenum (D), jejunum (J), ilium (I), cecum (C), proximal colon (P) and distal colon (Dist) were collected using clean technique and snap-frozen in liquid nitrogen then stored at -80°C.

**16S rRNA gene sequencing and sequence analysis.** Total DNA was extracted from approximately 250 mg of tissue using DNeasy PowerSoil Kit (Cat: 12888; Qiagen, Valencia, CA) following the standard protocol. Sequencing libraries were created by the Mayo Clinic Genome Analysis Core (Rochester, MN). Briefly, the V3-V5 region of the 16S rRNA gene was amplified with multiplexing barcodes using PCR (V3-341F:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGAGGCAGCAG;

V5-926R:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGTCAATTCMTTTRAGT). The libraries were then pooled and size-selected between 700 and 730 bp using a LabChip XT (PerkinElmer, Waltham, MA). Sequencing was performed on a single lane of an MiSeq sequencer (Illumina) using paired-end mode. On average, 64,937 quality reads (between 9,901 and 118,288) were generated per library. The sequencing results were analyzed using the gopher-pipelines (metagenomics-pipeline) developed and maintained by the University of Minnesota Informatics Institute. Briefly, the adapters and low quality reads were first trimmed using Trimmomatic v0.33. Then, the forward and reverse read pairs were merged using PandaSeq v2.8 [33]. OTUs were then picked using QIIME v1.9.1 “pick_open_reference_otus.py” script against Greengenes 16S database (May, 2013 release), allowing 97% similarity [34, 35]. The unfiltered OTU table is available in **Supplementary Table 1**.

**Metabolites extraction.** Metabolites were extracted from the immediate adjacent tissue that was used to generate 16S rRNA sequencing. There were insufficient amount of duodenum tissue from animal B09 to perform untargeted metabolomics, and were thus not analyzed. Approximately 15 mg of tissue was used to extract metabolites. The tissues were first grind into fine powder using CryoGrinder (OPS Diagnostics) on dry ice. The tissues were then suspended in 20 ul of 80% methanol per 1 mg tissue weight. The mixture were then homogenized using a probe sonicator at 10% amplitude for 15 s, with 1 minute rest on ice in between every 5 s of sonication. The sonicated samples

were then centrifuged at 14,000x g for 10 min at 4 °C. The supernatant from the centrifugation contain the metabolites and was saved in -80 °C before dry-down. The tissue pellets were then further processed for additional metabolites extraction. They were first suspended in 10 ul of 80% methanol per 1 mg of original tissue weight and sent through high pressure cycling on a Barocycler NEP2320 (Pressure Biosciences). The high pressure cycling protocol include 60 cycles of 20 s of 35,000 psi pressure, followed by 10 s of 0 psi for at 4 °C. After pressure cycling, the samples were again centrifuged at 14,000x g for 10 min at 4 °C and the supernatant were pooled with the previously extracted metabolites. Finally, the metabolites were dried under nitrogen stream.

**Untargeted metabolomics.** The dried metabolites were first suspended in 15 ul of 0.1% formic acid per 1 mg tissue weight. The suspensions were then separated for analysis using a C18 reverse-phase column and hydrophilic interaction liquid chromatography (HILIC) column. The reverse-phase analysis results in separation of larger non-polar molecules such as steroid-like compounds, certain amino acids, phospholipids and other lipids, while the HILIC analysis separates hydrophilic compounds such as amino acids and member of the citric acid cycle and glycolysis pathways. The samples were analyzed using reverse-phase positive mode (non-polar interaction) separation and HILIC analysis (polar interaction) separation before analyzed by Q Exactive LC-MS/MS quadrupole Orbitrap (Thermo Scientific). The reverse-phase analysis was performed in positive mode ionization with an additional proton (+1.0073) added. For HILIC analysis, the negative ionization mode was used with one additional

proton (-1.0073) removed. Since salts are present, compounds may occasionally form as a sodium salt (neutral mass plus 21.9944) for positive or a chloride salt (neutral mass plus 34.9688) for negative mode. Samples were loaded and analyzed in random orders and quality control samples were analyzed in regular intervals to eliminate extraneous signals. The untargeted metabolomics were performed by the University of Minnesota Center for Mass Spectrometry and Proteomics.

**Metabolomics data analysis.** The data was processed using Progenesis QI software (Thermo). The software first aligns all the features obtained in all the runs and then assigns intensity measures for features found in all the runs. The raw data were further processed by filtering for fidelity of individual feature detection using the quality control samples. Only features with a CV less than 10% over all quality control sampled were accepted. Features showing high intensity in background samples relative to the quality control samples and features not present in at least 67% of all samples were removed from analysis as per U.S. Food and Drug Administration recommendation. Each feature is uniquely identified with the mass to charge ratio (m/z) and the elution time from the column. Features were then assigned to metabolites identified by searching the Human Metabolome DataBase (HMDB) and using databases developed by the University of Minnesota **(Supplementary Table 3)**.

**Microbiome-Metabolome correlation analysis.** The Spearman's ranked correlation test with false discovery rate (FDR) adjustment were used to test the microbiome-metabolome correlation [McHardy]. The microbiome OTU data and metabolomics data were first combined and filtered to remove low abundance OTUs

and metabolites (appearing in less than 50% of samples). The Spearman's ranked correlation were calculated using *cor.test* function in R v3.4.4. The p-values were then adjusted using *p.adjust* function before filtering for significant correlations.

## Declarations

### Ethics approval and consent to participate

All protocols and procedures were approved by the University of Minnesota Institutional Animal Care and Use Committee, conducted in compliance with the Animal Welfare Act, and animals were housed and cared for according to the standards detailed in the Guide for the Care and Use of Laboratory Animals.

### Consent for publication

Not applicable

### Availability of data and material

All data generated or analysed during this study are included in this published article and its supplementary information files.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work is supported by Norman Wells Memorial Colorectal Cancer Fellowship (C.Y), Healthy Foods Healthy Lives Institute Graduate and Professional Student Research Grant (C.Y), the MnDrive-University of Minnesota Informatics Institute Graduate Fellowship (C.Y), Mezin-Koats Colon Cancer Research Award (S.S).

### Authors' contributions

CY and SS developed the concept, CY carried out data analysis with assistance from SS. CY, MG, and SS wrote the manuscript.

**Acknowledgements**

The authors thank the members of the Subramanian lab and the University of Minnesota Center for Mass Spectrometry and Proteomics (CMSP) for helpful discussions. We thank the Preclinical Research team and Research Animal Resources veterinarians for providing outstanding animal care and for providing tissue via the NHP tissue sharing program in an effort to reduce the overall number of NHPs used in research. We thank Dr. John Garbe of University of Minnesota Research Informatics Service for developing the gopher-pipelines. We also thank the Medical Genome Facility Genome Analysis Core at Mayo Clinic (Rochester, MN) for performing library prep and sequencing and CMSP for performing the untargeted metabolomics. This work was carried out, in part, using computing resources at the Minnesota Supercomputing Institute.

## Figure Legend

**Figure 1:** Microbiota along the non-human primate gastrointestinal tract. Stacked bar plot of bacterial phyla showing the relative abundance of **a.** each of the 6 tissue locations for each sample and **b.** the average relative abundance at each of the 6 tissue locations. **c.** Unweighted UniFrac Principal Coordinates Analysis (PCoA) of the samples.

**Figure 2:** Microbiota alpha diversity along the GI tract. **a.** Phylogenetic diversity, **b.** Shannon Index, and **c.** Chao1 Index.

**Figure 3:** Microbiota-metabolome similarity. Principal components analysis (PCA) of the **a.** tissue metabolome and **b.** microbiota. **c.** Procrustes analysis of the microbiota PCA against the metabolome PCA. Longer line length indicate less within sample similarities.

**Figure 4:** Microbiota-metabolites correlations. Significant Spearman's correlations between the microbiota and 6-Hydroxypentadecanedioic acid, 1-Isothiocyanato-7-(methylthio)heptane, and 3H-1,2-Dithiole-3-thione were visualized using Cytoscape. Red line indicates positive correlations and blue line indicate negative correlations. The thickness of the link indicate significance, where thicker link indicate more significant correlations.

**Supplementary Data**

**Supplementary Table 1:** Unfiltered OTU table used for analysis.

**Supplementary Table 2:** Unfiltered untargeted metabolomics data used for analysis.

**Supplementary Table 3:** Phyla level microbiota-metabolome correlations.

**Supplementary Figure 1:** Stacked bar plot of bacterial phyla showing the relative abundance comparing the Human (PRJNA284355), Baboon and Mouse microbiota samples.

**Supplementary Figure 2:** Principal Coordinates Analysis (PCoA). **a.** Unweighted UniFrac PCoA showing PC1 vs. PC3. Weighted UniFrac PCoA showing **b.** PC1 vs. PC2 and **c.** PC1 vs. PC3.

**Supplementary Figure 3:** Box plot of bacterial taxa with differential abundance. Seventeen bacterial taxa have differential abundance in tissue locations highlighted in the red-dotted box. Statistical significance are indicated by a. $p<0.1$; b. $p<0.05$; c. $p<0.01$; d. $p<0.005$; e. $p<0.001$.

**Supplementary Figure 4:** Tissue specific procrustes analysis. **a.** Jejunum, **b.** Duodenum, **c.** Ileum, **d.** Cecum, **e.** Proximal Colon, and **f.** Distal Colon.

**Supplementary Figure 5:** Global microbiota-metabolome correlation network of significant Spearman's correlations visualized using Cytoscape. Red line indicate positive correlations and blue line indicate negative correlations. The thickness of the link indicate significance, where thicker link indicate more significant correlations.

**Supplementary Figure 6:** Microbiota-metabolome correlation network. Significant Spearman's correlations of the **a.** small and **b.** large intestine, were visualized using

Cytoscape. Red line indicates positive correlations and blue line indicate negative correlations. The thickness of the link indicate significance, where thicker link indicate more significant correlations.

## References

1. Yuan C, Burns MB, Subramanian S, Blekhman R. Interaction between Host MicroRNAs and the Gut Microbiota in Colorectal Cancer. mSystems. 2018;3. doi:10.1128/mSystems.00205-17.

2. Abdollahi-Roodsaz S, Abramson SB, Scher JU. The metabolic role of the gut microbiota in health and rheumatic disease: mechanisms and interventions. Nat Rev Rheumatol. 2016;12:446–55. doi:10.1038/nrrheum.2016.68.

3. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157:121–41. doi:10.1016/j.cell.2014.03.011.

4. Vangay P, Ward T, Gerber JS, Knights D. Antibiotics, pediatric dysbiosis, and disease. Cell Host Microbe. 2015;17:553–64. doi:10.1016/j.chom.2015.04.006.

5. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med. 2015;7:55. doi:10.1186/s13073-015-0177-8.

6. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol. 2014;12:661–72. doi:10.1038/nrmicro3344.

7. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. Science. 2012;336:1262–7. doi:10.1126/science.1223813.

8. Said HM, Mohammed ZM. Intestinal absorption of water-soluble vitamins: an update. Curr Opin Gastroenterol. 2006;22:140–6. doi:10.1097/01.mog.0000203870.22706.52.

9. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027–31. doi:10.1038/nature05414.

10. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, et al. Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci USA. 2007;104:10643–8. doi:10.1073/pnas.0704189104.

11. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med. 2013;19:576–85. doi:10.1038/nm.3145.

12. Scheppach W. Effects of short chain fatty acids on gut morphology and function. Gut. 1994;35 1 Suppl:S35-8.

13. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian

colon. Cell Metab. 2011;13:517–26. doi:10.1016/j.cmet.2011.02.018.

14. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30. doi:10.3322/caac.21442.

15. Lupton JR. Microbial degradation products influence colon cancer risk: the butyrate controversy. J Nutr. 2004;134:479–82. doi:10.1093/jn/134.2.479.

16. O'Keefe SJD, Ou J, Aufreiter S, O'Connor D, Sharma S, Sepulveda J, et al. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. J Nutr. 2009;139:2044–8. doi:10.3945/jn.109.104380.

17. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol. 2007;73:1576–85. doi:10.1128/AEM.01996-06.

18. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, et al. Bacterial biogeography of the human digestive tract. Sci Rep. 2011;1:170. doi:10.1038/srep00170.

19. Gu S, Chen D, Zhang J-N, Lv X, Wang K, Duan L-P, et al. Bacterial community mapping of the mouse gastrointestinal tract. PLoS ONE. 2013;8:e74957. doi:10.1371/journal.pone.0074957.

20. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome. 2013;1:17. doi:10.1186/2049-2618-1-17.

21. Kong F, Hua Y, Zeng B, Ning R, Li Y, Zhao J. Gut microbiota signatures of longevity. Curr Biol. 2016;26:R832–3. doi:10.1016/j.cub.2016.08.015.

22. Gorbach S. Probiotics and gastrointestinal health. The American journal of gastroenterology. 2000;95:S2–4. doi:10.1016/S0002-9270(99)00806-0.

23. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308:1635–8. doi:10.1126/science.1110591.

24. McKenna P, Hoffmann C, Minkah N, Aye PP, Lackner A, Liu Z, et al. The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. PLoS Pathog. 2008;4:e20. doi:10.1371/journal.ppat.0040020.

25. Harrell L, Wang Y, Antonopoulos D, Young V, Lichtenstein L, Huang Y, et al. Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. PLoS ONE. 2012;7:e32545. doi:10.1371/journal.pone.0032545.

26. Stanley D, Hughes RJ, Moore RJ. Microbiota of the chicken gastrointestinal tract: influence on health, productivity and disease. Appl Microbiol Biotechnol. 2014;98:4301–10. doi:10.1007/s00253-014-5646-2.

27. Gong J, Si W, Forster RJ, Huang R, Yu H, Yin Y, et al. 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: from crops to ceca. FEMS Microbiol Ecol. 2007;59:147–57. doi:10.1111/j.1574-6941.2006.00193.x.

28. Suchodolski JS, Camacho J, Steiner JM. Analysis of bacterial diversity in the canine duodenum, jejunum, ileum, and colon by comparative 16S rRNA gene analysis. FEMS Microbiol Ecol. 2008;66:567–78. doi:10.1111/j.1574-6941.2008.00521.x.

29. Mao S, Zhang M, Liu J, Zhu W. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. Sci Rep. 2015;5:16116. doi:10.1038/srep16116.

30. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. Genome Biol. 2015;16:191. doi:10.1186/s13059-015-0759-1.

31. Blekhman R, Perry GH, Shahbaz S, Fiehn O, Clark AG, Gilad Y. Comparative metabolomics in primates reveals the effects of diet and gene regulatory variation on metabolic divergence. Sci Rep. 2014;4:5809. doi:10.1038/srep05809.

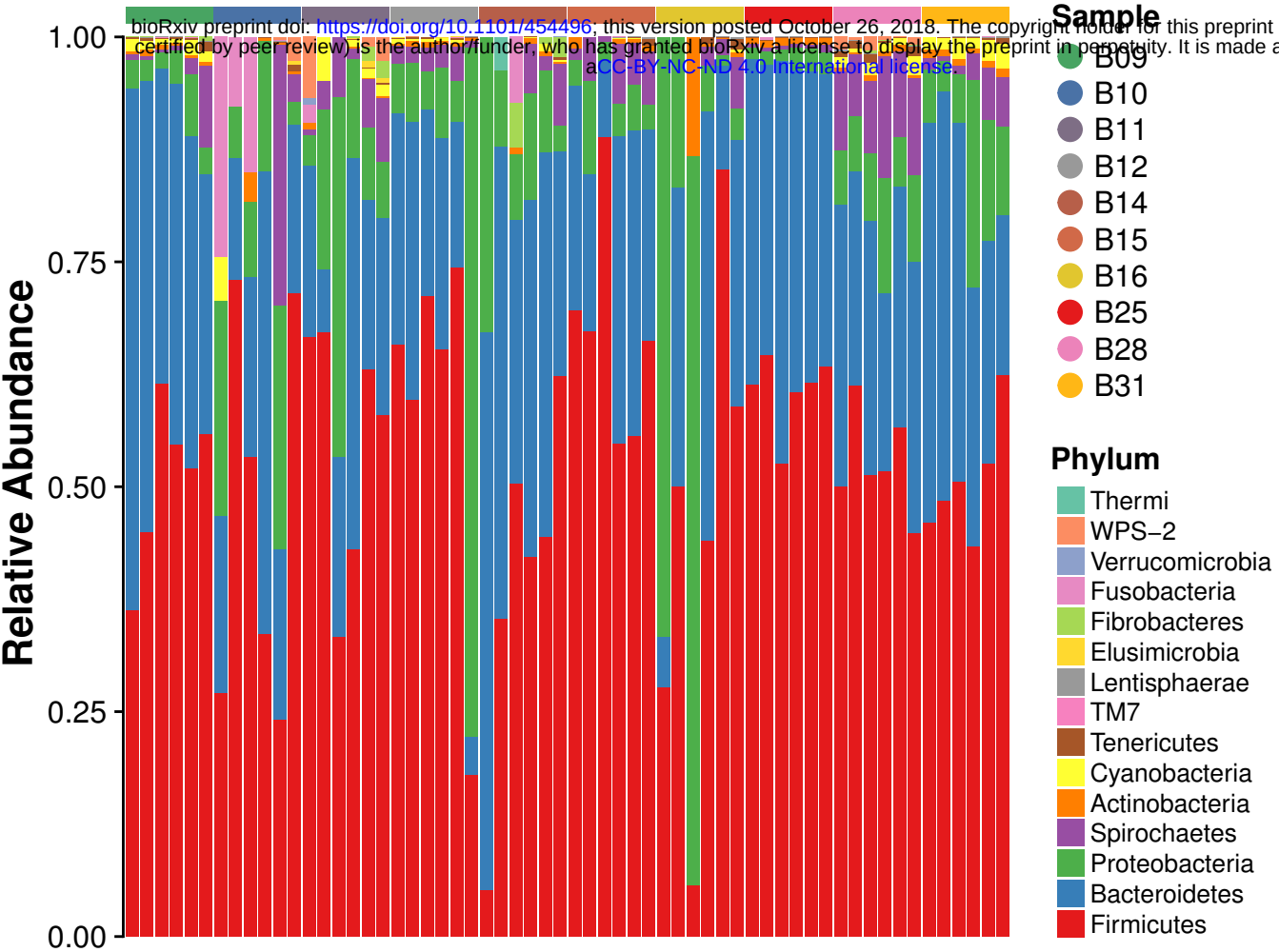
32. Kwak M-K, Kensler TW. Targeting NRF2 signaling for cancer chemoprevention. Toxicol Appl Pharmacol. 2010;244:66–76. doi:10.1016/j.taap.2009.08.028.

33. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012;13:31. doi:10.1186/1471-2105-13-31.

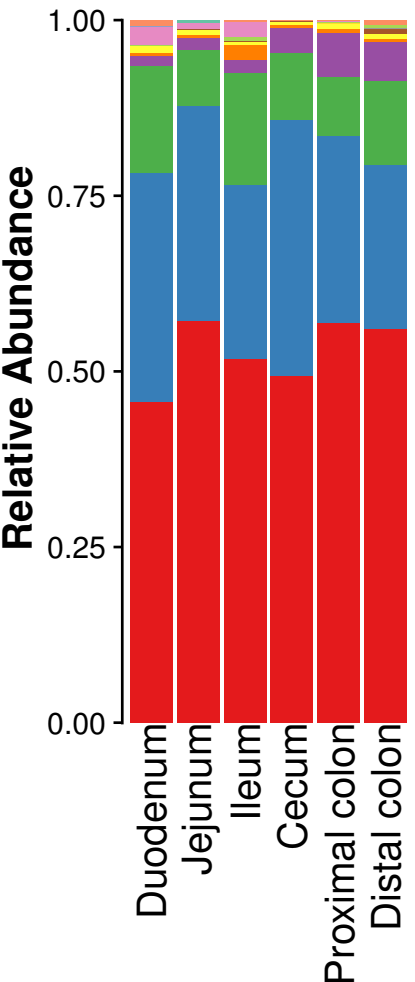
34. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6. doi:10.1038/nmeth.f.303.

35. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72. doi:10.1128/AEM.03006-05.

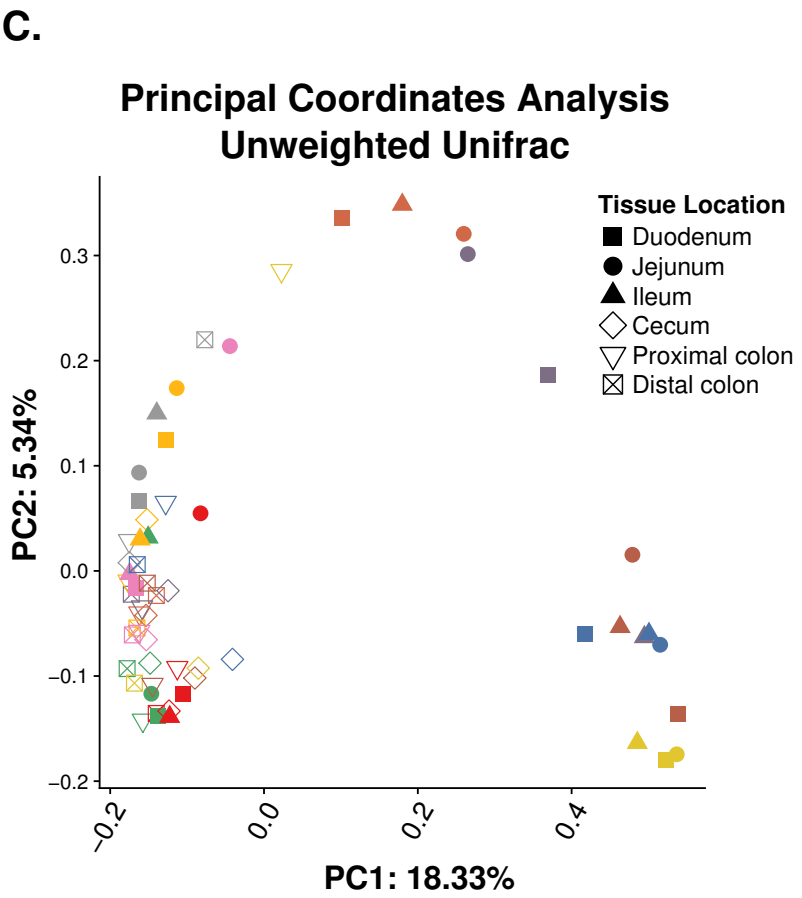
A.

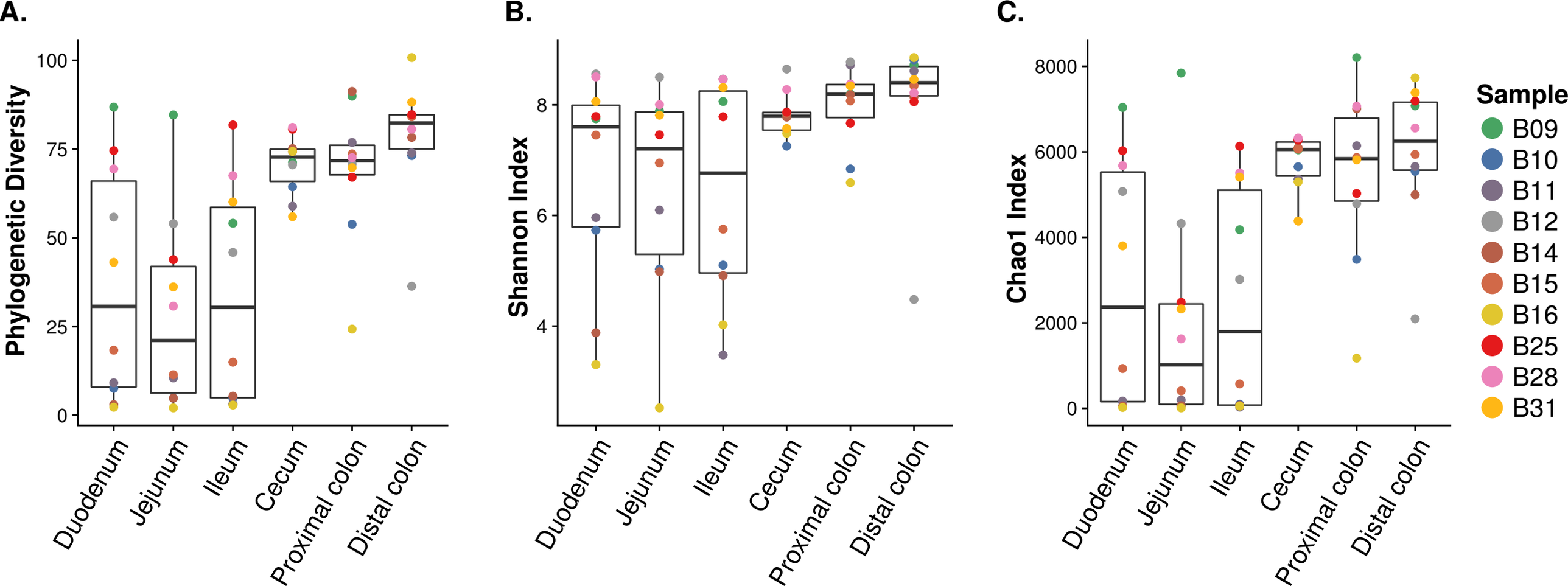


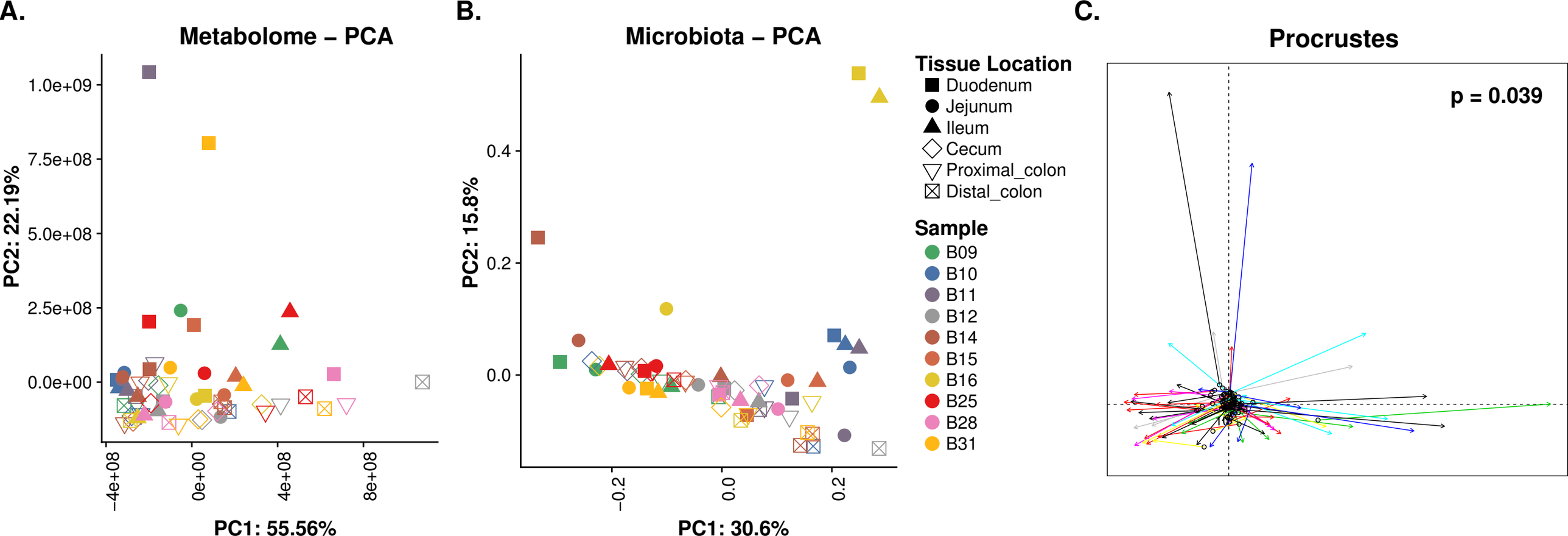

B.

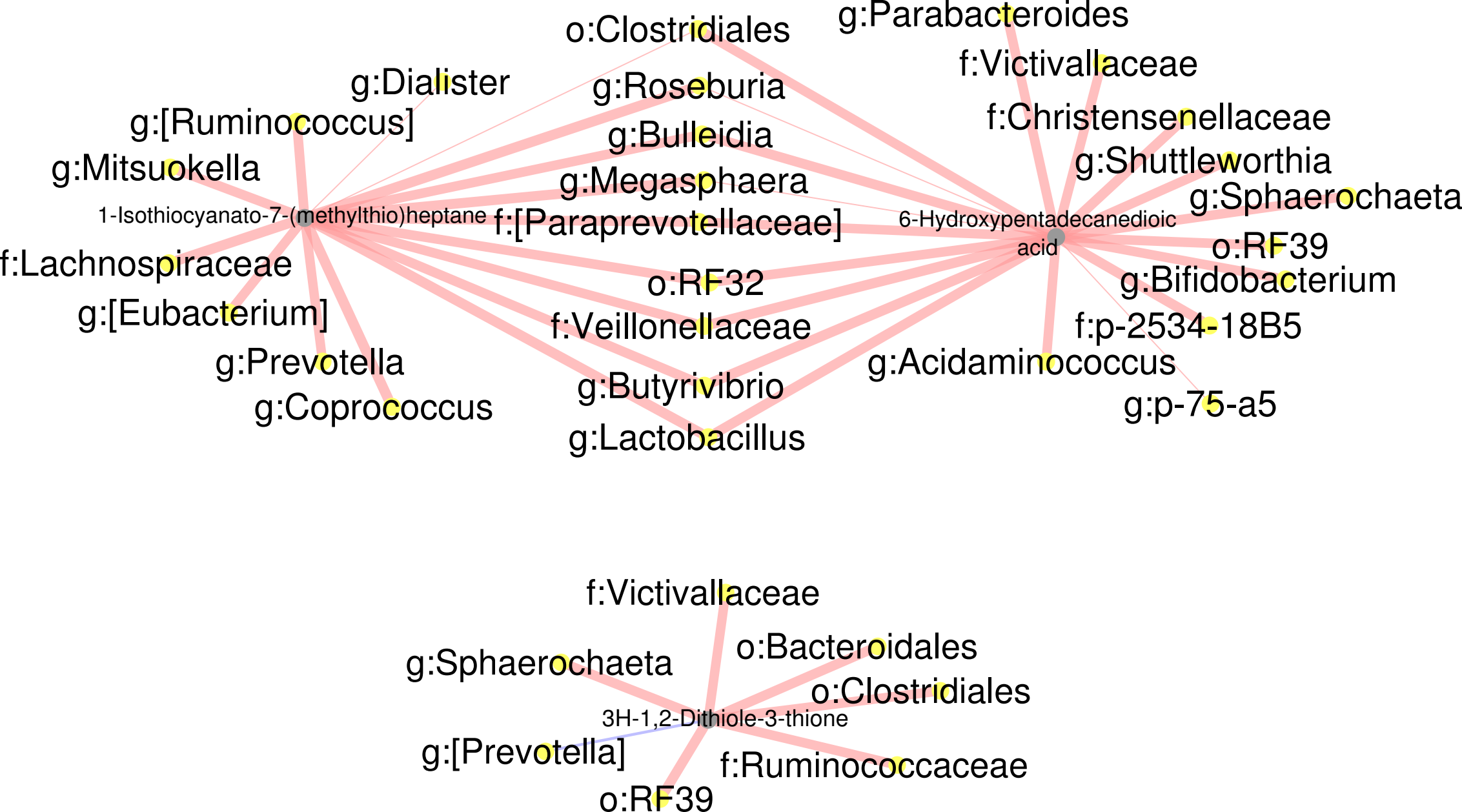



C.




A.
Phylogenetic Diversity
100
75
50
25
0
Duodenum
Jejunum
Ileum
Cecum
Proximal colon
Distal colon
B.
Shannon Index
8
6
4
Duodenum
Jejunum
Ileum
Cecum
Proximal colon
Distal colon
C.
Chao1 Index
8000
6000
4000
2000
0
Duodenum
Jejunum
Ileum
Cecum
Proximal colon
Distal colon
Sample
B09
B10
B11
B12
B14
B15
B16
B25
B28
B31


A.
Metabolome – PCA
1.0e+09
7.5e+08
5.0e+08
2.5e+08
0.0e+00
PC2: 22.19%
−4e+08
0e+00
4e+08
8e+08
PC1: 55.56%
B.
Microbiota – PCA
0.4
0.2
0.0
PC2: 15.8%
−0.2
0.0
0.2
PC1: 30.6%
Tissue Location
Duodenum
Jejunum
Ileum
Cecum
Proximal_colon
Distal_colon
Sample
B09
B10
B11
B12
B14
B15
B16
B25
B28
B31
C.
Procrustes
p = 0.039


o:Clostridiales
g:Parabacteroides
g:Dialister
f:Victivallaceae
g:Roseburia
g:[Ruminococcus]
f:Christensenellaceae
g:Bulleidia
g:Mitsuokella
g:Shuttleworthia
g:Megasphaera
g:Sphaerochaeta
1-Isothiocyanato-7-(methylthio)heptane
f:[Paraprevotellaceae]
6-Hydroxypentadecanedioic acid
o:RF39
f:Lachnospiraceae
g:Bifidobacterium
o:RF32
g:[Eubacterium]
f:Veillonellaceae
f:p-2534-18B5
g:Prevotella
g:Acidaminococcus
g:Butyrivibrio
g:Coprococcus
g:p-75-a5
g:Lactobacillus
f:Victivallaceae
g:Sphaerochaeta
o:Bacteroidales
o:Clostridiales
3H-1,2-Dithiole-3-thione
g:[Prevotella]
f:Ruminococcaceae
o:RF39