NIST Special Publication 260
NIST SP 260-252

Characterization of Human Fecal Material (NIST RM 8048) from Omnivore and Vegetarian Donors

Stephanie L. Servetas
Samuel P. Forry
Monique Hunter
Tracey B. Schock
W. Clay Davis
Amanda L. Bayless
Fabio Casu
Debra L. Ellisor
Aaron A. Urbas
Sandra M. Da Silva
Kirsten Parratt
Kristine Gierz
Hariharan Iyer
Scott A. Jackson
Nancy J. Lin

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-252

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Special Publication 260
NIST SP 260-252

Characterization of Human Fecal Material (RM 8048) from Omnivore and Vegetarian Donors

Stephanie Servetas
Samuel P. Forry
Monique Hunter
Sandra M. Da Silva
Kirsten Parratt
Nancy J. Lin
Biosystem and Biomaterials Division
Material Measurement Laboratory

Tracey B. Schock
W. Clay Davis
Amanda L. Bayless
Fabio Casu
Debra L. Ellisor
Aaron A. Urbas
Chemical Sciences Division
Material Measurement Laboratory

Kristine Gierz
Hariharan Iyer
Statistical Engineering Division
Information Technology Laboratory

Scott A. Jackson*
Amanda L. Bayless*
**Former NIST employee; all work for this publication was done while at NIST*
Material Measurement Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-252

April 2026

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Craig S. Burkhardt, Acting NIST Director and Under Secretary of Commerce for Standards and Technology

NIST SP 260-252
April 2026

Publication History
Approved by the NIST Editorial Review Board on 2026-03-23

How to Cite this NIST Technical Series Publication
Jackson S, Bayless AL, Casu F, Da Silva SM, Davis WC, Ellisor DL, Gierz K, Iyer H, Lin NJ, Parratt KH, Schock TB, Urbas AA, Forry SP, Hunter M, Servetas S. (2026) Characterization of Human Fecal Material (RM 8048) from Omnivore and Vegetarian Donors. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 260-252. https://doi.org/10.6028/NIST.SP.260-252

Author ORCID iDs
Scott A. Jackson: 0000-0003-4891-8323
Stephanie Servetas: 0000-0002-4924-4511
Samuel Forry: 0009-0000-8200-0327
Monique Hunter: 0009-0005-5490-2557
Tracey B. Schock: 0000-0002-1808-7816
W. Clay Davis: 0000-0001-9076-2620
Amanda Bayless: 0000-0002-9073-5582
Fabio Casu: 0000-0003-2378-375X
Debra L. Ellisor 0000-0002-5980-1887
Aaron. A. Urbas: 0000-0003-1535-5376
Sandra M. Da Silva: 0000-0002-7121-7544
Kirsten Parratt: 0000-0002-2640-3276
Nancy J. Lin: 0000-0002-6980-3288
Kristine Gierz: 0000-0002-4919-4841
Hariharan Iyer: 0000-0001-5198-9945

Contact Information
Please address technical questions you may have about these reference materials to srms@nist.gov where they will be assigned to the appropriate Technical Project Leader responsible for their support. For sales and customer service inquiries, please contact srminfo@nist.gov

Abstract

This report describes the development and characterization of RM 8048, a Human Fecal Reference Material, designed to support gut microbiome research. The purpose of this project was to create a well-characterized reference material that can be used to compare results across laboratories, instruments, and methods, ultimately enabling the advancement of our understanding of the complex interactions between the gut microbiome and human health.

The material consists of two sets of pooled, homogenized human stool, with one set from omnivore donors and one set from vegetarian donors. After the material was aliquoted into vials, homogeneity and stability studies were performed. Characterization of the material included next-generation sequencing (NGS)-based metagenomics, liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) and nuclear magnetic resonance (NMR) metabolomics, and flow cytometry. Taxonomic identification revealed DNA from a diverse microbial community, with a range of bacterial taxa present. Metabolomics analysis detected a range of metabolites, including those involved in energy metabolism, amino acid metabolism, and other key biological processes. Flow cytometry evaluated the material based on total cell number per vial per donor pool as well as intensity distributions from fluorescent probes that bind nucleic-acids. Collectively, results from these methods demonstrated that the reference material is suitable for use as a reference standard based on its homogeneity and stability.

Overall, the development and characterization of the Human Fecal Reference Material is a significant step towards standardizing microbiome research. This common, stable, and well-characterized material will enable researchers to evaluate experimental and measurement variability, leading to more accurate and reliable results. Moreover, the use of multiple characterization technologies demonstrates potential use cases across various application spaces and measurement fields, not only to human health, but also to other areas where complex microbiome reference materials are needed.

Keywords

Complex microbial community; Fecal material; Human whole stool reference material; Fluorescence flow cytometry; Gut microbiome; Liquid chromatography; Mass spectrometry; Metabolomics; Metagenomics; Microbiome; Next-generation sequencing; Nuclear magnetic resonance spectroscopy

Table of Contents

List of Tables

List of Figures

Acknowledgments

We would like to express our gratitude to the Institute for the Advancement of Food and Nutrition Sciences (IAFNS, formerly International Life Sciences Institute North America (ILSI NA)) for providing financial support for the development of the RM 8048. Additionally, we extend our thanks to Katrice Lippa, Christina Jones, and Paulina Piotrowski for initiating this effort.

1. Introduction

Over the past two decades, hundreds of clinical studies have demonstrated the critical role that the human microbiome plays in a vast and disparate set of health and disease states including cancer, diabetes, autoimmunity, and mental illness[1]. As such, the human microbiome has emerged as a new target for diagnostics and therapeutics. The human microbiome contains over 10^{14} microbial cells, including bacteria, archaea, fungi, phages, and viruses, that colonize the skin, mouth, genitals, and digestive tract (i.e., the gut)[2]. Most of the human microbiome resides in the gut, thus it is the gut microbiome that has been the focus of many clinical studies.

To identify new biomarkers that may serve as disease indicators and to understand clinically relevant attributes of the human gut microbiome, the field needs validated analytical measurements that accurately describe various properties of these complex microbial ecosystems, both quantitatively and qualitatively. Commonly applied measurements to describe the gut microbiome include 1) next generation sequencing (NGS)-based metagenomics and 2) metabolomics. In either case, no fit-for-purpose reference materials exist that enable researchers to compare results generated across different laboratories, instruments, and methods and to assess the impact of the multitude of methodological variables on the measurement results.

Reference materials comprising human whole stool can provide a common, stable, well-characterized material for use in gut microbiome studies. Stool reference materials are directly applicable to 1) comparison of methods and protocols used in microbiome-based DNA, metabolomics, and flow cytometry measurements, and 2) evaluation of experimental and measurement variability throughout workflows.

Here we describe the development and characterization of a Human Fecal Material Reference Material that is demonstrated to be sufficiently homogeneous, stable, and fit for purpose with respect to the microbial taxa (DNA) and key metabolites.

[1] Afzaal, M., et al. (2022). "Human gut microbiota in health and disease: Unveiling the relationship." Frontiers in Microbiology Volume 13 - 2022.
[2] FAQ: Human Microbiome. Washington (DC): American Society for Microbiology; 2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK562894/ doi: 10.1128/AAMCol.1-2013

2. Material Production

The NIST Human Fecal Material Reference Material is composed of two different parts corresponding to two dietary donor pools, part A: omnivore and part B: vegetarian. In its final form, the reference material comprises homogenized stool from multiple donors resuspended in water to a concentration of 100 mg/mL wet weight. The material production process involved collection of stool donations at BioIVT; cryohomogenization at NIST Charleston; resuspension, bacterial spike-in, and aliquoting into vials at BioIVT; and sample storage at NIST Gaithersburg, as described in detail below.

2.1. Donors, Diet Logs, and Sample Collection

Human fecal material collections were coordinated through BioIVT (Hicksville, NY), a third-party laboratory that specializes in biospecimen collection under informed consent. This work was reviewed and approved by the U. S. National Institute of Standards and Technology Research Protections Office. Donors were recruited by BioIVT to represent two dietary groups, omnivore and vegetarian; all donors were required to complete a food diary corresponding to the week around the donations (Appendix A). A snapshot of the dietary profiles for total fat, protein, carbohydrates, fiber and sugar for each donor group can be found in Figure 2-1 and is based on standard serving sizes unless the journal entry specifically stated an amount. Donors were self-reported healthy with no antibiotic use within three months of donation. All donors were screened for the following infectious agents via blood draw: hepatitis B/C virus, human immunodeficiency virus (types 1 and 2), West Nile virus, syphilis, and Chagas.

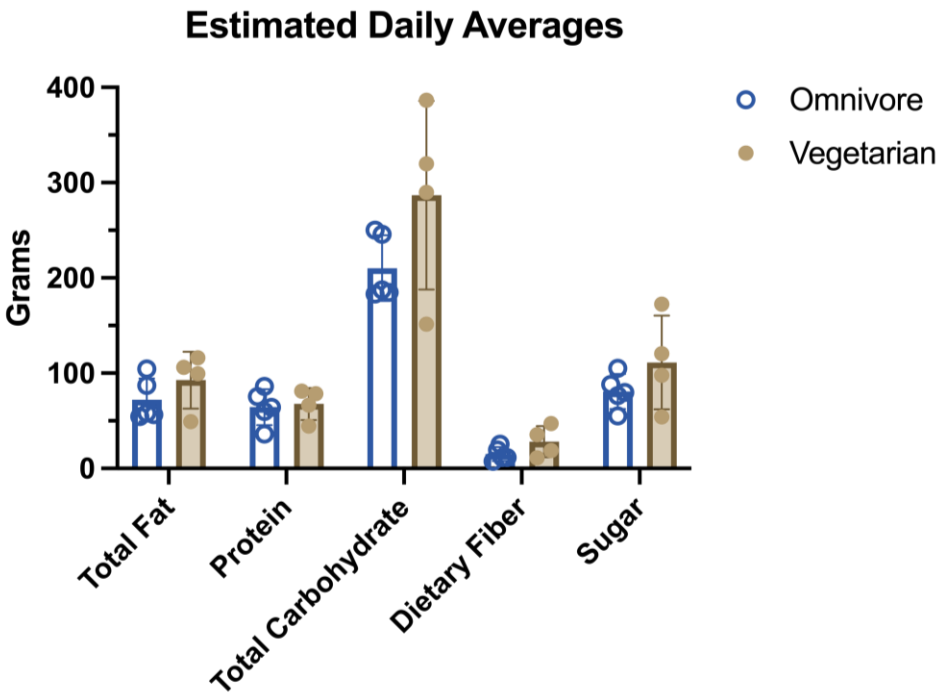


Figure 2-1: Dietary profiles for total fat, protein, carbohydrates, fiber and sugar per donor group. Open symbols/bars indicate omnivores, filled represent vegetarian data. Circles indicate results from individual donors, bars show the mean of donors with error bars representing standard deviation.

Samples were collected on site and immediately stored at -80 ˚ C as whole stool samples in the collection vessels. All donations were collected between December 2022 and January 2023. The omnivore group was composed of 5 donors (n = 3 bowel movements per donor) for a total of 1241 g of stool (approximately 200 g to 350 g collected per donor). The vegetarian group was composed of 4 donors (n = 3 bowel movements per donor) for a total of 1293 g of stool (approximately 200 g to 350 g collected per donor). Material was stored at BioIVT until all donations were received and then shipped to NIST Charleston on dry ice.

2.2. Cryogenic Homogenization

2.2.1. Sample Preparation

Samples were processed in an ISO Class 5 clean room, and personnel wore fitted N95 masks at all times. Large polytetrafluoroethylene (PTFE) disk mills were used for cryohomogenization. The mills can only effectively blend up to 500 g of material, thus the donations for each donor pool were subsampled into three groups (Omnivore 1 to 3 (Table 2-1), and Vegetarian 1 to 3 (Table 2-2). The intention was to ensure that each group comprised material from all donors of the respective diet type. On March 6, 2023, samples were sorted in a pre-cooled cryocart to prevent thawing. For the omnivore donations, sample masses had been written on the

collection containers, so donations were split into groups of similar mass (Table 2-1). Each group was collected in a 2.54 cm x 2.54 cm fluorinated ethylene propylene (FEP) bag, labeled and placed in a liquid nitrogen (LN_2) vapor-phase freezer (≤ −150 °C) overnight to cool.

No masses were provided for the vegetarian donations, so samples were qualitatively split into three groups. Additionally, donor 22-0611 only provided two donations, so donation #2 from this individual was split between groups Vegetarian 2 and Vegetarian 3 to ensure all donors were represented in each group (Table 2-2). Each group was collected in a FEP bag, labeled and placed in a LN_2 vapor-phase freezer overnight to cool. The fourth vegetarian donor's samples were omitted from the material processing.

Nalgene collection bottles (500 mL) were pre-cleaned by rinsing three times with Milli-Q water (minimum resistivity 18.2 MΩ-cm) and one time with ultra-high purity isopropanol and allowed to dry in a high efficiency particulate air (HEPA) hood. Each bottle intended to be shipped to BioIVT was labeled with the RM number, diet type, and bottle ID using temperature-appropriate pre-printed labels. Additional bottles were also prepared and labeled to store any extra material that was not required for RM production at the time of preparation. These bottles were labeled with the RM number, diet type, bottle ID, and "NIST" using an indelible marker. All homogenization equipment and collection bottles were also pre-cooled overnight.

Table 2-1: Omnivore donations from five separate donors and their inclusion in the omnivore homogenization groups.

Group ID	Donation ID	Mass (g)	Total Mass (g)
Omnivore 1	A3	58	411
	B3	53	
	C1	123	
	D2	75	
	E1	102	
Omnivore 2	A1	15	398
	B2	105	
	C3	53	
	D3	72	
	E2	153	
Omnivore 3	A2	111	432
	B1	20	
	C2	188	
	D1	63	
	E3	50	

Table 2-2: Vegetarian donations from three separate donors and their inclusion in the vegetarian homogenization groups.

Group ID	Donation ID	Date	Time	Homogenization Time (s)
Vegetarian 1	RM 10018 #2	11/12/2022	12:05 AM	120 s
	RM 10017 #3	11/13/2022	9:38 AM	
	22-0611 #1	na	na	
Vegetarian 2	RM 10018 #3	11/12/2022	10:00 AM	120 s
	RM 10017 #1	11/11/2022	2:15 PM	
	22-0611 #2 *	na	na	
Vegetarian 3	RM 10018 #1	11/11/2022	9:30 AM	120 s
	RM 10017 #2	11/12/2022	9:40 AM	
	22-0611 #2 *	na	na	

* indicates the sample that was split between Vegetarian groups 2 and 3

Group Omnivore 1 was poured into a large PTFE disk mill, and a PTFE spatula was used to manipulate the material between the ring and puck to promote effective homogenization (Figure 2-2). Any pieces that did not fit into the mill were broken apart using an FEP-covered mallet and added. After placing the lid on the mill, the mill was placed in the disk mill shaker (Siebtechnik TEMA, Cincinnati, OH) and homogenized for 90 s (Figure 2-3). The resultant fresh frozen homogenate (Figure 2-4) was bottled using a PTFE scoop, and collection bottles were filled in order (Table 2-3). Each bottle contained approximately 110 g of material yielding 4 to 5 bottles per group. Masses were measured using a Mettler Toledo balance with readability to 0.1 g (Mettler Toledo, Columbus, OH). Once all the homogenate from Omnivore 1 was bottled, Omnivore 2 and 3 were cryomilled in the same large PTFE disk mill in succession, with bottling occurring after each milling.

Figure 2-2: Omnivore 1 being loaded into the large PTFE disk.

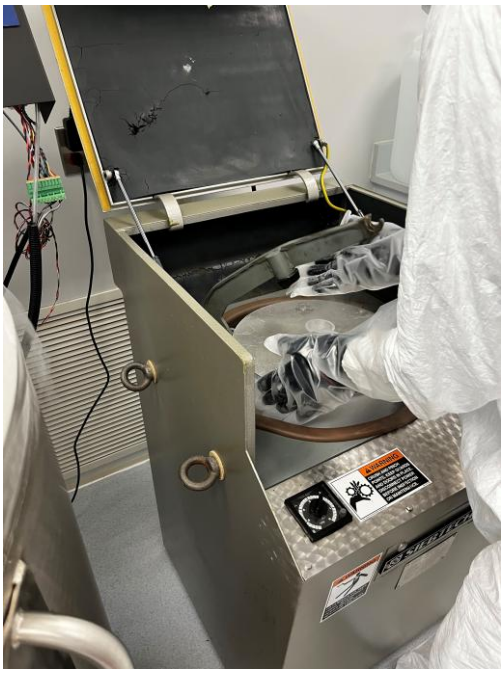


Figure 2-3: Large disk mill being placed in the Siebtechnik disk mill.

Figure 2-4: Cryohomogenized material ready for bottling.

Table 2-3: Distribution of omnivore (left) and vegetarian (right) groups into associated storage bottles.

Group ID	Bottle ID	Mass Contributed (g)
Omnivore 1	Jar 1	107.7
	Jar 2	113.3
	Jar 3	112.8
	Jar 4	66.4
Omnivore 2	Jar 4	41.4
	Jar 5	111.5
	NIST Jar 1	211.4
	NIST Jar 2	27.2
Omnivore 3	NIST Jar 2	203.6
	NIST Jar 3	231.5

Group ID	Bottle ID	Mass Contributed (g)
Vegetarian 1	Jar 1	111.5
	Jar 2	110.6
	Jar 3	52.0
Vegetarian 3	Jar 3	61.1
	Jar 4	115.4
	Jar 5	111.7
	NIST Jar 1	48.7
Vegetarian 2	NIST Jar 1	106.1
	NIST Jar 2	212.6

A similar process was undertaken with the vegetarian donations using a clean large PTFE disk mill to avoid cross-contamination. Vegetarian 1 was homogenized first, followed by Vegetarian 3 and Vegetarian 2. Because a large portion of the Vegetarian 3 group comprised a single donation, the intention was to ensure that this donation was split between the production batch and the batch for storage at NIST, promoting more homogeneity across the batches.

The final masses for each collection bottle are shown in Table 2-4. Following the bottling procedure, 10 bottles were stored at -80 °C in Freezer 09 in the NIST Biorepository until shipped overnight to BioIVT for final production.

Table 2-4: Final masses of each bottle of stool material.

Diet Type	Bottle ID	Total Mass (g)
Omnivore	Jar 1	107.7
	Jar 2	113.3
	Jar 3	112.8
	Jar 4	107.8
	Jar 5	111.5
Vegetarian	Jar 1	111.5
	Jar 2	110.6
	Jar 3	113.1
	Jar 4	115.4
	Jar 5	111.7

2.3. Bacterial, Whole Cell, Internal Control Strains

Three bacterial strains, candidates for the NIST Microbial Strain Collection, were selected for inclusion in the human gut microbiome reference material as internal controls, or spike-ins: *Deinococcus radiodurans* (NIST 0032), *Brenneria nigrifluens* (NIST 0080), and *Deltia acidovorans* (NIST 0136). These three strains were selected as they are unlikely to occur in a fecal sample naturally and can be selectively cultured and quantified via CFU after combination with the fecal material. The target final concentration for each spike-in strain in the final stool material was 1 % of the microbial cell concentration within the stool. The microbial cell concentration in the stool was estimated via flow cytometry to be on the order of ≈ 1 x 10^{11} microbes/gram stool, or ≈ 5 x 10^{13} per donor pool for the reference material, resulting in a target value of ≈ 5 x 10^{11} cells per spike-in strain per donor pool material. To maintain consistency between the omnivore and vegetarian materials, a combined spike-in sample was generated comprising ≈ 1 x 10^{12} cells per strain.

From a cryopreserved glycerol stock, strains were streak plated on tryptic soy agar (TSA) at 30 ° C for 48 h. After 48 h, a single colony was resuspended in 500 µL of tryptic soy broth (TSB), and a swab was used to inoculate ten TSA plates for each organism to generate a bacterial lawn. Plates were incubated at 30 ° C for 72 h. After growth, plates were visually inspected for potential contamination, then all ten plates per strain were swabbed, combined, and resuspended in 10 mL of 0.1 x phosphate buffered saline (PBS), to make one 10 mL tube of cell suspension per strain. Each suspension was mixed vigorously, by both vortexing and pipetting, to help disperse any aggregates. Samples were counted in duplicate using the Multisizer 4 Particle Counter (Beckman Coulter) with a 20 µm aperture tube to estimate the total number of cells. The combined spike-in sample was prepared in a single tube by adding ≈ 1 x 10^{12} cells per strain, with the volume added per strain based on cell concentration measured by the Multisizer. The combined sample was then pelleted, and the pellet was resuspended in 10 mL of 0.1 x PBS and split into two 5 mL aliquots, one for the omnivore cohort and one for vegetarian. These aliquots were pelleted, and the supernatant was removed for storage at 4 ° C before shipping (on ice) for addition to the final material during bottling.

2.4. Slurry Production, Aliquoting and Storage

Cryohomogenized materials (prepared by NIST Charleston) were returned to BioIVT on dry ice the week of March 15, 2023 and stored at -80 ° C until processing. The cryohomogenized material was prepared in the following manner: first, all bottles for a donor pool were combined with sterile water in a 5 L bottle to a concentration of 100 mg/mL (cryohomogenized stool/water). The bottle was placed on a cooling plate set to 12 ° C and packed in ice packs; the set-up was all contained within a styrofoam cooler. A Pro Scientific Pro250 Homogenizer was affixed to the top of the bottle to mix a homogenous slurry with the cryohomogenized powder and water. The pellet containing the whole cell bacterial control strains (described in section 2.3) was resuspended in 5 mL sterile water and mixed vigorously to disperse clumps then added to the fecal slurry. With constant mixing, a continuous pipettor was used to distribute 1 mL volumes into each sample vial. Vials were racked in boxes of 100, and each box was labeled in sequential order. Once a box (or set of 100 vials) was completed, the box was stored at 4 ° C. A

total of 5,000 vials were produced per donor pool. Once all vials from a single donor pool were filled, the resulting 50 boxes were transferred from 4 ˚ C to -80 ˚ C for storage until shipping on dry ice. Omnivore material was bottled on April 21, 2024, and vegetarian material was bottled the week of April 24, 2024. All materials were shipped on dry ice to NIST Gaithersburg in early May 2024 and stored at -80 ˚ C. A subset of samples from both donor pools were compiled at NIST Gaithersburg and sent to NIST Charleston for metabolomic analyses; materials were shipped overnight on dry ice.

2.5. Moisture Determination

2.5.1. Methods

Five vials from both the omnivore and vegetarian donor pools were analyzed for moisture content. All frozen samples were weighed on a five-place analytical balance (Mettler Toledo, Columbus, OH, accurate to four places) and returned to the freezer. In preparation for lyophilization, the sample vials were removed from the freezer, the labelled caps were quickly removed, and the frozen samples were placed into a pre-chilled glass freeze dryer flask. The flask was immediately connected to a lyophilizer, and the vacuum was slowly pulled. The mass of the samples was checked every 24 h until the change was negligible in relationship to the accuracy of the balance (< 0.0001 g). After obtaining the final sample masses, the vials were cleaned, dried under nitrogen, and weighed to obtain empty vial masses for the final calculations. The expanded uncertainty was calculated, and the statistical information along with the R script used to perform the computation is attached as (Appendix B).

2.5.2. Results

The ten stool samples were dried for 72 h, at which time the masses did not change by more than 0.0001 g from the previous mass check that was performed 24 h prior. The individual masses for each vial along with means and standard deviations can be found in Table 2-5. The water fraction was converted to a percentage and is represented as mean values ± expanded uncertainty (U). The percent water in RM 8048 vegetarian donor pool was 97.397 % ± 0.420 % (U) and the percent water in the omnivore donor pool was 97.329 % ± 0.426 % (U) (Appendix B).

Table 2-5: Stool masses with mean and standard deviation (SD) for moisture content determination.

Vegetarian

Vial	Starting Mass (g)	Mass 24 h (g)	Mass 48 h (g)	Mass 72 h (g)	Empty Vial (g)	Total water loss (g)	Dry stool (g)	Wet Stool (g)	Water (%)
1	2.57197	1.59499	1.59485	1.59476	1.56970	0.97721	0.02506	1.00227	97.500
2	2.56536	1.59014	1.58996	1.58992	1.56456	0.97544	0.02536	1.00080	97.466
3	2.57435	1.59462	1.59439	1.59431	1.56810	0.98004	0.02621	1.00625	97.395
4	2.56362	1.59130	1.59114	1.59107	1.56446	0.97255	0.02661	0.99916	97.337
5	2.57893	1.59349	1.59320	1.59316	1.56599	0.98577	0.02717	1.01294	97.318
					Mean	**0.97820**	**0.02608**	**1.00428**	**97.403**
					SD				**0.079**

Omnivore

Vial									
6	2.56359	1.58602	1.58567	1.58563	1.55789	0.97796	0.02774	1.00570	97.242
7	2.57437	1.59119	1.59096	1.59091	1.56373	0.98346	0.02718	1.01064	97.311
8	2.57308	1.59635	1.59607	1.59599	1.56884	0.97709	0.02715	1.00424	97.296
9	2.56608	1.58292	1.58277	1.58269	1.55731	0.98339	0.02538	1.00877	97.484
10	2.57411	1.59697	1.59676	1.59672	1.5698	0.97739	0.02692	1.00431	97.320
					Mean	**0.97986**	**0.02687**	**1.00673**	**97.330**
					SD				**0.091**

3. Assessments of Homogeneity and Stability

The homogeneity and stability of RM 8048 for each measurement was determined via statistical analysis by quantifying the level of discrepancy between any pair of samples (depending on the type of measurement, samples may be vials or replicate aliquots taken from a vial). This metric was defined as the coefficient of disagreement (COD). The procedure for calculating a COD was defined as follows:

1) For each measurement, a method for finding the discrepancy in an attribute of interest between two samples is defined. For example, for metagenomics we may wish to characterize how different the relative abundances are from one another, while for metabolomics we may wish to find discrepancies of presence/absence of a peak. This measure of discrepancy is described for each measurement type in the section describing the measurement.
2) Suppose we have a total of K corresponding pairs of attributes between the two samples. Each pair of "matching" attributes has a corresponding measure of discrepancy that can be found between the two samples. The 99^{th} percentile of this collection of measures of discrepancy is collected and used as an overall measurement of disagreement between the two samples. The 99^{th} percentile value rather than the maximum value is selected to account for occasional outliers.
3) Thus, the coefficient of disagreement (COD) is defined as follows:

   COD = 99^{th} percentile of the measure of discrepancy values from all pairs of corresponding attributes in the two samples.

The COD value is a measure of disagreement between specific attributes from a pair of samples. As an example, if the COD value for a given pair of samples is 0.09 it should be interpreted as follows: The attributes from two samples being considered satisfy the property that 99 % of the measured attribute values are within 9 % of each other.

We can use quantile confidence intervals (or, alternatively, an appropriate bootstrapping procedure) to find an approximate upper bound for the COD values. Since these are pairwise comparisons, we clearly do not have independent samples and cannot find an exact confidence limit without deriving the dependence structure. Therefore, the upper limits from the quantile confidence intervals can be interpreted as approximate limits only.

This approach can be tailored to a number of measurements, with the measure of discrepancy adapted to appropriately address the level of difference in an attribute of choice. Care should be taken to ensure that the chosen measure of discrepancy has interpretable value.

Data analysis on RM 8048 included homogeneity, stability, and non-certified value assignment using multiple methods; however, recognizing the limitations of current methods, not all data was used the same way. Each method will be described in detail in the following sections. Table 3-1 provides an overview of the type of analyses you can expect in each section. Briefly, homogeneity was determined evaluated using 16S amplicon sequencing, LC-MS/MS, NMR 700 MHz, and flow cytometry data; stability was evaluated using 16S amplicon sequencing, NMR

700 MHz, and flow cytometry data; and non-certified values were assigned using evaluated using 16S amplicon sequencing, LC-MS/MS, NMR 700 MHz, and NMR 600 MHz data.

Table 3-1: Method and associated data analyses during the characterization of RM 8048.

Method	Homogeneity	Stability	Non-certified values
Metagenomic (16S amplicon Sequencing)	X	X	X
LC-MS/MS	X		X
NMR 700MHz	X	X	X
NMR 600 MHz			X
Flow Cytometry	X	X	

4. Metagenomics

4.1. Methods

4.1.1. Material and Sample Selection

Three replicate vials from each of 10 different boxes were selected for homogeneity and stability testing. For homogeneity testing, one vial from each box was analyzed; for stability, two vials from different boxes were tested at each subsequent time point except for the 12-week time point when six vials were tested (Table 4-1). Molecular grade water was included as a negative control starting at the extraction step and carried through the entire workflow.

Table 4-1: Biological replicates and time points (in weeks) for metagenomics.

vial #	T0	T1 (+1w)	T2 (+2w)	T3 (+4w)	T4 (+8w)	T5[a] (+12w)	T6 (+16w)	T7 (+20w)	T8 (+24w)
1	Box 1	Box 1	Box 20	Box 6	Box 15	Box 1	Box 20	Box 15	Box 25
2	Box 6	Box 25	Box 50	Box 30	Box 44	Box 6	Box 50	Box 44	Box 36
3	Box 11					Box 11			
4	Box 15					Box 25			
5	Box 20					Box 30			
6	Box 25					Box 36			
7	Box 30								
8	Box 36								
9	Box 44								
10	Box 50								

[a] For the omnivore donor pool, six vials were sampled at T5 (Box 1, 6, 11, 25, 30, 36); for the vegetarian donor pool only two vials were sampled: one from box 11 and one from box 36.

4.1.2. DNA Extraction

Vials were thawed for approximately 5 min at room temp with vortexing approximately every minute prior to extraction. For each vial, four replicate DNA extractions (250 µL each) were carried out using the Zymo Research Quick-DNA Fecal/Soil Microbe Miniprep Extraction Kit [cat #D6010] per protocol with the following modification - mechanical disruption was carried out using the Vortex Genie Microtube adapter (Scientific Industries, Inc. Cat. No. S5001-7) with the vortex genie set to max r/min for 40 min (full protocol Appendix C). DNA was eluted in 100 µL of DNA Elution Buffer and stored at 4 ˚ C. Following extraction, DNA was quantified by fluorescence on a DeNovix DS-11 using the DeNovix High Sensitivity dsDNA Assay.

4.1.3. Amplicon Sequencing Library Preparation

16S rRNA library preps were carried out as followed: 1 µL of genomic DNA was amplified in a 25 µL reaction comprised of 2x Kapa HiFi Master Mix (Roche Cat # KK2601), and 1 µL of 10 µmol/L V4 515F/806R pre-mixed primers with Illumina adapters (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA -3', 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT -3'), under the following conditions: 95.0 ˚ C (3 min), 18 cycles of amplification 98.0 ˚ C (30 s), 55 ˚ C (15 s), 72 ˚ C (20 s), and a final extraction at 72 ˚ C for 5 min (full protocol Appendix C). 16S amplification products were purified with SPRIselect Beads (Beckman Coulter Cat # B23317) in a 0.8:1 (bead to product) ratio. The cleaned 16S products were indexed using IDT® for Illumina® DNA/RNA UD Indexes with 8 amplification cycles with the same thermocycling conditions used for the 16S amplification. The indexed 16S amplicons were purified with SPRIselect Beads in a 1:1 (bead to product) ratio. Purified indexed amplicons were pooled in approximately equal quantities by mass (DeNovix). The concentration of the final pool was measured by fluorescence, and average amplicon size was verified on the TapeStation 4200 (Agilent High Sensitivity D1000), and then the concentration was adjusted to a 4 nmol/L pool for sequencing preparation following the MiSeq Systems Denature and Dilute DNA Libraries Guide, Protocol A[3]. Paired-end sequencing was carried out on the Illumina MiSeq with the 600-cycle MiSeq Reagent Kit v3 (MS-102-3003). Raw data can be found in the metagenomics folder located here: doi:10.18434/mds2-3633.

4.1.4. Read Processing and Taxonomic Assignment

Read processing and taxonomic assignment were carried out in R (4.1.0) using CutAdapt[4] and DADA2 (1.20.0)[5] with the silva database (v138)[6], respectively. The complete R code is included as an R markdown in the metagenomics folder located here: doi:10.18434/mds2-3633.

[3] Document # 15039740 v10, February 2019. https://support.illumina.com/sequencing/sequencing_instruments/miseq/documentation.html
[4] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011)
[5] B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583 (2016)
[6] C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).

4.1.5. Statistical Analysis

Levels of homogeneity and stability were evaluated using a variety of methods based on the amplicon sequence variants (ASVs), or at the sequence level rather than based on a taxonomic classification. First, a global exploratory approach was taken to view all sequences collectively using Multi-Dimensional Scaling (MDS) based on Bray-Curtis and Hellinger distances. Characterization of the homogeneity of individual sequences was then carried out using a Bayesian model (described in Appendix I), to assess the coefficient of variation and the credible intervals for the top 300 most abundant sequences. The credible interval is a range of values within which there is a 95 % probability that the true value lies. For subsequent analyses, omnivore and vegetarian sequences were analyzed separately, and only the sequences with relative abundance greater than 0.0008 for each group were included. As the relative abundance of the DNA sequence decreases, so does the reliability of the measurement, and the threshold for relative abundance was chosen to include information from as many sequences as was reasonable based on the spread of biological and machine variability and the coefficient of variation found from the Bayesian modeling approach. Finally, for both homogeneity and stability, pairwise comparisons of all samples were conducted using a coefficient of disagreement defined in Section 3, with adjustments to make it appropriate for the compositional nature of metagenomic data. A detailed description and methods for the statistical assessment of homogeneity and stability on the metagenomics data set can be found in Appendix I, Section I.3.3.

4.2. Results

4.2.1. Library Preparation and Sequencing

The average number of reads generated per sample for the homogeneity samples was around 1.4×10^5 reads per sample (range 2.4×10^3 to 3.5×10^5). Data from one sample (V11) were removed as there were no usable reads coming off the sequencer (fastq file from MiSeq was empty). Negative controls were included with both vegetarian and omnivore DNA extractions and generated approximately 100 reads total. After quality filtering and processing through Cutadapt and DADA2 the average number of reads remaining in the homogeneity dataset for taxonomic assignment was 1.2×10^5 reads per sample (range 1.8×10^3 to 3.2×10^5).

The DADA2 analysis resulted in a total of 1067 ASVs. A total of 600 unique ASVs were found across the omnivore samples, ranging from 201 to 392 per sample. A total of 478 unique ASVs were found across the vegetarian samples, ranging from 198 to 349.

4.2.2. Taxonomic Identification

Amplicon Sequence Variants (ASV)s were assigned to a taxonomic designation in DADA2 using the Silva database (v138); data shown herein represent genus-level identification. A subset of the genera identified are shown in Figure 4-1(Homogeneity data) and Figure 4-1(Stability data),

and a complete list of genera identified is given in Table D.1 (Homogeneity data) and Table D.2 (Stability data).

For the homogeneity data, a total of 151 genera were identified in the omnivore material and 145 genera were identified in the vegetarian material. Approximately 3 % of reads in both materials could not be assigned to a genus. When both data sets are combined, a total of 175 genera were identified; 30 genera, accounting for approximately 4 % of a sample on average, were unique to omnivore samples (present at < 0.001 % in vegetarians), and 24 genera, accounting for approximately 4 % of a sample on average, were unique to vegetarian samples (present at < 0.001 % in omnivores). For the stability data, a total of 189 genera were reported in the omnivore material and 168 in the vegetarian, with a total of 209 when the materials were combined. Similar to the homogeneity data, about 3 % of the reads in both materials were grouped into the NA (or unclassified) designation. Thirty-seven (37) genera, accounting for approximately 4 % of a sample on average, were unique to omnivore samples (present at < 0.001 % in vegetarians), while 19 genera, accounting for approximately 1 % of a sample on average, were unique to vegetarian samples (present at < 0.001 % in omnivores). Discrepancy between total number of genera identified in the homogeneity and stability datasets was due to genera reported at very low abundance (around 0.001 %); notably these did not affect the conclusions regarding the stability of the material. ASVs corresponding to the bacterial spike-ins (*Deinococcus, Brenneria, and Delftia)* were observed in all samples. Additional characterization and evaluation of these spike-ins is ongoing.

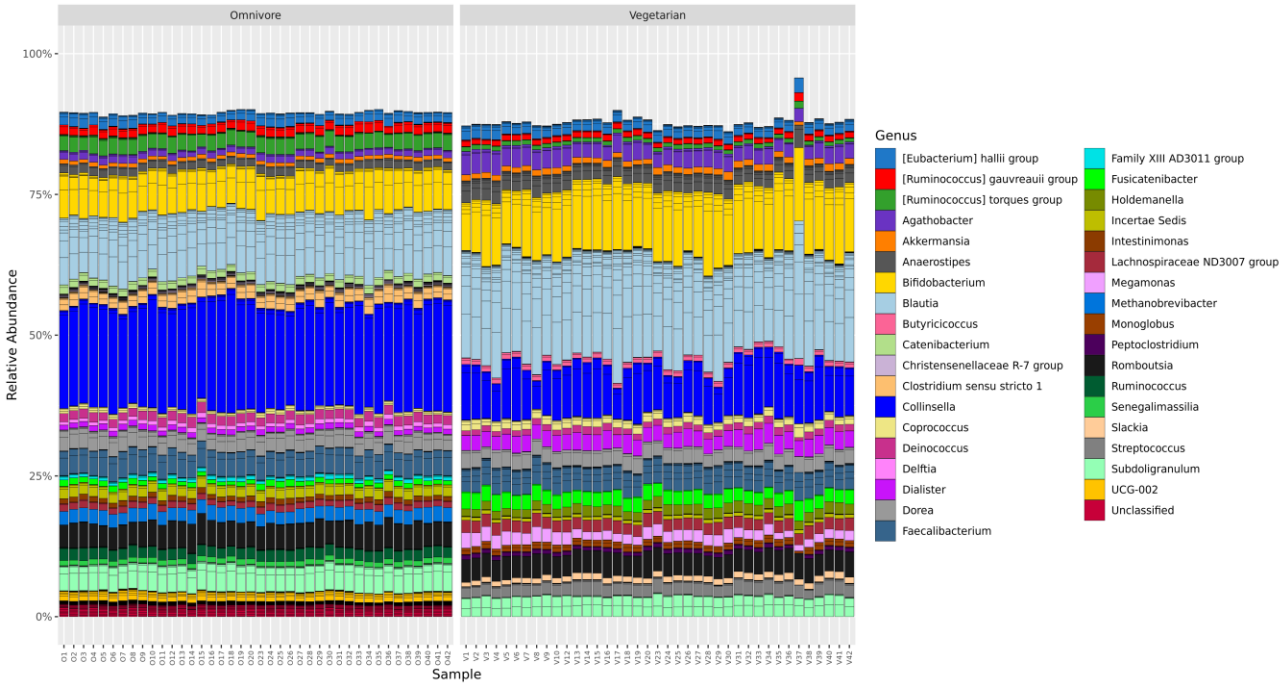


Figure 4-1: Subset of microbial genera identified in omnivore and vegetarian homogeneity data. Stacked bar charts show relative abundance for genera present at an average relative abundance of ≥ 0.006 (0.6 %) in either omnivore or vegetarian donor pools. *X*-axis names correspond to the "Sample Names" column in, Table D.1.

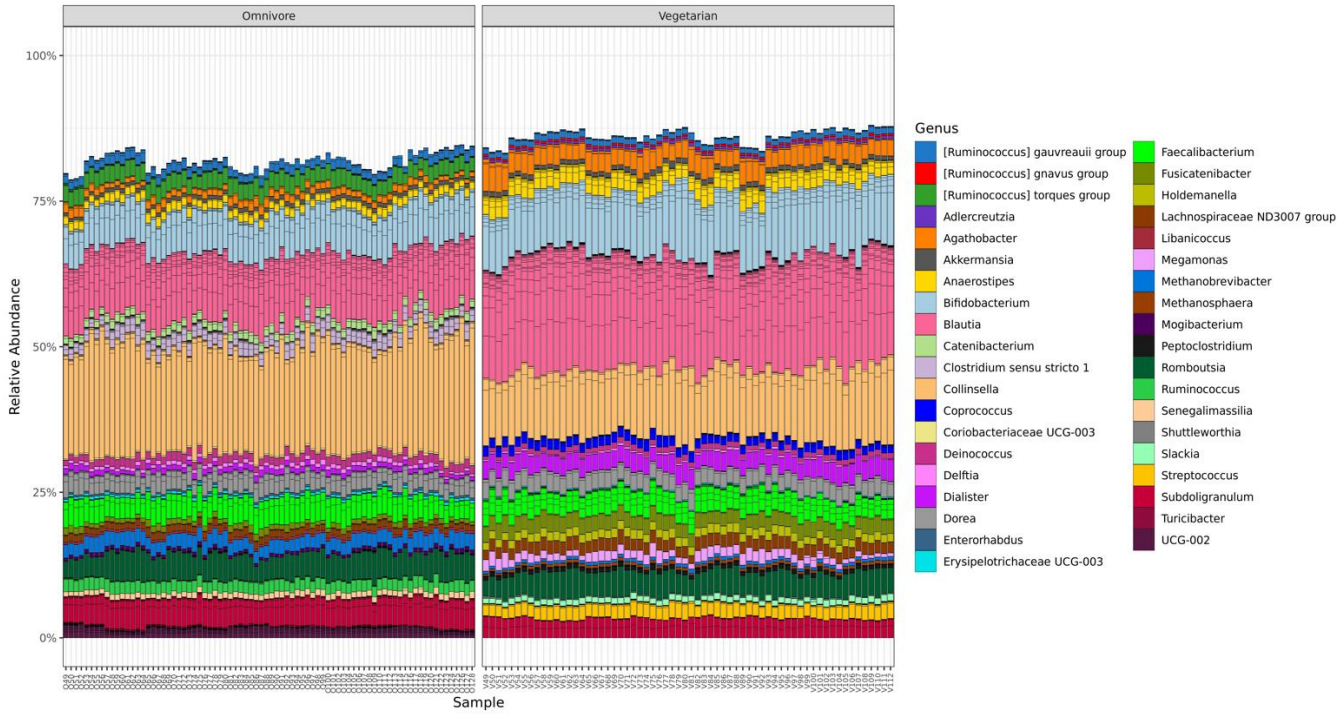


Figure 4-2: Subset of microbial genera identified in omnivore and vegetarian stability data.
Stacked bar charts show relative abundance for genera present at an average relative abundance of ≥ 0.0056 (0.56 %) in either omnivore or vegetarian donor pools. *X*-axis names correspond to the "Genus" column in Table D.2.

4.2.3. Homogeneity

There are a number of methods that are applicable as a summary metric of homogeneity for the metagenomic data, including the Bray-Curtis, Hellinger, or other applicable measures of distance between samples. These distances can then be assessed using clustering methods. Figure 4-3 shows the Multi-Dimensional Scaling (MDS) plot based on Bray-Curtis (left) and Hellinger (right) distances between each of the homogeneity samples. Specifically, for this data one vial was taken from one of ten boxes as described in Table 4-1. Each vial was separated into four aliquots, and each of these aliquots can be considered a sample. In this way, we have four samples from box 1, four samples from box 6, and so on. We can see that the omnivore and vegetarian samples are clustered together with more variability spread along the horizontal axis (between donor pools) than on the vertical axis (within a donor pool).

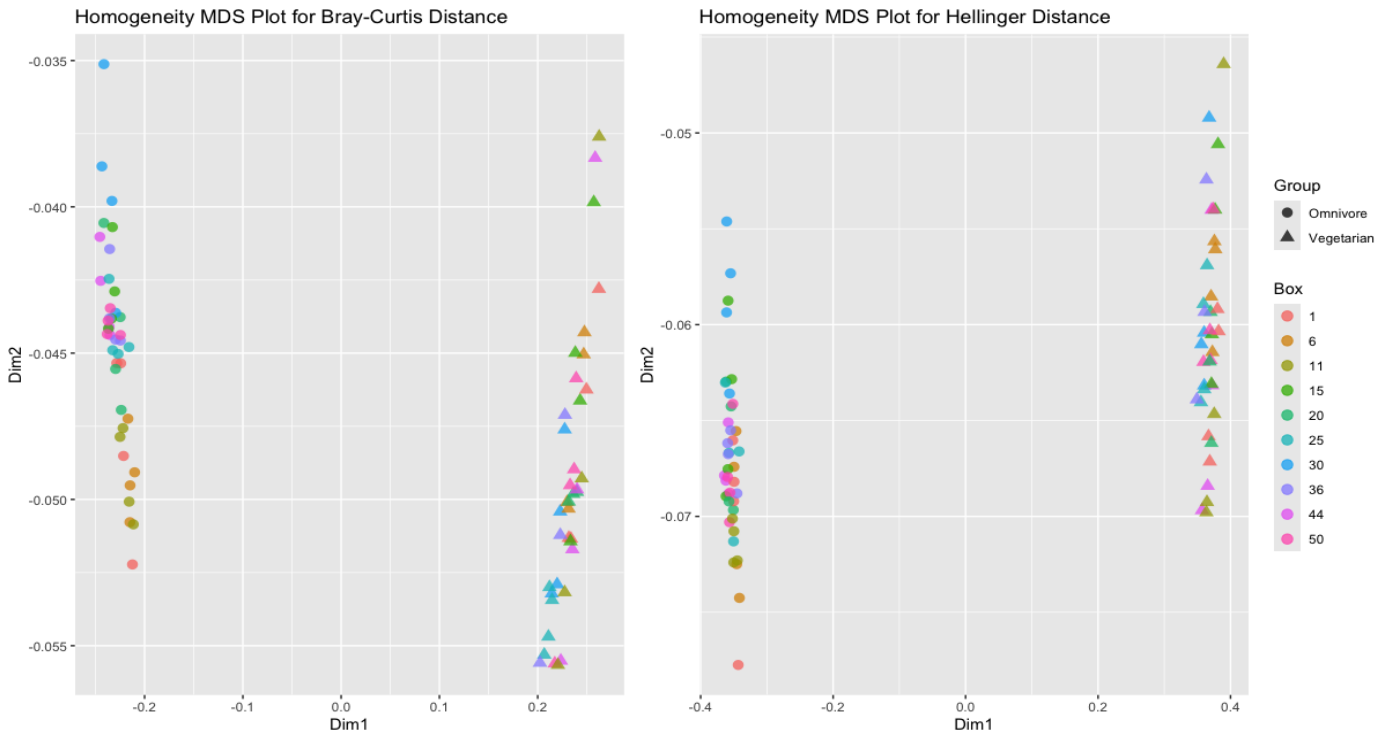


Figure 4-3: MDS plot based on Bray-Curtis (left) and Hellinger (right) distances for homogeneity of metagenomic data.
The data for the homogeneity analysis are made up of measurements taken from ten omnivore and ten vegetarian boxes. For each box, one vial was taken and separated into four aliquots as replicate measurements. Two of the vegetarian samples (box 20, replicate 3 and box 36, replicate 3) had experimental errors resulting in a lack of data (< 2000 reads), so the samples were not considered. As a result, there are 78 (40 omnivore and 38 vegetarian) rather than 80 total samples.

While summary metrics provide a useful overview, for the purposes of assessing homogeneity we chose to take a more granular approach and evaluated the samples by comparing individual ASVs. As a first step we examined the coefficient of variation (CV) estimated from a Bayesian modeling approach for each of the top 300 ASVs across the 40 measurements (four replicates each from ten samples) for omnivores (Figure 4-4A) and 38 measurements vegetarians (four replicates each from ten samples, minus 2 replicates that were excluded due to low read counts after sequencing) (Figure 4-4B). For the 300 most abundant sequences for omnivores, relative abundances across samples were ≥ 0.000044; for vegetarians, the 300 most abundant sequences had relative abundances across samples of ≥ 0.000024. Unsurprisingly, as the relative abundance decreases, the CV increases. For the 50 most abundant taxa, we observed that the average CV from the model was 5.7 % ± 4.3 % for the omnivore donor pool and 7.8 % ± 4.5 % for the vegetarian donor pool. However, one limitation of considering the CV is that it is not possible to distinguish technical variability from biological or sample to sample variability. Therefore, we sought a method that would enable us to tease apart technical from biological variability.

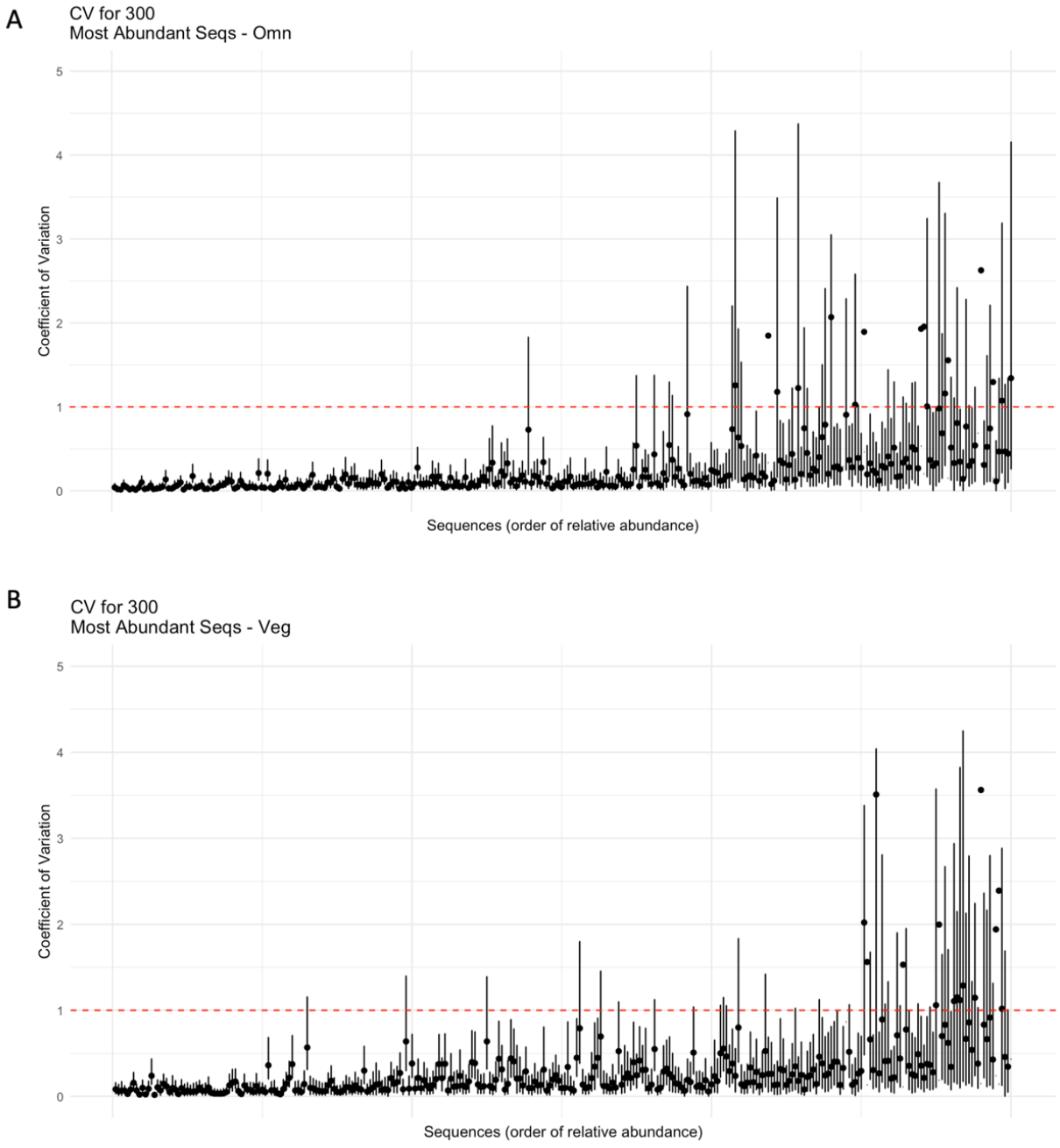


Figure 4-4: Coefficient of variation with 95 % credible intervals for the top 300 ASVs in the (A) omnivore and (B) vegetarian metagenomic data.
Sequences are plotted in order of decreasing relative abundance from left to right. The red dashed line indicates a CV of 1 or 100 %.

We explored the use of a Bayesian model for comparing ratios of model variation components for within vials (across four repeat extractions) to between vials (described in detail in Appendix I, I.3.3). In this case, if the biological variability (between vials) is less than or equal

to technical variability (within vials), we can consider the material to be sufficiently homogeneous. These data, presented in Appendix I, Figure 3, show that for omnivores only five of the top 300 sequences have credible intervals > 1, which amounts to 1.67 % of the most abundant sequences. For the vegetarians, 13 of the top 300 sequences have credible intervals > 1, which is 4.33 % of the most abundant sequences. In both groups, less than 5 % of the most abundant sequences have a biological variability that is greater than instrument and experimental variability, which could be expected based on chance alone. In general, this approach worked well for assessing homogeneity; however, it was not sufficient when trying to compare multiple time points for determining stability due to the additional complication of variation components due to time and sequencing effects. In order to have a common metric for homogeneity and stability, we chose instead to quantify a Coefficient of Disagreement (COD) for each pair of samples (Figure 4-5). The COD metric was based on the 99th percentile of disagreement between the samples for all the individual taxa observed within that donor pool and is described in detail in Appendix I.

Figure 4-5: Homogeneity coefficient of disagreement (COD) for metagenomic data of (A) omnivores and (B) vegetarians.

4.2.3.1. Omnivore COD – Homogeneity

Homogeneity for the omnivore material was assessed using a total of 40 samples. A COD was calculated between each pair of samples (Figure 4-5A). Each sample is identified by its label on the *x*-axis (the samples corresponding to the vial from box 1 are labeled "V1R1" through "V1R4" and so on until the samples corresponding to the vial from box 50 are "V10R1" through "V10R4"). Along the *y*-axis for each sample are the CODs for each pairwise comparison. For example, for "V1R1" there are 39 points along the vertical axis corresponding to a COD for each of the 39 other samples. The points are colored by the vial that each pairwise comparison is made from (a red circle on the vertical axis for "V1R1" signifies a comparison with one of the other three replicates from the vial from box 1, while an orange triangle would signify a comparison with one of the replicates from the vial taken from box 6). In this way, we can assess whether there are any boxes or replicates that are systematically

different from the others. From the plot, we conclude that 99 % of the most abundant taxa for the omnivore homogeneity samples exhibit relative abundances that are within 8 % of each other (Figure 4-5A). Using a quantile confidence interval, we can find an approximate 95 % upper confidence interval for the COD values of 7.267 %.

4.2.3.2. Vegetarian COD – Homogeneity

For assessing homogeneity of the vegetarian material, we had a total of 38 samples (replicates), as two samples were removed due to poor data quality (< 2000 total reads per sample). The plot (Figure 4-5B) is labeled and colored similarly to the plot for the omnivore samples, with the exception that since we have 38 CODs, each sample on the *x*-axis has 37 points that correspond to the 37 other samples. From the plot, we conclude that ≥ 99 % of the most abundant taxa for the vegetarian samples exhibit relative abundances that are within 13 % of each other. Using a quantile confidence interval to find an approximate 95 % upper confidence bound for the COD values. Using this procedure, we find that an approximate 95 % upper confidence limit is 10.570 %.

4.2.4. Stability

Similar to the homogeneity assessment, we began to assess stability by looking at the Bray-Curtis and Hellinger distances on MDS plots (Figure 4-6). As expected, the distance between donor pools is greater than the distance between measurements within a donor pool. Furthermore, there is no obvious grouping or distribution observed across time; rather, each donor pool appears as a distinct cluster.

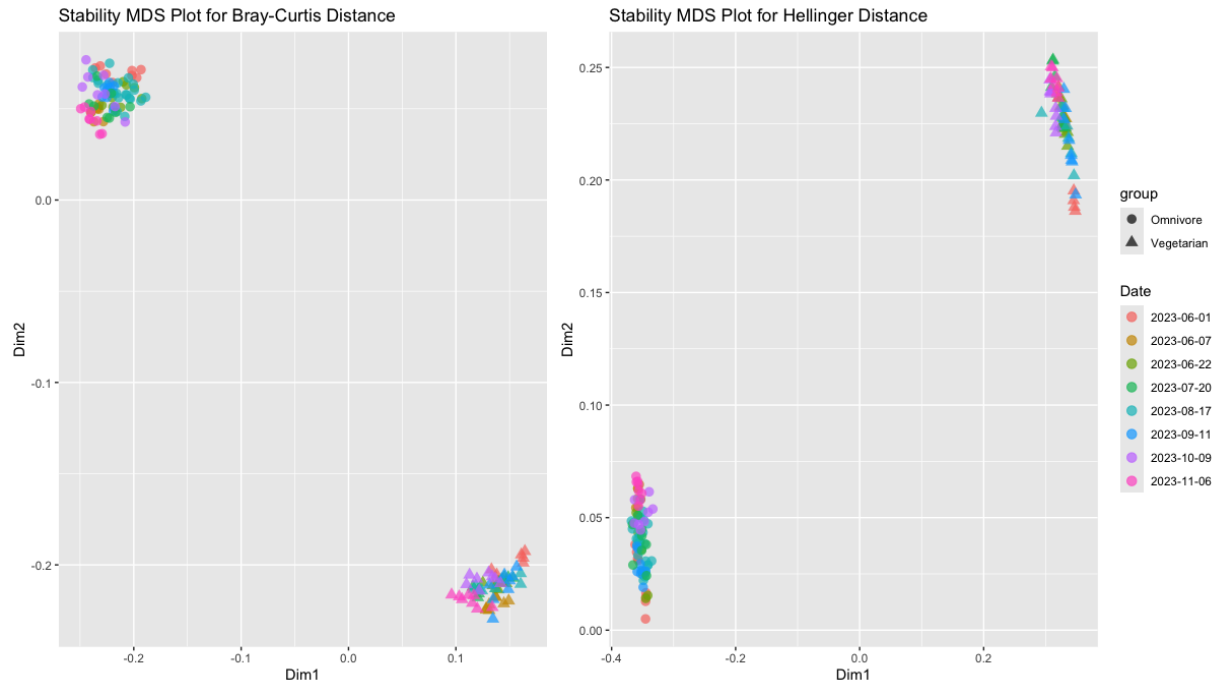


Figure 4-6: MDS plot for Bray-Curtis (left) and Hellinger (right) distances for metagenomic stability data. The data for the stability analysis are made up of measurements taken from ten omnivore and ten vegetarian boxes, sampled randomly over six months. At each time point, a minimum of two vials were taken from distinct boxes and each separated into four aliquots as replicate measurements. A total of 80 omnivore and 64 vegetarian data points are included.

Due to the more complex nesting structure of the stability data, deconvoluting potential sources of variability in order to properly assess the variability due solely to the stability of the material in a Bayesian model is not inconsequential. Additionally, the data from different time points were sequenced separately, meaning that we are unable to parse out variability introduced due to any potential sequencing effect (i.e., day-to-day measurement variability).

However, we can assess the level of stability of the material using the same COD procedure applied for homogeneity, with one minor adjustment. For the stability dataset, we define a sample as aggregated (averaged) technical replicates from each vial at each time point (n = 4) since these can be considered as repeated measurements. The homogeneity dataset described above was included in the stability analysis as time point 0.

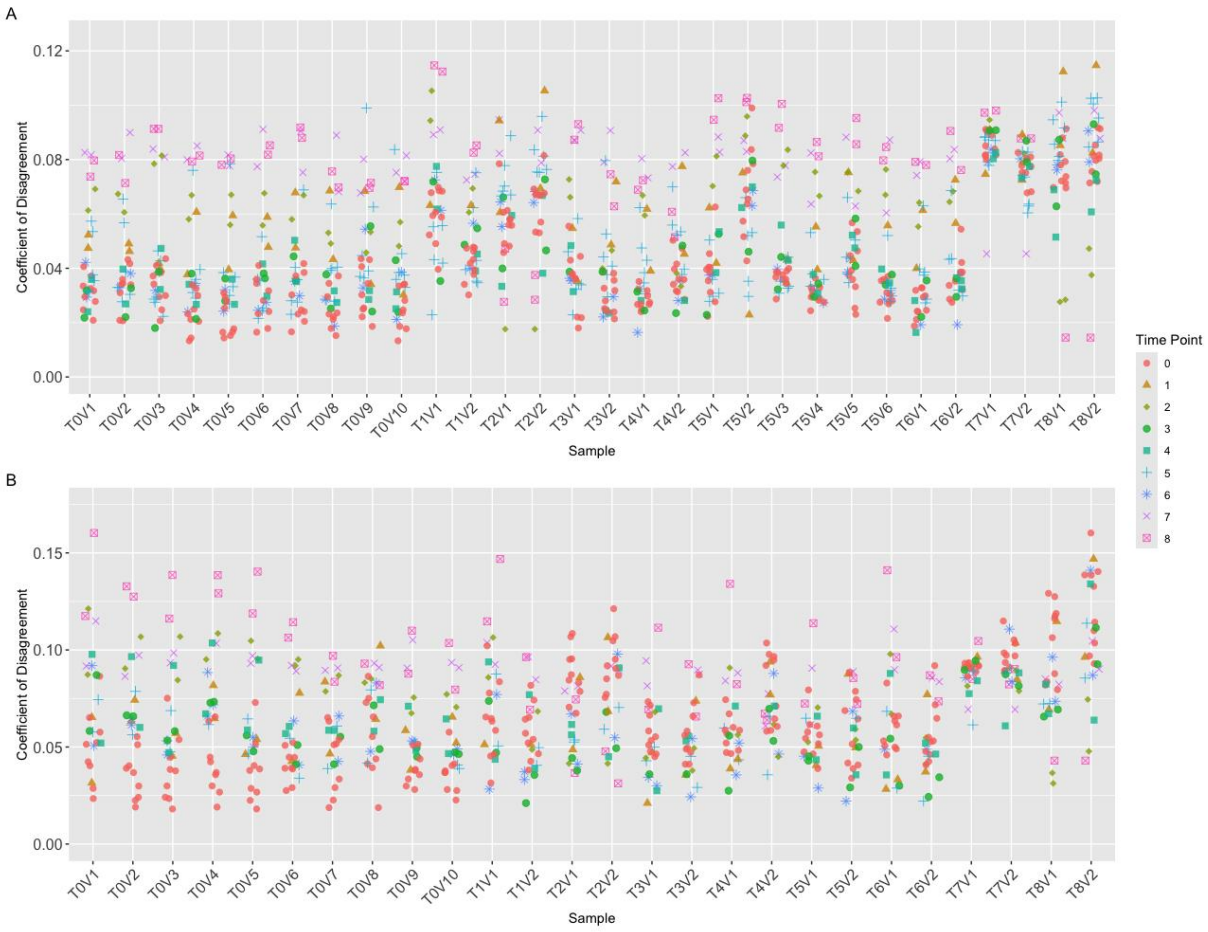


Figure 4-7: Coefficient of disagreement (COD) for metagenomic stability datasets (A) omnivores and (B) vegetarians.

4.2.4.1. Omnivores COD – Stability

In Figure 4-7A, we have visualized the CODs quantified for the stability data within the omnivore donor pool. For the stability data, there are a total of 30 samples, where sample is defined as above. Each sample is identified by its label on the *x*-axis (from time point 0, vial 1, labeled "T0V1", through time point 8, vial 2, labeled "T8V2"). Each vial along the *x*-axis has 29 points corresponding to the comparisons with other samples. The points are colored by day, so that a red circle on the vertical axis along 'T0V1' signifies a comparison to another vial measured at time point 0, while an orange triangle signifies a comparison to the vials measured at time point 1. In this way, we can check for any systematic patterns over time. From the plot, we conclude that ≥ 99 % of the most abundant taxa for the omnivore samples exhibit relative abundances that are within 12.5 % of each other (Figure 4-7A). Using a quantile confidence interval, an approximate 95 % upper confidence limit for the COD values is 11.401 %.

4.2.4.2. Vegetarians COD – Stability

The CODs quantified for the stability data within the vegetarian donor pool are shown in Figure 4-7B. The plot is labeled and colored similarly to Figure 4-7A for the omnivore samples, with the exception that the vegetarian dataset contains a total of 26 samples. We conclude from these data that ≥ 99 % of the most abundant taxa for the vegetarian samples exhibit relative abundances that are within 17.5 % of each other (Figure 4-7B). An approximate 95 % upper confidence limit for these COD values using a quantile confidence interval is 15.04 %. In this case, we can see that the time point 0 (or Day 1) samples are among those with the highest discrepancies from all other time points (the red circles for each non-time point 0 sample). There may be several explanations for this difference, and we are currently unable to determine whether this is attributed to day-to-day variability or to another cause.

4.3. Metagenomics Summary

Based on the COD analysis presented herein, RM 8048 is considered sufficiently homogeneous and stable with the measured relative abundances of ≥ 99 % of the most abundant individual taxa varying by ≤ 17.5 % between either individual vials (homogeneity) or different days (stability over ≥ 168 days).

5. LC-MS/MS Metabolomics

5.1. Methods

Ten vials from each of the omnivore and vegetarian donor pools (20 vials total) were characterized for metabolites and assessed for homogeneity. Samples were prepared using a Bligh & Dyer extraction (See Appendix E for detailed methods). The polar and non-polar layers were separated, dried, and weighed. The polar fraction was resuspended in 600 µL of 2 % methanol in water (volume fraction) prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. The non-polar fraction was resuspended in 250 µL of isopropanol prior to LC-MS/MS analysis.

For annotation of polar metabolites, samples were run on two different instruments: a Vanquish ultra-high performance liquid chromatography (UHPLC) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) and a Vanquish UHPLC coupled to the Q Exactive (QE) Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Sample injections (2 µL) were separated by an Acquity HSS T3 (1.8 µm, 2.1 mm inner diameter x 150 mm length; Waters, Milford, MA, USA) C18 column at 350 µL/min and 45 °C with the gradient program listed in Appendix E Table E.1.2. On the Fusion Lumos, samples from each donor pool were acquired in Acquire X data acquisition mode to provide additional MS^2 spectra using both higher energy collision dissociation (HCD) and collision-induced dissociation (CID) fragmentation data spectra for spectral library matching and annotation.

For annotation of non-polar metabolites, samples were run on a Vanquish UHPLC coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer. Sample injections (1 µL) were separated by an Acclaim C30 column (3 µm, 2.1 mm id x 100 mm length; Thermo Scientific Sunnyvale, CA, USA) at 300 µL/min and 45 °C with the gradient program listed in Appendix E Table E.1.3. Samples from each donor pool were acquired in Acquire X data acquisition mode to provide additional MS^2 spectra using both HCD and CID fragmentation data for spectral library matching and annotation.

Resulting raw files were processed and searched with Compound Discoverer (CD, version 3.3.2) using both the local NIST23 library and the online database mzCloud (https://www.mzcloud.org/). A search was performed for all ten omnivore samples and a separate search was performed for all ten vegetarian samples. AcquireX ID samples were included in the Lumos data searches. Data from Compound Discover were sorted and combined according to donor pool and ionization mode, where compounds were removed if not found across all three algorithms and across both instruments (polar only). The three algorithms used were NIST, HighChem-HighRes (HCHR), and HighChem Dot Product (HCDP).

The initial mass features from the CD output (prior to the sorting mentioned above) were exported in excel and were assessed for homogeneity across vials by calculating the relative standard deviation (RSD) of the relative abundance of each feature across vials. An average RSD was calculated from the large number of features obtained for each sample to show a representative group RSD. The average RSDs were calculated for both the polar and non-polar

fractions for each donor pool and for each analytical method applied, including both positive and negative ionization mode on two different instruments.

See Appendix E for detailed methods.

5.2. Results

Table The relative abundance of each feature across vials had an average RSD of ≤ 20.2 % for each sample fraction (polar and non-polar) in both donor pools and in all analytical methods (Table).

Table 5-1: The average relative standard deviations (RSDs) of feature intensity for the ten samples per donor pool analyzed by each LC-MS/MS method.

Method	Omnivore (Avg RSD %)	Vegetarian (Avg RSD %)
Polar positive ionization Lumos	17.9	18.7
Polar negative ionization Lumos	15.1	16.9
Polar positive ionization QE	18.3	17.6
Polar negative ionization QE	15.2	20.1
Non-polar positive ionization Lumos	14.7	15.4
Non-polar negative ionization Lumos	15.5	18.2

In addition to sample processing, sample size, and instrumentation bias, one major cause for differences in identified compounds across metabolomics studies is the utilization of different software and/or algorithms for annotation. For the untargeted characterization for RM 8048, the software and libraries remained the same for each search, but three different algorithms were used to produce an annotation list. Results were different among the NIST, HCHR, and HCDP algorithms. As an example, the number of compounds for the Lumos instrument in positive ionization mode was 369 (NIST), 666 (HCHR), and 474 (HCDP) in the independent algorithm searches but these lists were merged to 181 concordant compounds across all three algorithms. This list was further reduced when compared across instruments (removing compounds not found across all instruments). The final characterization lists were combined to contain polar and non-polar compounds, resulting in 112 compounds for the omnivore donor pool (Appendix F Table F.1) and 148 compounds for the vegetarian donor pool (Appendix F Table F.2). The lists also include the InChIKey associated with each common name.

5.2.1. Omnivore COD – LC-MS/MS Homogeneity Assessment

For assessing homogeneity, the aligned and unfiltered feature peak areas for the ten omnivore vials (replicates) were exported from Compound Discoverer for both the polar and non-polar LC-MS/MS analysis and for both positive and negative polarities. A COD was then calculated for the peak areas of all features between each pair of samples in each polarity mode for both the

polar and non-polar analysis (Table 5-2). As described in the metagenomics section above, an assessment of vials that are systematically different from the others can be visualized (Figure 5-1). Each sample is identified by its label on the *x*-axis corresponding to the eight vials analyzed from boxes across the material production. Along the *y*-axis for each sample are the CODs for each pairwise comparison. From the plot, we conclude that ≥ 95 % of the features for the omnivore homogeneity samples exhibit peak areas that are within 15.1 % of one another.

Table 5-2: Uncertainty intervals for the COD values corresponding to the eight different LC-MS/MS measurement scenarios (four different measurement methods for omnivores and four for vegetarians).

Experiment	COD Lower Limit	COD Upper Limit	Confidence
Polar Omnivore Positive Lumos	0.143	0.151	95 %
Polar Omnivore Negative Lumos	0.113	0.123	95 %
Non-polar Omnivore Positive Lumos	0.117	0.129	95 %
Non-polar Omnivore Negative Lumos	0.087	0.094	95 %
Polar Vegetarian Positive Lumos	0.091	0.119	95 %
Polar Vegetarian Negative Lumos	0.130	0.161	95 %
Non-polar Vegetarian Positive Lumos	0.141	0.151	95 %
Non-polar Vegetarian Negative Lumos	0.112	0.149	95 %

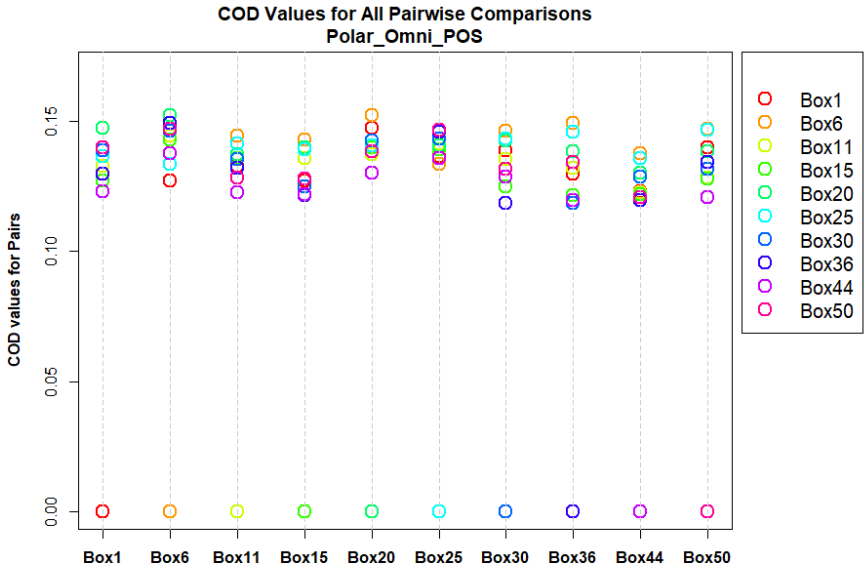


Figure 5-1: Example COD plot for omnivore LC-MS/MS analysis (polar extract, positive polarity).

5.2.2. Vegetarian COD – LC-MS/MS Homogeneity Assessment

For assessing homogeneity, the aligned and unfiltered feature peak areas for ten vegetarian vials (replicates) were exported from Compound Discoverer for both the polar and non-polar LC-MS/MS analysis and for both positive and negative polarities. A COD was then calculated for the peak area of all features between each pair of samples in each polarity mode for both the polar and non-polar analysis (Table 5-2) and plotted (Figure 5-2). Each sample is identified by its label on the *x*-axis corresponding to the eight vials analyzed from boxes across the material production. Along the *y*-axis for each sample are the CODs for each pairwise comparison. From the plot, we conclude that ≥ 95 % of the features for the vegetarian homogeneity samples exhibit peak areas that are within 16.1 % of one another.

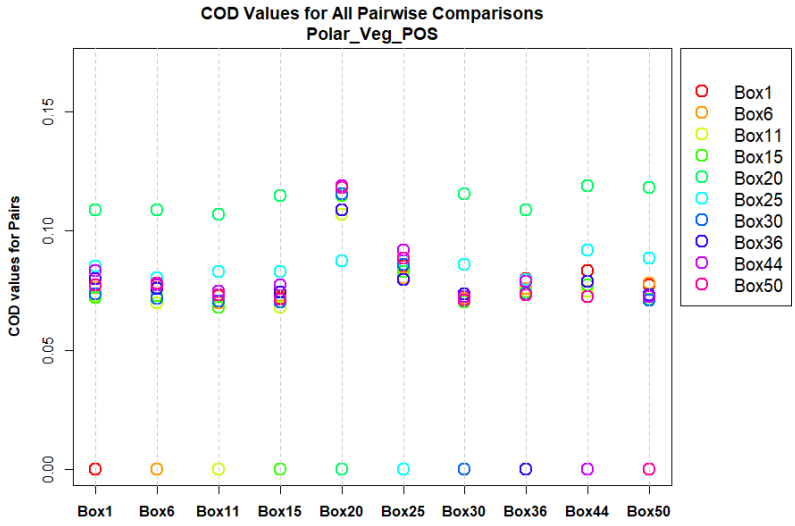


Figure 5-2: Example COD plot for vegetarian LC-MS/MS analysis (polar extract, positive polarity).

5.3. LC-MS/MS Metabolomics Summary

Based on the COD analysis, RM 8048 is considered sufficiently homogeneous for the purpose of untargeted metabolomics using LC-MS/MS with ≥ 95 % of the features exhibiting peak area variability of ≤ 15.1 % and ≤ 16.1 % between replicates for omnivore and vegetarian, respectively.

6. NMR Metabolomics

An untargeted NMR-based metabolomics approach was used to provide a qualitative characterization of RM 8048 polar metabolites for both vegetarian and omnivore materials. Homogeneity and stability of the material production was assessed by two different sample preparation methods and analyzed by two different NMR field strengths: 1) a biphasic organic solvent extraction to broadly characterize the polar chemical constituents (700 MHz NMR) and 2) an aqueous extract ("fecal water") that is commonly used within the metabolomics community (600 MHz NMR). The annotation data for both methods are provided to deliver the confidently identified polar metabolites of RM 8048 by NMR spectroscopy for the two donor pools (omnivore and vegetarian). Data from the two methods were combined in a table in the Reference Material Information Sheet (RMIS) for RM 8048.

6.1. Organic Solvent Extraction – Polar Metabolites – 700 MHz

6.1.1. Methods

Homogeneity and stability for the RM 8048 materials were tested via nuclear magnetic resonance (NMR) spectroscopy using different boxes across the whole production. For homogeneity testing, one vial from each of ten different boxes was analyzed (T0); for stability testing, one vial from each of four different boxes was tested at each subsequent time point, T1-T6 (Table 6-1).

Table 6-1: Vial selection and time points (in months) for homogeneity and stability analysis of ^{1}H NMR spectra of extracted polar metabolites.

vial #	T0	T1 (+2mo)	T2 (+4mo)	T3 (+6mo)	T4 (+8mo)	T5 (+10mo)	T6 (+13mo)
1	Box 1	Box 1	Box 1	Box 1	Box 6	Box 6	Box 6
2	Box 6	Box 15	Box 15	Box 15	Box 20	Box 20	Box 20
3	Box 11	Box 30	Box 30	Box 30	Box 25	Box 25	Box 25
4	Box 15	Box 50	Box 50	Box 50	Box 36	Box 36	Box 36
5	Box 20						
6	Box 25						
7	Box 30						
8	Box 36						
9	Box 44						
10	Box 50						

Bligh-Dyer Extraction of Polar Metabolites. For homogeneity testing, polar metabolites were extracted in two separate batches in parallel by two different analysts independently focusing on ten vials from one donor pool (one analyst for omnivore and one for vegetarian). For each stability time point, one analyst extracted a single batch containing four vials of omnivore samples and four vials of vegetarian samples. Each extraction batch also included quality control samples (QC) consisting of three replicates of a NIST standard reference material (SRM 1946 – Lake Superior Fish Tissue, 100 mg ± 3 mg per replicate) in addition to one procedural blank. These QC samples were extracted and analyzed alongside the experimental samples within each batch to assess measurement reproducibility as well as contamination. All RM 8048 samples and SRM 1946 replicates were kept at -80 °C until analysis. Samples were processed using a Bligh-Dyer biphasic solvent extraction (See Appendix G for detailed methods). The polar layer was separated, dried, and weighed. Dried metabolites were then rehydrated in 600 µL of 100 mmol/L sodium phosphate NMR buffer in D_2O (pH = 7.3, measured) containing 1.0 mmol/L sodium 3-Trimethylsilyl 2,2,3,3-d_4 propionate (TMSP, CAS: 24493-21-8) as the chemical shift reference standard. The rehydrated samples were vortexed for ≈ 30 s and centrifuged at 14,000 x *g* at 4 °C for 5 min to pellet any particulates. Samples (550 µL) were then transferred into 5 mm NMR tubes (Sample Vault Series™, Norell®, Morganton, NC), spun with a manual centrifuge, and randomized for subsequent NMR spectroscopy analysis.

NMR Analysis and Data Processing. All spectra were acquired at a temperature of 298 K on a Bruker Avance II 700 MHz NMR spectrometer equipped with a 5 mm triple-resonance, z-gradient TXI room temperature probe with a SampleJet automatic sample changer (Bruker BioSpin, Inc., Billerica, MA). NMR experiments included one-dimensional (1D) ^{1}H NMR spectra (Bruker pulse program '*noesygppr1d*'), as well as the following 2D experiments: edited ^{1}H,^{13}C-HSQC spectra with adiabatic ^{13}C decoupling (Bruker pulse program '*hsqcedetgpsisp2.2'*), phase-sensitive Total Correlation Spectroscopy (TOCSY) spectra (Bruker pulse program '*dipsi2esgpph'*), and phase-sensitive gradient-enhanced ^{1}H, ^{13}C -HSQC-TOCSY (Bruker pulse program '*hsqcdietgpsisp.2'*). The 2D experiments were acquired on selected samples to aid in metabolite identification. See Appendix G for detailed methods. Raw and processed NMR data can be found in the NMR 700 MHz folder located here https://doi.org/10.18434/mds2-3633.

Statistical Analysis. NMR spectra were binned from 0.20 ppm to 10.0 ppm with a resolution factor of 0.5 and a signal-to-noise ratio threshold of 3 using NMRProcFlow v1.4[7]. The following spectral regions were excluded from the analysis since they contained artifacts either due to water suppression or the presence of contaminants detected in blank spectra: acetate (1.91 ppm to 1.93 ppm), methanol (3.35 ppm to 3.37 ppm), water (4.71 ppm to 4.90 ppm), chloroform (7.67 ppm to 7.69 ppm), and formate (8.45 ppm to 8.47 ppm). Spectral binning resulted in a total of 695 bins for omnivore and 652 bins for vegetarian donor pools for subsequent homogeneity and stability analysis. For multivariate analysis the binned data were normalized by summing of intensities, mean-centered, and scaled using Pareto scaling in

[7] Jacob D, Deborde C, Lefebvre M, Maucourt M, Moing A (2017) NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. Metabolomics 13(4):36. https://doi.org/10.1007/s11306-017-1178-y

MetaboAnalyst 6.0 software[8]. The technical variability was visualized using a principal component analysis (PCA) scores plot for all samples and QC materials. Homogeneity assessment using spectrum-wide median Relative Standard Deviation (RSD)[9] was performed on omnivore and vegetarian samples. For characterization of both stability and homogeneity, a COD procedure as detailed in Section 3 was completed with a measure of discrepancy based on peak presence / absence described in Appendix I.

Metabolite Identification. Metabolites were assigned using the 1D ^{1}H, and 2D ^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY spectra and were putatively identified at a Level 2 where possible[10]. Metabolite assignments were based on chemical shift comparisons using reference spectra from the Human Metabolome Database (HMDB)[11], the Biological Magnetic Resonance Bank (BMRB)[12], the spectral library in Chenomx NMR Analysis software (Chenomx Inc. version 8.5), as well as an in-house database (See Appendix G for detailed methods). Additionally, the COLMARm web tool[13] was used for automated metabolite IDs using the 2D NMR spectra including ^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY. Input files for the COLMARm tool were generated using MestReNova v.14.3.3 and included NMR data files for each spectrum (^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY).

Metabolomics Storage Stability Analysis. Initially, PCA scores plots were calculated using MetaboAnalyst 6.0 software for RM 8048 omnivore and vegetarian donor pools at seven time points (T0 through T6) over 13-months of storage at -80 °C. Each time point is approximately two-months apart with T6 analysis occurring at three months from the previous time point. PCA was used to evaluate data clustering and potential time-dependent changes in metabolite profiles. Subsequently, Partial Least Squares Discriminant Analysis (PLSDA) scores plots were calculated to identify specific metabolites that correlate with time. PLSDA Variable Importance in Projection (VIP) scores (VIP > 1.0) were used to select spectral bins that showed significant time-dependent changes. Metabolites corresponding to the individual bins were subsequently identified in 1D ^{1}H and 2D ^{1}H, ^{13}C HSQC NMR spectra using Chenomx NMR Analysis software (1D spectra) and an in-house 2D HSQC spectral library (2D spectra).

Box plots were generated from normalized bin intensities for individual metabolites at the seven time points using MetaboAnalyst 6.0. One way Analysis of Variance (ANOVA) followed by a Tukey-Kramer post-hoc multiple comparison test (JMP v.11.1.1, SAS Institute Inc.) (significance set for $p < 0.05$) was used to compare every pair of means at each time point to

[8] Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics* 68(1):e86. https://doi.org/https://doi.org/10.1002/cpbi.86
[9] Parsons HM, Ekman DR, Collette TW, Viant MR (2009) Spectral relative standard deviation: a practical benchmark in metabolomics. *Analyst* 134(3):478-485.
[10] Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3(3):211-221. https://doi.org/10.1007/s11306-007-0082-2
[11] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y, Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A (2017) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* 46(D1):D608-D617. https://doi.org/10.1093/nar/gkx1089
[12] Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2007) BioMagResBank. *Nucleic Acids Research* 36(suppl_1):D402-D408. https://doi.org/10.1093/nar/gkm957
[13] Bingol K, Li D-W, Zhang B, Brüschweiler R (2016) Comprehensive Metabolite Identification Strategy Using Multiple Two-Dimensional NMR Spectra of a Complex Mixture Implemented in the COLMARm Web Server. *Analytical Chemistry* 88(24):12411-12418. https://doi.org/10.1021/acs.analchem.6b03724

determine whether there were significant differences in bin intensities over the time period from T0 through T6.

6.1.2. Results

6.1.2.1. Homogeneity

NMR metabolomics characterization of RM 8048 homogeneity utilized a biphasic extraction method, Bligh-Dyer, that effectively captures the full breadth of polar metabolites by removing protein and non-polar compounds. While this method is not necessarily used in fecal metabolomics workflows, it provides a deeper chemical analysis of the reference materials. Stool extract mass (dried) and corresponding mean % RSD values are summarized in Table 6-2**.**

Table 6-2: Stool polar metabolite extract mass (dried) with mean, standard deviation (SD), and mean mass % RSD for ten RM 8048 omnivore and ten vegetarian samples.

Omnivore			
Sample	**Empty Vial (g)**	**Vial + Extract (g)**	**Polar Extract (g)**
O_31	1.17528	1.17917	0.00389
O_32	1.17437	1.17822	0.00385
O_33	1.18131	1.18547	0.00416
O_34	1.17955	1.18359	0.00404
O_35	1.18224	1.18647	0.00423
O_36	1.17536	1.17955	0.00419
O_37	1.17177	1.17612	0.00435
O_38	1.17487	1.17886	0.00399
O_39	1.17443	1.17896	0.00453
O_40	1.18210	1.18535	[a]0.00325
		Omni Mean	0.00414
		SD	0.00022
		Mass % RSD	5.31
Vegetarian			
Sample	**Empty Vial (g)**	**Vial + Extract (g)**	**Polar Extract (g)**
V_21	1.18051	1.18379	0.00328
V_22	1.17940	1.18254	0.00314
V_23	1.17533	1.17823	0.00290
V_24	1.18046	1.18350	0.00304
V_25	1.18030	1.18355	0.00325
V_26	1.17509	1.17867	0.00358
V_27	1.17838	1.18140	0.00302
V_28	1.18263	1.18638	0.00375
V_29	1.18169	1.18539	0.00370
V_30	1.17993	1.18354	0.00361
		Veg Mean	0.00333
		SD	0.00031
		Mass % RSD	9.31

[a] Not included in the calculations due to partial sample loss.

Sample O_40 was not included in the RSD calculations due to partial sample loss that occurred during sample processing but was included with normalization in the metabolomics analysis for homogeneity assessment and metabolite identification. The average polar metabolite extract mass for omnivore samples was 0.00414 g ± 0.00022 g, while the average mass for vegetarian samples was 0.00333 g ± 0.00031 g. The higher mass for omnivore polar metabolite extracts compared with vegetarian samples is likely due to the different chemical composition of the stool material as a result of the different dietary regimes. For example, a plant-based diet (vegetarian) is known to contain substantially larger amounts of complex starch and dietary

fibers compared with non-vegetarian diets[14]. Consequently, the vegetarian stool may have a higher fiber content compared with omnivore stool. Since we performed a solvent extraction to isolate low-molecular weight metabolites, the high-molecular weight fibers (complex carbohydrates) would have been removed during extraction, thus resulting in a lower average mass of the corresponding vegetarian stool extract compared with the omnivore stool extract.

For homogeneity testing, 1D ^{1}H NMR spectra for 10 omnivore and 10 vegetarian samples are shown in Figure 6-1 and Figure 6-2, respectively.

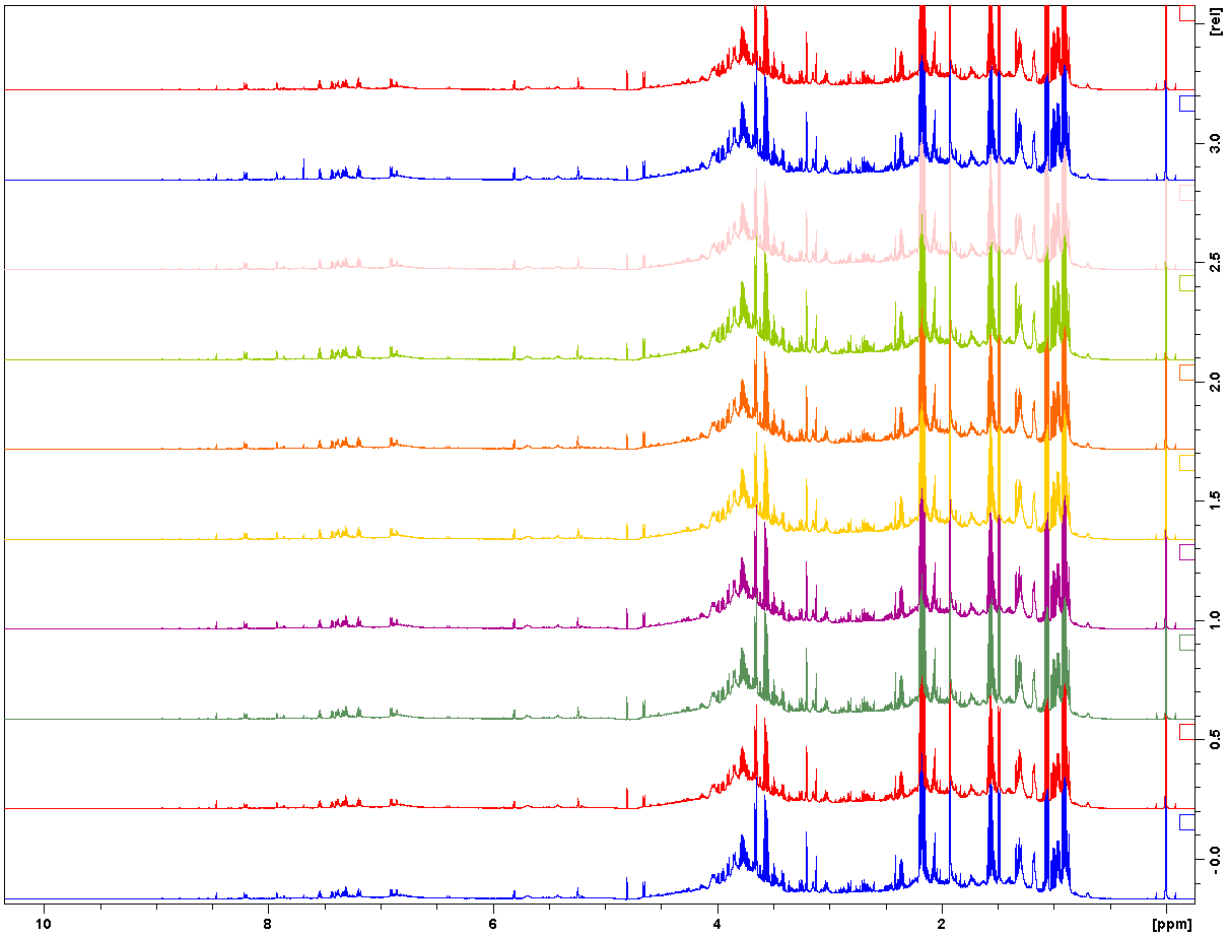


Figure 6-1: ^{1}H NMR full spectra comparison of the polar metabolites **(Bligh-Dyer extraction) from** ten omnivore RM 8048 vials.

[14] Craig WJ, Mangels AR, Fresán U, Marsh K, Miles FL, Saunders AV, Haddad EH, Heskey CE, Johnston P, Larson-Meyer E, Orlich M (2021) The Safe and Effective Use of Plant-Based Diets with Guidelines for Health Professionals. *Nutrients* 13(11). https://doi.org/10.3390/nu13114144

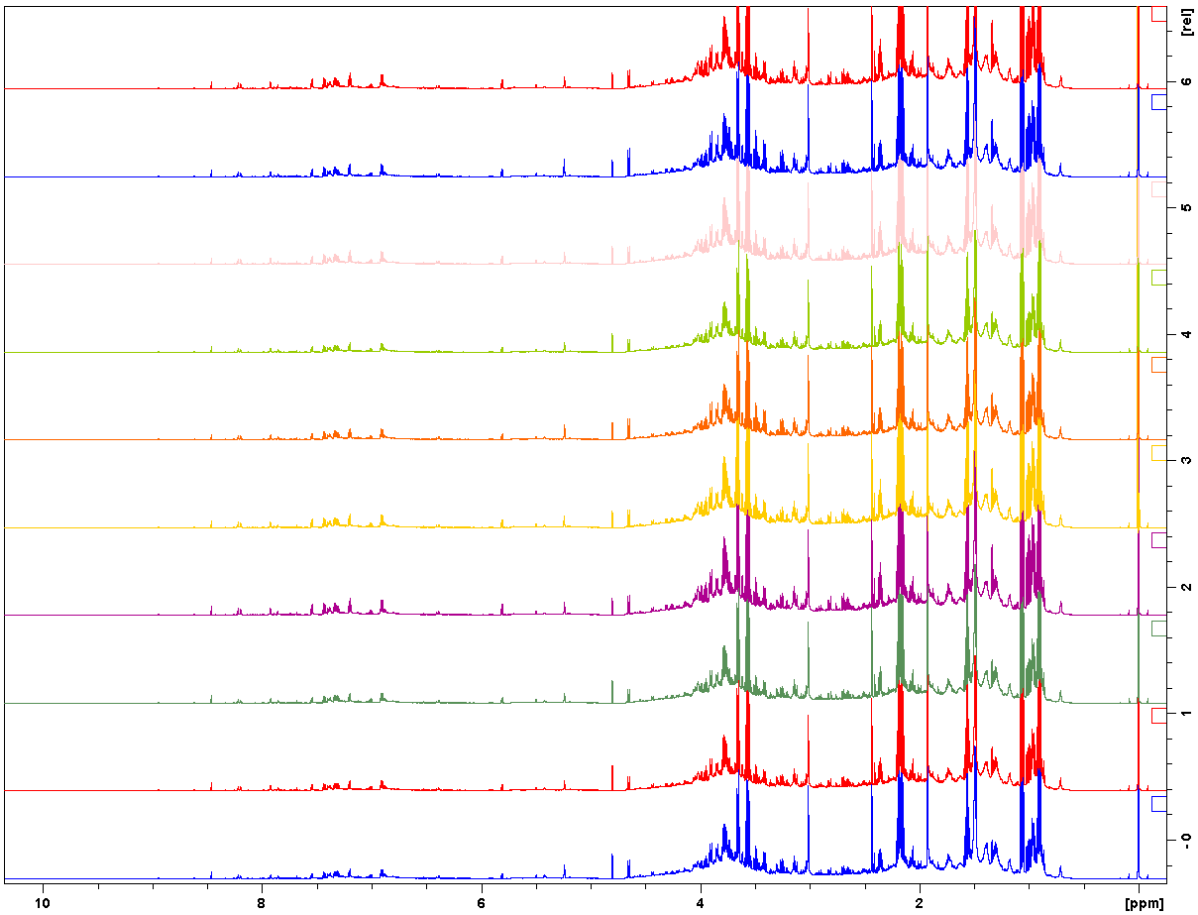


Figure 6-2: ^{1}H NMR full spectra comparison of the polar metabolites **(Bligh-Dyer extraction) from** ten vegetarian RM 8048 vials.

^{1}H NMR metabolome differences are displayed in an overlay of representative omnivore and vegetarian spectra is shown in Figure 6-3.

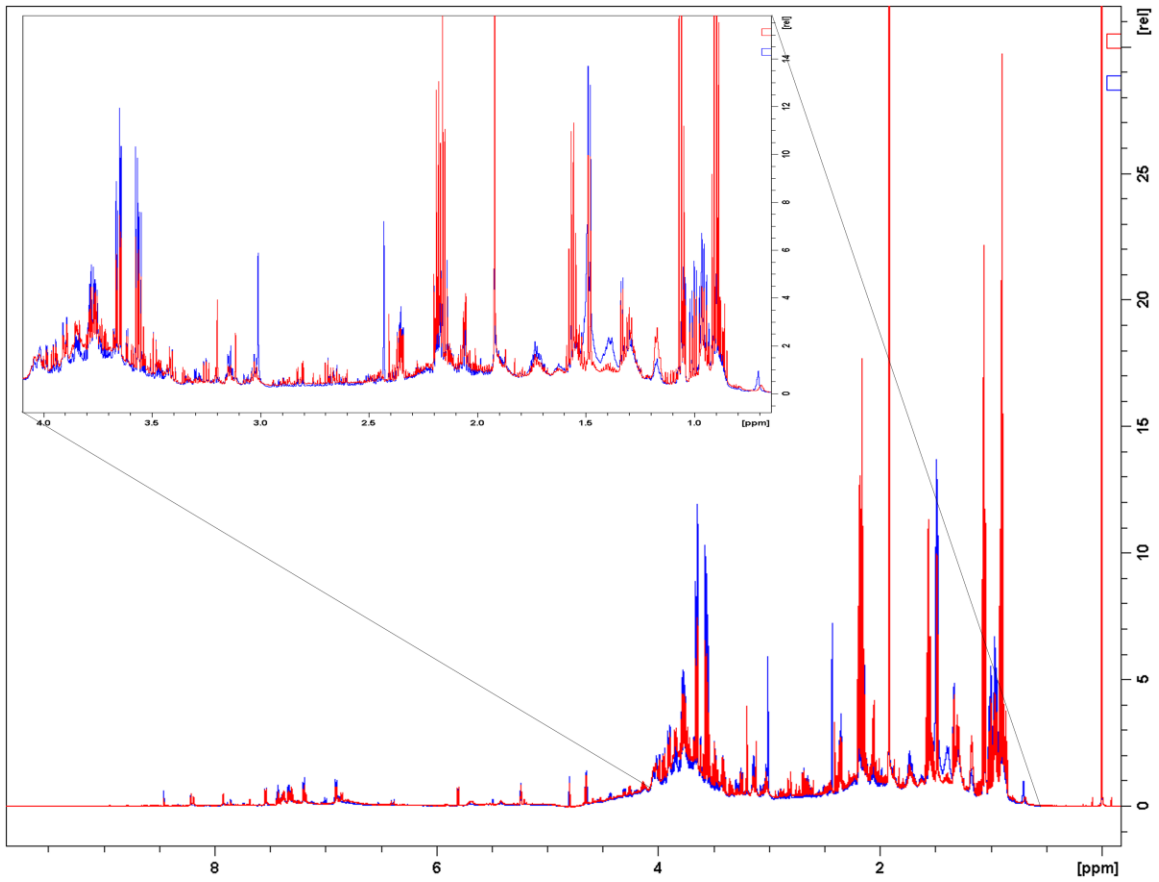


Figure 6-3: Overlay of representative ^{1}H NMR spectra of polar metabolites **(Bligh-Dyer extraction) from** vegetarian (blue) and omnivore (red) RM 8048 stool materials.
Data show differences in metabolite composition between the two materials. Inset shows expanded spectral region between 0.6 ppm and 4.1 ppm.

The polar metabolite differences in donor pools are further elucidated in a PCA scores plot (Figure 6-4).

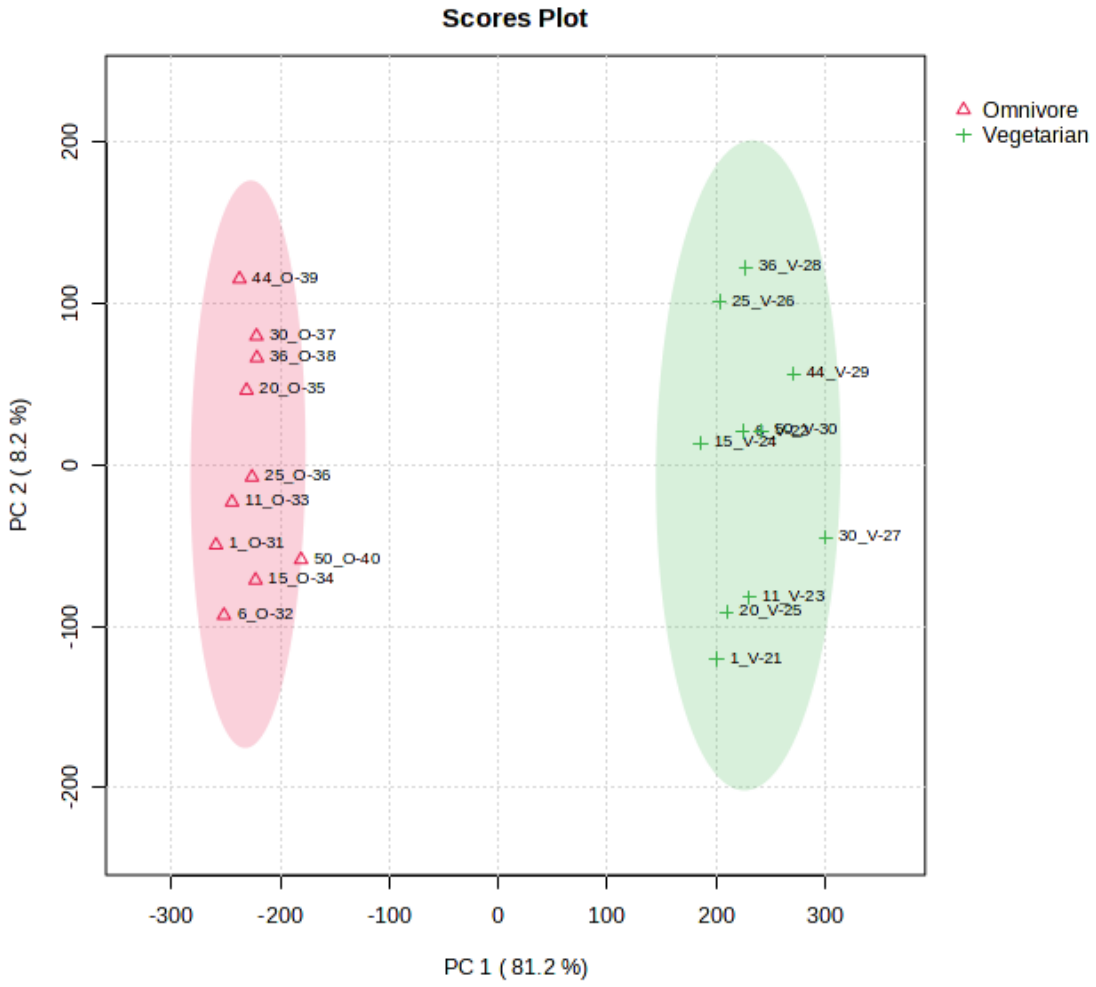


Figure 6-4: Differential comparison of RM 8048 omnivore and vegetarian stool ^{1}H NMR spectra using PCA. The first number in the Sample IDs indicates a specific box number for that vial. Ellipses represent 95 % confidence regions for each group.

As expected, the largest variation in metabolite profiles along the *x*-axis (PC1, explained variance (EV) 81.2 %) was associated with donor pool. The intra-donor pool variability was observed along PC2 (EV 8.2 %), corroborating that the variance measured with metabolite extracts (Table 6-2) is larger for the vegetarian material.

6.1.2.1.1. Omnivore COD

For assessing homogeneity, 1D ^{1}H NMR pre-processed and binned spectra (695 bins) for ten omnivore vials were used. A COD measuring the disagreement in peak presence/absence between any pair of samples, based on potential peaks within each spectral bin was then calculated (Figure 6-5). Additional details on calculation of the COD can be found in Section 3, and each sample is identified by its label on the *x*-axis corresponding to the 10 vials analyzed from different boxes selected across the material production. Along the *y*-axis for each sample are the CODs for each pairwise comparison. They are colored such that an orange circle along

the vertical axis above ‘B01T0’ is a comparison with ‘B06’ while a pink circle is a comparison with ‘B50,’ for example. Overall, we conclude that ≤ 11 % of the peaks differ between any pair of omnivore samples. Using a percentile bootstrapping procedure with 10,000 replicates, we find an upper 95 % confidence limit for the COD values of 10.250 %.

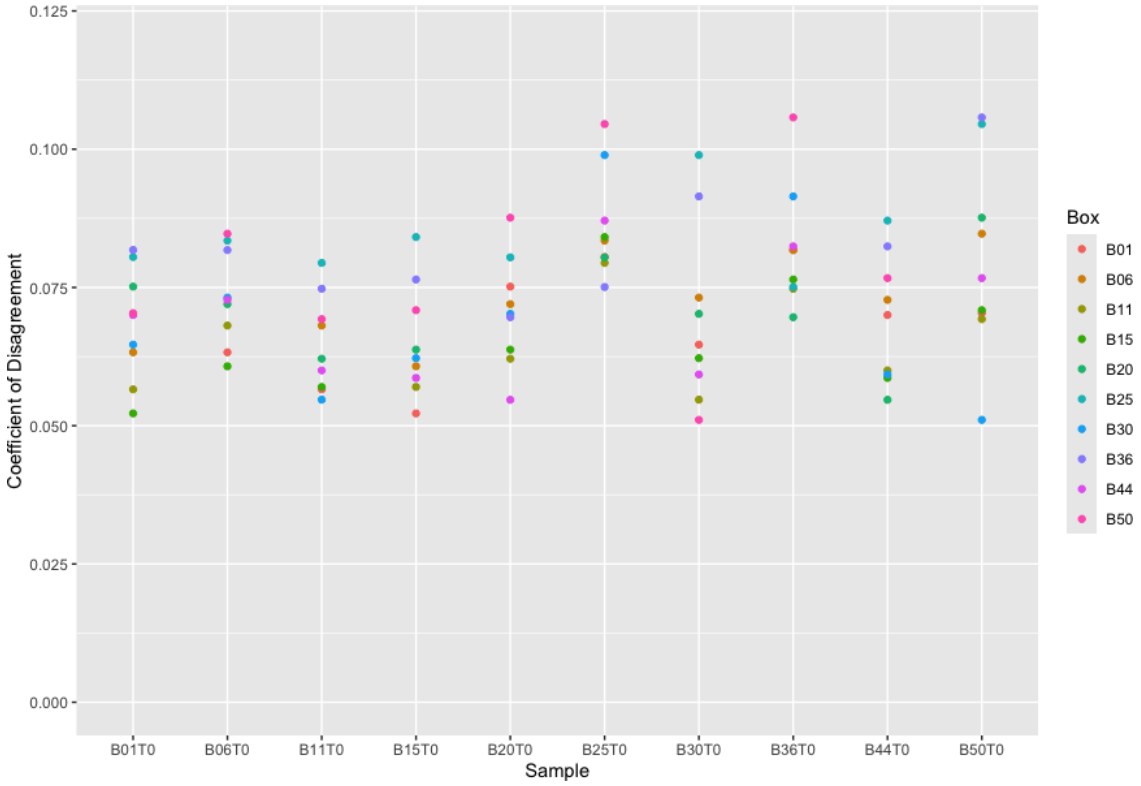


Figure 6-5: COD point plot for ^{1}H NMR polar metabolite extract homogeneity analysis of omnivore samples. Points are colored by the paired box.

6.1.2.1.2. Vegetarian COD

For assessing homogeneity, 1D ^{1}H NMR pre-processed and binned (652 bins) spectra for the 10 vegetarian vials were used. A COD was then calculated between each pair of samples (Figure 6-6). Each sample is identified by its label on the *x*-axis corresponding to the 10 vials analyzed from different boxes selected across the material production. Along the *y*-axis for each sample are the CODs for each pairwise comparison. They are colored such that an orange circle along the vertical axis above ‘B01T0’ is a comparison with ‘B06’ while a pink circle is a comparison with ‘B50,’ for example. Overall, we conclude that ≤ 8 % of the peaks differ between any pair of vegetarian samples. Using a percentile bootstrapping procedure with 10,000 replicates, we find an upper 95 % confidence limit for the COD values of 6.820 %.

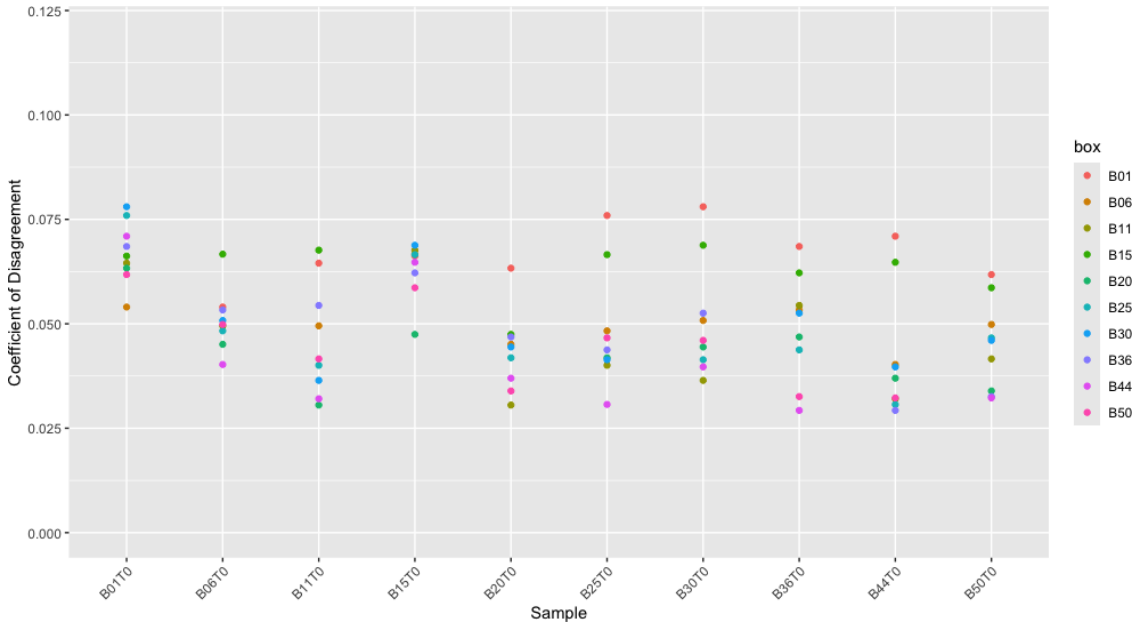


Figure 6-6: COD point plot for ^{1}H NMR polar metabolite extract homogeneity analysis of vegetarian samples. Points are colored by the paired box.

6.1.2.2. Metabolite Identification

Representative annotated 1D ^{1}H NMR spectra for both omnivore and vegetarian materials are shown in Figure 6-7 and Figure 6-8, respectively.

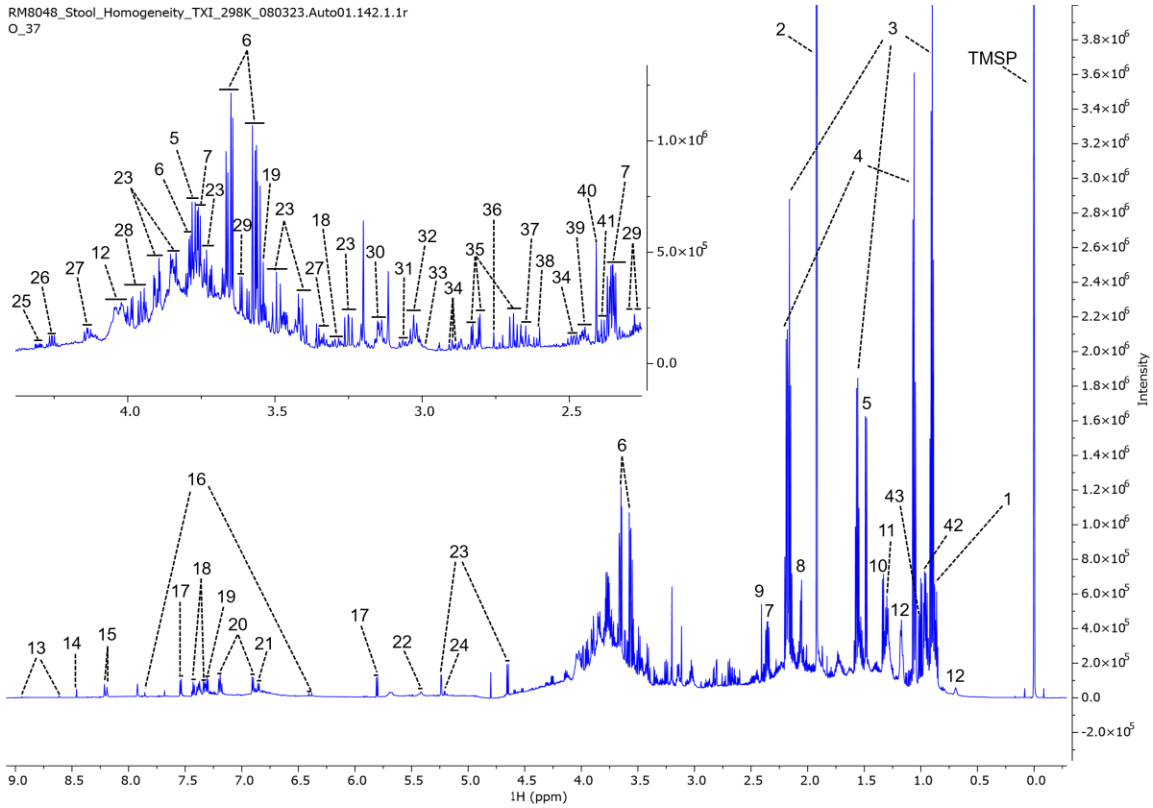


Figure 6-7: Representative 700 MHz ^{1}H-NMR spectrum of RM 8048 omnivore material polar metabolite extract. Numbers indicate the following annotated metabolites: 1: valerate; 2: acetate; 3: butyrate; 4: propionate; 5: alanine; 6: glycerol; 7: glutamate; 8: isovalerate; 9: succinate; 10: threonine; 11: caprylate; 12: bile acid; 13: nicotinate; 14: formate; 15: hypoxanthine; 16: urocanate; 17: uracil; 18: phenylalanine; 19: phenylacetate; 20: tyrosine; 21: 4-hydroxyphenylacetate; 22: maltose; 23: glucose; 24: N-acetylglucosamine; 25: malate; 26: threonine; 27: proline; 28: serine; 29: valine; 30: citrulline; 31: ornithine; 32: lysine; 33: pipecolic acid; 34: 3-phenylpropionate; 35: aspartate; 36: methionine sulfoxide; 37: methionine; 38: methylamine; 39: glutamine; 40: succinate; 41: isobutyrate; 42: leucine; 43: isoleucine.

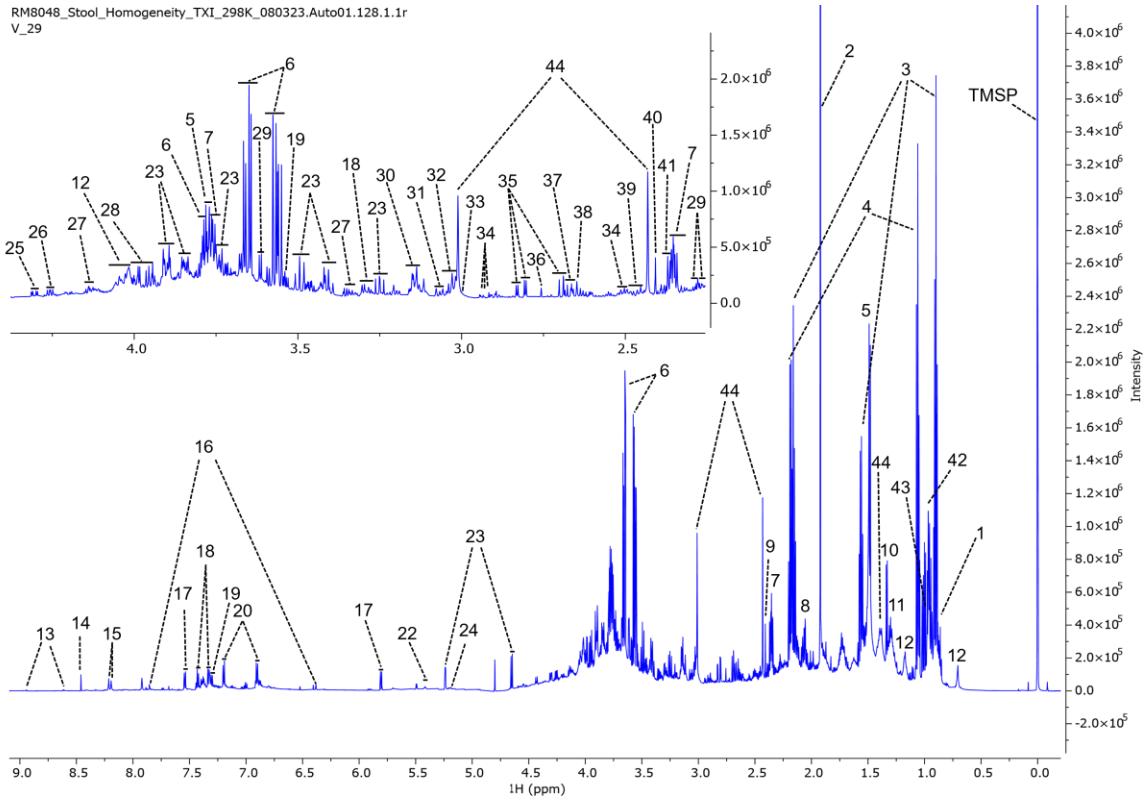


Figure 6-8: Representative 700 MHz ^{1}H-NMR spectrum of RM 8048 vegetarian material polar metabolite extract. Numbers indicate the following annotated metabolites: 1: valerate; 2: acetate; 3: butyrate; 4: propionate; 5: alanine; 6: glycerol; 7: glutamate; 8: isovalerate; 9: succinate; 10: threonine; 11: caprylate; 12: bile acid; 13: nicotinate; 14: formate; 15: hypoxanthine; 16: urocanate; 17: uracil; 18: phenylalanine; 19: phenylacetate; 20: tyrosine; 22: maltose; 23: glucose; 24: N-acetylglucosamine; 25: malate; 26: threonine; 27: proline; 28: serine; 29: valine; 30: citrulline; 31: ornithine; 32: lysine; 33: pipecolic acid; 34: 3-phenylpropionate; 35: aspartate; 36: methionine sulfoxide; 37: methionine; 38: methylamine; 39: glutamine; 40: succinate; 41: isobutyrate; 42: leucine; 43: isoleucine; 44: Gabapentin.

Comprehensive lists of confidently annotated metabolites identified using both 1D and 2D NMR spectra in the RM 8048 polar extracts are provided in Appendix G "Detailed results for NMR metabolomics" (Appendix G Tables G.1.1. for omnivore and G.1.2. for vegetarian). A total of 81 metabolites were confidently annotated in both omnivore and vegetarian RM 8048. The two lists include the InChIKey associated with each common metabolite name, as well as ^{1}H and ^{13}C chemical shifts from 1D and 2D NMR experiments, and ^{1}H signal multiplicities.

6.1.2.3. Stability

Metabolomics Stability Analysis. RM 8048 omnivore and vegetarian materials stored at -80 °C clearly show time-dependent changes in metabolite profiles over a 13-month time period (T0-T6) from August 2nd, 2023 to September 16th, 2024 as shown in the corresponding PLSDA score plots (Figure 6-9 and Figure 6-10).

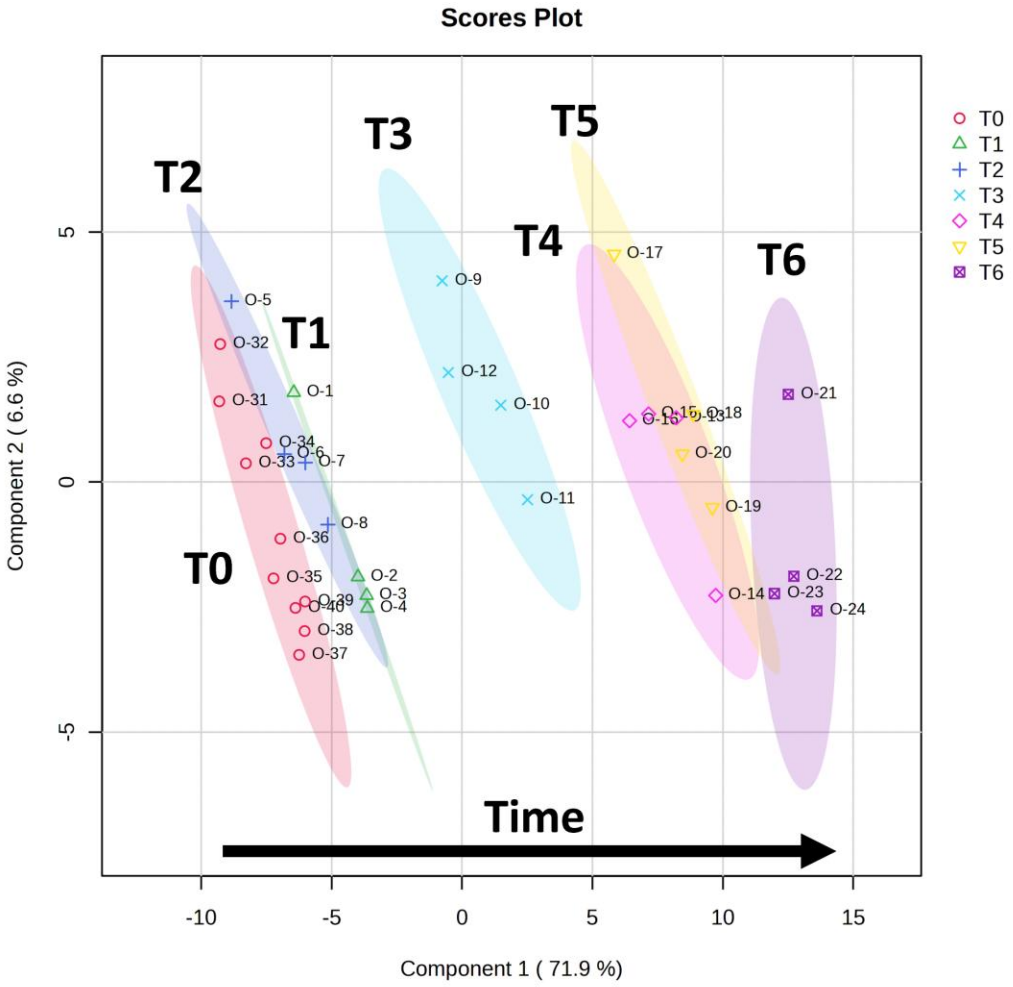


Figure 6-9: PLSDA scores plot of ^{1}H NMR spectra for RM 8048 omnivore polar metabolite extracts over 13 months of storage at -80 °C from T0 through T6.
N = 10 for T0, and N = 4 for each subsequent time point. Ellipses represent 95 % confidence regions for each time point.

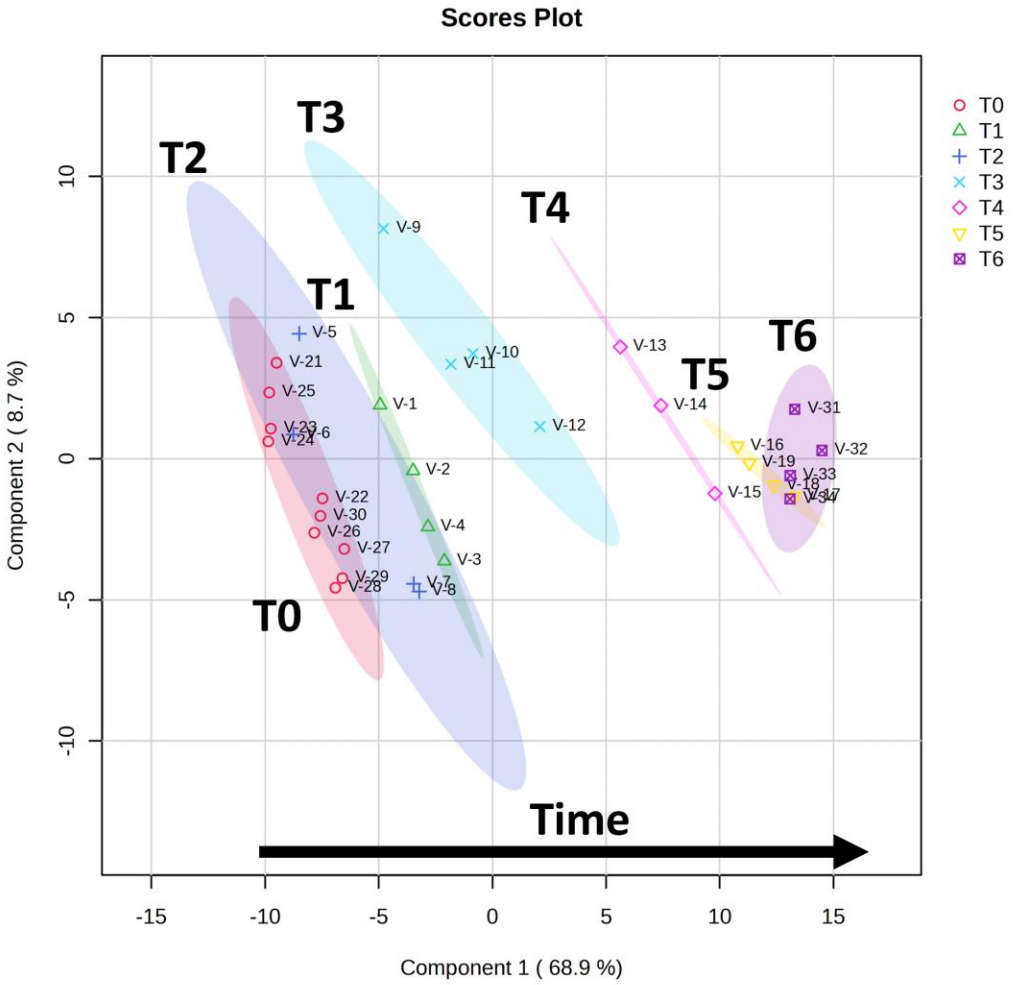


Figure 6-10: PLSDA scores plot of ^{1}H NMR spectra for RM 8048 vegetarian polar metabolite extracts over 13-months of storage at -80 °C from T0 through T6.
N = 10 for T0, and N = 4 for each subsequent time point, except T4 (N = 3). Ellipses represent 95 % confidence regions for each time point.

A total of 19 metabolites (omnivore) and 20 metabolites (vegetarian) were identified as showing significant time-dependent variation over 13-months of storage at -80 °C from T0 through T6 (Table 6-3 and Table 6-4). Seventeen of these metabolites were common to both materials. Metabolites include the following chemical classes: bile acids, short chain fatty acids, carbohydrates and polyols, amino acids and derivatives, organic acids, dicarboxylates, and drugs.

Table 6-3: List of 19 metabolites in the omnivore material with significant time-dependent changes over 13-months of storage at -80 °C.
Peak chemical shifts and corresponding VIP scores from PLSDA are shown, along with an upward/downward arrow indicating increasing/decreasing levels of the corresponding metabolite over time.

Peak (ppm)	VIP Score	Change	Metabolite
1.29	4.0217	↓	Bile acid
0.91	3.6414	↓	Isovalerate
0.90	3.3822	↓	Butyrate
1.06	2.7699	↓	Isobutyrate
1.07	2.7352	↓	Propionate
2.76	2.3952	↑	Methionine sulfoxide
2.17	2.3550	↓	Valerate
2.14	2.0051	↓	Methionine
3.89	1.8843	↑	Glucose
3.03	1.8367	↑	Lysine
3.66	1.7706	↑	Glycerol
3.91	1.4080	↑	Aspartate
3.95	1.3987	↑	Serine
3.59	1.3767	↑	Threonine
3.76	1.3741	↑	Glutamate
3.56	1.2737	↑	Glycine
3.62	1.2360	↑	Valine
4.00	1.1506	↑	Phenylalanine
3.07	1.0165	↑	Tyrosine

Table 6-4: List of 20 metabolites in the vegetarian material with significant time-dependent changes over 13-months of storage at -80 °C.
Peak chemical shifts and corresponding VIP scores from PLSDA are shown, along with an upward/downward indicating increasing/decreasing levels of the corresponding metabolite over time.

Peak (ppm)	VIP Score	Change	Metabolite
0.90	3.7046	↓	Butyrate
1.50	3.6169	↑	Gabapentin
1.06	3.5512	↓	Propionate
1.30	3.1152	↓	Valerate
0.92	2.6269	↓	Isovalerate
0.71	2.5630	↓	Bile acid
2.75	2.5226	↑	Methionine sulfoxide
3.49	2.1777	↑	Glucose
2.14	2.1010	↓	Methionine
3.03	2.0613	↑	Lysine
1.05	2.0309	↑	Isobutyrate
3.65	1.6426	↑	Glycerol
3.76	1.4682	↑	Glutamate
3.41	1.2130	↑	Proline
3.95	1.1741	↑	Serine
3.56	1.1726	↑	Glycine
2.41	1.1701	↑	Succinate
3.91	1.1317	↑	Aspartate
3.62	1.1007	↑	Valine
3.59	1.0158	↑	Threonine

Box plots derived from normalized bin intensities displayed for individual metabolites show significant differences among samples over time based on one-way ANOVA (Figure 6-11 and Figure 6-12).

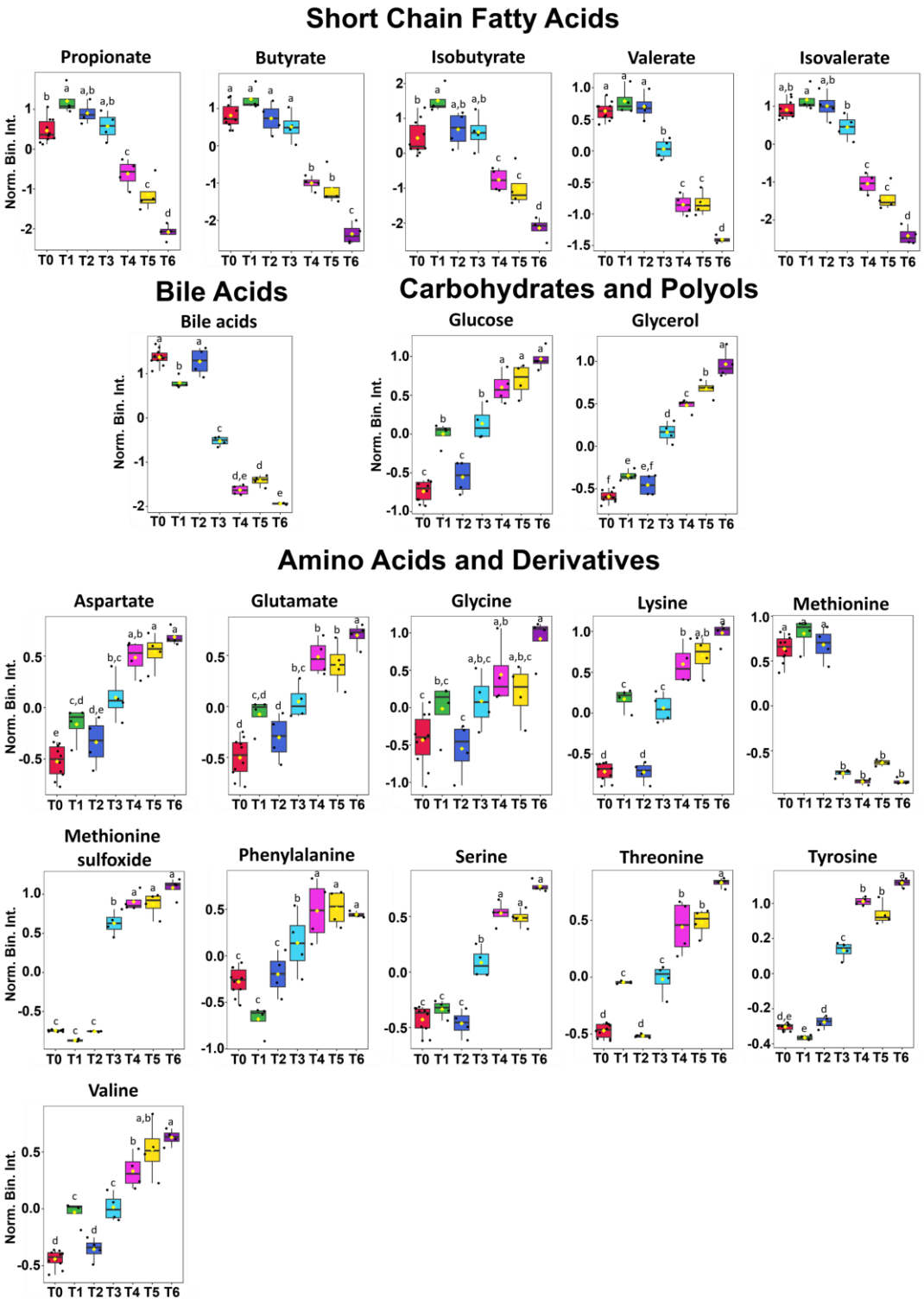


Figure 6-11: Box plots of relative metabolite levels over 13 months of storage at -80 °C from T0 through T6 for the 19 significantly changing metabolites in RM 8048 omnivore samples.
The *x*-axis lists the time points, and the *y*-axis provides the normalized bin intensity. Relative metabolite levels were estimated using normalized spectral bin densities. Different lowercase letters within the plots indicate significantly different time points (ANOVA, Tukey-Kramer, $p < 0.05$).

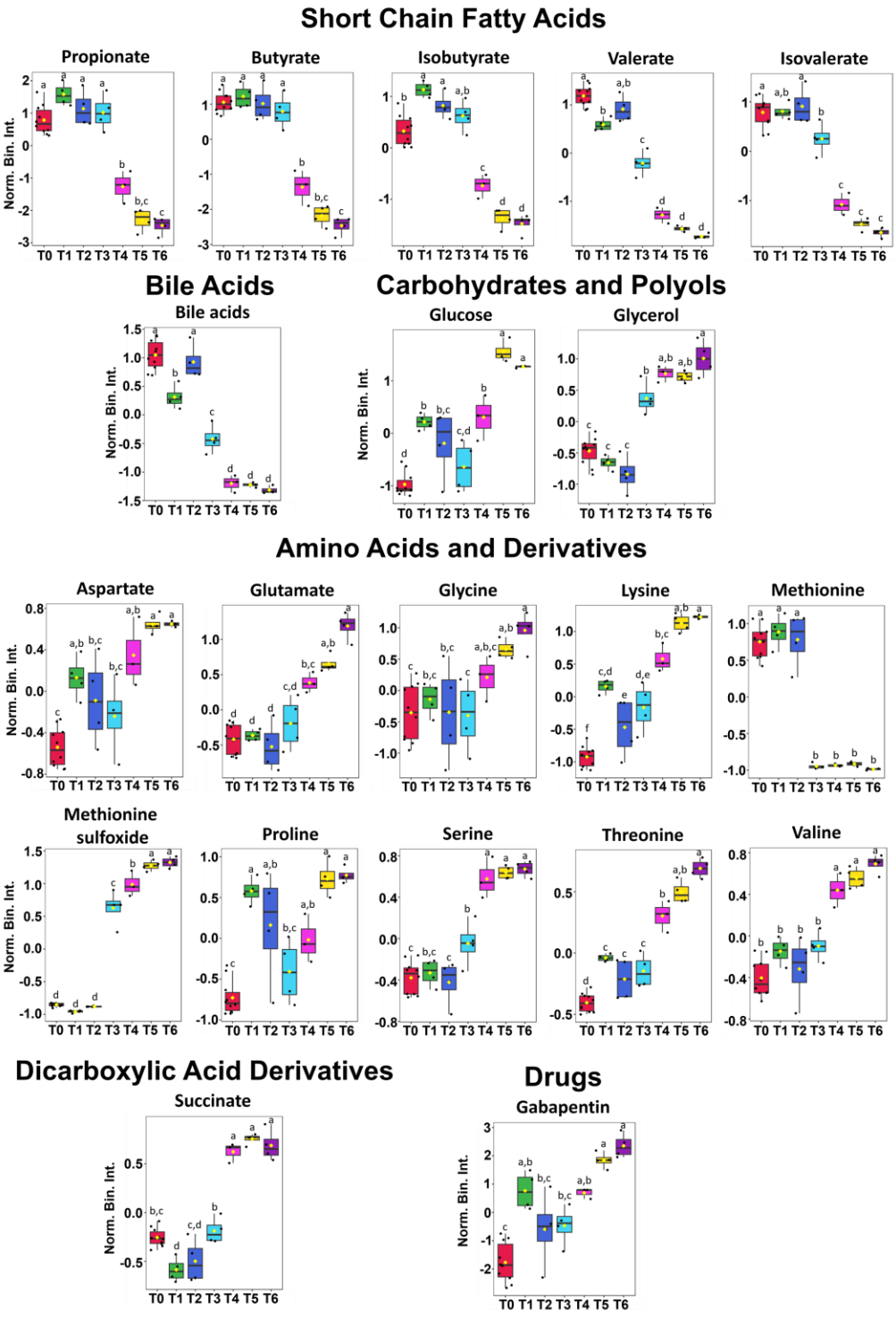


Figure 6-12: Box plots of relative metabolite levels over 13-months of storage at -80 °C from T0 through T6 for the 20 significantly changing metabolites in RM 8048 vegetarian samples.
The *x*-axis lists the time points, and the *y*-axis provides the normalized bin intensity. Relative metabolite levels were estimated using normalized spectral bin densities. Different lowercase letters within the plots indicate significantly different time points (ANOVA, Tukey-Kramer, $p<0.05$).

Out of the hundreds of metabolites expected to be present in a human stool sample, only a small fraction (≈ 20) has been identified by NMR to show time-dependent changes during storage at -80 °C. Importantly, metabolites that decreased over time were still detectable by NMR after ≈13 months from the initial time point (T0), with some reaching stability within ≈ 6 months from T0.

Measuring time-dependent trends of extracted polar metabolite ^{1}H NMR spectra helped identify specific metabolites and chemical classes (e.g., short chain fatty acids, bile acids, amino acids) that displayed different degrees of stability within the complex matrix that characterizes a stool material during storage at -80 °C. Additionally, stability differences for the same metabolite but in the two different materials (omnivore and vegetarian) highlight the importance of matrix composition on the overall stability of individual compounds.

6.1.2.3.1. Omnivore COD

CODs calculated for the stability data within the omnivore donor pool (34 samples across seven time points, T0-T6) are shown in the point plot in Figure 6-13. Each sample is identified by its label on the *x*-axis (e.g., sample from box 1, time zero (T0), labeled “B01T0”). Each sample along the *x*-axis has 34 points corresponding to the comparisons with other samples. The points are colored by time point, so that a red point for T0 signifies a comparison to another vial measured at T0, while a magenta point signifies a comparison to the vials measured at T6. In this way, we can check for any systematic trends over time. From the plot, we conclude that samples at T5 are those with the highest discrepancies both intra- and inter-time points. These discrepancies were not observed in the case of the vegetarian samples, which suggests that the source of this variation was not associated with the methods.

With the inclusion of T5 samples, ≤ 23 % of the peaks differ between any pair of omnivore samples (an upper 95 % confidence limit of 22.880 % using a percentile bootstrap with 10,000 replicates) over 13-months of storage at -80 °C. However, with the exclusion of T5 samples (Figure 6-14), a lower degree of variation is observed, and ≤ 15 % of the peaks differ between any pair of omnivore samples (an upper 95 % confidence limit of 14.350 % using a percentile bootstrap with 10,000 replicates) over 13-months of storage at -80 °C, similar to the variation observed in the vegetarian donor pool.

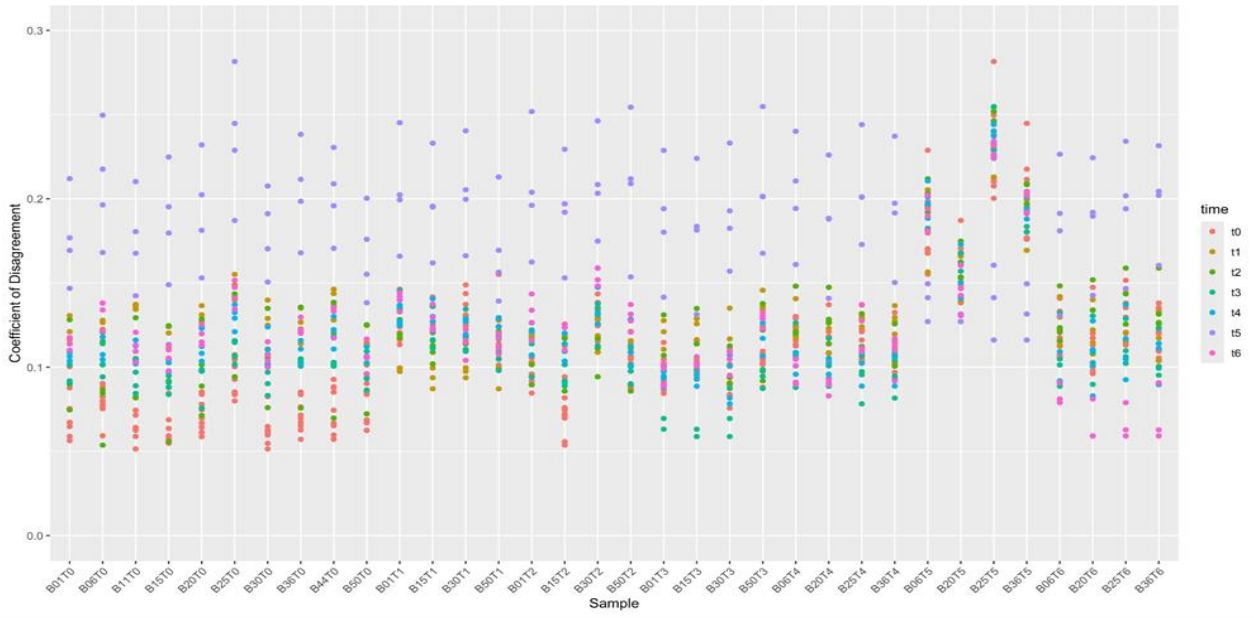


Figure 6-13: COD point plot of ^{1}H NMR stability data for omnivore samples over 13 months of storage at -80 °C (T5 included). Points are colored by the time point to which the sample on the *x*-axis was compared.

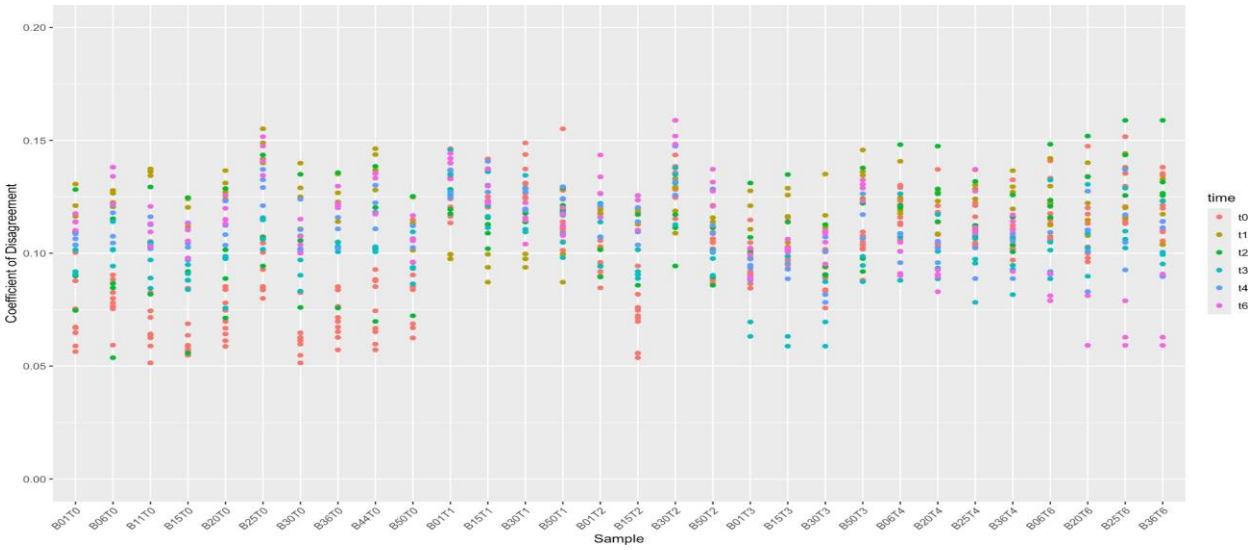


Figure 6-14: COD point plot of stability data for omnivore samples over 13 months of storage at -80 °C (T5 excluded). Points are colored by the time point to which the sample on the *x*-axis was compared.

6.1.2.3.2. Vegetarian COD

Figure 6-15 shows the COD point plot for the stability data within the vegetarian donor pool (34 samples across seven time points, T0-T6). The plot is labeled and colored similarly to the omnivore samples. No significant dissimilarities were observed for the T5 samples, in this case; however, the sample from box 1, time point 1 (B01T1) shows a larger discrepancy from all other

samples including other samples from T1. From these data, and including B01T1, we conclude that ≤ 20 % of the peaks differ between any pair of vegetarian samples (an upper 95 % confidence limit of 16.320 % using a percentile bootstrap with 10,000 replicates) over 13-months of storage at -80 °C from T0 through T6.

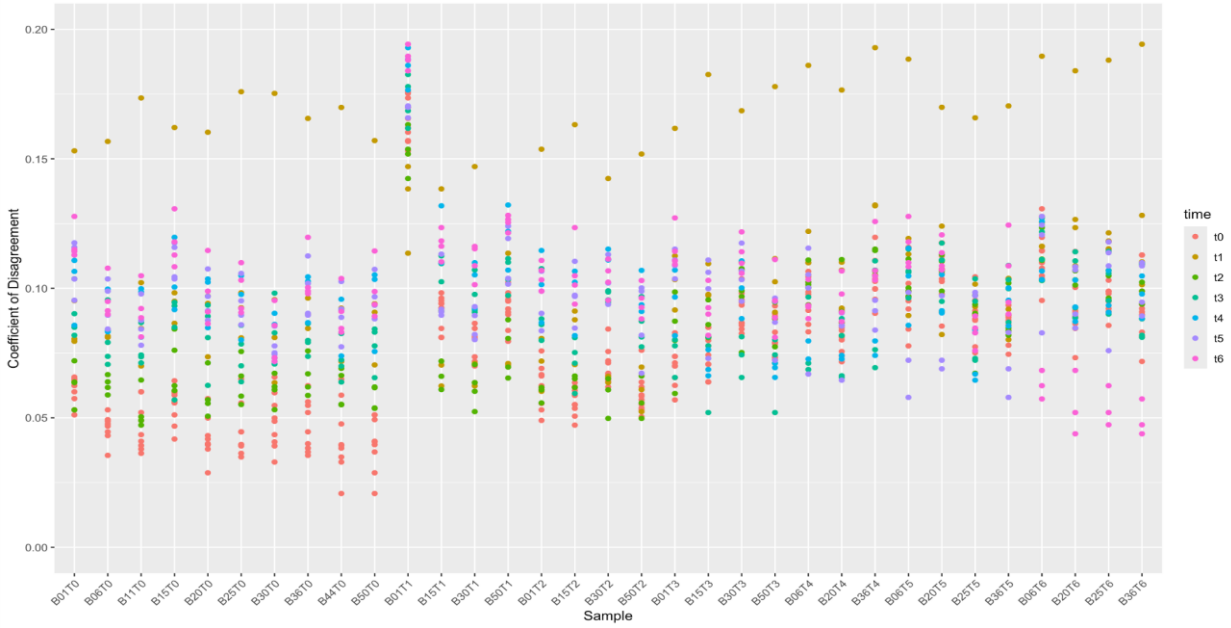


Figure 6-15: COD point plot of ^{1}H NMR stability data for vegetarian samples over 13-months of storage at -80 °C. Points are colored by time point.

6.1.3. Summary

Based on the COD analysis, RM 8048 is deemed sufficiently homogeneous and stable for both omnivore and vegetarian donor pools for use in untargeted metabolomics workflows. A total of 81 metabolites were confidently annotated in both omnivore and vegetarian parts of RM 8048 using 1D (^{1}H) and 2D NMR experiments (^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY).

6.2. Aqueous Extraction (Fecal water) – Polar Metabolites - 600 MHz

6.2.1. Methods

D_2O-DSS phosphate buffer: Prepared by mixing 0.5323 g of anhydrous disodium hydrogenate phosphate (Na_2HPO_4, M.W. 141.96 g/mol, 75 mmol/L), 0.1229 g of monosodium phosphate (NaH_2PO_4, M.W. 119.98 g/mol, 20.5 mmol/L), 0.03033 g of sodium 2,2-dimethyl-2-silapentane-5-sulfonate-D_6 (DSS-D_6, M.W. 224.4 g/mol, 5 mmol/L), and 0.2 g of sodium azide (NaN_3, 0.4 %

w/v) into 50 mL of D_2O. The solution (pH = 7.2) was aliquoted into 1 mL high performance liquid chromatography (HPLC) vials and stored at -20 °C.

10x phosphate buffer: Prepared by mixing 5.325 g of Na_2HPO_4 (750 mmol/L) and 1.2298 g of NaH_2PO_4 (250 mmol/L) into 50 mL of HPLC-grade water.

10x sodium azide: Prepared by dissolving 2 g of NaN_3 in HPLC-grade water.

Experimental Design: A total of 22 vials of each RM 8048 donor pool were used to characterize the materials. Four vials were used to assess homogeneity, and all 22 vials were used to assess the storage stability of metabolites (Table 6-5). The stability was monitored at five time points across 274 days. Additionally, six vials from a BioIVT pilot batch were used as experimental control. The control samples were prepared by mixing ten vials from a pilot batch of stool material (BioIVT), splitting them into ten vials, and storing them at -80 °C. In each experiment, a control sample was included.

Table 6-5: Vial selection and time points (in days) for homogeneity and stability analysis of ^1H NMR spectra from an aqueous extraction (fecal water) protocol.

Time Point	Days Elapsed	Vial	Time Point	Days Elapsed	Vial
1	0	omni_15-1	1	0	veg_15-1
1	0	omni_25-1	1	0	veg_25-1
1	0	omni_36-1	1	0	veg_36-1
1	0	omni_6-1	1	0	veg_6-1
2	78	omni_15-2	2	79	veg_15-2
2	78	omni_25-2	2	79	veg_25-2
2	78	omni_36-2	2	79	veg_36-2
2	78	omni_6-2	2	79	veg_6-2
3	105	omni-15-3	3	105	veg-15-3
3	105	omni-25-3	3	105	veg-25-3
3	105	omni-36-3	3	105	veg-36-3
3	105	omni-6-3	3	105	veg-6-3
4	160	omni-15-4	4	160	veg-15-4
4	160	omni-25-4	4	160	veg-25-4
4	160	omni-36-4	4	160	veg-36-4
4	160	omni-6-4	4	160	veg-6-4
5	274	omni-1-3	5	280	veg-1-3
5	274	omni-1-4	5	280	veg-1-4
5	274	omni-15-5	5	280	veg-15-5
5	274	omni-25-5	5	280	veg-25-5
5	274	omni-44-3	5	280	veg-44-3
5	274	omni-44-4	5	280	veg-44-4

Extraction Protocol for Polar Metabolites for Homogeneity and Stability: Each day of an experiment, four stool samples plus one control sample were analyzed. Samples were removed intermittently from -80 °C storage such that the next sample would be thawed and ready for processing when the previous sample was placed in the NMR for analysis. Vials were thawed at 4 °C for 1.5 h after removal from -80 °C storage. Thawed samples were vortexed for 15 s at high speed, transferred to a 1.5 mL microfuge tube, and centrifuged for 30 min at 14,000 × *g* and 4 °C. A 500 µL aliquot of the supernatant, 50 µL of cold phosphate buffer, and 50 µL of cold sodium azide were weighed into a 1.5 mL microfuge tube and vortexed for 15 s. Gravimetrically, 500 µL aliquot of that mixture plus 50 µL of D_2O/DSS (Stock = 5 mmol/L) were combined in a microcentrifuge tube, vortexed for 15 s, and transferred to a 5 mm NMR tube for analysis. For omnivore samples, the reagent and extract masses ranged from 0.93 % to 2.73 %, while for vegetarian samples, the range was from 0.84 % to 2.05 % (Appendix G Table G.2.1 and Table G.2.2). Although the masses were recorded, they were not used in further analyses but were used to monitor the mass consistency throughout the sample preparation.

NMR Analysis and Processing. ^{1}H NMR spectra were acquired at 298 K on a Bruker Avance II 600 MHz NMR spectrometer equipped with a room-temperature broadband inverse (BBI) probe. ^{1}H spectra were obtained using a nuclear Overhauser effect spectroscopy (NOESY) water presaturation sequence (Bruker sequence noesypr1d) with the following acquisition parameters: 8403.361 Hz (14.0025 ppm) sweep width, 32,768 complex data points, 3.899 second acquisition time, 4-second delay (D1), 256 scans (NS), 8 dummy scans (DS), 2822.41 Hz (4.703 ppm) transmitter offset (O1). The total acquisition time per spectrum was approximately 34 min . The transmitter offset had previously been adjusted for optimal water suppression with fecal water.

The ^{1}H NMR spectra were processed using MATLAB (R2023b, MathWorks). Line shape correction was performed using the Reference Deconvolution (RD) method available in the General NMR Analysis Toolbox (GNAT) for MATLAB[15]. For this, FIDs were imported and processed in GNAT which included zero-filling (SI = 2 x TD), apodization (1.5 Hz Lorentzian line shape), Fourier transformation, and manual phase correction. Spectra were referenced to the DSS-d6 internal standard at 0.00 ppm for ^{1}H with RD. In addition, a peak alignment procedure was applied using an interval-based spectral shifting algorithm (icoshift)[16] that uses correlation to optimize alignment. The chemical shift axis was segmented into 212 spectral intervals for alignment that were manually selected based on the examination of the ^{1}H NMR spectra from the two RM 8048 samples. Due to differences in the spectra from the two donor pools, global alignment introduced undesirable artifacts in certain regions. Consequently, the alignment procedure was performed on spectra from the same sample group, i.e., the omnivore and vegetarian data sets were aligned independently to their respective median spectra. The median of each respective dataset was used as the alignment reference in each case.

Statistical Analysis: To assess homogeneity, the COD method described previously (Appendix I section 3) was used to perform pairwise comparisons of the NMR spectra. The

[15] Castañar L, Dal Poggetto G, Colbourne AA, Morris GA, Nilsson M. The GNAT: A new tool for processing NMR data. *Magn Reson Chem*. 2018; 56: 546–558. https://doi.org/10.1002/mrc.4717
[16] Savorani F, Tomasi G, Engelsen SB (2010) icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 202(2):190-202. https://doi.org/10.1016/j.jmr.2009.11.012

spectral regions below 0.5 ppm (including the DSS-d6 peak), between 4.67 ppm and 4.9 ppm (residual water signal), and above 9.5 ppm (no observable peaks) were removed from the data set prior to statistical analysis. Following this protocol, twenty-two omnivore vials and twenty-two vegetarian vials were analyzed. The data collected from these vials were used to assess whether two randomly selected vials measured at two different time points agreed. The analysis provided insight into the stability of the metabolites. Additionally, four vials out of 22 were used to examine homogeneity (Table 6-5).

Filtration Protocol used for Metabolite ID: This protocol used 3 kDa filters which contain glycerin as a wetting agent to maintain the integrity of the filter. Before use, the filters were rinsed three times by adding 500 µL of HPLC-grade water each time, centrifuging for 30 min at 13,000 x *g*, and placing them in a 2 L beaker filled with distilled water. The water was stirred at low speed on a magnetic stirrer for six h. The water was then replaced with another 2 L of distilled water, and the tubes were stirred at low speed overnight (approximately 16 h). Before use, the tubes were shaken to remove excess water.

Two samples (omni_50-1 and veg_50-1) were removed from -80 °C storage and placed in a refrigerator (4 °C) to thaw. Thawed samples were kept on ice until processed. Samples were first vortexed for 15 s at high speed, transferred to 1.5 mL centrifuge tubes, and centrifuged for 30 min at 14,000 × *g* at 4 °C. An aliquot of 800 µL of fecal water was transferred and weighed into a 1.5 mL microfuge tube along with 100 µL of cold 10x phosphate buffer (100 mmol/L) and 100 µL of cold 10 x sodium azide (0.4 %, 6.15 mmol/L), and vortexed. The mixture was then transferred to a 0.1 µm centrifugal filter and centrifuged for 60 min at 10,000 × g at 4 °C. The omnivore samples required two filtration steps with the 0.1 µm filter, each allowing approximately half the volume to pass through before the filter clogged. The filtrate was transferred to a rinsed 3 kDa filter tube, vortexed, and centrifuged for 60 min at 10,000 × g at 4 °C. For omnivore samples, two 3 kDa filters were required to collect enough volume for the next step. After filtration, 500 µL of weighed filtrate and 50 µL of weighed D_2O-DSS (Stock = 5 mmol/L) were combined in a 1.5 mL microfuge tube, vortexed for 15 s, and transferred to a 5 mm NMR tube for analysis with a 900 MHz NMR.

NMR Analysis and Processing. To aid in metabolite identification, multiplicity edited 2D ^{1}H-^{13}C HSQC spectra (Bruker sequence hsqcedetgpsisp2.3) were acquired on a Bruker Avance III 900 MHz spectrometer equipped with a triple resonance cryogenically-cooled TCI probe. Acquisition parameters were: 88 scans in 512 increments, 32 dummy scans, 2 second delay, 14,423 Hz (16 ppm) spectral width in F2, 37,350 Hz (165 ppm) spectral width in F1, 2018 complex data points in F2, and acquisition times of 140 ms and 6.85 ms in the direct (F2) and indirect (F1) dimensions, respectively. Spectra were referenced to the DSS-d6 internal standard at 0.00 ppm for ^{1}H and ^{13}C.

Metabolite Identification. For metabolite identification, deconvolution was carried out using the Chenomx NMR Suite (V10.1, Chenomx Inc., Edmonton, Canada). The DSS concentration served as an internal standard to determine metabolite concentrations. In the processor module, the FID spectra were automatically phased, and baseline and line broadening (1.0 Hz) were applied. The pH was adjusted to between 7.0 and 7.5, and the DSS

was calibrated to the Chemical Shift Indicator (CSI) for 0.54073 mmol/L. Metabolite assignments by two dimensional NMR ^{1}H-^{13}C HSQC spectra were based on chemical shift comparisons using reference spectra from the Human Metabolome Database (HMDB)[17] the Biological Magnetic Resonance Bank (BMRB)[18], the spectral library in Chenomx NMR Analysis software (Chenomx Inc. Version 8.5), as well as an in-house database.

6.2.2. Results

An aqueous extraction protocol was used to identify polar metabolites in RM 8048. The common and simple protocol was used to minimize sample variability during preparation. Additionally, each vial was allowed to fully thaw at 4 °C for the same amount of time (1 h) after removal from -80 °C. Once thawed, each vial was processed and analyzed individually before the next vial was processed.

The COD was computed between all pairs of vials regardless of being from the omnivore or vegetarian donor pool. Forty-four vials were analyzed in total, as shown in Figure 6-16. This analysis results in a 44 x 44 matrix with the diagonal element of the matrix as zero because the disagreement between a sample and itself is zero. The larger COD values come from a comparison of an omnivore sample with a vegetarian sample. For instance, the very first vial (o25.1) along the horizontal axis is an omnivore sample. When compared to other omnivore samples, for red points along the vertical corresponding to o25.1, the values remain below 0.15 on the vertical scale. However, COD values as compared to vegetarian samples (green) are all above 0.15 on the vertical scale (Figure 6-16). The omnivore material differs greatly from the vegetarian material, as evident by a big gap between the lower set of points (*y*-axis values below 0.15) and the upper set of points (*y*-axis values above 0.15).

[17] Wishart DS, Guo AC, Oler E, et al., *HMDB 5.0: the Human Metabolome Database for 2022.* Nucleic Acids Res. 2022. Jan 7;50(D1):D622–31
[18] "Biological Magnetic Resonance Data Bank", Jeffrey C Hoch; Kumaran Baskaran; Harrison Burr; John Chin; Hamid R Eghbalnia; Toshimichi Fujiwara; Michael R Gryk; Takeshi Iwata; Chojiro Kojima; Genji Kurisu; Dmitri Maziuk; Yohei Miyanoiri; Jonathan R Wedell; Colin Wilburn; Hongyang Yao; Masashi Yokochi; **Nucleic Acids Research**, Volume 51, Issue D1, 6 January 2023, Pages D368–D376 doi: 10.1093/nar/gkac1050

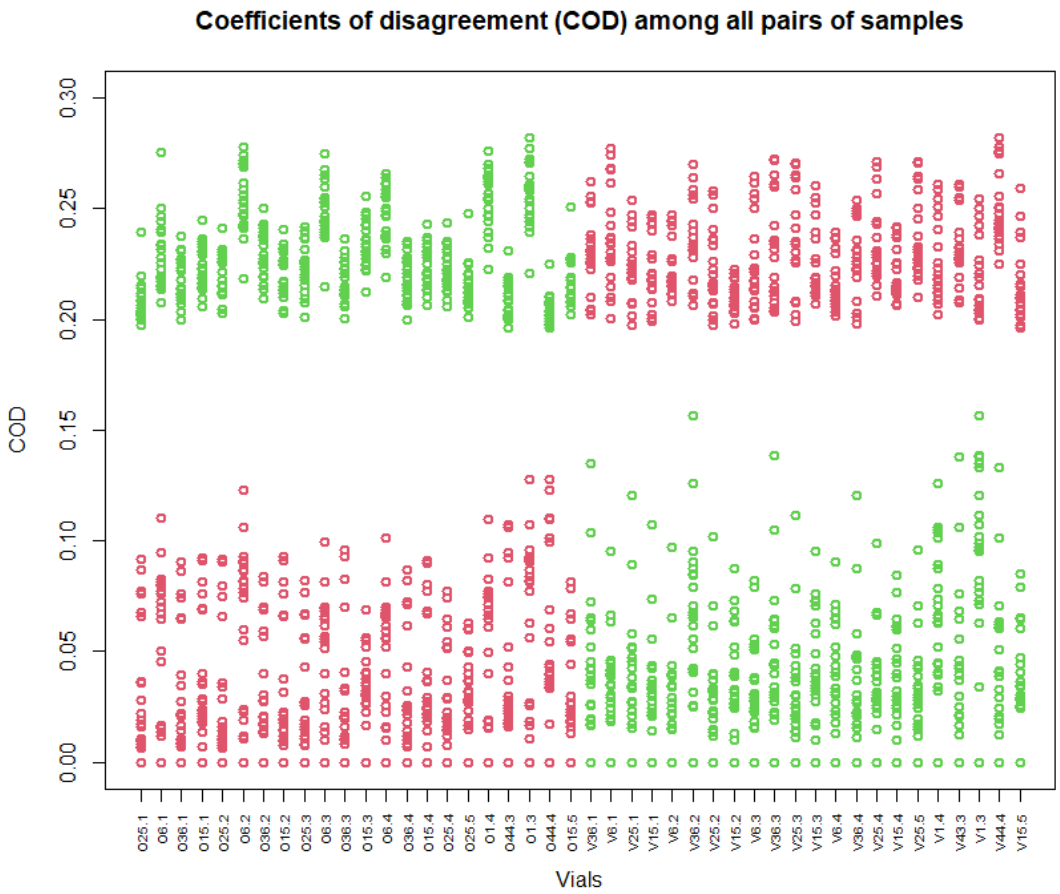


Figure 6-16: COD for each ^{1}H NMR sample when compared to all other samples of the unfiltered aqueous extract. Red indicates omnivore while green indicates vegetarian.

Further analysis looked at the COD values separately for each donor pool (i.e., twenty-two vials). The 22 vials analyzed per donor pool represent the entire collection of samples analyzed over a nine-month time frame. The 22 by 22 matrix of COD values among omnivore samples and vegetarian samples is shown in Appendix G G.2.4 and G.2.5.

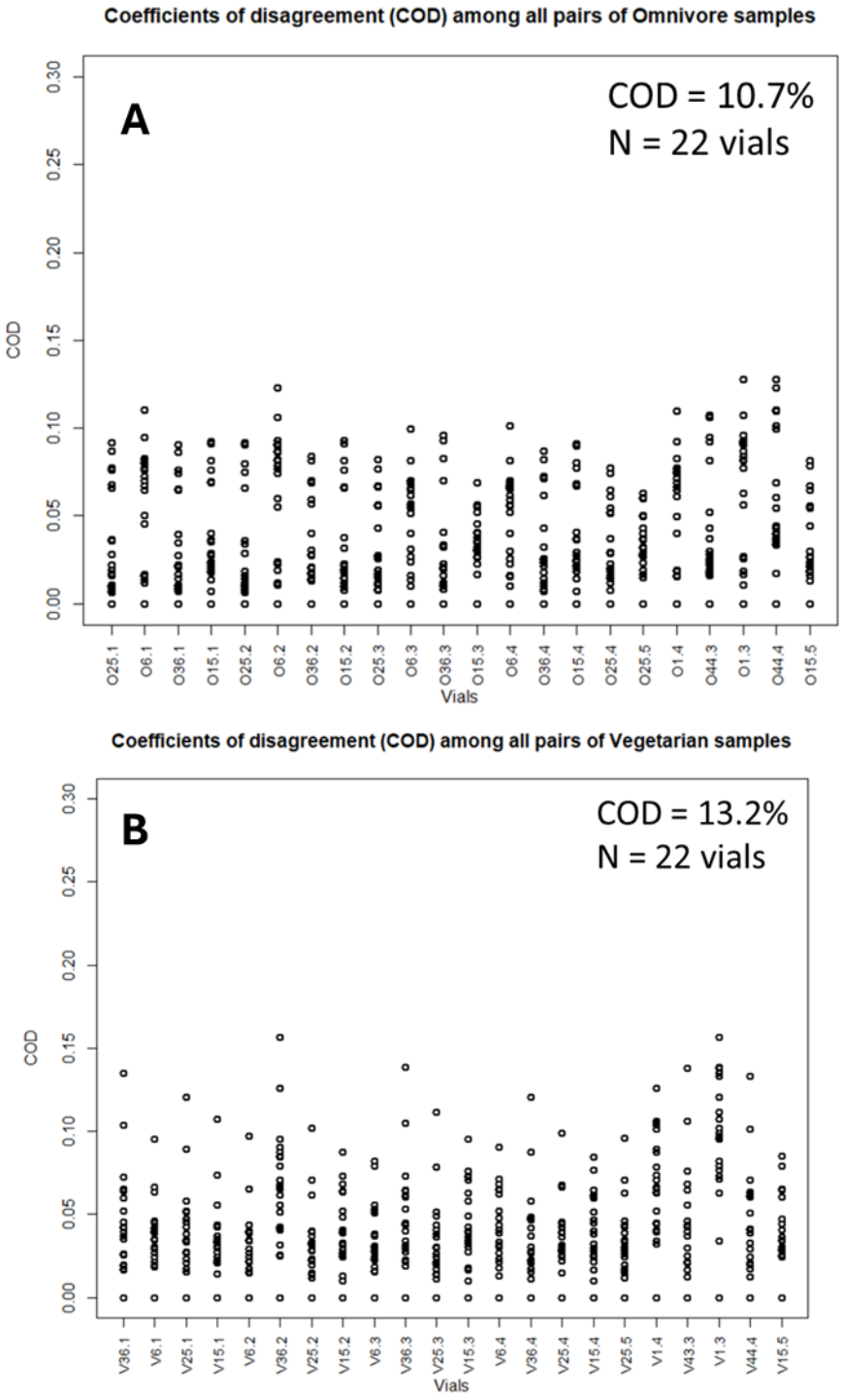


Figure 6-17: Visualization of the COD (*y*-axis) for each ^{1}H NMR spectra from omnivore (A) and vegetarian (B) samples in relation to all other omnivore and vegetarian samples, respectively, as labeled along the *x*-axis.

The analysis shows that we can be approximately 95 % confident that the difference in peak intensity of the NMR spectra for both omnivore and vegetarian vials, measured on the same

day or at different times within 280 days, will be at most 10.69 % for omnivore vials and 13.23 % for vegetarian vials (Figure 6-17). These findings quantify the homogeneity and stability of the samples in relation to their NMR spectra using the unfiltered protocol to extract hydrophilic metabolites.

Given the differences between the two donor pools, we have listed confident metabolite annotations for each donor pool in tables found in Appendix G Tables G.2.6 and G.2.7. We annotated 49 metabolites in the vegetarian samples and 44 metabolites in the omnivore samples using Chenomx and further confirmed the identity with ^{1}H ^{13}C HSQC. Desaminotyrosine was only identified in the omnivore donor pool, while gabapentin, lactate, and sarcosine were identified in the vegetarian donor pool. It should be noted that compounds detected in only one donor pool may be present in the other donor pool but at concentrations too low to be detected by ^{1}H NMR. Several peaks in the ^{1}H NMR spectra could not be identified using Chenomx nor HMBD libraries, possibly because the databases do not have the NMR spectra for all chemical standards encompassing the chemical constituents in a stool sample, or the concentration of the compounds cannot be detected by NMR analysis. As a result, we focused on the confidently identifiable peaks assigned with the 2D NMR analysis.

6.2.3. Summary

Using the unfiltered process, we characterized the metabolites present in RM 8048 and determined that the material is fit for purpose. Results from the omnivore and vegetarian materials analyzed over a nine-month period indicated a high degree of confidence in the homogeneity and stability of the samples using the unfiltered extraction protocol. The difference in NMR peak intensities across omnivore and vegetarian samples measured on the same day or at different time points was less than 15 %. We confidently annotated 50 metabolites in the omnivore samples and 51 metabolites in the vegetarian samples. Overall, the characterization of RM 8048 using a simple aqueous extraction protocol to assess polar metabolites highlighted metabolic distinctions between the two donor pools and suitable homogeneity and stability.

7. Flow Cytometry Enumeration and Intensity Measurements

Flow cytometers analyze particles suspended in liquid using multiple lasers and detectors designed to measure scattered light and emitted fluorescence. Flow cytometry (FCM) has recently been applied to the characterization of complex microbial communities, such as stool, though these methods vary widely between experimenters. This technique was used to quantify (1) total number of cells per vial and (2) intensity distributions from scattered light and fluorescent emissions from added probes, e.g., for nucleic acid content.

7.1. Methods

Experimental Design

Homogeneity experiments were carried out on two days, one per donor pool, and consisted of eight vials of material from different box numbers (1, 6, 15, 25, 30, 36, 44). Stability experiments used one to two vials from each donor pool on each date selected from different box numbers. Five stability experiments were conducted over six months (August 2023 to January 2024), noting that the third stability timepoint (October) was not completed.

Several different types of control samples were used. Extra material from each vial used in the homogeneity experiments and two entire vials of material (from boxes 20, 50) were combined to generate donor pool-specific pooled control materials. Aliquots of the pooled material were stored in the freezer and analyzed in each stability experiment. Buffer was collected as a cell-free control. QC beads were used to qualify the instrument before collecting data. Beads with multiple known fluorescent intensities were used to relate the arbitrary fluorescence intensities to standardized units. Beads with a known particle concentration were used to estimate the volumes that were acquired for each sample.

Sample Handling

Vials of material (biological replicates = box replicates) were removed from the -80 °C freezer at the beginning of each experiment. Samples were moved to the prepared biosafety cabinet and allowed to come to room temperature. During the homogeneity experiment for each material a “pooled” sample was generated by combining 250 µL from each biological replicate with an entire vial from Boxes 20 and 50. This pooled sample (≈ 4 mL) was pipetted into 300 µL aliquots which were stored at -80 °C. One pooled aliquot (250 µL) from each donor pool was used at each timepoint in the stability experiments.

Fluorescent Probes

While RM 8048 vials were defrosting, fluorescent probe stock solution was prepared [Syto24 stock (S24) – 5 mmol/L, Invitrogen, Cat# B34956, Lot# 2068240]. CellBriteFix640 stock (CBF640) – Prepared at 1000X solution in DMSO per manufacturer’s instructions [Biotium, Cat# 30089A, Lot# 17C1115-1145191 and 17C1115-1201116]. These probes target nucleic acids and reactive amines, respectively, and were selected to distinguish cells from the background matrix. Phosphate buffered saline [PBS; Gibco, Cat# 14190] was filtered using a 0.22 µm syringe filter, and 5.25 mL PBS was pipetted into a 15 mL centrifuge tube. S24 and CBF640 stock solutions were defrosted, and 2.1 µL of each was added to generate the cell staining solution. The

solution was vortexed, and 360 µL was pipetted into 2 mL microcentrifuge tubes and protected from light.

Sample Filtration and Dilution

Samples were filtered to remove large debris in preparation for flow cytometry. Briefly, RM 8048 samples were mixed, and a 250 µL aliquot of material was pipetted directly onto a cell strainer [Corning, 40 µm Nylon, Ref# 431750] that had been placed in a 50 mL centrifugation tube. The material captured by the filter was rinsed using a serological pipette containing 5 mL of PBS. The process of filtering 250 µL aliquots was repeated three times per biological replicate to generate a total of three separate sample replicates. During the homogeneity assessment, samples were filtered in numerical order (thus confounding the box number with time). This issue was corrected in the stability studies in which biological replicate filtration order was randomized. For both homogeneity and stability studies, the technical replicates from a single sample vial were prepared together (meaning the three replicate filtrations from a single vial were performed before moving onto a new vial).

After all RM 8048 filtrations had been performed, each tube of cell suspension (diluted with the 5 mL PBS rinse) was vortexed, 40 µL was pipetted into a 2 mL microcentrifuge tube containing 360 µL staining solution, and the microcentrifuge tube was vortexed. All tubes were incubated at room temperature in the biosafety cabinet while protected from light for 50 min. After the incubation, the tube was vortexed again, and 200 µL was pipetted into a flow cytometry tube. The volume calibrant bead stock [CountBrightPlus Absolute Count Beads (CBPbeads), Invitrogen, Cat# C36995, Lot# 2707082] was vortexed for at least 30 s, and 20 µL was mixed into each flow tube.

Flow Cytometry

Detailed methods can be found in Appendix H. A MoFlo Astrios EQ cell sorter (Beckman Coulter) was used. Fluorescence standardization was performed using Rainbow Calibration Particles [8peakRCP - SpheroTech, RCP-30-5A, Lot AN04]. Five intensities were evaluated: Forward scatter (FSC1), Side scatter (SSC), Syto24 (nucleic acids, FL11), count calibration beads (FL20), and CellBriteFix640 (cell body, FL25). Data from each sample were acquired for a total of two min or ≈ 490,000 events depending on which occurred first. Each set of RM 8048 files in an experiment was processed as a batch including fluorescence calibration, quality control, clustering, and quantification.

Counts:

The total normalized number of events classified as “cells” for each file was calculated from the sum of the identified metaclusters and exported for statistical analysis (detailed analysis in Appendix HI.3.5). Omnivore and vegetarian data were evaluated separately. Homogeneity was evaluated using the homogeneity dataset that was collected from multiple boxes. Stability was evaluated using only the pooled samples since they were generated during the homogeneity experiments (discussed further in Section 7.2.2.2). For homogeneity, Bayesian models were

used to calculate confidence intervals that encompassed the expected normalized counts for any vial in the donor pool. For stability, pooled data were evaluated to test for the significance of any trend. Trends were not significant.

Intensities:

Raw *.FCS data files were used for statistical analysis of the fluorescent intensities (detailed analysis in Appendix H). It is important to note that (1) these intensity distributions represent all events in the file and not specifically events that were classified as cells, and (2) no fluorescence calibration was applied. For each file, the peak(s) of the intensity distribution in the five channels of interest were identified, and the intensity (*x*-axis) and peak height (*y*-axis) were identified for each peak. For homogeneity, Bayesian models were used to calculate confidence intervals for peak positions and peak heights that encompassed the expected log10-transformed intensities for any vial in the donor pool. For stability, pooled data were evaluated to test for the significance of any trend.

7.2. Results

Results are presented as qualitative interpretations of the raw data (which would be available to the operator during an experiment) and quantitative conclusions for the counts and intensity distributions. Additional results can be found in Appendix H.

7.2.1. Raw Data

7.2.1.1. Fluorescence Repeatability

The on-instrument fluorescent alignment resulted in intensities that were similar across experiments (Figure 7-1). This was expected as the 8-peakRCP were used to set instrument acquisition conditions on each date such that the resulting intensities would align with previous dates. As laser and fluidics alignment is a manual process on this instrument, some deviation was observed.

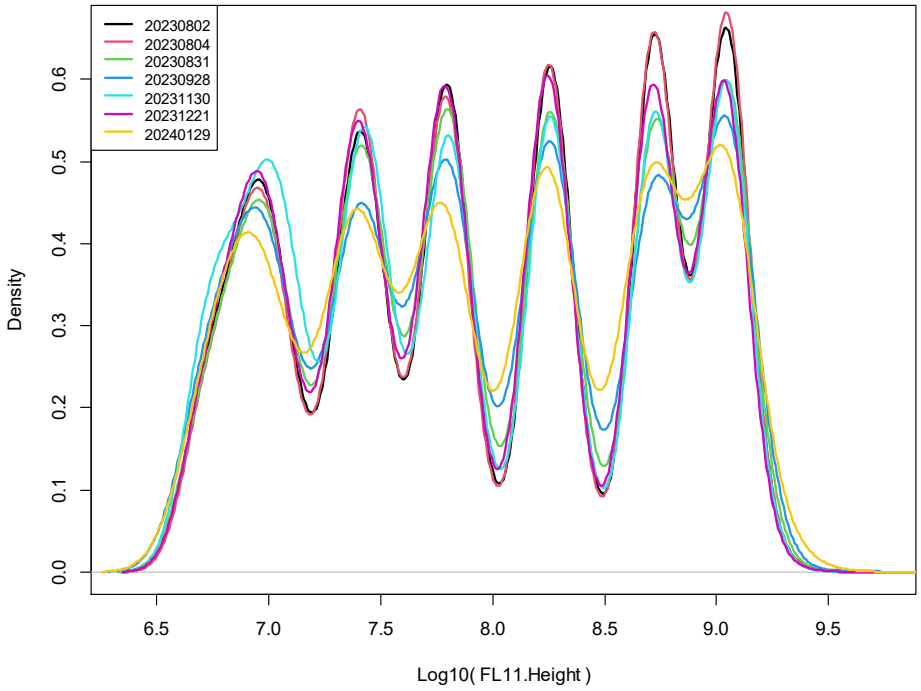


Figure 7-1: Alignment of 8-peakRCP across FCM experiments.
Intensity data from the 8-peakRCP are overlaid to illustrate fluorescent standardization across experiments. The *x*-axis is the log10 intensity of the FL11-Height channel, the *y*-axis is the subpopulation relative frequency, colors correspond to different dates.

7.2.1.2. Distinguishing Omnivore and Vegetarian Samples

The two different donor pools were easily distinguished by eye while running an experiment (Figure 7-2). These differences were maintained across dates such that all samples could manually be classified by donor pool (if this were necessary). Experimenters would likely find this valuable when working with the material. These differences were also apparent in an MDS plot summarizing the peaks in the intensity data (Figure 7-3). However, there were noticeable differences in the staining profiles across dates which is described further in Section 0.

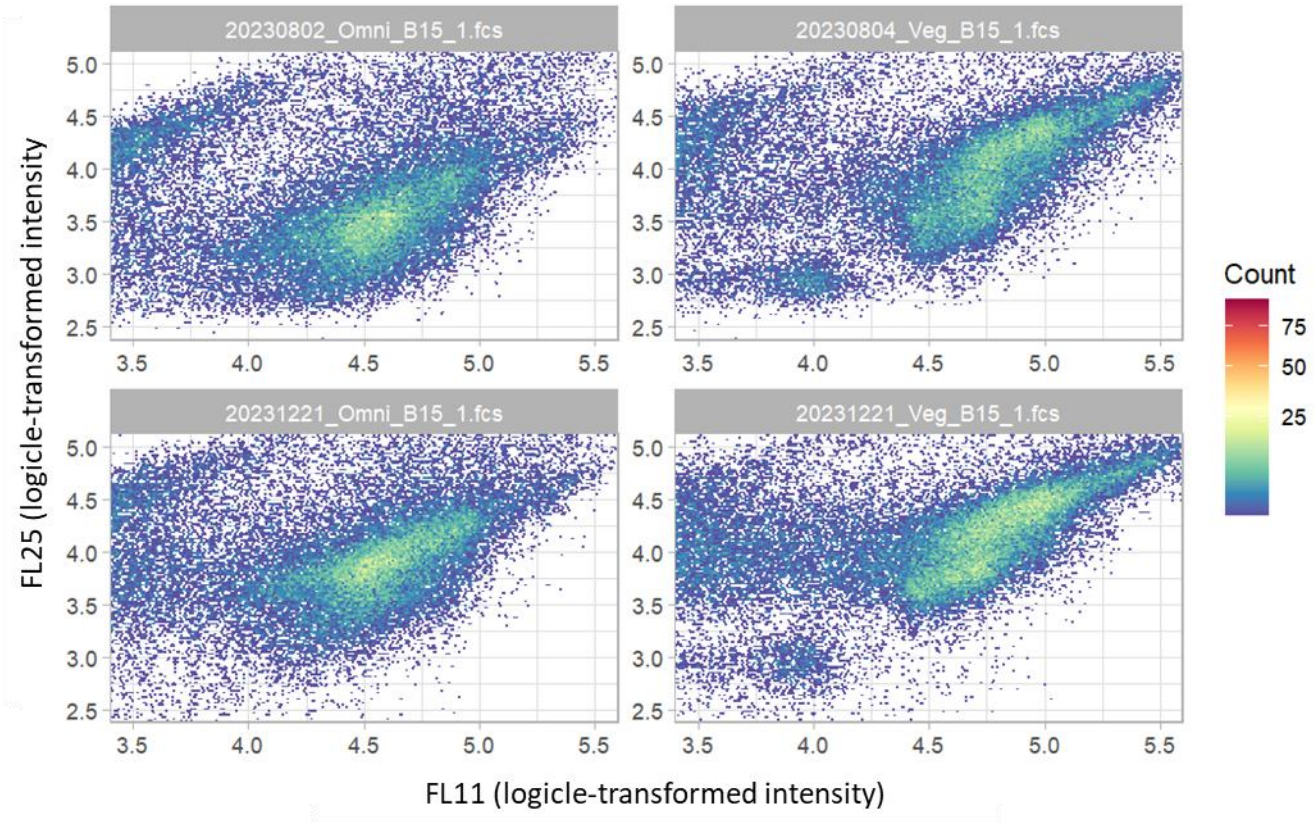


Figure 7-2: Comparison of staining profiles across experiments and donor pools for FCM data.
Samples from Box15 of each donor pool were assayed on two dates (once for homogeneity and once for stability). Here, the rows correspond to experimental dates and the columns to donor pools. Plots are shown on the same scale (*x*-axis: FL11, nucleic acid probe, *y*-axis: FL25, cell body probe). The color corresponds to the count for each area of the plot such that blue areas are sparse and yellow areas are dense, thus each yellow region can be understood as a sub-population.

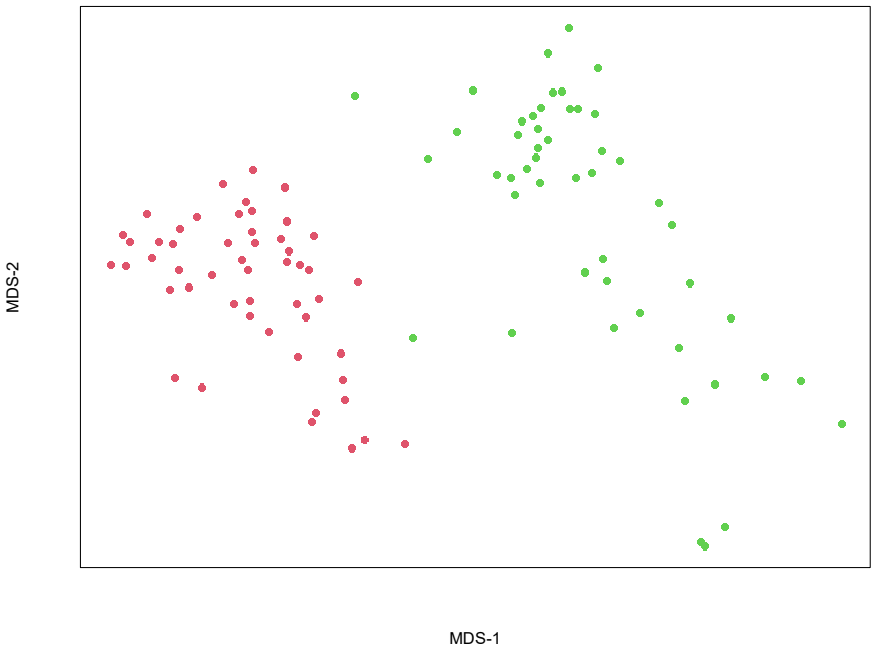


Figure 7-3: MDS plot using peak intensities across donor pools for FCM data.
An MDS plot calculated from the five key intensity channels shows that the two donor pools could be distinguished. Red corresponds to the omnivore samples and Green to the vegetarian samples.

7.2.2. Count Data

Count data are presented as a normalized number of cells per bead in each sample (Figure 7-4) and were split into homogeneity (Figure 7-5) and stability (Figure 7-6) datasets for evaluation. There are two critical caveats to consider when interpreting this data: bead loss and changes in the mixing protocol.

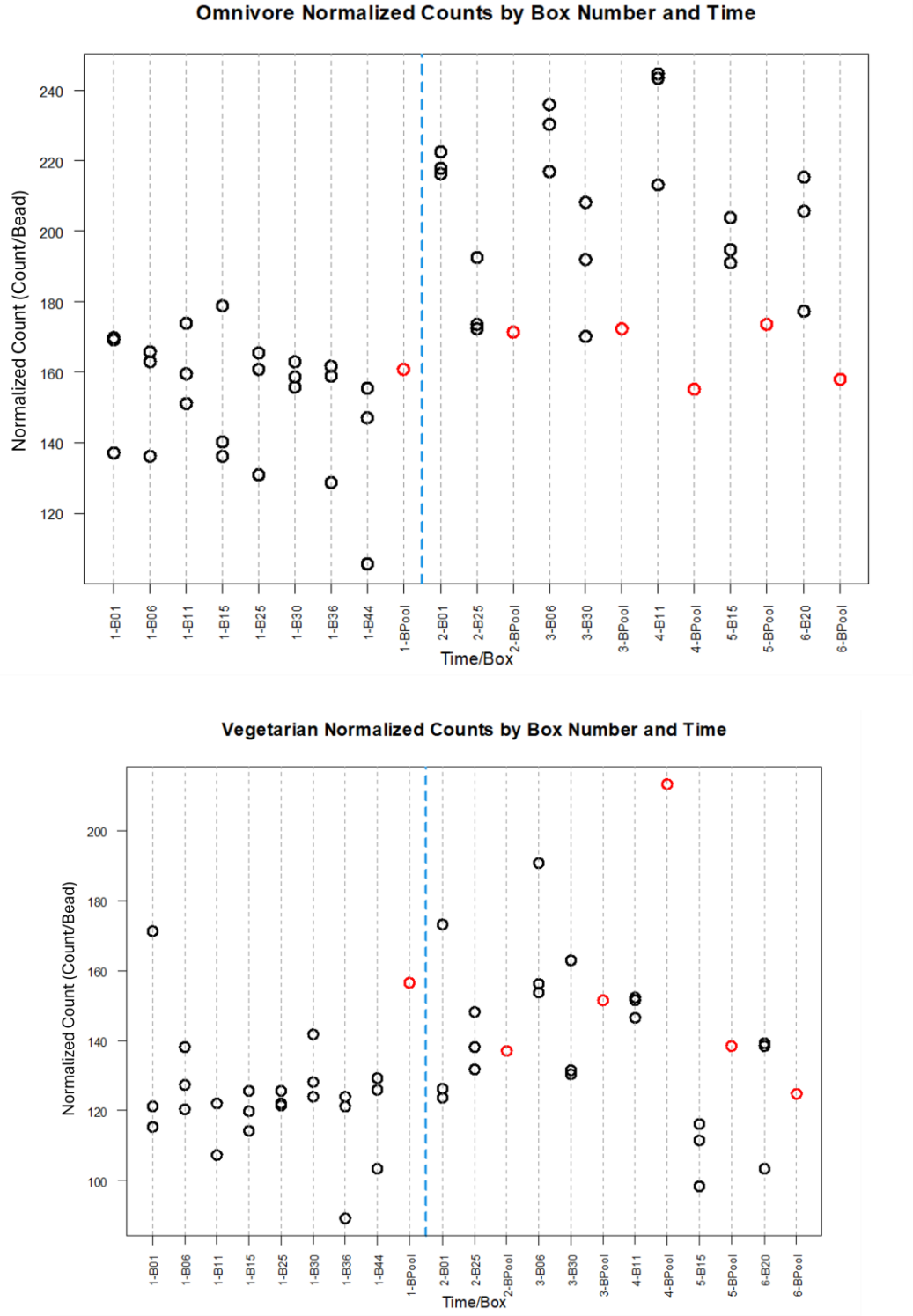


Figure 7-4: Normalized cell counts per donor pool via FCM.
The *x*-axis lists all the experiments as date/box number with a dashed blue line separating the homogeneity (left) and stability (right) data. The *y*-axis represents the number of cells normalized by the number of beads. Red data points represent the "pooled" samples. (top – omnivore, bottom – vegetarian).

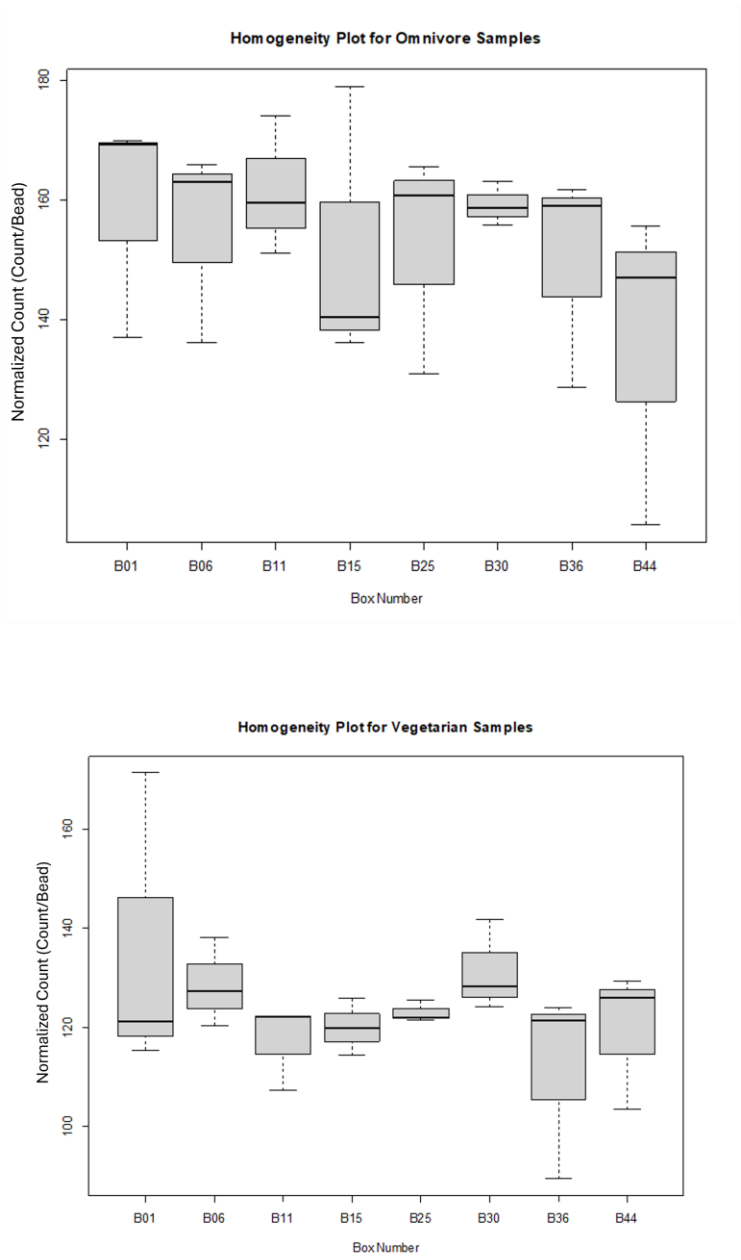


Figure 7-5: Homogeneity plots for normalized cell count via FCM.
The range of expected normalized cell count values are summarized for the homogeneity data. The *x*-axis lists box numbers. The *y*-axis represents the number of cells normalized by the number of beads. (top – omnivore, bottom – vegetarian).

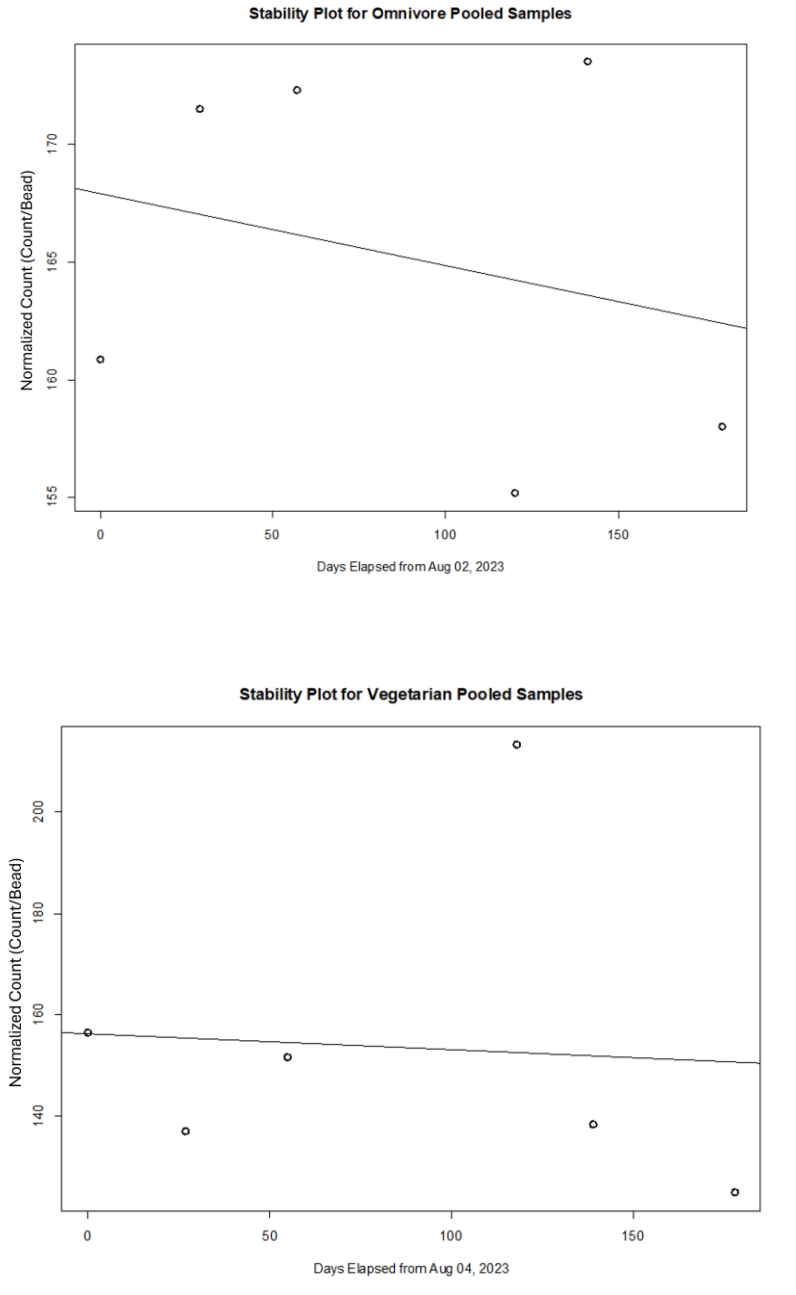


Figure 7-6: Stability plots for normalized cell counts via FCM.
The normalized cell count values for each of the pooled samples were evaluated for trends as stability data. The *x*-axis gives the elapsed time since the pooled samples were generated. The *y*-axis represents the number of cells normalized by the number of beads. P-values for trends in the normalized cell count limits of 0.61 and 0.89 for the omnivore and vegetarian materials, respectively. (top – omnivore, bottom – vegetarian).

7.2.2.1. Bead Loss

Ideally, the number of beads would have been used to calculate an absolute number of cells per starting vial of material. However, the number of beads per sample was shown to be lower than expected for an acquired volume (using another cytometer), which violates assumptions necessary to undertake this calculation. Since the percentage loss of beads was presumed to be constant across dates and sample types, a normalized count could be used to evaluate homogeneity and stability, but an absolute count could not be calculated.

7.2.2.2. Changes in Mixing Protocol

There was one large protocol change between the homogeneity and stability experiments. When samples were defrosted, the protocol called for mixing before subsampling. During the homogeneity experiment mixing was done using a vortex and during the stability experiments it was done via pipette. It is hypothesized that pipetting provides more mixing due to the viscosity of the material and this explains the increase in normalized cell number for both donor pools in the stability data (Figure 7-4). Mixing is expected to be especially important when subsampling a complex sample with large aggregates and acellular material that could impact the homogeneity of subsamples taken from a single vial. Thus, data from the first two time points (homogeneity experiments) were compared separately than data from all subsequent time points (stability experiments). The only crossover data are from “pooled” samples (shown in red in Figure 7-4) as they were generated during homogeneity experiments then refrozen and measured during each stability experiment. Normalized cell counts from the pooled samples measured during the stability studies do appear to be more similar to normalized counts from the homogeneity experiments as compared to other samples in the stability studies.

7.2.2.3. Homogeneity and Stability

The two donor pools appear to have different normalized cell counts, but this is difficult to conclude because (1) normalized cell count data had wide ranges for all replicate samples and (2) there were protocol changes across homogeneity and stability experiments. The wide variability in replicates was expected because the experimental design called for technical replicates to be generated early in the protocol to incorporate uncertainty generated from the filtration and staining steps, rather than running multiple aliquots from the same vial (without variability due to filtration and staining). Technical replicates were generated by sub-sampling each vial and filtering the sample through a 50 µm cell strainer. This was followed by fluorescent probe incubation, dilution, and running the sample on the instrument. The filtration step was expected to introduce large variability in the count data and was likely compounded by the issue described in Section 7.2.2.2.

Thus, samples were deemed sufficiently homogeneous for the omnivore and vegetarian materials. For the omnivore samples, any randomly chosen vial from a randomly chosen box will have a normalized count between 135.5 and 170.7 (95 % confidence). For the vegetarian samples, any randomly chosen vial from a randomly chosen box will have a normalized count

between 106.3 and 140.4 (95 % confidence). Pooled samples were deemed stable with P-values for trends in the normalized cell count limits of 0.61 and 0.89 for the omnivore and vegetarian materials, respectively.

7.2.3. Intensity Data

Intensity data represent the intensity of the scattered light or fluorescence emission associated with a particular laser and detector pair. Note that these are arbitrary intensities specific to the dataset but could be converted with the 8-peakRCP to standardized values. Data are presented as density plots with a logarithmically scaled *x*-axis. Five primary channels of interest were evaluated to determine if the intensity of the primary peak(s) was sufficiently homogeneous and stable (Figure 7-7). The FL20 channel corresponded primarily to the spiked-in concentration control beads and was not used in conclusions. The full intensity distributions for each of the four remaining channels and each material were overlaid for each of the dates (Appendix H).

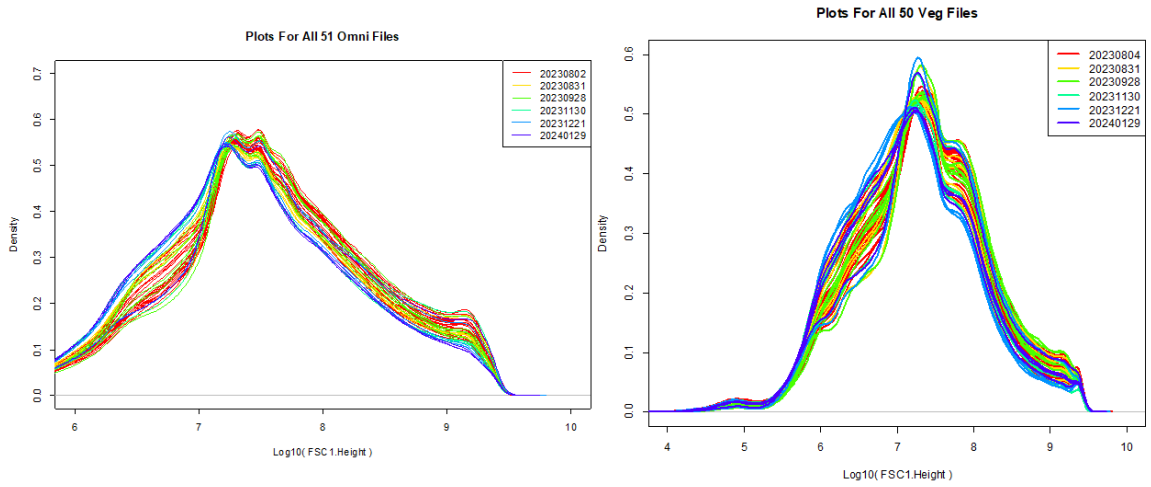


Figure 7-7: Density plots of FSC1 intensity distribution across donor pools for FCM data. Representative data for the FSC1 cytometer channel (left – omnivore, right – vegetarian). Each plot shows the log10-transformed intensity of the channel on the *x*-axis and the density on the *y*-axis with color-coding corresponding to experiment date. *X*-axis and *y*-axis scales vary across plots.

As with the count data, the intensity data were divided into two parts for determination of homogeneity and stability. Starting with the homogeneity data, for each peak, the location of the peak (intensity and height) was recorded (Figure 7-8). For some channels, there were multiple peaks which were calculated separately. To determine the extent of variation that could be expected in the peak intensity and peak height for each channel, lower and higher limits of $\geq$ 95 % confidence intervals around these values were calculated. The coverage of these limits applies to the entire reference material, not only tested samples (Figure 7-9, Appendix H)

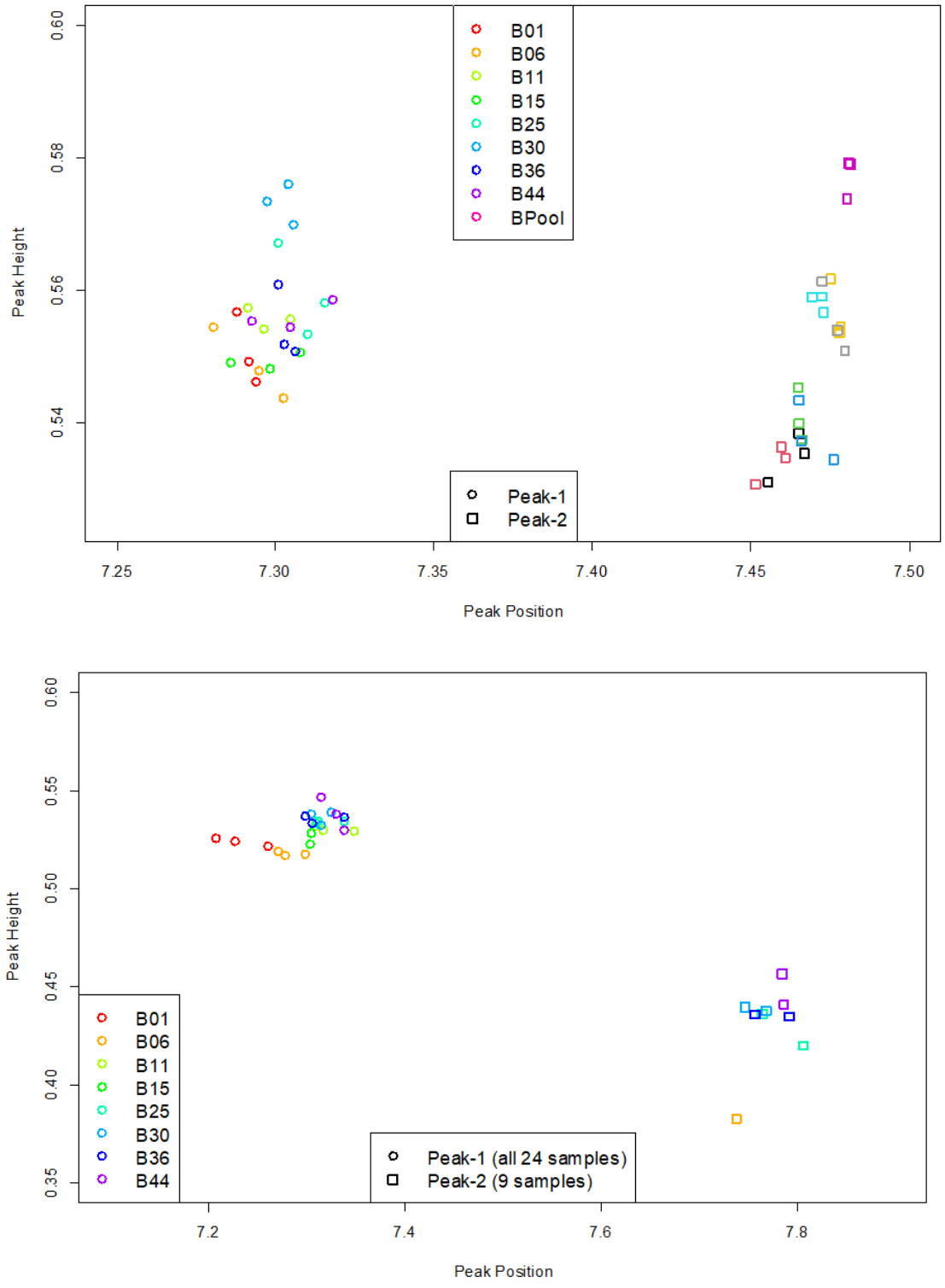


Figure 7-8: Intensity peak position and height across donor pools for channel FSC1 in FCM data.
Each plot (top – omnivore, bottom – vegetarian) shows the location of the identified peak in terms of the log10-transformed intensity (*x*-axis) and the density (*y*-axis) with color-coding corresponding to box number. *X*-axis and *y*-axis scales vary across plots.

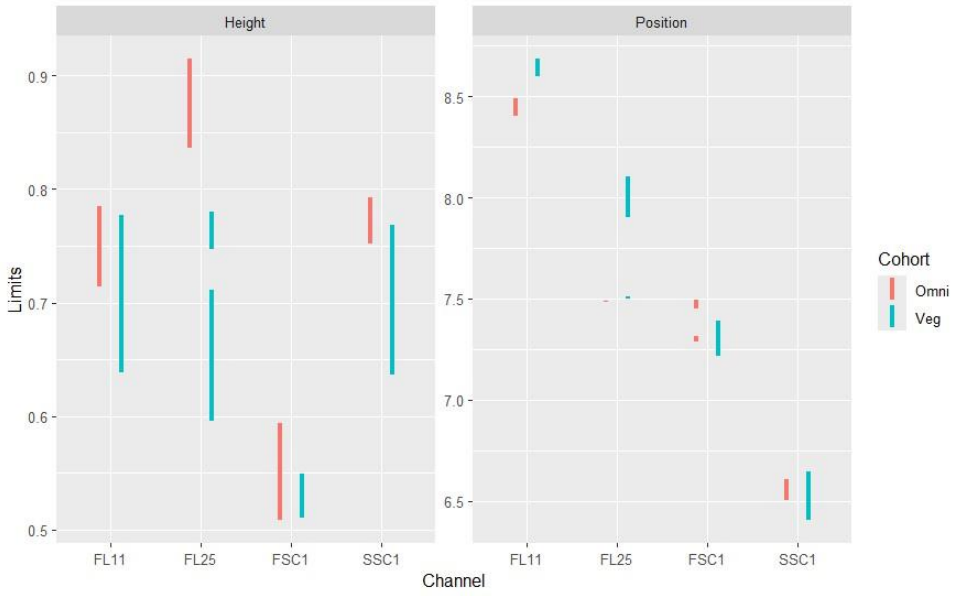


Figure 7-9: Limits across peak height and peak position for each donor pool and all channels via FCM.
The estimated lower and upper limit of the peak height (left panel) and peak position as log10-transformed intensities (right panel) and are shown for each channel (*x*-axis) and material (color). FL20 was excluded since it was intended for the bead control and not cell characterization.

For the stability analysis, similar to the count data only the pooled samples were used to evaluate if there were trends in the peak positions (Figure 7-10, Appendix H).

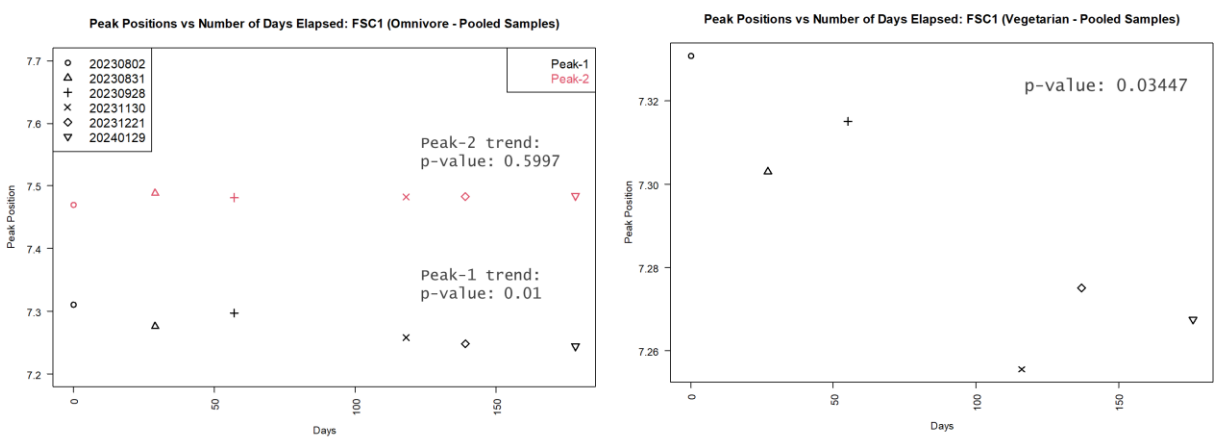


Figure 7-10: Stability plots for FSC1 intensities in FCM data.
The peak intensity values for each of the pooled samples in FSC1 were evaluated for trends as stability data. The *x*-axis gives the elapsed time since the pooled samples were generated. The *y*-axis represents the peak position (left – omnivore, right – vegetarian).

7.2.3.1. Issues in Cell Body Staining

Analysis of the FL25 channel demonstrated batch effects, which was expected based on day-to-day staining, but the magnitude was much greater than expected, and there was a noticeable trend in the data associated with study duration (Appendix H). Interestingly, the overall intensity shift over time seemed to be to higher values, and one noticeable peak decreased in intensity across experiments. Most likely this is due to changes in the fluorescent probe reactivity because of repeated freeze-thaw cycles. Literature from the manufacturer indicates that the reconstituted probe is only stable for one month at -20 °C which was greatly exceeded by the duration of this study and freeze-thaw cycles are not recommended.

Homogeneity and Stability

Samples were deemed sufficiently homogeneous and stable with the $\geq$ 95 % confidence intervals for values that could be expected in the peak intensity and peak height for each channel (apart from FL25). P-values for some analyses were $\leq$ 0.1 but we note the small absolute changes in peak positions.

The location of the peak differs between the two donor pools for FSC1, FL11, and FL25 but not FL20 (which corresponds to the spike-in beads) or SSC. Two peaks were observed in FSC for the omnivore material and for the unstable FL25 channel.

Interestingly, the FSC vegetarian samples had peak positions distinct from the other samples in that material production. FL11 and FL20 did not have a trend in peak position, as was the goal based on daily calibration. FL25 was unstable for both donor pools and is discussed in detail in Section 7.2.3.1.

The peak heights had greater relative variability compared to the peak positions. This is likely because the peak height of a density plot is partially dependent on the amount of low intensity signal (debris) that was acquired and might be more subject to batch effects due to daily alignment. To remove this potential source of variability, future analyses could instead only characterize the subpopulation classified as cells.

7.3. Flow Cytometry Summary

Two flow cytometry measurements were evaluated for homogeneity and stability: (1) the normalized number of cells per mL and (2) the intensity of scattered light and added fluorescent probes.

Normalized cell count was determined to be sufficiently homogeneous for each RM 8048 material and stable (Section 7.2.2.3), though the homogeneity and stability counts could not be compared to one another due to a protocol change (detailed in Section 7.2.2.2).

The intensity distributions of the two materials were easily distinguished by eye or quantitatively (Section 7.2.3). Intensities were determined to be sufficiently homogeneous and stable within the stated coverage limits with ≥ 95 % confidence for all channels except FL25

(Section 0). The instability in FL25 is hypothesized to be due to an expired reagent, because the shifts in intensity distribution are apparent after the expiration date of the reagent.

Based on the data collected and analyses described in this report, RM 8048 was determined to be sufficiently homogeneous and stable within the stated bounds for the fluorescence flow cytometry measurements performed. Therefore, RM 8048 is appropriate for harmonization of flow cytometry measurements of particle count and nucleic acid fluorescent staining.

8. Summary and Future Directions

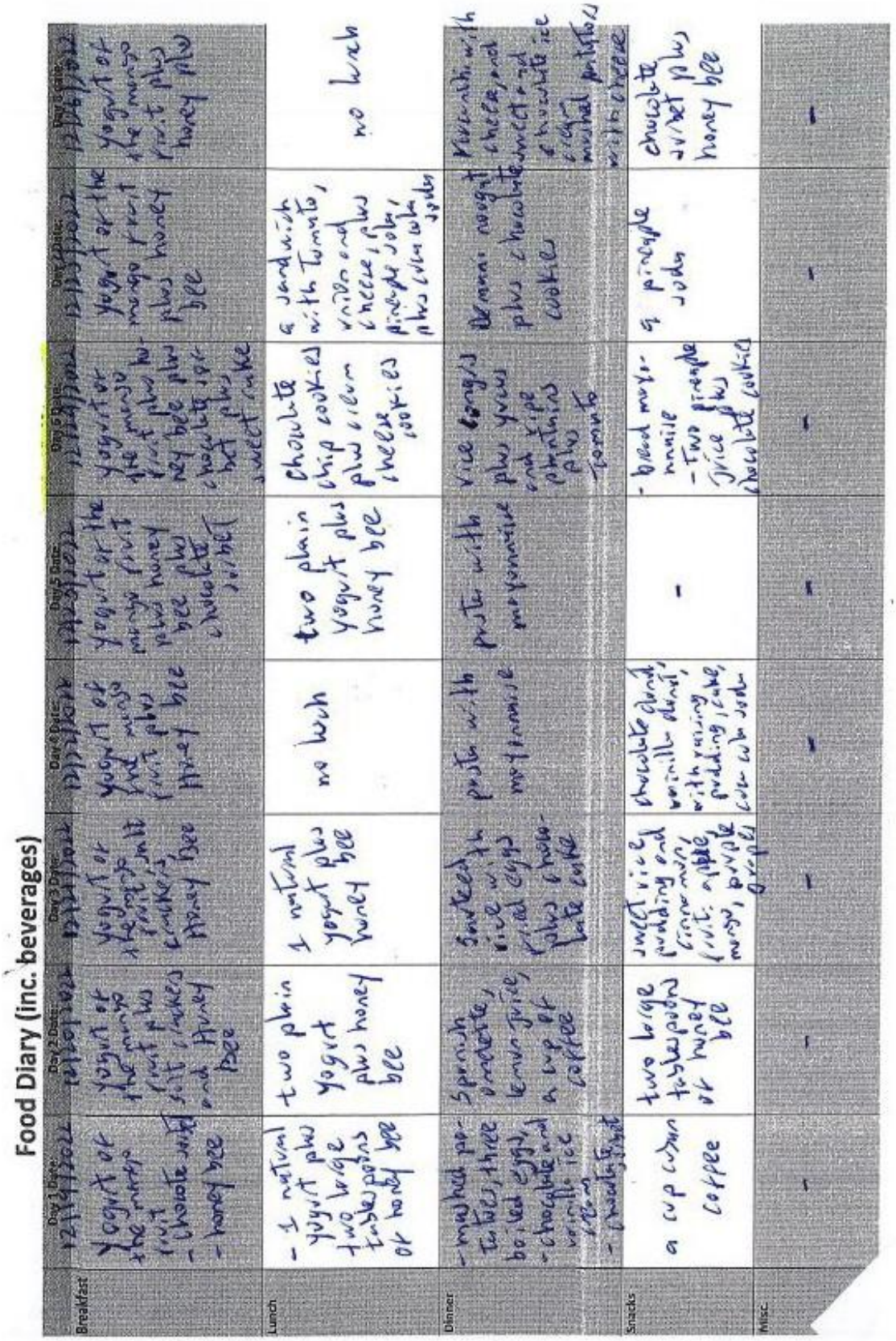
The work presented in this publication provides a comprehensive overview of the development and characterization of RM 8048, a Human Fecal Reference Material. The initial material characterization included homogeneity assessments (metagenomics, LC/MS, NMR, and flow cytometry) and stability monitoring over 6 time points encompassing approximately 6 months (metagenomics, NMR, and flow cytometry). In addition, information on taxonomic assignments and metabolite identities and the corresponding methods has been included. It is important to remember that microbiome measurements are the result of complex workflows, and while these attributes are shared, they are not meant to provide a ground truth. The purpose of this RM is to provide a quality assurance/quality control (QA/QC) material for routine metagenomic and metabolomic measurements, differential studies, interlaboratory comparisons, laboratory and instrument qualification, and training by the metagenomic and metabolomic communities. RM 8048 can be used to assess comparability within a laboratory, between laboratories, or among different measurement approaches.

RM 8048 is a viable biological material, as such, NIST will make annual measurements to monitor for any changes and is engaging with end-users to learn how the material is being used, and to monitor for potential issues or ways to improve future similar materials. In addition to monitoring stability, NIST continues its characterization of this material and is currently in the process of working out protocols for proteomic analyses and virome measurements. Finally, NIST is developing complementary resources including a Fecal Metabolite Calibrant Solution (RGTM 10212), a metagenome assembled genome (MAG) database, and a collection of microbial isolates.

Appendix A. Diet Logs

A.1. Handwritten Omnivore Diet Log

S10022014228 : HMN1084508 (Collections 1-3)



S10022014228 : HMN1084508 (Collections 1-3) con't

Food Diary (inc. beverages)

	Day 1 Date: 12/27/2022	Day 2 Date: 12/28/2022	Day 3 Date: 12/29/2022	Day 4 Date: 12/30/2022
Breakfast	yogurt or the mango fruit plus honey bee	Condensed milk plus cuban coffee	Condensed milk plus cuban coffee or cookies	yogurt or the mango fruit plus honey bee
Lunch	no lunch	a natural yogurt plus honey bee	a natural yogurt	no lunch
Dinner	Condensed milk plus cuban coffee	Chinese rice plus fried eggs plus crackers	rice with vegetables and pizza	bread with cream cheese plus mango yogurt and honey bee
Snacks	no snacks	no snacks	no snack	cookies, vanilla ice cream, honey bee and chocolate cake
Misc.	three 500-milligram ibuprofen pills	-	-	-

S10022014228 : HMN1084508 (Collections 4-6)

Food Diary (inc. beverages)

Donor #

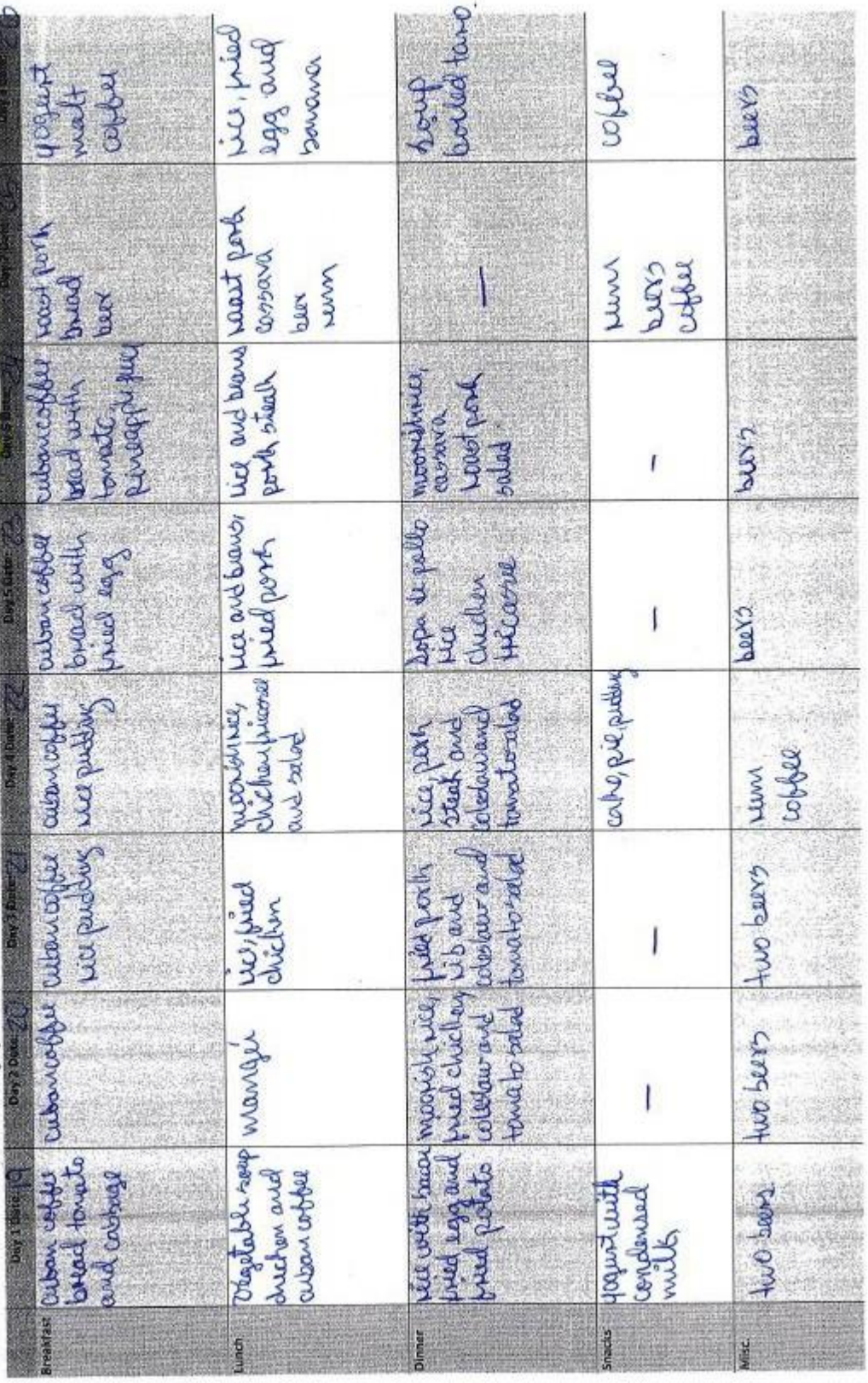
	Day 1 Date: 19	Day 2 Date: 20	Day 3 Date: 21	Day 4 Date: 22	Day 5 Date: 23	Day 6 Date: 24	Day 7 Date: 25	Day 8 date: 26
Breakfast	Soda cracker Cream cheese Pineapple juice	Cuban coffe Soda cracker Butter strawberry yogurt.	Soursop yogurt Donal	Soda cracker with cheese and soursop yougurt.	Cuban coffee malt cracker cream cheese.	scrambled egg, tomatoes, chips, bread, orange juice. Cuban coffe	Bread with butter, oregat malt. Cuban coffee	Bread with cream cheese, malt Cuban coffee
Lunch	Pizza cheese lemon spray	spaghetti cheese	White rice fried pork ribs salad tomato spinach, lettuce.	Rice, peas croquettes and avocado	Rice beans stew pork steak coke soda	Rice with grey; tamales, roast pork, yucca, milk flan	Roast white rice, yucca tomato salad	bread with suckling pig cola soda
Dinner	taro cream Fried egg	White rice Fried Pork rib Salad tomato spinach, lettuce. Vanilla ice cream cone	White rice. fried pork lettuce spinach salad and lettuce	Chickpeas soup popsicle ice cream.	coke soda White rice salad tomato spinach, lettuce steak pork	Rice with grey, tamale, Roast pork, salad avocado, tomato spinach and lettuce	Fried yucca, tamale. Pork meat. Lemon soda	Fish fillet Salad tomato onion spinach lettuce. Pistachios cookie Filled with coconut cream
Snacks	juice guava Pistachio Cuban coffe	Candy lemon spray	Fruit apple, mango.	cocacola malta	yogurt.	Yogurt.	tea	Fritter sweet coconut almonds mango yogurt
Misc.		Enalapril 20mg	Enalapril 20mg	Enalapril 20mg	Enalapril 20mg	Enalapril 20mg	Enalapril 20mg	Enalapril 20mg

S10022014228 : HMN1084508 (Collections 4-6) con't

Food Diary (inc. beverages)

	Day 1 Date: 27	Day 2 Date: 28	Day 3 Date: 29
Breakfast	Butter bread malt coban coffee	bread with spicy meat puff. Cola sock coban coffe	Cuban coffee bread with butter malt
Lunch	Bread with sucking pig cola soda	Grilled white Meat chicken and pasta in a garlic Sauce with broccoli carrots and corn	
Dinner	Fried fish salad tomato cabbage & lettuce and spinach	White Rice potage chicken Fries pork Rib and tomato salad	
Snacks	Yogurt Almond popcorn	Yogurt coconut salad	
Misc.			Acetaminophen 50mg melatonin aspirin

S10022014228 : HMN1084508 (Collections 7-9)

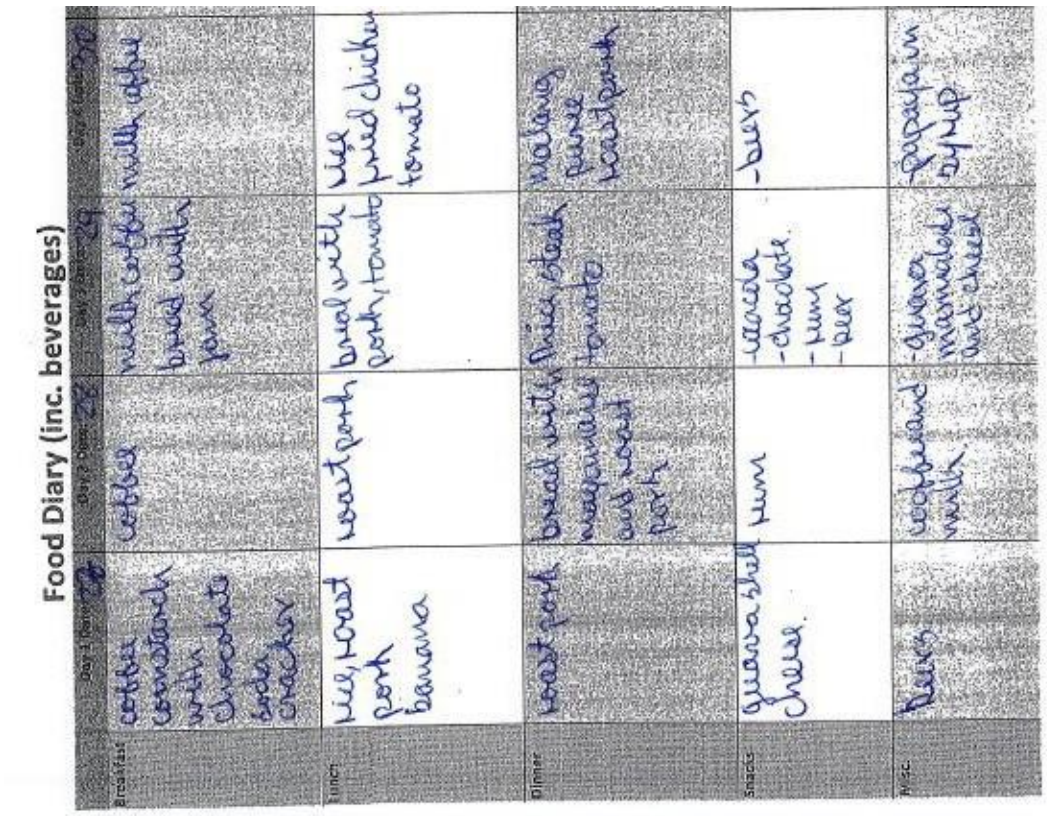


Food Diary (inc. beverages)

Donor #

	Day 1 Date: 19	Day 2 Date: 20	Day 3 Date: 21	Day 4 Date: 22	Day 5 Date: 23	Day 6 Date: 24	Day 7 Date: 25	Day 8 Date: 26
Breakfast	Cuban coffee bread tomato and cabbage	Cuban coffee	Cuban coffee rice pudding	Cuban coffee bread with fried eggs	Cuban coffee bread with tomato pineapple juice	Cuban coffee bread with tomato pineapple juice	roast pork bread beer	yogurt malt cereal
Lunch	Vegetable soup chicken and Cuban coffee	mangu	rice, fried chicken	moondongo, chicken, piccadillo and salad	rice and beans fried pork	rice and beans pork steak	roast pork cassava beer rum	rice, fried egg and banana
Dinner	rice with bacon fried egg and fried potato	moondongo, rice, fried chicken and coleslaw and tomato salad	pork ribs and fried pork rib and steak and coleslaw and tomato salad	rice, pork steak and coleslaw and tomato salad	bistec de palito rice chicken and macaroni	moondongo, cassava, roast pork salad	—	soup boiled taro
Snacks	yogurt with condensed milk	—	—	cake, pie pudding	—	—	rum beer coffee	coffee
Misc.	two beers	two beers	two beers	rum coffee	beers	beers		beers

S10022014228 : HMN1084508 (Collections 7-9) con't



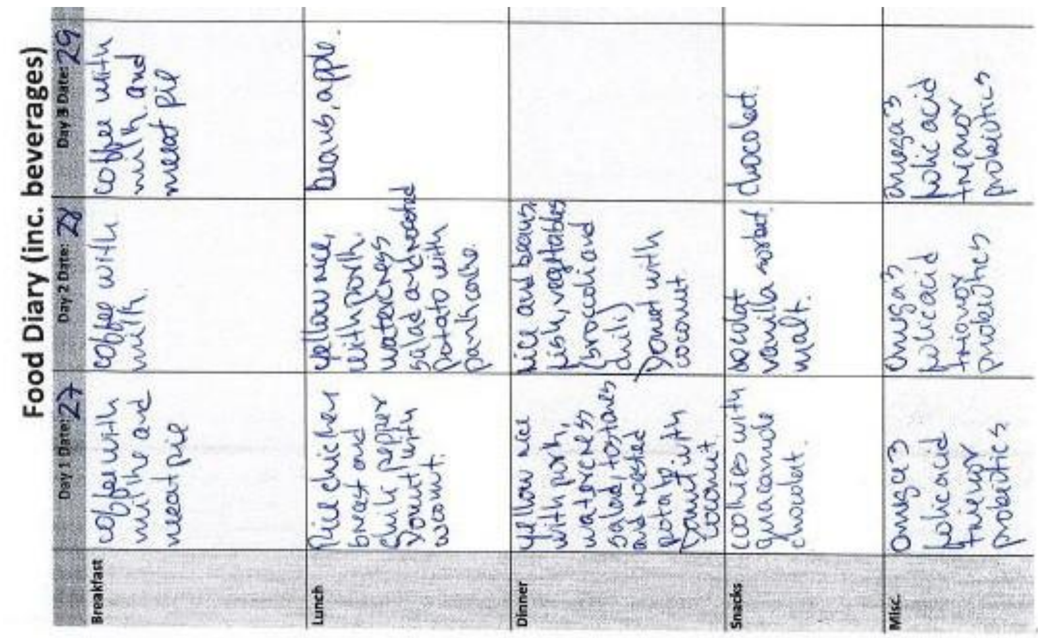
S10022014228 : HMN1084508 (Collections 10-12)

Food Diary (inc. beverages)

Donor #

	Day 1 Date: 19	Day 2 Date: 20	Day 3 Date: 21	Day 4 Date: 22	Day 5 Date: 23	Day 6 Date: 24	Day 7 Date: 25	Day 8 date: 26
Breakfast	coffee with milk and meat pie	Coffee with milk	coffee with milk and meat pie	coffee with milk	coffee	pancake, nutella passion fruit juice and coffee	yogurt with grapes, strawberries blackberries and honey. Cofee	coffee
Lunch	beef with vegetables, rice and avocado	chicharron rice and salad muffin.	Beans and rice, beef hash	Rice, peas, croquettes and avocado	chinese soup	lasagna	moorish rice cassava, roast pork, donut with coconut.	Hot dog with hot chocolat
Dinner	chicharron rice and salad	Beans and rice, beef hash, yucca and watermelon	Rice, peas, spinach croquettes and avocado	chicken breast rice and salad (lettuce, tomato spinach and grapes)	beef with potatoes, rice and avocado.	moorish rice, cassava, roast pork, donut with coconut.	Pizza and guava smoothie	corn meal with fried egg. strawberries yogurt.
Snacks	chocolat.	candy, pringles grapes, strawberries	apple mango, grapes	cocacola		chocolat. tequila whiskey ham and cheese olives		whiskey ham.
Misc.	Omega 3 folic acid trienor (birth control pill) probiotics	Omega 3 folic acid trienor probiotics	Omega 3 folic acid trienor Probiotics	Omega 3 folic acid trienor Probiotics	Omega 3 folic acid trienor probiotics	Omega 3 folic acid trienor probiotics	omega 3 folic acid trienor probiotics	Omega 3 folic acid trienor probiotics

S10022014228 : HMN1084508 (Collections 10-12) con't



Food Diary (inc. beverages)

	Day 1 Date: 27	Day 2 Date: 28	Day 3 Date: 29
Breakfast	coffee with milk and meat pie	coffee with milk.	coffee with milk and meat pie
Lunch	Rice chicken breast and chili pepper Donut with coconut.	yellow rice, with pork. watercress salad and roasted potato with pankcake.	beans, apple.
Dinner	yellow rice with pork, watercress salad, tostones and roasted potato. Donut with coconut.	rice and beans, fish, vegetables (broccoli and chili) Donut with coconut	
Snacks	cookies with guacamole chocolat.	chocolat vanilla sorbet. malt.	chocolat.
Misc.	Omega 3 folic acid thienor probiotics	Omega 3 folic acid thienor probiotics	Omega 3 folic acid thienor probiotics

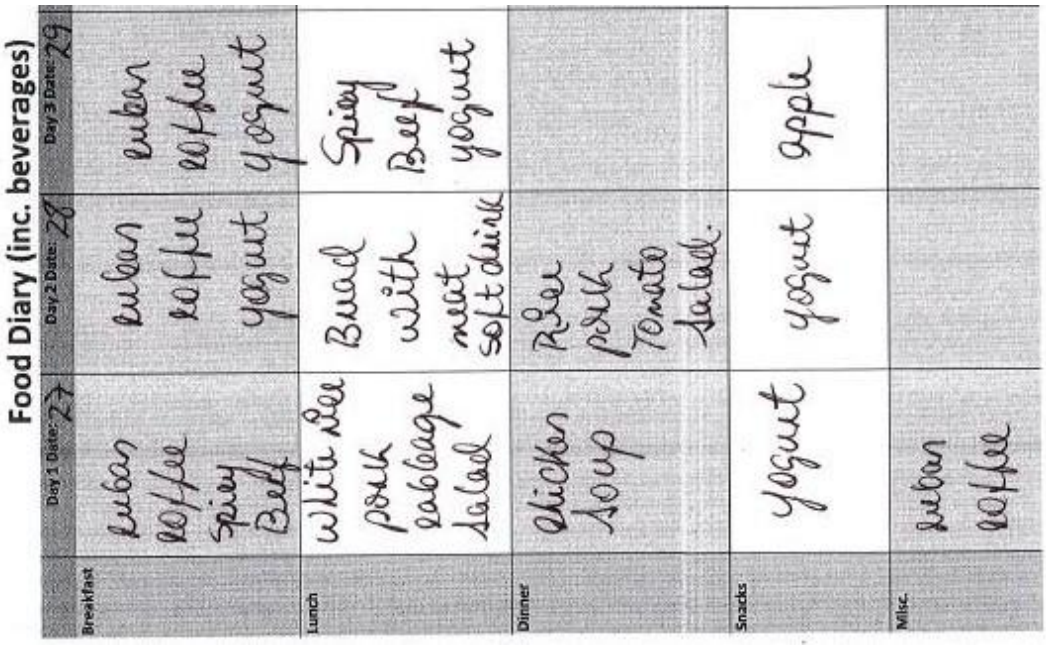
S10022014228 : HMN1084508 (Collections 13-15)

Food Diary (inc. beverages)

Donor #

	Day 1 Date: 19	Day 2 Date: 20	Day 3 Date: 21	Day 4 Date: 22	Day 5 Date: 23	Day 6 Date: 24	Day 7 Date: 25	Day 8 Date: 26
Breakfast	cuban coffee, roadwad, spicy Beef	cuban coffee, yogurt	cuban coffee	cuban coffee	cuban coffee	cuban coffee	cuban coffee, chocolate milk	cuban coffee, yogurt
Lunch	Spicy Beef, Kool-aid, chocolate, cuban coffee	Mango, mango juice	cheese baguette, guava and cheese pastry	cuban sandwich	pizza chicken	milkshake, mimy pizza	Black rice, roast pork, cabbage salad.	Rice, pork steak
Dinner	White rice, fried egg, cabbage and Tomato salad, fried potato	Black rice, fried chicken, cabbage and Tomato salad	white rice, pork rib, cabbage salad	spaghetti	white rice, pork rib, cabbage salad.	Black rice, Roast pork, Yuaca, Tomato salad.	white rice, fried egg, Tomato salad.	Rice, pork rib, Tomato salad.
Snacks	chocolat, pork rind	chocolat, yogurt, mango juice	rice pudding	yogurt	pudding	yogurt, malt with milk	Mamoncha	Bread with Tomato
Misc.	sweet potato, yogurt, cuban coffee	fruit, grapes, mango, orange			coca cola			yogurt

S10022014228 : HMN1084508 (Collections 13-15) con't

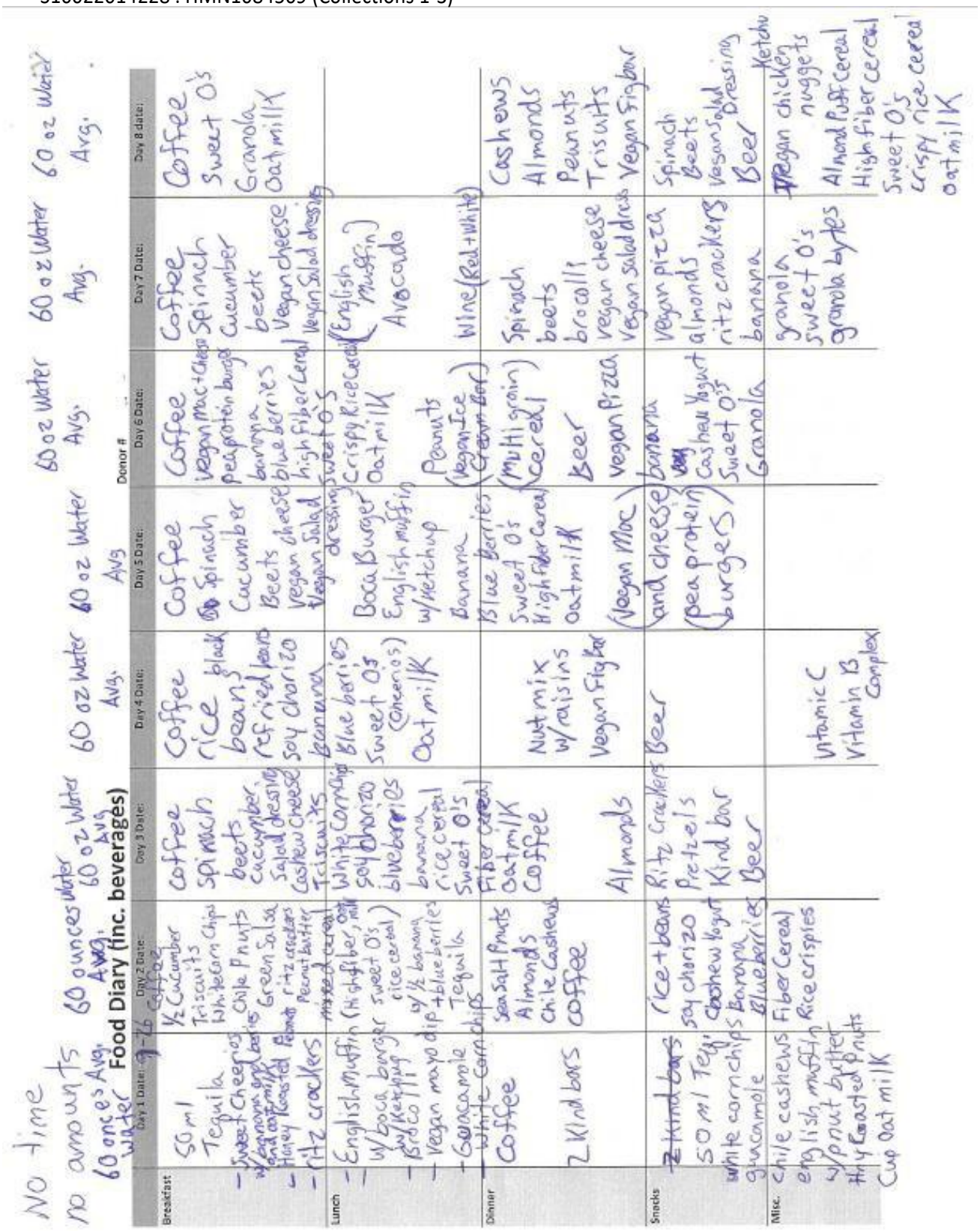


Food Diary (inc. beverages)

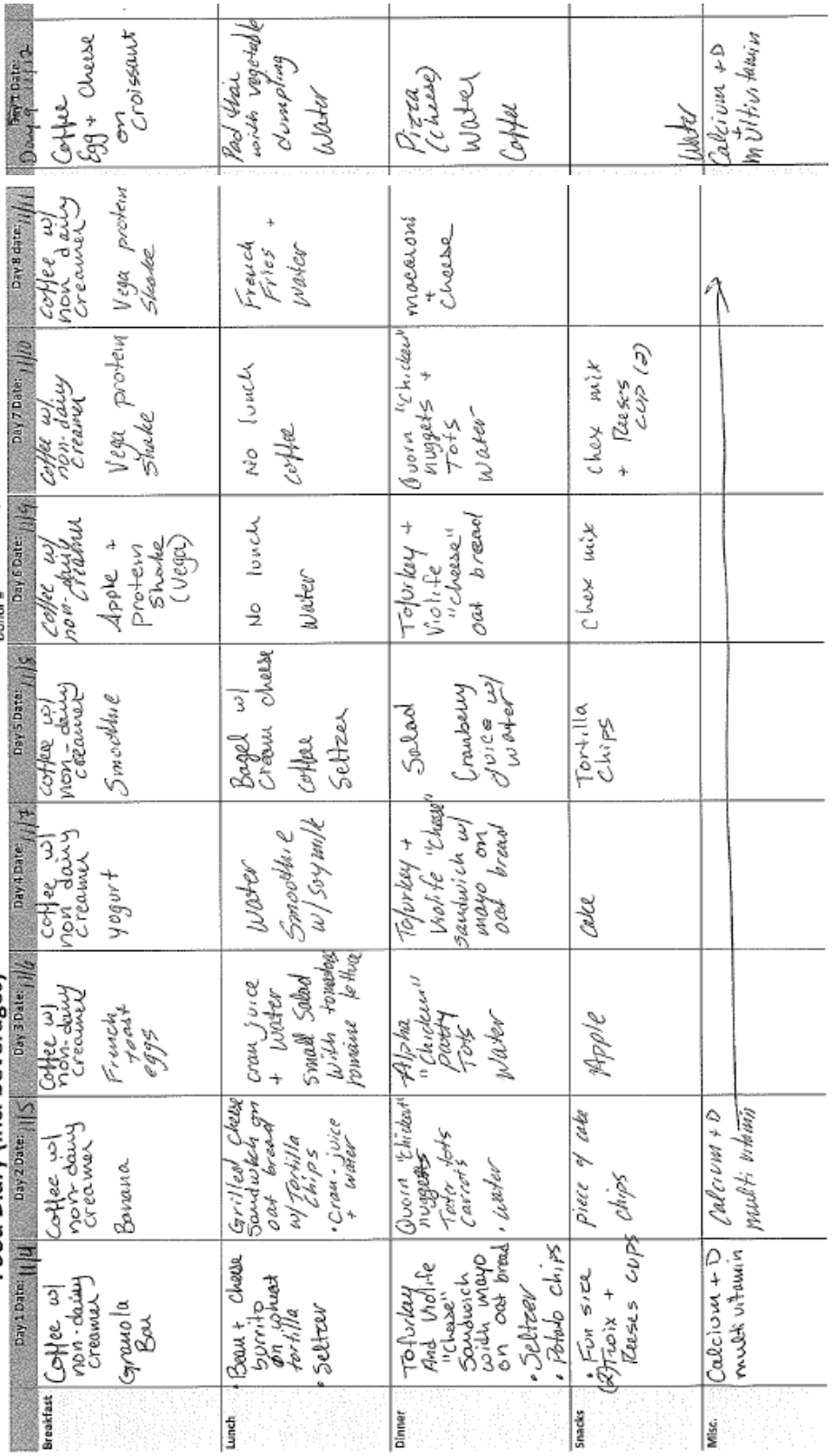
	Day 1 Date: 27	Day 2 Date: 28	Day 3 Date: 29
Breakfast	Cuban coffee Spicy Beef	Cuban coffee yogurt	Cuban coffee yogurt
Lunch	White rice pork cabbage salad	Bread with meat soft drink	Spicy Beef yogurt
Dinner	Chicken soup	Rice pork Tomato salad.	
Snacks	yogurt	yogurt	apple
Misc.	Cuban coffee		

A.2. Handwritten Vegetarian Diet Log

S10022014228 : HMN1084509 (Collections 1-3)



S10022014228 : HMN1084509 (Collections 4-6)

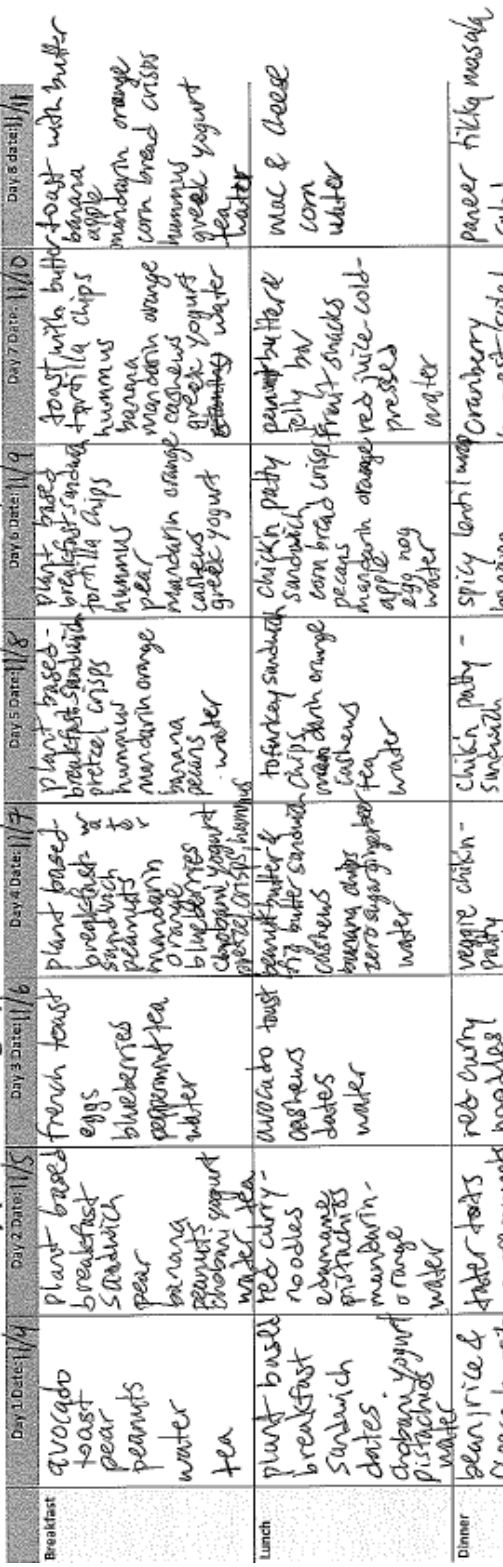


Food Diary (inc. beverages)

Donor #

	Day 1 Date: 11/4	Day 2 Date: 11/5	Day 3 Date: 11/6	Day 4 Date: 11/7	Day 5 Date: 11/8	Day 6 Date: 11/9	Day 7 Date: 11/10	Day 8 date: 11/11	Day 9 Date: Day 9 11/12
Breakfast	Coffee w/ non-dairy creamer Granola Bar	Coffee w/ non-dairy creamer Banana	Coffee w/ non-dairy creamer French toast eggs	coffee w/ non dairy creamer Yogurt	coffee w/ non-dairy creamer Smoothie	coffee w/ non-dairy creamer Apple + Protein Shake (Vega)	coffee w/ non-dairy creamer Vega protein Shake	coffee w/ non dairy creamer Vega protein Shake	Coffee Egg + Cheese on Croissant
Lunch	• Bean + cheese burrito in tortilla • Seltzer	Grilled cheese Sandwich on oat bread w/ Tortilla Chips • Cran-juice + water	cran juice + water small Salad with tomatoes & house vinaigrette	Water Smoothie w/ Soy milk	Bagel w/ cream cheese coffee Seltzer	No lunch Water	No lunch coffee	French Fries + Water	Pad thai with vegetable dumpling Water
Dinner	Tofurkey and Violife "cheese" Sandwich with mayo on oat bread • Seltzer • Potato chips	Quorn "chicken" nuggets Tater tots Carrots • water	Alpha "Chicken" Patty Tots Water	Tofurkey + Violife "cheese" sandwich w/ mayo on oat bread	Salad Cranberry juice w/ water	Tofurkey + Violife "cheese" oat bread	Quorn "chicken" nuggets + Tots Water	macaroni + cheese	Pizza (cheese) Water Coffee
Snacks	• Fun size (3) Twix + Reeses cups	piece of cake chips	Apple	cake	Tortilla Chips	Chex mix	Chex mix + Reeses cup (2)		
Misc.	Calcium + D multi vitamin	Calcium + D multi vitamin							Water Calcium + D Multi vitamin

S10022014228 : HMN1084509 (Collections 7-9)



Food Diary (inc. beverages)

Donor ff

	Day 1 Date: 11/4	Day 2 Date: 11/5	Day 3 Date: 11/6	Day 4 Date: 11/7	Day 5 Date: 11/8	Day 6 Date: 11/9	Day 7 Date: 11/10	Day 8 date: 11/11
Breakfast	avocado toast pear peanuts water tea	plant based breakfast sandwich pear banana peanuts Greek yogurt water tea	french toast eggs blueberries peppermint tea water	plant based breakfast sandwich w peanuts & mandarin orange & blueberries cranberry yogurt pretzel crisps hummus	plant based breakfast sandwich pretzel crisps hummus mandarin orange banana pecans water	plant based breakfast sandwich tortilla chips hummus pear mandarin orange cashews greek yogurt	toast with butter tortilla chips hummus banana mandarin orange cashews greek yogurt drinking water	toast with butter banana apple mandarin orange corn bread crisps hummus greek yogurt tea water
Lunch	plant based breakfast sandwich dates cranberry yogurt pistachios water	red curry noodles edamame pistachios mandarin orange water	avocado toast cashews dates water	peanut butter & fig butter sandwich cashews banana chips zero sugar gingerbeer water	tofu turkey sandwich chips mandarin orange cashews tea water	chicken patty sandwich corn bread crisps pecans mandarin orange apple egg roll water	peanut butter jelly bar fruit snacks red juice-cold-pressed water	mac & cheese corn water
Dinner	brown rice & cheese burrito chips mandarin-orange cucumber water	tater tots veggie nuggets carrots banana chips water	red curry noodles apple bread water	veggie chicken-patty apple pistachios water	chicken patty - sandwich apple mandarin orange seltzer crackers water	spicy lentil wrap banana clif bar water	cranberry harvest-salad chicken patty sandwich water	paneer tikka masala salad peanuts seltzer water
Snacks	almond butter chocolate oat bar cake snack	90 macro-bar cake snack	cookies 90 macro bar	90 macro bar party snack mix	90 macro bar party snack mix	fruit snacks clif bar	90 macro bar fruit snacks	90 macro bar fruit snacks
Misc.	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium	multi-vitamin B-complex turmeric supplement magnesium

S10022014228 : HMN1084509 (Collections 10-11)

Food Diary (inc. beverages)

Donor #

	Day 1 Date: 9/19/22	Day 2 Date: 9/20/22	Day 3 Date: 9/21/22	Day 4 Date: 9/22/22	Day 5 Date: 9/23/22
Breakfast	Kettlecorn Popcorners (95g) Greek yogurt (150g) Spicy Queso Popcorners (80 pieces)	Kettlecorn Popcorners (198g) Greek yogurt (150g) Water (20 fl oz)	Green tea (8 fl oz) Multigrain crispbread (39g) Lingonberry spread (2 Tbsp) Greek yogurt (150g) Water (46 fl oz)	Multigrain crispbread (39g) Lingonberry spread (2 Tbsp) Greek yogurt (150g) Old fashioned cake donut (1/2 piece) Coffee (12 fl oz) Water (26 floz)	Multigrain crispbread (39g) Lingonberry spread (2 Tbsp) Coffee (12 fl oz) Unsweetened applesauce (200 g)
Lunch	Diet Coke (20 fl oz) Onion straws (5 oz) Beyond Burger (1) Texas toast (2 slices) Red onion (1/4 cup) Jalapenos (2) Pepper jack cheese (2 slices) Ranch Dressing (2 Tbsp) French fries (4 oz) Shredded cheddar cheese (1/2 cup)	Cheese Pizza Lunchable w/ Airhead and Capri Sun (1 pkg)	1/4 white potato 1/2 medium carrot 1/4 white onion Golden Curry sauce mix (18g) Tofu (3 1/2 oz) Brown rice (160g) Honey (1/2 Tbsp) Diet Coke (12 fl oz)	1/4 white potato 1/2 medium carrot 1/4 white onion Golden Curry sauce mix (18g) Tofu (3 1/2 oz) Brown rice (160g) Honey (1/2 Tbsp) Diet Coke (12 fl oz)	1/4 white potato 1/2 medium carrot 1/4 white onion Golden Curry sauce mix (18g) Tofu (3 1/2 oz) Brown rice (160g) Honey (1/2 Tbsp) Diet Coke (12 fl oz)
Dinner	White Cheddar Popcorners (80g) Kettlecorn Popcorners (23 pieces)	Philly Veggie Sub, Lenny's (1) Diet Coke (20 fl oz) Jalapeno kettle cooked chips (1 pkg)	Celcius Energy Drink (12 fl oz) French baguette (6 inch) Jalapeno poppers (5 pieces) Shredded cheddar cheese (1/2 cup) Diet Coke (20 fl oz)	Diet Pepsi (30 fl oz) Spicy potato taco, Taco Bell (1) Cheesy Gordita Crunch, Taco Bell (1) Cheese quesadilla, Taco Bell (1) Chips & cheese, Taco Bell (1)	Celcius Energy Drink (12 fl oz) French baguette (6 inch) Jalapeno poppers (5 pieces) Shredded cheddar cheese (1/2 cup) Diet Coke (20 fl oz) Bang Energy Drink (12 fl oz) Soft dutch cocoa cookies (2) Kettlecorn Popcorners (46g)
Snacks	Water (50 fl oz) Coffee (10 fl oz) Diet Coke (24 fl oz)	Old fashioned cake donut (1) Glazed lemon donut holes (4) Diet Coke (14 fl oz) Saltine crackers (15 pieces)	Diet Coke (10 fl oz) Glazed lemon donut holes (2 pieces) Old fashioned cake donut (1 piece) Spree candy (135 pieces) Water (25 fl oz)	Water (26 fl oz)	Water (39 fl oz) Coffee (12 fl oz)
Misc.		Women's Multi-vitamin Iron (65mg) Vitamin D (2000 IU) Flax Seed Oil (1200mg)		Women's Multi-vitamin Iron (65mg) Vitamin D (2000 IU) Flax Seed Oil (1200mg)	

	Day 6 Date: 9/24/22	Day 7 Date: 9/25/22	Day 8 Date: 9/26/22	Day 9 Date: 9/27/22	Day 10 Date: 9/28/22
Breakfast	Saltine crackers (26 pieces) Soft dutch cocoa cookies (2) Chocolate protein shake (8 fl oz) Coffee (8 fl oz)	Guava Goddess kombucha (16 fl oz) S'mores greek yogurt (150g) Multigrain crispbread (39g) Lingonberry spread (2 Tbsp) Unsweetened applesaurce (191g)	Guava Goddess kombucha (16 fl oz) S'mores greek yogurt (150g) Egg whites (1/3 cup) Garlic powder (1 tsp) Onion powder (1 tsp) Unsweetened applesauce (228g) Water (13 fl oz)	S'mores greek yogurt (150g) Unsweetened applesauce (162 g) Water (13 fl oz)	Water (10 fl oz) Moose Tracks ice cream (2/3 cup) Vegetarian Buffalo Chik patty (1) Coffee (12 fl oz)
Lunch	Snapple Elements Rain (12 fl oz) White bread (2 slices) American cheese (2 slices) Butter (2 Tbsp) Seasoned french fries (6 oz) Fruit flavored jelly slices (224 1/2 g)	Pesto Tortellini microwave meal (1)	Mexican-style Caesar salad (150g) Vegetarian Buffalo Chik patty (1)	Mushroom veggie burger (1) American cheese (1 slice) Hamburger bun (1) Lettuce (1/10 cup) Red onion (1/4 cup) Mayo (1 Tbsp) Ketchup (1 Tbsp) Ranch dressing (2 Tbsp) Seasoned french fries (6 oz) Diet Coke (40 fl oz)	Light Life veggie dog (2) Apple Mountain Dew (12 fl oz) Ketchup (2 1/2 Tbsp)
Dinner	Reeses mini pumpkin (1) Cherry Coke Zero (12 fl oz) Vegan white cheddar chickpea puffs (224g)	S'mores greek yogurt (150g) Mountain Dew (12 fl oz) Mexican-style Caesar salad (150g) Vegetarian Buffalo Chik patty (1)	Four Cheese Hot Pocke (1) Buffalo bleu cheese puffs (84g) Cherry Coke Zero (24 fl oz)	Peanut M&Ms (93 g) Four Cheese Hot Pocket (1) Cherry Coke Zero (24 oz)	Peach Bellini (8 fl oz) Breadsticks, Olive Garden (5) Five cheese marinara dipping sauce (5 Tbsp) Fettuccine alfredo, Olive Garden (1) House salad, Olive Garden (3 cups) Water (20 fl oz)
Snacks	Water (26 fl oz) Sugar-free Gatorade (828 ml) Mountain Dew (12 fl oz) Soft dutch cocoa cookies (2) S'mores greek yogurt (150g)	Fruit flavored jelly slices (102g) S'mores greek yogurt (150g) Cherry Coke Zeros (24 fl oz) Water (13 fl oz)	Coffee (10 fl oz) Water (26 flo oz) Mountain Dew (12 fl oz) Moose Tracks ice cream (1 cup)	Mountain Dew (24 fl oz) Moose Tracks ice cream (1/2 cup) Water (16 fl oz)	Diet Coke (12 fl oz) Mountain Dew (12 fl oz) Water (13 fl oz) Mini Reeses cup (1)
Misc.	Olli Multivitamin (2 pieces)	Olli Multivitamin (2 pieces)	Olli Multivitamin (2 pieces)	Olli Multivitamin (2 pieces)	Olli Multivitamin (2 pieces)

Appendix B. Moisture Determination Methods

Appendix C. Detailed Methods for Metagenomic

C.1. DNA Extraction

Zymo Research Quick DNA Fecal/Soil Microbe Mini Prep Kit Cat# D6010.

1) Add 200 µL of fecal sample to ZR BashingBead™ Lysis Tube (0.1 & 0.5 mm).
2) Add 750 µl BashingBead™ Buffer to each tube.
3) Secure tubes in Vortex Genie Microtube adapter (Scientific Industries, Inc. Cat. No. S5001-7). NOTE: adapter should not spin
4) Set Vortex Genie to max and "on" and let run for 40 min continuously
5) Centrifuge the ZR BashingBead™ Lysis Tube (0.1 & 0.5 mm) in a microcentrifuge at ≥ 10,000 x *g* for 1 minute.
6) Transfer up to 400 µl supernatant to a Zymo-Spin™ III-F Filter in a Collection Tube and centrifuge at 8,000 x *g* for 1 minute.
7) Add 1,200 µl of Genomic Lysis Buffer to the filtrate in the Collection Tube from Step 4. Mix well.
8) Transfer 800 µl of the mixture from Step 5 to a Zymo-Spin™ IICR Column4 in a Collection Tube and centrifuge at 10,000 x *g* for 1 minute.
9) Discard the flow through from the Collection Tube and repeat Step 6.
10) Add 200 µl DNA Pre-Wash Buffer to the Zymo-Spin™ IICR Column in a new Collection Tube and centrifuge at 10,000 x *g* for 1 minute.
11) Add 500 µl g-DNA Wash Buffer to the Zymo-Spin™ IICR Column and centrifuge at 10,000 x *g* for 1 minute.
12) Transfer the Zymo-Spin™ IICR Column to a clean 1.5 ml microcentrifuge tube and add 100 µl (50 µl minimum) DNA Elution Buffer directly to the column matrix. Centrifuge at 10,000 x *g* for 30 s to elute the DNA
13) Place a Zymo-Spin™ III-HRC Filter in a clean Collection Tube and add 600 µl Prep Solution. Centrifuge at 8,000 x *g* for 3 min .
14) Transfer the eluted DNA to a prepared Zymo-Spin™ III-HRC Filter in a clean 1.5 ml microcentrifuge tube and centrifuge at exactly 16,000 x *g* for 3 min .
15) Store samples at 4 ° C.
16) Quantify samples using DeNovix HS dsDNA assay

C.2. 16S Amplicon Sequencing Library Preparation

16S rRNA V4 amplicon sequencing with Illumina unique dual indexes can be done all together

- Add 1 µL of DNA for each sample for the 16S PCR
- Make a master mix of the 16 PCR reaction since all of the input DNA volumes are the same, directions are below

- Primer pairs: 10 µmol/L pre-mixed V4 515F/806R primers with Illumina adapters
  - 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA-3'
  - 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT-3'

1. 16s PCR

a. Reaction: Prepare master mix (Kapa, water, primers) for **primer pair V4** (tube contains both F and R primers)

component	Volume (µL) per 25 µL reaction	Volume needed for 100 reaction Master Mix
Kapa HiFi 2x Master Mix	12.5	1.25 mL
primer pair (V4 II)	1	100 µL
sample	1	N/A
molecular grade water	10.5	1.05 mL

a. Thermocycler conditions -in B255 labeled as KapaV13_16s
   - 95.0 °C – 3 min
   - 18 cycles (17*x repeats*)
     a. 98.0 °C – 30 s
     b. 55.0 °C – 15 s
     c. 72.0 °C – 20 s
   - 72.0 °C – 5 min
   - 4.0 °C – hold

b. Clean-up- Use SPRIselect beads to remove primers
   i. Add beads at a 0.8:1 ratio (20 µL of beads for 25 µL reaction).
   ii. Let sit 1 minute.
   iii. Place on magnet, let sit 1 minute.
   iv. Remove liquid, make sure not to remove any beads (orange stuff).
   v. Keep plate on magnet; add 180ul of 80 % molecular grade ethanol.
   vi. Let sit for 30 s; remove ethanol - first with 200 µL pipette (green tip) set to 180 µL, then with 20 µL (red tip) pipette set to 20 µL.
   vii. Remove from magnet, and add 40 µL of RNase/DNase free molecular biology grade water.
   viii. Allow to sit for 1 minute, gently tap to make sure everything is mixed
   ix. Place on magnet and allow to sit for 1 minute.
   x. Carefully remove water without any beads **(safe stopping point).**

2. Indexing PCR using unique dual index plate from **Illumina** (ex: IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)) (not NEB).
   a. PCR with KapaHiFi
      a. Reaction

component	µL	Notes
Kapa HiFi 2x Master Mix	12.5	

N7** / S5** (adapters)	10	
DNA	2.5	From cleaned-up 16S PCR reaction

b. Thermocycler conditions
   95.0 °C – 3 min
   8 cycles (7x repeats)
      98.0 °C – 30 s
      55.0 °C – 15 s
      72.0 °C – 20 s
   72.0 °C – 5 min
   4.0 °C – hold

b. Clean-up- Use SPRIselect beads to remove primers
   a. Add beads at a 1:1 ratio (25 µL of beads for 25 µL reaction).
   b. Let sit 1 minute.
   c. Place on magnet, let sit 1 minute.
   d. Remove liquid; make sure not to remove any beads (orange stuff).
   e. Keep plate on magnet; add 180 µL of 80 % molecular grade ethanol.
   f. Let sit for 30s; remove ethanol- first with 200 µL pipette (green tip) set to 180 µL, then with 20 µL (red tip) pipette set to 20 µ.L
      i. Remove from magnet, and add 40 µL of RNase/DNase free molecular biology grade water.
      ii. Allow to sit for 1 minute; gently tap to make sure everything is mixed,
      iii. Place on magnet, and allow to sit for 1 minute.
      iv. Carefully remove water without any beads **(safe stopping point).**

Appendix D. Summary of Taxonomic Data

D.1. Table. Summary of Genera Identified in Homogeneity Data Set[1]

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Blautia	11.423 %	9.609 %-12.476 %	18.533 %	14.866 %-24.446 %
Collinsella	18.769 %	16.513 %-22.117 %	9.236 %	5.041 %-11.820 %
Bifidobacterium	7.484 %	6.514 %-9.127 %	11.963 %	8.425 %-14.871 %
Faecalibacterium	4.461 %	4.072 %-4.833 %	4.592 %	4.085 %-5.164 %
Romboutsia	4.450 %	3.742 %-5.443 %	4.050 %	3.607 %-4.907 %
Subdoligranulum	4.756 %	4.373 %-5.224 %	3.694 %	3.188 %-4.237 %
Dorea	3.361 %	2.798 %-3.775 %	2.849 %	1.995 %-3.208 %
Unclassified	2.716 %	2.427 %-2.946 %	2.980 %	0.443 %-3.305 %
Anaerostipes	1.534 %	1.323 %-1.751 %	3.231 %	2.651 %-4.096 %
Agathobacter	1.368 %	1.040 %-1.650 %	3.271 %	2.326 %-4.313 %
[*Eubacterium*] *hallii* group	2.014 %	1.596 %-2.632 %	2.102 %	1.784 %-2.689 %
Dialister	0.929 %	0.781 %-1.094 %	2.916 %	2.359 %-3.665 %
[*Ruminococcus*] *torques* group	2.943 %	2.346 %-3.300 %	0.812 %	0.734 %-1.275 %
Fusicatenibacter	0.915 %	0.649 %-1.250 %	2.608 %	2.030 %-3.603 %
Lachnospiraceae ND3007 group	1.233 %	0.922 %-1.485 %	2.127 %	1.895 %-2.452 %
Methanobrevibacter	2.445 %	2.080 %-2.834 %	0.529 %	0.292 %-0.725 %
Streptococcus	0.408 %	0.319 %-0.455 %	2.465 %	1.691 %-2.963 %
[*Ruminococcus*] *gauvreauii* group	1.727 %	1.270 %-1.940 %	1.131 %	0.865 %-1.497 %
Deinococcus	1.644 %	1.377 %-1.908 %	1.133 %	0.554 %-1.407 %
Incertae Sedis	2.047 %	1.901 %-2.224 %	0.697 %	0.584 %-0.806 %
Clostridium sensu stricto 1	2.568 %	2.084 %-3.317 %	0.044 %	0.025 %-0.061 %
Coprococcus	0.765 %	0.563 %-1.007 %	1.621 %	1.278 %-2.439 %
Ruminococcus	2.114 %	1.858 %-2.390 %	0.221 %	0.159 %-0.263 %
Holdemanella	0.414 %	0.355 %-0.465 %	1.697 %	1.298 %-2.047 %
Megamonas	ND	ND	1.929 %	1.109 %-2.749 %
UCG-002	1.784 %	1.573 %-2.129 %	0.008 %	0.003 %-0.018 %
Akkermansia	0.701 %	0.561 %-0.808 %	1.059 %	0.610 %-1.453 %
Monoglobus	0.337 %	0.283 %-0.405 %	1.182 %	0.388 %-1.352 %
Butyricicoccus	0.435 %	0.358 %-0.495 %	1.022 %	0.905 %-1.167 %
Catenibacterium	1.429 %	1.229 %-1.656 %	ND	ND

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Delftia	0.724 %	0.557 %-0.908 %	0.523 %	0.358 %-0.742 %
Slackia	0.137 %	0.104 %-0.172 %	1.051 %	0.770 %-1.292 %
Christensenellaceae R-7 group	0.643 %	0.526 %-0.748 %	0.527 %	0.436 %-0.618 %
Roseburia	0.505 %	0.322 %-0.634 %	0.547 %	0.299 %-0.907 %
Prevotella	0.441 %	0.321 %-0.610 %	0.568 %	0.307 %-0.963 %
Family XIII AD3011 group	0.824 %	0.594 %-1.076 %	0.104 %	0.037 %-0.144 %
Senegalimassilia	0.925 %	0.851 %-1.054 %	ND	ND
Methanosphaera	0.315 %	0.258 %-0.353 %	0.564 %	0.326 %-0.805 %
Intestinimonas	0.864 %	0.756 %-0.999 %	0.011 %	0.011 %-0.011 %
Phascolarctobacterium	0.272 %	0.185 %-0.352 %	0.588 %	0.443 %-0.713 %
Peptoclostridium	ND	ND	0.819 %	0.684 %-0.921 %
Bacteroides	0.426 %	0.324 %-0.536 %	0.320 %	0.179 %-0.618 %
Marvinbryantia	0.559 %	0.451 %-0.658 %	0.167 %	0.080 %-0.238 %
Dickeya	0.403 %	0.286 %-0.556 %	0.290 %	0.191 %-0.401 %
NK4A214 group	0.467 %	0.385 %-0.594 %	0.152 %	0.121 %-0.187 %
Erysipelotrichaceae UCG-003	0.401 %	0.334 %-0.463 %	0.197 %	0.128 %-0.260 %
Mogibacterium	0.575 %	0.487 %-0.694 %	ND	ND
Lactobacillus	0.037 %	0.013 %-0.053 %	0.533 %	0.333 %-0.647 %
Lachnospiraceae FCS020 group	0.357 %	0.233 %-0.462 %	0.188 %	0.136 %-0.252 %
Adlercreutzia	0.039 %	0.026 %-0.056 %	0.487 %	0.381 %-0.593 %
Lachnoclostridium	0.076 %	0.016 %-0.110 %	0.403 %	0.286 %-0.592 %
Libanicoccus	0.417 %	0.328 %-0.471 %	ND	ND
Enterorhabdus	0.332 %	0.268 %-0.387 %	0.005 %	0.004 %-0.006 %
Peptococcus	0.326 %	0.284 %-0.380 %	ND	ND
[*Ruminococcus*] *gnavus* group	ND	ND	0.306 %	0.233 %-0.377 %
Parabacteroides	0.273 %	0.188 %-0.382 %	0.030 %	0.008 %-0.065 %
Faecalitalea	0.291 %	0.214 %-0.348 %	0.011 %	0.007 %-0.019 %
Turicibacter	0.266 %	0.177 %-0.370 %	0.025 %	0.017 %-0.037 %
Coriobacteriaceae UCG-003	0.269 %	0.218 %-0.323 %	ND	ND
UCG-005	0.202 %	0.154 %-0.249 %	0.032 %	0.012 %-0.063 %
Alloprevotella	0.128 %	0.091 %-0.172 %	0.079 %	0.048 %-0.118 %
Eggerthella	0.070 %	0.042 %-0.114 %	0.113 %	0.071 %-0.333 %
Shuttleworthia	0.007 %	0.006 %-0.008 %	0.174 %	0.129 %-0.224 %
Lachnospiraceae NK4A136 group	0.067 %	0.030 %-0.107 %	0.113 %	0.058 %-0.164 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Negativibacillus	0.098 %	0.060 %-0.147 %	0.071 %	0.037 %-0.104 %
Actinomyces	0.035 %	0.022 %-0.060 %	0.119 %	0.071 %-0.169 %
Howardella	0.081 %	0.061 %-0.110 %	0.068 %	0.045 %-0.092 %
Tuzzerella	ND	ND	0.147 %	0.102 %-0.277 %
CAG-352	ND	ND	0.140 %	0.105 %-0.179 %
Gordonibacter	0.008 %	0.003 %-0.015 %	0.124 %	0.082 %-0.159 %
Erysipelatoclostridium	0.032 %	0.015 %-0.050 %	0.081 %	0.043 %-0.113 %
Lachnospiraceae UCG-010	0.062 %	0.038 %-0.082 %	0.043 %	0.021 %-0.060 %
Alistipes	0.096 %	0.055 %-0.126 %	0.006 %	0.002 %-0.014 %
[*Eubacterium*] *siraeum* group	0.045 %	0.023 %-0.066 %	0.051 %	0.037 %-0.068 %
CHKCI002	0.086 %	0.053 %-0.134 %	ND	ND
[*Eubacterium*] *fissicatena* group	0.050 %	0.033 %-0.078 %	0.031 %	0.011 %-0.049 %
Colidextribacter	0.036 %	0.013 %-0.053 %	0.042 %	0.012 %-0.061 %
[*Eubacterium*] *ruminantium* group	0.076 %	0.049 %-0.104 %	ND	ND
UCG-003	0.022 %	0.008 %-0.036 %	0.049 %	0.026 %-0.081 %
Lachnospira	0.015 %	0.004 %-0.021 %	0.046 %	0.016 %-0.084 %
UBA1819	0.020 %	0.010 %-0.036 %	0.037 %	0.024 %-0.049 %
Lachnotalea	0.057 %	0.036 %-0.087 %	ND	ND
Intestinibacter	0.037 %	0.022 %-0.064 %	0.018 %	0.011 %-0.038 %
DTU089	0.044 %	0.026 %-0.066 %	0.007 %	0.004 %-0.014 %
Escherichia-Shigella	0.040 %	0.027 %-0.058 %	0.010 %	0.003 %-0.019 %
Eisenbergiella	0.033 %	0.021 %-0.056 %	0.008 %	0.006 %-0.010 %
Family XIII UCG-001	0.014 %	0.006 %-0.023 %	0.027 %	0.013 %-0.039 %
Olsenella	0.041 %	0.027 %-0.056 %	ND	ND
Tyzzerella	0.034 %	0.020 %-0.054 %	0.004 %	0.002 %-0.007 %
Rothia	0.013 %	0.007 %-0.020 %	0.024 %	0.012 %-0.039 %
Allisonella	0.008 %	0.005 %-0.018 %	0.028 %	0.005 %-0.064 %
Lactococcus	ND	ND	0.036 %	0.021 %-0.061 %
CAG-56	0.014 %	0.009 %-0.024 %	0.022 %	0.006 %-0.038 %
Fournierella	0.021 %	0.008 %-0.033 %	0.015 %	0.008 %-0.021 %
Acidaminococcus	0.011 %	0.009 %-0.013 %	0.024 %	0.009 %-0.048 %
Lachnospiraceae AC2044 group	0.016 %	0.010 %-0.027 %	0.019 %	0.009 %-0.037 %
[*Eubacterium*] *eligens* group	0.016 %	0.007 %-0.034 %	0.015 %	0.007 %-0.023 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
[*Bacteroides*] *pectinophilus* group	ND	ND	0.031 %	0.011 %-0.047 %
Oscillibacter	0.008 %	0.003 %-0.012 %	0.021 %	0.013 %-0.034 %
Angelakisella	0.020 %	0.011 %-0.036 %	0.008 %	0.004 %-0.018 %
[*Eubacterium*] *brachy* group	0.009 %	0.004 %-0.014 %	0.015 %	0.005 %-0.030 %
Flavonifractor	0.011 %	0.007 %-0.020 %	0.013 %	0.007 %-0.021 %
Lachnospiraceae UCG-004	0.011 %	0.006 %-0.017 %	0.013 %	0.009 %-0.019 %
[*Clostridium*] *innocuum* group	0.009 %	0.005 %-0.021 %	0.013 %	0.005 %-0.022 %
Holdemania	0.009 %	0.003 %-0.019 %	0.013 %	0.004 %-0.022 %
Veillonella	0.008 %	0.003 %-0.012 %	0.013 %	0.007 %-0.032 %
[*Eubacterium*] *ventriosum* group	0.008 %	0.004 %-0.015 %	0.013 %	0.004 %-0.023 %
Granulicatella	0.005 %	0.003 %-0.009 %	0.015 %	0.006 %-0.026 %
[*Eubacterium*] *xylanophilum* group	0.007 %	0.005 %-0.015 %	0.011 %	0.003 %-0.017 %
Sellimonas	ND	ND	0.018 %	0.005 %-0.038 %
Rikenellaceae RC9 gut group	0.006 %	0.002 %-0.011 %	0.011 %	0.003 %-0.023 %
Candidatus Stoquefichus	ND	ND	0.017 %	0.010 %-0.025 %
Butyrivibrio	ND	ND	0.016 %	0.009 %-0.032 %
Paraprevotella	0.008 %	0.005 %-0.013 %	0.008 %	0.002 %-0.024 %
Hungatella	0.015 %	0.011 %-0.022 %	ND	ND
Raoultibacter	0.009 %	0.005 %-0.014 %	0.006 %	0.004 %-0.007 %
Lachnospiraceae NK4B4 group	ND	ND	0.015 %	0.006 %-0.028 %
Christensenella	0.010 %	0.004 %-0.014 %	0.004 %	0.002 %-0.007 %
Defluviitaleaceae UCG-011	0.006 %	0.002 %-0.013 %	0.008 %	0.003 %-0.013 %
Bilophila	0.011 %	0.005 %-0.021 %	0.003 %	0.001 %-0.005 %
Desulfovibrio	0.012 %	0.004 %-0.026 %	0.002 %	0.001 %-0.002 %
Solobacterium	0.010 %	0.006 %-0.021 %	0.003 %	0.001 %-0.005 %
Acinetobacter	0.006 %	0.001 %-0.012 %	0.007 %	0.001 %-0.034 %
Haemophilus	0.008 %	0.003 %-0.014 %	0.005 %	0.005 %-0.005 %
Enorma	0.013 %	0.003 %-0.026 %	ND	ND
Anaerofustis	0.003 %	0.002 %-0.005 %	0.010 %	0.004 %-0.020 %
Corynebacterium	0.004 %	0.002 %-0.006 %	0.009 %	0.004 %-0.016 %
Terrisporobacter	0.013 %	0.009 %-0.019 %	ND	ND
DNF00809	0.013 %	0.008 %-0.017 %	ND	ND
Methylobacterium-Methylorubrum	0.006 %	0.001 %-0.019 %	0.006 %	0.002 %-0.013 %

Genus	Omnivore	Vegetarian	Genus	Omnivore
	Average Relative Abundance	Range		Average Relative Abundance
Anaerotruncus	0.005 %	0.004 %-0.006 %	0.007 %	0.002 %-0.013 %
Terrabacter	0.005 %	0.004 %-0.007 %	0.007 %	0.007 %-0.007 %
Paludicola	0.008 %	0.004 %-0.013 %	0.004 %	0.004 %-0.004 %
UCG-009	0.008 %	0.004 %-0.012 %	0.003 %	0.003 %-0.003 %
Catenibacillus	ND	ND	0.011 %	0.004 %-0.025 %
Cloacibacillus	0.010 %	0.006 %-0.016 %	ND	ND
Candidatus Soleaferrea	0.006 %	0.004 %-0.010 %	0.005 %	0.002 %-0.007 %
Mitsuokella	0.010 %	0.003 %-0.022 %	ND	ND
Denitrobacterium	0.010 %	0.004 %-0.017 %	ND	ND
Erysipelotrichaceae UCG-006	0.010 %	0.005 %-0.016 %	ND	ND
Sphingomonas	0.005 %	0.002 %-0.012 %	0.005 %	0.002 %-0.008 %
Gemella	0.002 %	0.002 %-0.002 %	0.007 %	0.003 %-0.014 %
Eubacterium	0.009 %	0.004 %-0.017 %	ND	ND
Roseomonas	0.005 %	0.002 %-0.010 %	0.004 %	0.001 %-0.008 %
Sutterella	0.005 %	0.004 %-0.007 %	0.002 %	0.002 %-0.004 %
Oxalobacter	0.008 %	0.002 %-0.015 %	ND	ND
Pygmaiobacter	ND	ND	0.008 %	0.003 %-0.012 %
Catabacter	0.008 %	0.006 %-0.010 %	ND	ND
Barnesiella	0.004 %	0.002 %-0.008 %	0.003 %	0.001 %-0.004 %
Lachnospiraceae UCG-001	0.007 %	0.007 %-0.007 %	ND	ND
TM7x	0.003 %	0.002 %-0.005 %	0.004 %	0.002 %-0.005 %
Odoribacter	0.004 %	0.001 %-0.012 %	0.002 %	0.002 %-0.004 %
Prevotellaceae UCG-001	0.006 %	0.003 %-0.011 %	ND	ND
Prevotellaceae NK3B31 group	0.006 %	0.003 %-0.012 %	ND	ND
Butyricimonas	0.003 %	0.002 %-0.005 %	0.003 %	0.002 %-0.004 %
Varibaculum	ND	ND	0.006 %	0.002 %-0.008 %
Lawsonella	0.005 %	0.005 %-0.005 %	ND	ND
Lactonifactor	ND	ND	0.005 %	0.004 %-0.007 %
Parasutterella	0.003 %	0.002 %-0.005 %	0.003 %	0.001 %-0.005 %
Atopobium	0.002 %	0.002 %-0.003 %	0.003 %	0.002 %-0.006 %
UCG-007	ND	ND	0.005 %	0.005 %-0.005 %
Leuconostoc	ND	ND	0.005 %	0.002 %-0.008 %
Enterococcus	ND	ND	0.005 %	0.005 %-0.005 %
UC5-1-2E3	0.004 %	0.004 %-0.005 %	ND	ND
Fusobacterium	0.003 %	0.003 %-0.003 %	0.002 %	0.001 %-0.002 %

Genus	Omnivore	Vegetarian	Genus	Omnivore
	Average Relative Abundance	Range		Average Relative Abundance
Dielma	0.004 %	0.002 %-0.006 %	ND	ND
Acetanaerobacterium	ND	ND	0.004 %	0.002 %-0.007 %
Faecalicoccus	0.003 %	0.003 %-0.003 %	ND	ND
Porphyromonas	0.002 %	0.001 %-0.002 %	0.001 %	0.001 %-0.001 %
F0332	ND	ND	0.003 %	0.002 %-0.004 %
Coprobacter	0.002 %	0.002 %-0.002 %	ND	ND
Pseudopropionibacterium	ND	ND	0.002 %	0.002 %-0.002 %
[*Eubacterium*] *nodatum* group	ND	ND	0.002 %	0.002 %-0.002 %
Propionibacterium	ND	ND	0.002 %	0.002 %-0.002 %
Peptostreptococcus	ND	ND	0.002 %	0.002 %-0.002 %
Cystobacter	0.001 %	0.001 %-0.001 %	ND	ND

[1]ND corresponds to a relative abundance value of less than 0.001 %
[2] Purple filled rows indicate genera present at higher relative abundance (≥1 % difference between average relative abundance for omnivore and vegetarian samples) in omnivores.
[3] Green filled rows indicate genera present at higher relative abundance (≥1 % difference between average relative abundance for omnivore and vegetarian samples) in vegetarians

D.2. Table. Summary of Genera Identified in Stability Data Set[1]

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Collinsella	19.084 %	14.717 %-25.049 %	12.122 %	9.466 %-15.452 %
Blautia	11.435 %	7.676 %-12.921 %	19.524 %	16.040 %-22.175 %
Bifidobacterium	7.558 %	5.914 %-9.488 %	10.907 %	8.378 %-14.582 %
Romboutsia	4.844 %	2.590 %-5.967 %	4.168 %	2.774 %-5.300 %
Faecalibacterium	4.247 %	3.057 %-5.634 %	4.133 %	3.273 %-4.902 %
Subdoligranulum	4.891 %	3.951 %-5.740 %	3.386 %	2.866 %-4.014 %
Dorea	3.292 %	2.516 %-4.128 %	3.003 %	2.503 %-3.626 %
Unclassified	2.964 %	2.303 %-3.841 %	2.772 %	2.203 %-3.345 %
Agathobacter	1.367 %	0.730 %-2.059 %	3.520 %	2.570 %-5.549 %
Anaerostipes	1.482 %	1.081 %-1.817 %	3.200 %	2.451 %-3.961 %
[*Eubacterium*] *hallii* group	2.153 %	1.495 %-2.892 %	2.165 %	1.618 %-2.702 %
Dialister	0.892 %	0.555 %-1.192 %	3.017 %	2.362 %-3.800 %
[*Ruminococcus*] *torques* group	2.931 %	1.867 %-3.379 %	0.886 %	0.708 %-1.042 %
Fusicatenibacter	0.893 %	0.550 %-1.434 %	2.646 %	2.131 %-3.169 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Lachnospiraceae ND3007 group	1.230 %	0.891 %-1.515 %	2.070 %	1.358 %-2.500 %
Catenibacterium	1.535 %	1.217 %-2.038 %	ND	ND
Streptococcus	0.421 %	0.302 %-0.517 %	2.549 %	1.920 %-3.050 %
[*Ruminococcus*] *gauvreauii* group	1.711 %	1.080 %-2.133 %	1.201 %	0.988 %-1.476 %
Methanobrevibacter	2.385 %	1.667 %-3.040 %	0.523 %	0.361 %-0.770 %
Clostridium sensu stricto 1	2.634 %	1.703 %-3.426 %	0.041 %	0.023 %-0.063 %
Incertae Sedis	2.017 %	1.696 %-2.257 %	0.645 %	0.518 %-0.731 %
Coprococcus	0.761 %	0.497 %-1.306 %	1.657 %	1.317 %-2.094 %
Ruminococcus	2.125 %	1.587 %-2.407 %	0.203 %	0.126 %-0.265 %
Deinococcus	1.402 %	0.959 %-1.703 %	0.918 %	0.589 %-1.110 %
Senegalimassilia	0.994 %	0.766 %-1.216 %	ND	ND
Holdemanella	0.446 %	0.324 %-0.634 %	1.506 %	1.220 %-1.834 %
UCG-002	1.776 %	1.005 %-2.532 %	0.004 %	0.001 %-0.011 %
Peptoclostridium	ND	ND	0.834 %	0.574 %-1.119 %
Akkermansia	0.704 %	0.533 %-0.915 %	0.816 %	0.661 %-1.214 %
Megamonas	0.013 %	0.002 %-0.051 %	1.505 %	0.577 %-2.678 %
Monoglobus	0.334 %	0.252 %-0.401 %	1.178 %	0.985 %-1.318 %
Butyricicoccus	0.423 %	0.310 %-0.515 %	1.018 %	0.798 %-1.182 %
Slackia	0.138 %	0.091 %-0.182 %	1.004 %	0.842 %-1.243 %
Christensenellaceae R-7 group	0.656 %	0.377 %-0.804 %	0.461 %	0.379 %-0.592 %
Family XIII AD3011 group	0.865 %	0.584 %-1.162 %	0.118 %	0.059 %-0.168 %
Delftia	0.659 %	0.399 %-0.973 %	0.294 %	0.100 %-0.450 %
Libanicoccus	0.432 %	0.318 %-0.557 %	ND	ND
Intestinimonas	0.851 %	0.600 %-1.139 %	0.005 %	0.002 %-0.010 %
Roseburia	0.365 %	0.219 %-0.518 %	0.480 %	0.189 %-0.881 %
Methanosphaera	0.291 %	0.187 %-0.402 %	0.527 %	0.308 %-0.659 %
Marvinbryantia	0.555 %	0.269 %-0.721 %	0.190 %	0.082 %-0.249 %
Prevotella	0.440 %	0.140 %-0.844 %	0.259 %	0.094 %-0.597 %
Phascolarctobacterium	0.227 %	0.085 %-0.367 %	0.441 %	0.291 %-0.620 %
Erysipelotrichaceae UCG-003	0.431 %	0.325 %-0.568 %	0.197 %	0.147 %-0.276 %
NK4A214 group	0.466 %	0.287 %-0.656 %	0.145 %	0.092 %-0.207 %
Mogibacterium	0.599 %	0.423 %-0.798 %	0.003 %	0.002 %-0.006 %
Coriobacteriaceae UCG-003	0.280 %	0.195 %-0.396 %	ND	ND
Dickeya	0.367 %	0.137 %-0.522 %	0.187 %	0.104 %-0.270 %
Lactobacillus	0.037 %	0.012 %-0.062 %	0.506 %	0.343 %-0.653 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Bacteroides	0.417 %	0.238 %-0.622 %	0.124 %	0.037 %-0.254 %
Lachnospiraceae FCS020 group	0.353 %	0.173 %-0.513 %	0.186 %	0.104 %-0.252 %
Adlercreutzia	0.040 %	0.019 %-0.059 %	0.483 %	0.348 %-0.630 %
Lachnoclostridium	0.070 %	0.014 %-0.131 %	0.393 %	0.229 %-0.664 %
Enterorhabdus	0.351 %	0.280 %-0.528 %	0.002 %	0.002 %-0.004 %
[*Ruminococcus*] *gnavus* group	0.004 %	0.002 %-0.010 %	0.343 %	0.272 %-0.415 %
Faecalitalea	0.307 %	0.188 %-0.411 %	0.011 %	0.005 %-0.024 %
Peptococcus	0.315 %	0.224 %-0.418 %	0.000 %	0.000 %-0.000 %
Turicibacter	0.258 %	0.100 %-0.450 %	0.025 %	0.013 %-0.045 %
Parabacteroides	0.247 %	0.152 %-0.388 %	0.012 %	0.001 %-0.027 %
Tuzzerella	ND	ND	0.124 %	0.063 %-0.199 %
UCG-005	0.208 %	0.111 %-0.357 %	0.032 %	0.007 %-0.085 %
Lachnospiraceae NK4A136 group	0.071 %	0.020 %-0.118 %	0.125 %	0.025 %-0.184 %
Shuttleworthia	0.004 %	0.003 %-0.009 %	0.191 %	0.122 %-0.325 %
CHKCI002	0.089 %	0.053 %-0.143 %	ND	ND
Eggerthella	0.069 %	0.036 %-0.101 %	0.099 %	0.039 %-0.154 %
Alloprevotella	0.128 %	0.052 %-0.222 %	0.038 %	0.010 %-0.078 %
Howardella	0.088 %	0.051 %-0.116 %	0.068 %	0.040 %-0.096 %
Actinomyces	0.039 %	0.018 %-0.068 %	0.111 %	0.075 %-0.252 %
Negativibacillus	0.091 %	0.049 %-0.132 %	0.059 %	0.021 %-0.105 %
[*Eubacterium*] *ruminantium* group	0.068 %	0.019 %-0.142 %	ND	ND
Gordonibacter	0.009 %	0.003 %-0.020 %	0.120 %	0.070 %-0.180 %
Erysipelatoclostridium	0.035 %	0.012 %-0.056 %	0.081 %	0.044 %-0.102 %
CAG-352	0.003 %	0.003 %-0.003 %	0.110 %	0.070 %-0.174 %
Alistipes	0.102 %	0.023 %-0.193 %	0.004 %	0.001 %-0.009 %
Lachnospiraceae UCG-010	0.060 %	0.018 %-0.104 %	0.035 %	0.010 %-0.057 %
Terrabacter	0.052 %	0.002 %-0.645 %	0.041 %	0.001 %-0.644 %
[*Eubacterium*] *siraeum* group	0.042 %	0.017 %-0.072 %	0.032 %	0.015 %-0.067 %
Colidextribacter	0.037 %	0.011 %-0.081 %	0.036 %	0.011 %-0.065 %
Intestinibacter	0.037 %	0.018 %-0.065 %	0.020 %	0.010 %-0.034 %
Lachnospira	0.016 %	0.005 %-0.034 %	0.039 %	0.014 %-0.092 %
[*Bacteroides*] *pectinophilus* group	ND	ND	0.027 %	0.013 %-0.056 %
UBA1819	0.020 %	0.010 %-0.037 %	0.033 %	0.018 %-0.057 %
Methylobacterium-Methylorubrum	0.020 %	0.001 %-0.153 %	0.031 %	0.001 %-0.300 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
DTU089	0.044 %	0.024 %-0.076 %	0.005 %	0.001 %-0.011 %
UCG-003	0.022 %	0.007 %-0.046 %	0.026 %	0.009 %-0.051 %
Escherichia-Shigella	0.039 %	0.018 %-0.068 %	0.008 %	0.003 %-0.016 %
Microbacterium	0.042 %	0.001 %-0.237 %	0.002 %	0.001 %-0.002 %
Olsenella	0.040 %	0.019 %-0.064 %	0.002 %	0.001 %-0.004 %
Eisenbergiella	0.034 %	0.015 %-0.051 %	0.006 %	0.002 %-0.010 %
Family XIII UCG-001	0.013 %	0.003 %-0.040 %	0.024 %	0.007 %-0.047 %
CAG-56	0.015 %	0.005 %-0.035 %	0.021 %	0.004 %-0.057 %
Rothia	0.012 %	0.004 %-0.040 %	0.023 %	0.008 %-0.044 %
Fournierella	0.021 %	0.006 %-0.045 %	0.013 %	0.006 %-0.021 %
Lactococcus	0.002 %	0.001 %-0.003 %	0.033 %	0.015 %-0.047 %
Staphylococcus	0.003 %	0.001 %-0.005 %	0.028 %	0.028 %-0.028 %
[*Eubacterium*] *brachy* group	0.010 %	0.005 %-0.019 %	0.019 %	0.009 %-0.029 %
Leptotrichia	0.014 %	0.014 %-0.014 %	ND	ND
Acinetobacter	0.017 %	0.002 %-0.131 %	0.012 %	0.001 %-0.070 %
Candidatus Stoquefichus	ND	ND	0.014 %	0.006 %-0.027 %
Lachnospiraceae AC2044 group	0.014 %	0.007 %-0.035 %	0.013 %	0.004 %-0.032 %
Acidaminococcus	0.011 %	0.005 %-0.018 %	0.015 %	0.003 %-0.044 %
Enorma	0.013 %	0.003 %-0.024 %	ND	ND
[*Eubacterium*] *eligens* group	0.013 %	0.004 %-0.038 %	0.012 %	0.003 %-0.026 %
Angelakisella	0.018 %	0.008 %-0.031 %	0.007 %	0.003 %-0.015 %
Allisonella	0.010 %	0.003 %-0.025 %	0.015 %	0.001 %-0.044 %
Terrisporobacter	0.012 %	0.004 %-0.019 %	ND	ND
Butyrivibrio	ND	ND	0.012 %	0.004 %-0.041 %
Holdemania	0.008 %	0.002 %-0.018 %	0.014 %	0.005 %-0.084 %
Oscillibacter	0.008 %	0.002 %-0.019 %	0.015 %	0.005 %-0.033 %
[*Clostridium*] *innocuum* group	0.009 %	0.004 %-0.023 %	0.013 %	0.004 %-0.032 %
Lachnospiraceae UCG-004	0.010 %	0.003 %-0.018 %	0.011 %	0.003 %-0.025 %
Granulicatella	0.004 %	0.002 %-0.009 %	0.017 %	0.007 %-0.026 %
Flavonifractor	0.010 %	0.003 %-0.020 %	0.011 %	0.003 %-0.028 %
Sphingomonas	0.012 %	0.001 %-0.042 %	0.008 %	0.001 %-0.021 %
Sellimonas	0.003 %	0.001 %-0.004 %	0.018 %	0.003 %-0.050 %
Cloacibacillus	0.010 %	0.003 %-0.029 %	ND	ND
Eubacterium	0.010 %	0.005 %-0.021 %	ND	ND
Denitrobacterium	0.010 %	0.004 %-0.026 %	ND	ND

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Erysipelotrichaceae UCG-006	0.010 %	0.005 %-0.022 %	ND	ND
[*Eubacterium*] *fissicatena* group	0.005 %	0.002 %-0.010 %	0.014 %	0.008 %-0.026 %
DNF00809	0.010 %	0.003 %-0.022 %	ND	ND
Rikenellaceae RC9 gut group	0.010 %	0.003 %-0.023 %	0.009 %	0.003 %-0.021 %
Catenibacillus	ND	ND	0.009 %	0.003 %-0.020 %
Veillonella	0.007 %	0.002 %-0.018 %	0.011 %	0.004 %-0.019 %
Mitsuokella	0.009 %	0.003 %-0.025 %	ND	ND
[*Eubacterium*] *ventriosum* group	0.007 %	0.002 %-0.023 %	0.011 %	0.001 %-0.031 %
Defluviitaleaceae UCG-011	0.005 %	0.001 %-0.012 %	0.010 %	0.002 %-0.017 %
Hungatella	0.011 %	0.005 %-0.031 %	0.003 %	0.003 %-0.003 %
Enterococcus	0.007 %	0.005 %-0.012 %	ND	ND
Anaerofustis	0.002 %	0.001 %-0.004 %	0.012 %	0.003 %-0.018 %
Desulfovibrio	0.012 %	0.003 %-0.036 %	0.002 %	0.001 %-0.005 %
Bilophila	0.010 %	0.003 %-0.021 %	0.003 %	0.001 %-0.009 %
[*Eubacterium*] *xylanophilum* group	0.006 %	0.002 %-0.014 %	0.007 %	0.003 %-0.017 %
Candidatus Soleaferrea	0.007 %	0.002 %-0.014 %	0.005 %	0.002 %-0.013 %
Solobacterium	0.010 %	0.003 %-0.028 %	0.002 %	0.001 %-0.004 %
Paraprevotella	0.008 %	0.003 %-0.018 %	0.004 %	0.001 %-0.011 %
Pygmaiobacter	0.005 %	0.002 %-0.010 %	0.007 %	0.003 %-0.015 %
Lachnospiraceae NK4B4 group	0.002 %	0.002 %-0.002 %	0.010 %	0.002 %-0.030 %
Corynebacterium	0.003 %	0.001 %-0.009 %	0.008 %	0.003 %-0.015 %
Christensenella	0.009 %	0.003 %-0.021 %	0.003 %	0.001 %-0.005 %
Raoultibacter	0.009 %	0.004 %-0.017 %	0.002 %	0.001 %-0.003 %
Prevotellaceae NK3B31 group	0.005 %	0.002 %-0.016 %	ND	ND
UCG-009	0.009 %	0.002 %-0.025 %	0.002 %	0.001 %-0.004 %
Prevotellaceae UCG-001	0.005 %	0.002 %-0.013 %	ND	ND
Pseudobutyrivibrio	0.005 %	0.003 %-0.006 %	ND	ND
Paludicola	0.006 %	0.003 %-0.014 %	0.003 %	0.001 %-0.006 %
Gemella	0.002 %	0.001 %-0.005 %	0.007 %	0.003 %-0.016 %
Oxalobacter	0.009 %	0.003 %-0.017 %	0.000 %	0.000 %-0.000 %
Catabacter	0.006 %	0.002 %-0.015 %	0.002 %	0.002 %-0.002 %
Haemophilus	0.007 %	0.002 %-0.025 %	0.001 %	0.001 %-0.003 %
Anaerotruncus	0.004 %	0.002 %-0.006 %	0.004 %	0.001 %-0.008 %
Pseudomonas	0.004 %	0.004 %-0.004 %	ND	ND

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
Lactonifactor	0.004 %	0.002 %-0.012 %	0.004 %	0.002 %-0.008 %
Dielma	0.004 %	0.001 %-0.010 %	ND	ND
UC5-1-2E3	0.004 %	0.002 %-0.008 %	ND	ND
Lachnospiraceae UCG-001	0.003 %	0.003 %-0.003 %	0.004 %	0.002 %-0.007 %
Tyzzerella	ND	ND	0.003 %	0.001 %-0.007 %
Peptostreptococcus	0.002 %	0.001 %-0.005 %	0.004 %	0.002 %-0.006 %
Klebsiella	0.003 %	0.002 %-0.004 %	ND	ND
Leuconostoc	ND	ND	0.003 %	0.001 %-0.006 %
Faecalicoccus	0.003 %	0.002 %-0.005 %	ND	ND
Sutterella	0.004 %	0.002 %-0.008 %	0.002 %	0.001 %-0.004 %
Roseomonas	0.003 %	0.001 %-0.008 %	0.002 %	0.001 %-0.006 %
Papillibacter	0.005 %	0.001 %-0.016 %	0.001 %	0.001 %-0.001 %
Tannerella	0.003 %	0.003 %-0.003 %	ND	ND
GCA-900066575	0.003 %	0.001 %-0.007 %	0.002 %	0.002 %-0.003 %
Acetanaerobacterium	0.003 %	0.003 %-0.003 %	0.003 %	0.001 %-0.006 %
Frisingicoccus	0.003 %	0.001 %-0.005 %	ND	ND
Gardnerella	0.003 %	0.002 %-0.003 %	ND	ND
Odoribacter	0.005 %	0.002 %-0.013 %	0.001 %	0.001 %-0.001 %
UCG-007	0.003 %	0.003 %-0.003 %	ND	ND
Varibaculum	ND	ND	0.003 %	0.002 %-0.004 %
Porphyromonas	0.003 %	0.001 %-0.008 %	ND	ND
Cutibacterium	0.002 %	0.001 %-0.005 %	0.003 %	0.001 %-0.004 %
Anaerofilum	0.003 %	0.003 %-0.004 %	0.002 %	0.002 %-0.002 %
Butyricimonas	0.004 %	0.001 %-0.011 %	0.001 %	0.001 %-0.001 %
Fusobacterium	0.002 %	0.001 %-0.004 %	ND	ND
F0332	0.003 %	0.001 %-0.004 %	0.001 %	0.001 %-0.002 %
Coprobacillus	0.002 %	0.002 %-0.002 %	0.003 %	0.003 %-0.003 %
Parasutterella	0.003 %	0.001 %-0.005 %	0.001 %	0.000 %-0.003 %
Campylobacter	0.002 %	0.002 %-0.002 %	ND	ND
Atopobium	0.002 %	0.001 %-0.003 %	0.002 %	0.001 %-0.005 %
Alloscardovia	ND	ND	0.002 %	0.002 %-0.002 %
Victivallis	0.002 %	0.002 %-0.002 %	ND	ND
Pseudopropionibacterium	0.002 %	0.002 %-0.002 %	0.002 %	0.001 %-0.004 %
Syntrophococcus	0.002 %	0.001 %-0.003 %	ND	ND
S5-A14a	0.003 %	0.003 %-0.003 %	0.001 %	0.001 %-0.001 %
Phocea	ND	ND	0.002 %	0.001 %-0.002 %

Genus	Omnivore		Vegetarian	
	Average Relative Abundance	Range	Average Relative Abundance	Range
[*Eubacterium*] *nodatum* group	0.002 %	0.001 %-0.004 %	0.001 %	0.001 %-0.001 %
Barnesiella	0.003 %	0.001 %-0.013 %	0.001 %	0.000 %-0.001 %
Sanguibacteroides	0.002 %	0.001 %-0.004 %	ND	ND
TM7x	0.002 %	0.001 %-0.003 %	0.002 %	0.001 %-0.003 %
Oribacterium	0.002 %	0.001 %-0.003 %	0.001 %	0.001 %-0.003 %
Weissella	ND	ND	0.001 %	0.001 %-0.001 %
Neisseria	ND	ND	0.001 %	0.001 %-0.001 %
Alicyclobacillus	0.001 %	0.001 %-0.002 %	ND	ND
Abiotrophia	ND	ND	0.001 %	0.001 %-0.001 %
Criibacterium	0.001 %	0.001 %-0.001 %	ND	ND
Propionibacterium	0.001 %	0.001 %-0.002 %	0.001 %	0.001 %-0.002 %
Bulleidia	ND	ND	0.001 %	0.001 %-0.002 %
Lawsonella	0.001 %	0.001 %-0.001 %	0.001 %	0.001 %-0.001 %
Megasphaera	0.002 %	0.001 %-0.003 %	0.000 %	0.000 %-0.000 %
Flexilinea	ND	ND	0.001 %	0.001 %-0.001 %
Bacillus	ND	ND	0.001 %	0.001 %-0.001 %
Oceanobacillus	0.001 %	0.001 %-0.001 %	ND	ND
Fretibacterium	ND	ND	0.001 %	0.001 %-0.001 %
Herbaspirillum	ND	ND	0.001 %	0.001 %-0.001 %
Coprobacter	0.001 %	0.001 %-0.001 %	0.001 %	0.001 %-0.001 %

[1] ND corresponds to a relative abundance value of less than 0.001 %

[2] Purple filled rows indicate genera present at higher relative abundance (≥1 % difference between average relative abundance for omnivore and vegetarian samples in omnivores.

[3] Green filled rows indicate genera present at higher relative abundance (≥1 % difference between average relative abundance for omnivore and vegetarian samples) in vegetarians

Appendix E. Detailed Methods for LC-MS/MS Metabolomics

E.1. Detailed Methods

LC-MS/MS Metabolomics Sample Preparation

Ten vials from each the omnivore and vegetarian material types were characterized for metabolites. The water fraction and estimated wet mass of the materials can be found in section 2.5 Moisture Determination. Samples were prepared using a Bligh & Dyer extraction. Briefly, 30 µL of each Metabolomics QReSS Kit (Cambridge Isotope Laboratories, Inc.) and EquiSPLASH LIPIDOMIX Quantitative Mass Spec Internal Standard (Avanti, Birmingham, AL) were added to each labelled 10 mL glass tube, weighed, and placed in the freezer at -80 °C. The whole stool vial (1.007 g ± 0.003 g omnivore and 1.004 g ± 0.006 g vegetarian) was slightly thawed under cool running water for 30 s and subsequently transferred to the cooled glass tube containing internal standards. The sample was returned to -80 °C until extraction. Six aliquots of SRM 1946 Lake Superior Fish Tissue (103.61 mg ± 1.78 mg SD) were extracted alongside the samples to provide a sample processing QC. Two vials (1 mL volume each) of a pilot material from BioIVT- Human Feces Homogenate were extracted; one was used as an instrument QC with repeat injections throughout the instrument run and the other was used for column conditioning.

All extraction solvents and samples were kept on ice during the procedure and experimental samples were extracted in two batches at the same time by two different investigators. A 2:2:1.8 solvent system of methanol:chloroform:water was added to each sample. The volumes were added in a two-step method. Due to the water content in the stool, no water was added to the sample for the first step. All solutions were added using a pipet. First, 4.0 mL/g wet mass cold methanol (1100 µL) was added to each stool sample (410 µL for blanks and SRMs). Cold water was then added to SRMs (90 µL) and blanks (159 µL) only. All samples were vortexed for 1 min each. Next, 4.0 mL/g wet mass cold chloroform (1100 µL) was added to each stool sample followed by cold water (210 µL) in blanks and SRMs only. All samples were vortexed for 1 min each then allowed to sit on ice for 10 min. Samples were centrifuged (Eppendorf 5810R centrifuge, Eppendorf AG, Hamburg, Germany) at 2000 x *g* for 5 min at 4 °C for phase separation. Using a glass pipet and taking care not to disturb the protein layer and phase transition, the polar layer and non-polar layer were each removed and placed in their respective pre-weighed microcentrifuge tube for LC-MS/MS analysis. Samples were dried in a Vacufuge Concentrator plus (Eppendorf AG, Hamburg, Germany) at room temperature and weighed upon complete dryness. Dried samples were immediately stored at -80 °C. The polar fraction was resuspended in 600 µL of 2 % methanol in water volume fraction prior to LC-MS/MS analysis. The non-polar fraction was resuspended in 250 µL of isopropanol prior to LC-MS/MS analysis.

LC-MS/MS Instrument Method for Polar Metabolite Characterization

For polar annotations listed on the NIST RM 8048 reference material information sheet (RMIS), samples were run on two different instruments. Order of analysis is shown in the sample queue (Table E.1.1). First, the polar fraction samples were analyzed using a Vanquish UHPLC coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Samples were also analyzed using the Vanquish UHPLC but coupled to the Q Exactive (QE) Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Sample injections (2 µL) were separated by an Acquity HSS T3 (1.8 µm, 2.1 mm id x 150 mm length; Waters, Milford, MA, USA) C18 column at 350 µL/min and 45 °C with the gradient program listed in Table E.1.2. The Lumos mass spectrometer was operated in both positive and negative polarity separately with default source parameters used for the flow rate in data dependent mode (topN, 1 s cycle time) and a dynamic exclusion of 10 s (with 10 ppm error). The RF lens was set at 60 %. Full scan resolution using the orbitrap was set at 60,000 and the mass range was set to 60 *m/z* to 1000 *m/z*. Full scan ion target value was 5.0 x 10^5 allowing a maximum injection time of 118 ms. Monoisotopic peak determination was used, specifying small molecule and an intensity threshold of 2.5 x 10^4 was used for precursor selection. Data-dependent fragmentation was performed using higher-energy collisional dissociation (HCD) with a stepped collision energy of 20, 35, and 50 au and quadrupole isolation at 1.5 *m/z* width. The fragment scan resolution using the orbitrap was set at 15,000, 50 *m/z* as the first mass, ion target value of 5.0 x 10^4 and 22 ms maximum injection time.

Analysis on the QE was performed using the same sample queue, column, and gradient as used for the Lumos. For instrument settings on the QE, the mass spectrometer was also operated in both positive and negative polarity separately with default source parameters used for the flow rate in data dependent mode (top10) and a dynamic exclusion of 10 s. The RF lens was set at 50 %. Full scan resolution using the orbitrap was set at 35,000 and the mass range was set to 60 *m/z* to 900 *m/z*. Full scan ion target value was 3.0 x 10^6 allowing a maximum injection time of 100 ms. Monoisotopic peak determination was used, specifying small molecule and an intensity threshold of 5.0 x 10^4 was used for precursor selection. Data-dependent fragmentation was performed using higher-energy collisional dissociation (HCD) with a stepped collision energy of 20, 35, and 50 au and quadrupole isolation at 2.0 *m/z* width. The fragment scan resolution using the orbitrap was set at 17,500, 50 *m/z* as the first mass, ion target value of 2.0 x 10^5 and 100 ms maximum injection time.

On the Fusion Lumos, samples from each donor pool were acquired in Acquire X data acquisition mode to provide additional MS^2 spectra using both HCD and collisional induced dissociation (CID) fragmentation data spectra for spectral library matching and annotation. The MS^1 method for determining the initial exclusion list from the blank and inclusion list from the sample was run with an orbitrap resolution of 120,000, mass range of 67 *m/z* to 1000 *m/z*, and ion target value of 1.0 x 10^5 allowing a maximum injection time of 50 ms. For the subsequent data-dependent acquisition (DDA) runs, the mass spectrometer was operated in positive polarity, default source parameters for the flow rate, and data dependent mode (topN, 1 s cycle time) with a dynamic exclusion of 2.5 s (with 10 ppm error). The RF lens was set at 60 %. Full scan resolution using the orbitrap was set at 120,000 and the mass range was set to 67 *m/z* to 1000 *m/z*. Full scan ion target value was 1.0 x 10^5 allowing a maximum injection time of 50 ms.

Monoisotopic peak determination was used, specifying small molecule, and an intensity threshold of 2.5 x 10^4 was used for precursor selection. Data-dependent fragmentation was performed using HCD with a stepped collision energy of 20, 30, and 50 au and quadrupole isolation at 1.5 *m/z* width with a CID collision energy of 30 %. The fragment scan resolution using the orbitrap was set at 30,000, 60 *m/z* as the first mass, ion target value of 5.0 x 10^4 and 54 ms maximum injection time with the parallelizable option.

LC-MS/MS Instrument Method for Non-polar Metabolite Characterization

For non-polar annotations listed on the NIST RM 8048 reference material information sheet (RMIS), samples were run on a Vanquish UHPLC coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Order of analysis is shown in the sample queue (Table 1). Sample injections (1 µL) were separated by an Acclaim C30 column (3 µm, 2.1 mm id x 100 mm length; Thermo Scientific Sunnyvale, CA, USA) at 300 µL/min and 45 °C with the gradient program listed in Table E.1.3. The mass spectrometer was operated in both positive and negative polarity separately with default source parameters used for the flow rate in data dependent mode (topN, 1.5 s cycle time) and a dynamic exclusion of 10 s (with 10 ppm error). The RF lens was set at 35 %. Full scan resolution using the orbitrap was set at 120,000 and the mass range was set to 250 *m/z* to 1500 *m/z*. Full scan ion target value was 4.0 x 10^5 allowing a maximum injection time of 100 ms. Monoisotopic peak determination was used, specifying small molecule and an intensity threshold of 5.0 x 10^4 was used for precursor selection. Data-dependent fragmentation was performed using higher-energy collisional dissociation (HCD) with a stepped collision energy of 20, 30, and 35 au and quadrupole isolation at 1.5 *m/z* width. The fragment scan resolution using the orbitrap was set at 15,000, ion target value of 5.0 x 10^4 and a dynamic maximum injection time.

Samples from each donor pool were acquired in Acquire X data acquisition mode to provide additional MS^2 spectra using both HCD and collisional induced dissociation (CID) fragmentation data for spectral library matching and annotation. The MS^1 method for determining the initial exclusion list from the blank and inclusion list from the sample was run with an orbitrap resolution of 120,000, mass range of 250 *m/z* to 1500 *m/z*, and ion target value of 4.0 x 10^5 allowing a maximum injection time of 100 ms. For the subsequent data-dependent acquisition (DDA) runs, the mass spectrometer was operated in positive polarity, default source parameters for the flow rate, and data dependent mode (topN, 1 s cycle time) with a dynamic exclusion of 1 s (with 10 ppm error). The RF lens was set at 35 %. Full scan resolution using the orbitrap was set at 120,000 and the mass range was set to 250 *m/z* to 1500 *m/z*. Full scan ion target value was 5.0 x 10^5 allowing a maximum injection time of 54 ms. Monoisotopic peak determination was used, specifying small molecule, and an intensity threshold of 5.0 x 10^4 was used for precursor selection. Data-dependent fragmentation was performed using HCD with a stepped collision energy of 25, 30, and 35 au and quadrupole isolation at 1.5 *m/z* width. During each 1.5 s cycle of the untargeted dd-MS^2 profiling method, additional targeted product ion (m/z 170.0576 and 184.0733) or neutral loss (fatty acid + NH3, Table E.1.4) collisional induced dissociation (CID) MS^2 and MS^3 experiments were selectively triggered to provide higher quality characterization of phosphatidylcholine (PC) and triglyceride (TG) lipids. CID triggered MS^2 scans

from parent ions containing a HCD generated MS^2 containing either a m/z 170.0576 or a *m/z* 184.0733 fragment were collected with a CID collision energy of 32 %, 10 ms activation time, and 50 ms maximum injection time at 30,000 resolution. Additional MS^3 triggered spectra were collected from the three most intense (Top 3 mode) HCD MS^2 fragments of a fatty acid and + NH_3 neutral loss list (Table E.1.4). A collision energy of 35 %, 10 ms activation time, and 60 ms maximum injection time at 15,000 resolution was used for MS^3 spectra collection. An additional analysis utilizing the same separation method and mass spectrometer settings in negative mode was also performed.

LC-MS/MS Metabolomics Search Parameters

Resulting raw files were processed and searched with Compound Discoverer (CD, version 3.3.2) using both the local NIST23 library and the online database mzCloud (https://www.mzcloud.org/). A bulk search was performed for all ten vegetarian samples and a separate bulk search was performed for all ten omnivore samples. AquireX ID samples were included in the Lumos data search.

The base workflow used for all polar searches was the pre-programmed CD workflow "Untargeted Metabolomics with Statistics Detect Unknowns with ID using Online Databases.cdProcessingWF" with adjustments per algorithm. The base workflow used for non-polar searches was "NIST23_max ID_HighChemHighRes - Detect Unknowns with ID Using NIST Database and mzCloud Searches Bulk Sample" with adjustments per algorithm. Specifically, the following search parameters were used for Compound Discoverer searches: Retention time alignment was ChromAlign. The detect compounds node was set to a mass tolerance of 5 ppm, S/N threshold of 1.5, minimum peak intensity of 10,000, and base ions of [M+H]+1, [M-H]-1. The assign compound annotations node was used with 5 ppm mass tolerance with mzVault and mzCloud. All library searches of mzVault and mzCloud included a mass tolerance of 10 ppm of precursor and product ions and maximum retention time shift of 2 min. This search was performed using three different algorithms for comparison: NIST, HighChem-HighRes, and HighChem-HighRes Dot Product. Fill Gaps, Apply SERRF QC Correction, and Mark Background Compounds nodes were also applied. Total mass features were reduced by removing background features found in the solvent blank and removing any features with a blank annotation name. Annotations were only kept if found in at least three samples and had a match score ≥ 65 in at least one library. The library search results were exported in Excel for further sorting, and group CVs were obtained to assess homogeneity across vials.

Data Processing for Characterization

For each compound, the best match score was obtained in the CD search from ten vial replicates. For the polar characterization, this information was exported from CD as an Excel file for each donor pool, analytical instrument, ionization mode, and search algorithm (24 files). For the non-polar characterization, this information was exported from CD as an Excel file for each donor pool, ionization mode, and search algorithm (12 files). Files were then sorted and combined according to donor pool and ionization mode, where compounds were removed if

not found across all three algorithms and across both instruments (polar only). Compounds from positive and negative ionization mode were compiled to create the final full annotation lists per donor pool. The average score, standard deviation, and percent RSD shown with each compound was calculated from the searches performed for all three algorithms and on two different instruments for polar metabolites (n = 6) and all three algorithms for non-polar metabolites (n = 3). These calculations are intended to provide a value of uncertainty on each annotation.

E.1.1. Table - Sample run queue

Sample Name	Sample Name
Solvent_Blank_1_POS	Solvent_Blank_1_NEG
Solvent_Blank_2_POS	Solvent_Blank_2_NEG
CC_1_POS	CC_1_NEG
CC_2_POS	CC_2_NEG
CC_3_POS	CC_3_NEG
CC_4_POS	CC_4_NEG
Blank_1_POS	Blank_1_NEG
CC_5_POS	CC_5_NEG
CC_6_POS	CC_6_NEG
CC_7_POS	CC_7_NEG
CC_8_POS	CC_8_NEG
Stool_QC_1A_POS	Stool_QC_1A_NEG
Veg_7_POS	Veg_7_NEG
Veg_4_POS	Veg_4_NEG
Veg_5_POS	Veg_5_NEG
Veg_1_POS	Veg_1_NEG
Veg_10_POS	Veg_10_NEG
Stool_QC_1B_POS	Stool_QC_1B_NEG
Veg_6_POS	Veg_6_NEG
Veg_3_POS	Veg_3_NEG
Veg_8_POS	Veg_8_NEG
Veg_2_POS	Veg_2_NEG
Veg_9_POS	Veg_9_NEG
Stool_QC_1C_POS	Stool_QC_1C_NEG
Omnivore_3_POS	Omnivore_3_NEG
Omnivore_5_POS	Omnivore_5_NEG
Omnivore_2_POS	Omnivore_2_NEG
Omnivore_1_POS	Omnivore_1_NEG
Omnivore_9_POS	Omnivore_9_NEG
Stool_QC_1D_POS	Stool_QC_1D_NEG
Omnivore_8_POS	Omnivore_8_NEG
Omnivore_6_POS	Omnivore_6_NEG
Omnivore_10_POS	Omnivore_10_NEG
Omnivore_7_POS	Omnivore_7_NEG
Omnivore_4_POS	Omnivore_4_NEG
Stool_QC_1E_POS	Stool_QC_1E_NEG
Blank_2_POS	Blank_2_NEG

SRM_1946_1A_POS	SRM_1946_1A_NEG
SRM_1946_1B_POS	SRM_1946_1B_NEG
SRM_1946_1C_POS	SRM_1946_1C_NEG
SRM_1946_1D_POS	SRM_1946_1D_NEG
SRM_1946_1E_POS	SRM_1946_1E_NEG
SRM_1946_3_POS	SRM_1946_3_NEG
SRM_1946_2_POS	SRM_1946_2_NEG
SRM_1946_5_POS	SRM_1946_5_NEG
SRM_1946_6_POS	SRM_1946_6_NEG
SRM_1946_4_POS	SRM_1946_4_NEG
Blank_3_POS	Blank_3_NEG

E.1.2. Table - Reverse phase C18 separation gradient program for polar characterization

Time (min)	Flow Rate (µL/min)	Water with 0.1 % formic acid	Methanol with 0.1 % formic acid
0.0	350	98.0	2.0
1.0	350	98.0	2.0
10.5	350	50.0	50.0
11.0	350	2.0	98.0
16.0	350	2.0	98.0
16.5	350	98.0	2.0
22.0	0.35	98.0	2.0

E.1.3. Table - Reverse phase C30 separation gradient program for nonpolar characterization

Time (min)	Flow Rate (µL/min)	60 % acetonitrile and 40 % water with 10 mmol/L ammonium formate and 0.1 % formic acid	90 % isopropanol and 10 % acetonitrile with 10 mmol/L ammonium formate and 0.1 % formic acid
0.0	300	70.0	30.0
2.0	300	57.0	43.0
12.0	300	35.0	65.0
18.0	300	15.0	85.0

20.0	300	0.0	100
25.0	300	0.0	100
25.1	300	70.0	30.0
30.0	300	70.0	30.0

E.1.4. Table - Lipid neutral loss (Fatty Acid + NH_3) list used for Acquire X data

m/z
215.2042
217.2042
243.2198
245.2355
269.2355
271.2511
273.2668
285.2668
287.2824
295.2511
297.2668
299.2824
301.2981
315.3137
319.2511
321.2668
323.2824
325.2981
327.3137
329.3294
334.3450
345.2668
347.2824
349.2981
351.3137
353.3294
355.3450
357.3607
371.3763
383.3763
385.3920
413.4233

Appendix F. Detailed Results for LC-MS/MS Metabolomics

F.1. Table - Omnivore Metabolite Characterization List by LC-MS/MS

Metabolite ID	InChIKey
(1-Ethoxy-1-oxo-4-phenylbutan-2-yl)alanine	CEIWXEQZZZHLDM-UHFFFAOYSA-N
(2E,4E,6E)-11-(1,3-Benzodioxol-5-yl)-N-(2-methylpropyl)undeca-2,4,6-trienamide	ITXVWRWWWSDXOY-HNZFQPFYSA-N
(2S)-2-[(3-Methylbutanoyl)amino]pentanedioic acid	HCVIJONZMYVLAW-ZETCQYMHSA-N
(3-beta)-Allopregnanolone sulfate	MENQCIVHHONJLU-FZCSVUEKSA-N
(4-Methylphenyl)oxidanesulfonic acid	WGNAKZGUSRVWRH-UHFFFAOYSA-N
1,2-Di-(9Z,12Z,15Z-octadecatrienoyl)-sn-glycero-3-phosphocholine	XXKFQTJOJZELMD-JICBSJGISA-N
1,2-Dioctadecanoyl-sn-glycerol	UHUSDOQQWJGJQS-QNGWXLTQSA-N
1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine-N,N-dimethyl	XHPZRQBHFOVLEJ-UNUIOPIBSA-N
1,2-Dioleoyl-sn-glycerol	AFSHUZFNMVJNKX-LLWMBOQKSA-N
1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine	SLKDGVPOSSLUAI-PGUFJCEWSA-N
1,2-Dipalmitoyl-sn-glycero-O-ethyl-3-phosphatidylcholine cation	VBZSUWNESITEBV-KHBQWKRXSA-N
1,2-Dipentadecanoyl-sn-glycero-3-phosphocholine	LJARBVLDSOWRJT-PSXMRANNSA-N
1,2-Dipentadecanoyl-sn-glycero-3-phosphoethanolamine	SKVKIGSFTGVBOX-MGBGTMOVSA-N
1,3,7-Trimethyluric acid	BYXCFUMGEBZDDI-UHFFFAOYSA-N
1,4-Anhydro-6-O-[(9Z)-9-octadecenoyl]-D-glucitol	NWGKJDSIEKMTRX-AAZCQSIUSA-N
1,6-Hexamethylene-bis[3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate]	ZVVFVKJZNVSANF-UHFFFAOYSA-N
1,7-Dimethyluric acid	NOFNCLGCUJJPKU-UHFFFAOYSA-N
13(S)-HOTrE	KLLGGGQNRTVBSU-FQSPHKRJSA-N
13-Keto-9Z,11E-octadecadienoic acid	JHXAZBBVQSRKJR-BSZOFBHHSA-N
14-(acetyloxy)-6-(furan-3-yl)-12-hydroxy-1,7,11,15,15-pentamethyl-5-oxo-3-oxapentacyclo[8.8.0.02,4.02,7.011,16]octadecan-18-yl acetate	PVCLMZBWMWIHKX-UHFFFAOYSA-N
1a,1b-Dihomo prostaglandin E1	YMDDELUTDBQEMT-QZCLESEGSA-N
1-Behenoyl-2-hydroxy-sn-glycero-3-phosphocholine	UIINDYGXBHJQHX-GDLZYMKVSA-N
1-Dodecyl-2-pyrrolidinone	NJPQAIBZIHNJDO-UHFFFAOYSA-N
1-Hexadecyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine	DUUSFCFZBREELS-LXDDVLOTSA-N
1-Hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine	SIEDNCDNGMIKST-IYEJTHTFSA-N
1-Isopropyl-5-oxo-3-pyrrolidinecarboxylic acid	YKWIPHFVESIOGN-UHFFFAOYSA-N
1-Monolinolenin	GGJRAQULURVTAJ-PDBXOOCHSA-N
1-Monolinoleoyl-rac-glycerol	WECGLUPZRHILCT-HZJYTTRNSA-N
1-Myristoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	RPXHXZNGZBHSMJ-GOSISDBHSA-N
1-Oleoyl-2-hydroxy-sn-glycero-3-phospho-(1'-rac-glycerol)	FQQQKGAFQIIGLQ-VYFUBJQISA-N
1-Oleoyl-2-linoleoyl-rac-glycerol	BLZVZPYMHLXLHG-RQOIEFAZSA-N
1-O-Octadecyl-sn-glyceryl-3-phosphorylcholine	XKBJVQHMEXMFDZ-AREMUKBSSA-N
1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	YVYMBNSKXOXSKW-HXUWFJFHSA-N
1-Palmitoyl-2-linoleoyl-rac-glycerol	SVXWJFFKLMLOHO-BCTRXSSUSA-N
1-Palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine	JLPULHDHAOZNQI-ZTIMHPMXSA-N
1-Palmitoyl-2-myristoyl-sn-glycero-3-phosphocholine	UIXXHROAQSBBOV-PSXMRANNSA-N
1-Palmitoyl-2-oleoyl-sn-glycerol	YEJYLHKQOBOSCP-OZKTZCCCSA-N

Metabolite ID	InChIKey
1-Palmitoyl-3-oleoyl-sn-glycero-2-phosphoethanolamine	BLMYNJDXEOSYBR-OTMQOFQLSA-N
1-Palmitoyl-sn-glycero-3-phosphocholine	ASWBNKHCZGQVJV-HSZRJFAPSA-N
1-Pentadecanoyl-sn-glycero-3-phosphocholine	RJZVWDTYEWCUAR-JOCHJYFZSA-N
1-Stearoyl-2-arachidonyl-sn-glycero-3-phosphocholine	PSVRFUPOQYJOOZ-RYIOQAQXSA-N
1-Stearoyl-2-hydroxy-sn-glycero-3-phosphocholine	IHNKQIMGVNPMTC-RUZDIDTESA-N
1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	BBYWOYAFBUOUFP-JOCHJYFZSA-N
1-Stearoyl-2-oleoyl-sn-glycero-3-phosphocholine	ATHVAWFAEPLPPQ-VRDBWYNSSA-N
1-Stearoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine	JQKOHRZNEOQNJE-DJEJVYNPSA-N
2-((3aR,4S,7R,7aS)-1,3-Dioxohexahydro-1H-4,7-methanoisoindol-2(3H)-yl)propanoic acid	REFMTLIXGKZVDF-RQOWMOBWSA-N
2-(14,15-Epoxyeicosatrienoyl) glycerol	LPMVKZXODWQHGJ-ILYOTBPNSA-N
2-(1-Methyl-1H-pyrrol-2-yl)acetic acid	SYYOUHJJSOLSJD-UHFFFAOYSA-N
2-(Hexadecylamino)ethanol	SIFHZKKXWMJWOB-UHFFFAOYSA-N
2-(O-beta-D-Glucopyranosyl)atractyligenin	ZUDBDCBPZHGAEL-ICRASABDSA-N
2,3,4,9-Tetrahydro-1H-beta-carboline-3-carboxylic acid	FSNCEEGOMTYXKY-UHFFFAOYSA-N
2,3-Dihydroxypropyl octadecanoate	VBICKXHEKHSIBG-UHFFFAOYSA-N
2,4-Bis(1,1-dimethylethyl)phenol phosphate	GUDSEWUOWPVZPC-UHFFFAOYSA-N
2-[(2R,4aS,8S,8aS)-8-{2-[(4aS,7R,8aR)-7-(1-Carboxyvinyl)-1-hydroxy-4a-methyl-2-oxo-1,2,4a,5,6,7,8,8a-octahydro-1-naphthalenyl]ethyl}-4a-methyl-7-oxo-1,2,3,4,4a,7,8,8a-octahydro-2-naphthalenyl]acrylic acid	BSSWIRVAMSGQMG-RTTXQYISSA-N
2-[(3R,4S)-3-{[5-(Cyclohexylmethyl)-1,2-oxazol-3-yl]methyl}-4-piperidinyl]-N-[2-(4-morpholinyl)ethyl]acetamide	CBYSOGZECVBHQN-SFTDATJTSA-N
2-Arachidonoyl glycerol	RCRCTBLIHCHWDZ-DOFZRALJSA-N
2'-Deoxyadenosine	OLXZPDWKRNYJJZ-RRKCRQDMSA-N
2'-Deoxyinosine	VGONTNSXDCQUGY-JXBXZBNISA-N
2-Linoleoyl-1-palmitoyl-sn-glycero-3-phosphoethanolamine	HBZNVZIRJWODIB-NHCUFCNUSA-N
2-Oleoyl-1-palmitoyl-sn-glycero-3-phosphocholine	WTJKGGKOPKCXLL-VYOBOKEXSA-N
2-Palmitoyl-rac-glycerol	BBNYCLAREVXOSG-UHFFFAOYSA-N
3-(1H-Imidazol-4-yl)propanoic acid	ZCKYOWGFRHAZIQ-UHFFFAOYSA-N
3,4-Dihydroxyhydrocinnamic acid	DZAUWHJDUNRCTF-UHFFFAOYSA-N
3-[5-(2-Methylpropyl)-3,6-dioxopiperazin-2-yl]propanoic acid	CSRXCVLPMJBQIM-UHFFFAOYSA-N
3-beta-Hydroxy-5-cholestenoic acid	WVXOMPRLWLXFAP-AMQKJUDNSA-N
3-Hydroxybutyric acid	WHBMMWSBFZVSSR-UHFFFAOYSA-N
3-Hydroxymyristic acid	ATRNZOYKSNPPBF-UHFFFAOYSA-N
3-Hydroxypyridine	GRFNBEZIAWKNCO-UHFFFAOYSA-N
3-Hydroxyurs-12-en-23-oic acid	NBGQZFQREPIKMG-UHFFFAOYSA-N
3-Sulfopropanoic acid	OURSFPZPOXNNKX-UHFFFAOYSA-N
4,4-Dimethylcholest-8(9),24-dien-3.beta.-ol	CHGIKSSZNBCNDW-QGBOJXOESA-N
4-Cholestenone	NYOXRYYXRWJDKP-GYKMGIIDSA-N
4-Hydroxyomeprazole sulfide	LVUFHVGKGMRSQW-UHFFFAOYSA-N
4-Pyridoxic acid	HXACOUQIXZGNBF-UHFFFAOYSA-N

Metabolite ID	InChIKey
5-(Dimethylamino)pentanoic acid	UYZSNVLEDLCWGU-UHFFFAOYSA-N
5-[(10Z)-14-(3,5-dihydroxyphenyl)tetradec-10-en-1-yl]benzene-1,3-diol	PFIQXUYXVYYERO-ALCCZGGFSA-N
5-Acetylamino-6-amino-3-methyluracil	POQOTWQIYYNXAT-UHFFFAOYSA-N
5-Hydroxyindole-3-acetic acid	DUUGKQCEGZLZNO-UHFFFAOYSA-N
5-O-Feruloylquinic acid	RAGZUCNPTLULOL-KQJPBSFVSA-N
5-Oxotetrahydro-2-furancarboxylic acid	QVADRSWDTZDDGR-UHFFFAOYSA-N
6-Cyclopentacycloundecenecarboxylic acid, 1,2,3,3a,4,5,6,7,8,9,12,12a-dodecahydro-3a,10-dimethyl-1-(1-methylethenyl)-	FVDVQQJLECWRKH-NVNXTCNLSA-N
6-Methyl-1,2,3-oxathiazin-4(3H)-one 2,2-dioxide	YGCFIWIQZPHFLU-UHFFFAOYSA-N
8-(3-Octyl-2-oxiranyl)octanoic acid	IMYZYCNQZDBZBQ-UHFFFAOYSA-N
9-Nitrooleate	CQOAKBVRRVHWKV-SAPNQHFASA-N
Allyl 4-amino-1-piperidinecarboxylate	XKLGRJGBMKOQDK-UHFFFAOYSA-N
alpha,alpha-Dilaurin	KUVAEMGNHJQSMH-UHFFFAOYSA-N
alpha-Hyodeoxycholic acid methyl ester	BWDRDVHYVJQWBO-QWXHOCAMSA-N
alpha-Tocopheryl acetate	ZAKOWWREFLAJOT-PDNQPUDYSA-N
Anacardic acid	ADFWQBGTDJIESE-UHFFFAOYSA-N
Aspartic acid	CKLJMWTZIZZHCS-UHFFFAOYSA-N
Asp-Pro	UKGGPJNBONZZCM-WDSKDSINSA-N
Bacilysocin	DIGHTWUQPWHBPG-UHFFFAOYSA-N
Benzenesulfonic acid, 4-(acetylamino)-	ZQPVMSLLKQTRMG-UHFFFAOYSA-N
Benzyldimethyldodecylammonium cation	CYDRXTMLKJDRQH-UHFFFAOYSA-N
Benzyldimethylstearylammonium cation	FWLORMQUOWCQPO-UHFFFAOYSA-N
beta-Hyodeoxycholic acid	DGABKXLVXPYZII-MMTMODRTSA-N
beta-Muricholic acid	DKPMWHFRUGMUKF-CRKPLTDNSA-N
beta-Sitosterol	KZJWDPNRJALLNS-VJSFXXLFSA-N
Bilirubin	BPYKTIZUTYGOLE-IFADSCNNSA-N
Biliverdin	QBUVFDKTZJNUPP-BBROENKCSA-N
Bis-(3,4-dimethyldibenzylidenesorbitol)	YWEWWNPYDDHZDI-JJKKTNRVSA-N
CAY10683	HTOYBIILVCHURC-UHFFFAOYSA-N
Chenodeoxycholic acid	RUDATBOHQWOJDD-BSWAIDMHSA-N
Cholan-24-oic acid, 12-hydroxy-3-oxo-	WMUMZOAFCDOTRW-UHFFFAOYSA-N
Cholan-24-oic acid, 12-hydroxy-3-oxo-, methyl ester	LOJPGRHPJBGMMF-JJNSWQHQSA-N
Cholestan-3-ol, (3 alpha,5 beta)-	QYIXCDOBOSTCEI-VZNZRSCWSA-N
Cholestan-3-one, (5 alpha)-	PESKGJQREUXSRR-UXIWKSIVSA-N
Cholesterol	HVYWMOMLDIMFJA-DPAQBDIFSA-N
Cholesterol 3-sulfate	BHYOQNUELFTYRT-DPAQBDIFSA-N
cis,cis-9,12-Octadecadien-1-ol	JXNPEDYJTDQORS-HZJYTTRNSA-N
cis-10-Nonadecenoic acid	BBOWBNGUEWHNQZ-KTKRTIGZSA-N
cis-13-Eicosenoic acid	URXZXNYJPAJJOQ-FPLPWBNLSA-N

Metabolite ID	InChIKey
cis-Vaccenic acid	UWHZIFQPPBDJPM-FPLPWBNLSA-N
Citrulline	RHGKLRLOHDJJDR-UHFFFAOYSA-N
Cocamidopropyl betaine	MRUAUOIMASANKQ-UHFFFAOYSA-N
Crotonic acid	LDHQCZJRKDOVOX-NSCUHMNNSA-N
Decaethylene glycol	DTPCFIHYWYONMD-UHFFFAOYSA-N
Dehydropipernonaline	KAYVDASZRFLFRZ-PQECNABGSA-N
Deoxycholic acid	KXGVEGMKQFWNSR-LLQZFEROSA-N
D-erythro-Sphinganine	OTKJDMGTUTTYMP-ZWKOTPCHSA-N
D-Glucosyl-beta-1,1-N-palmitoyl-D-erythro-sphingosine	VJLLLMIZEJJZTE-NNTBDIJYSA-N
Didecyldimethylammonium	JGFDZZLUDWMUQH-UHFFFAOYSA-N
Dihomo-gamma-linolenic acid methyl ester	QHATYOWJCAQINT-JPFHKJGASA-N
Dimethyl azelate	DRUKNYVQGHETPO-UHFFFAOYSA-N
DL-beta-Hydroxypalmitic acid	CBWALJHXHCJYTE-UHFFFAOYSA-N
DL-Indole-3-lactic acid	XGILAAMKEQUXLS-UHFFFAOYSA-N
DL-Serine, succinate (ester)	ZAHSBRLHJRVFAU-UHFFFAOYSA-N
D-Mannosamine	FZHXIRIBWMQPQF-KVTDHHQDSA-N
Dodecanedioic acid	TVIDDXQYHWJXFK-UHFFFAOYSA-N
D-Sphingosine	WWUZIQQURGPMPG-KRWOKUGFSA-N
Enalapril	GBXSMTUPTTWBMN-XIRDDKMYSA-N
Enalaprilat	LZFZMUMEGBBDTC-QEJZJMRPSA-N
Epilupeol	MQYXUWHLBZFQQO-ISZJTHHZSA-N
Ergocalciferol	MECHNRXZTMCUDQ-ANGGWTPUSA-N
Erioglaucine	CTRXDTYTAAKVSM-UHFFFAOYSA-N
Erucamide	UAUDZVJPLUQNMU-KTKRTIGZSA-N
Erythrodiol	PSZDOEIIIJFCFE-KWVOMAKGSA-N
Ethanesulfonic acid, 2-[(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)amino]-	IZRUXXTXKZAGQQ-UHFFFAOYSA-N
Fucosterol	OSELKOCHBMDKEJ-JUGJNGJRSA-N
Geranyllinalool	IQDXAJNQKSIPGB-HQSZAHFGSA-N
Ginkgolic acid I	YXHVCZZLWZYHSA-FPLPWBNLSA-N
Glu-Ile	SNFUTDLOCQQRQD-LKEWCRSYSA-N
Glycerol monooleate	RZRNAYUHWVFMIP-KTKRTIGZSA-N
Glycerol trilaurate	VMPHSYLJUKZBJJ-UHFFFAOYSA-N
Glycodeoxycholic acid	WVULKSPCQVQLCU-UHFFFAOYSA-N
Glycyl-L-norleucine	XVUIZOUTLADVIH-UHFFFAOYSA-N
Gly-Val	STKYPAFSDFAEPH-LURJTMIESA-N
Guanine	UYTPUPDQBNUYGX-UHFFFAOYSA-N
Heptadecasphinganine	KFQUQCFJDMSIJF-DLBZAZTESA-N
Ile-Glu	KTGFOCFYOZQVRJ-ACLDMZEESA-N
Ile-Phe	WMDZARSFSMZOQO-UHTWSYAYSA-N
Ile-Ser	TWVKGYNQQAUNRN-ACZMJKKPSA-N

Metabolite ID	InChIKey
Isobutyric acid	KQNPFQTWMSNSAP-UHFFFAOYSA-N
Kaurane-17,18-dioic acid, 18-methyl ester, (4 alpha,16 alpha)-	YDAVVXOXSSFVPF-BBBHLOGDSA-N
Lactosyl-Ceramide(d18:1/22:0)	QYWVASPEUXEHSY-DYGYHFHFSA-N
Laurylsulfuric acid	MOTZDAYCYVMXPC-UHFFFAOYSA-N
Linoleic Acid	OYHQOLUKZRVURQ-HZJYTTRNSA-N
Linoleic acid-biotin	TXYWCIYPXRNWJU-MJDXKYDGSA-N
Lithocholic acid	SMEROWZSTRWXGI-HVATVPOCSA-N
Lithocholylglycine	XBSQTYHEGZTYJE-OETIFKLTSA-N
Lupeol	MQYXUWHLBZFQQO-QGTGJCAVSA-N
Methionine sulfoxide	QEFRNWWLZKMPFJ-YGVKFDHGSA-N
Methyl deoxycholate	ZHUOOEGSSFNTNP-JMKDMENQSA-N
Methyl hydrogen azelate	VVWPSAPZUZXYCM-UHFFFAOYSA-N
Methyl linolenate	DVWSXZIHSUZZKJ-YSTUJMKBSA-N
Myristyl sulfate	URLJMZWTXZTZRR-UHFFFAOYSA-N
N-(14-Methylpentadecanoyl)phenylalanine	WICAPTTWPBEVPW-UHFFFAOYSA-N
N-(6-Methyl-2-oxohexahydro-4-pyrimidinyl)urea	CZAUMIGWDFREBR-UHFFFAOYSA-N
N-(Octadecanoyl)sphing-4-enine-1-phosphocholine	LKQLRGMMMAHREN-NWBJSICCSA-N
N,N-Dimethylsphingosine	YRXOQXUDKDCXME-YIVRLKKSSA-N
N2-Isobutyryl-2'-deoxyguanosine	SIDXEQFMTMICKG-UHFFFAOYSA-N
N6-Palmitoyl-L-lysine	IWKZTTDNVPAHNP-FQEVSTJZSA-N
N-Acetyl-.beta.-D-mannosamine	OVRNDRQMDRJTHS-OZRXBMAMSA-N
N-Acetyldihydrosphingosine	CRJGESKKUOMBCT-VQTJNVASSA-N
N-Acetyl-D-norleucine	JDMCEGLQFSOMQH-SSDOTTSWSA-N
N-Acetylglutamine	KSMRODHGGIIXDV-YFKPBYRVSA-N
N-Acetyl-L-glutamic acid	RFMMMVDNIPUKGG-YFKPBYRVSA-N
N-alpha-Acetyl-L-lysine	VEYYWZRYIYDQJM-ZETCQYMHSA-N
N-Benzyl-N,N-dimethyl-1-tetradecanaminium cation	WNBGYVXHFTYOBY-UHFFFAOYSA-N
N-Docosanoyl-4-sphingenyl-1-O-phosphorylcholine	FJJANLYCZUNFSE-TWKUQIQBSA-N
N-epsilon-Acetyl-L-lysine	DTERQYGMUDWYAZ-ZETCQYMHSA-N
Nervonoyl ethanolamide	LISKWSFNVJTQKH-KTKRTIGZSA-N
N-Hexadecyl sulfate	LPTIRUACFKQDHZ-UHFFFAOYSA-N
N-Hexanoylsphingosine	NPRJSFWNFTXXQC-QFWQFVLDSA-N
Nitrobenzene	LQNUZADURLCDLV-UHFFFAOYSA-N
N-Lauroyl-D-erythro-sphinganine	UHWYQXNZIBLESO-URLMMPGGSA-N
N-L-gamma-Glutamyl-L-leucine	MYFMARDICOWMQP-UHFFFAOYSA-N
N-Methyl-L-phenylalanine	SCIFESDRCALIIM-VIFPVBQESA-N
N-Myristoylsphinganine	UDTSZXVRDXQARY-IOWSJCHKSA-N
N-Nervonoyl-D-erythro-sphingosylphosphorylcholine	WKZHECFHXLTOLJ-QYKFWSDSSA-N
N-Octanoylsphingosine	APDLCSPGWPLYEQ-WRBRXSDHSA-N
N-Oleoyl-4-sphingenine	OBFSLMQLPNKVRW-RHPAUOISSA-N

Metabolite ID	InChIKey
N-Oleoyl-D-erythro-sphinganine	MJQIARGPQMNBGT-WWUCIAQXSA-N
N-Oleoylethanolamine	BOWVQLFMWHZBEF-KTKRTIGZSA-N
Nonaethylene glycol	YZUUTMGDONTGTN-UHFFFAOYSA-N
Nonanedioic acid	BDJRBEYXGGNYIS-UHFFFAOYSA-N
N-Palmitoyl-D-sphingosine	YDNKGFDKKRUKPY-TURZORIXSA-N
N-Stearoylsphinganine	KZTJQXAANJHSCE-OIDHKYIRSA-N
N-Tetracosenoyl-4-sphingenine	VJSBNBBOSZJDKB-KPEYJIHVSA-N
Octaethylene glycol	GLZWNFNQMJAZGY-UHFFFAOYSA-N
Oleamide	FATBGEAMYMYZAF-KTKRTIGZSA-N
Oleoyl ethylamide	JZJYYCFYGXPUMF-QXMHVHEDSA-N
Oleoylserotonin	LCQKHZYXPCLVBI-KTKRTIGZSA-N
Palmitoyl sphingomyelin	RWKUXQNLWDTSLO-GWQJGLRPSA-N
Palmitoyleicosapentaenoyl phosphatidylcholine	KLTHQSWIRFFBRI-KOQZQRJKSA-N
Palmitoyl-L-carnitine	XOMRRQXKHMYMOC-OAQYLSRUSA-N
Pantothenic acid	GHOKWGTUZJEAQD-UHFFFAOYSA-N
PEG Monooleate n8	RRMPJCDMVAUQGF-MDZDMXLPSA-N
PEG14	AKWFJQNBHYVIPY-UHFFFAOYSA-N
PEG16	OWTQQPNDSWCHOV-UHFFFAOYSA-N
Pheophorbide a	NSFSLUUZQIAOOX-QEWKCGBTSA-N
Phe-Val	IEHDJWSAXBGJIP-RYUDHWBXSA-N
Piperchabamide B	CHOLQJRIMZGPNC-QHKWOANTSA-N
Pipernonaline	PKLGRWSJBLGIBF-JMQWPVDRSA-N
Piperolein A	MIWPBXQTBYPJEF-XBXARRHUSA-N
Poly THF n8	RPGZMBUUQBRFFX-UHFFFAOYSA-N
Pregnenolone sulfate	DIJBBUIOWGGQOP-QGVNFLHTSA-N
Prostaglandin K2	LGMXPVXJSFPPTQ-DJUJBXLVSA-N
Pyridoxine	LXNHXLLTXMVWPM-UHFFFAOYSA-N
Pyruvic acid	LCTONWCANYUPML-UHFFFAOYSA-N
Retrofractramide A	BPSWISYORIWKCT-FCGWLDPVSA-N
Riboflavin	AUNGANRZJHBGPY-SCRDCRAPSA-N
Sarsasapogenin	GMBQZIIUCVWOCD-WWASVFFGSA-N
Sebacic acid	CXMXRPHRNRROMY-UHFFFAOYSA-N
Stearamide	LYRFLYHAGKPMFH-UHFFFAOYSA-N
Stearic Acid	QIQXTHQIDYTFRH-UHFFFAOYSA-N
Stearoyl ethanolamide	OTGQIQQTPXJQRG-UHFFFAOYSA-N
Suberic acid	TYFQFVWCELRYAO-UHFFFAOYSA-N
Sucralose	BAQAVOSOZGMPRM-QBMZZYIRSA-N
Sucrose	CZMRCDWAGMRECN-UGDNZRGBSA-N
Tacrolimus	QJJXYPPXXYFBGM-LJIGMGMYSA-N
Tauroursodeoxycholic acid	BHTRKEVKTKCXOH-OGTVOWCVSA-N
Metabolite ID	InChIKey
Termitomycamide E	JTSNNDBVDQFVEG-HZJYTTRNSA-N

Testosterone sulfate	WAQBISPOEAOCOG-DYKIIFRCSA-N
Tetradecanedioic acid, 5,6-dihydroxy-	JCOCIXAHAFALHI-UHFFFAOYSA-N
Tetrahydroharman-3-carboxylic acid	ZUPHXNBLQCSEIA-UHFFFAOYSA-N
Thiamine cation	JZRWCGZRTZMZEH-UHFFFAOYSA-N
Threonine	AYFVYJQAPQTCCC-UHFFFAOYSA-N
Thr-Glu	BECPPKYKPSRKCP-JCGDXUMPSA-N
Thr-Leu	BQBCIBCLXBKYHW-BIIVOSGPSA-N
Tildipirosin	HNDXPZPJZGTJLJ-UEJFNEDBSA-N
Tomatidine	XYNPYHXGMWJBLV-KLPIGKMYSA-N
trans-Petroselinic acid	CNVZJPUDSLNTQU-OUKQBFOZSA-N
trans-Vaccenic acid	UWHZIFQPPBDJPM-BQYQJAHWSA-N
Tridemorph	YTOPFCCWCSOHFV-UHFFFAOYSA-N
Triethylene glycol bis(2-ethylhexanoate)	FRQDZJMEHSJOPU-UHFFFAOYSA-N
Trihydroxycholestanoic acid	CNWPIIOQKZNXBB-VCVMUKOKSA-N
Trilinolein	HBOQXIRUPVQLKX-BBWANDEASA-N
Tuberonic acid	RZGFUGXQKMEMOO-BSANDHCLSA-N
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N
Uric acid	LEHOTFFKMJEONL-UHFFFAOYSA-N
Urobilin	KDCCOOGTVSRCHX-UYMYUHGCSA-N
Urocanic acid	LOIYMIARKYCTBW-OWOJBTEDSA-N
Val-Ala	HSRXSKHRSXRCFC-WDSKDSINSA-N
Val-Glu	UPJONISHZRADBH-POYBYMJQSA-N
Val-Ile	PNVLWFYAPWAQMU-DJLDLDEBSA-N
Val-Leu	XCTHZFGSVQBHBW-UHFFFAOYSA-N
Val-Val	KRNYOVHEKOBTEF-UHFFFAOYSA-N
Xanthine	LRFVTYWOQMYALW-UHFFFAOYSA-N
Xanthurenic acid	FBZONXHGGPHHIY-UHFFFAOYSA-N

F.2. Table - Vegetarian Metabolite Characterization List by LC-MS/MS

Metabolite ID	InChIKey
alpha-Tocopherol	GVJHHUAWPYXKBD-IEOSBIPESA-N
(15:3)-Anacardic acid	QUVGEKPNSCFQIR-UTOQUPLUSA-N
(1S,2S,3aR,4S,5S,9R,11R,13aS)-4,9,11-tris(acetyloxy)-3a-hydroxy-2,5,8,8,12-pentamethyl-1H,2H,3H,3aH,4H,5H,8H,9H,10H,11H,13aH-cyclopenta[12]annulen-1-yl benzoate	BRVXVMOWTHQKHC-OTOOIOIQSA-N
(2-{4-[(R)-(4-chlorophenyl)(phenyl)methyl]-1-piperazinyl}ethoxy)acetic acid	ZKLPARSLTMPFCP-OAQYLSRUSA-N
(2E,4E,6E)-11-(1,3-Benzodioxol-5-yl)-N-(2-methylpropyl)undeca-2,4,6-trienamide	ITXVWRWWWSDXOY-HNZFQPFYSA-N
(2S)-2-[(3-Methylbutanoyl)amino]pentanedioic acid	HCVIJONZMYVLAW-ZETCQYMHSA-N
(2S-,cis)-5-Benzyl-3,6-dioxo-2-piperazineacetic acid	VNHJXYUDIBQDDX-UWVGGRQHSA-N
(3-beta)-Allopregnanolone sulfate	MENQCIVHHONJLU-FZCSVUEKSA-N
(3-Oxo-2,3-dihydro-4H-1,4-benzoxazin-4-yl)acetic acid	POGLODLVBYOXAO-UHFFFAOYSA-N
(4-Methylphenyl)oxidanesulfonic acid	WGNAKZGUSRVWRH-UHFFFAOYSA-N
(4S)-4-Amino-5-hydroxypentanoic acid	JPYGFLFUDLRNKX-BYPYZUCNSA-N
(6E,10Z,14E)-6,14-Dimethyl-3-methylidene-2-oxo-3a,4,5,8,9,12,13,15a-octahydrocyclotetradeca[b]furan-10-carboxylic acid	SORYERHBQFTRIK-AMVXUGKJSA-N
(Carboxymethyl)dodecyldimethylammonium cation	DVEKCXOJTLDBFE-UHFFFAOYSA-O
1,1-Dimethyl-2,3,4,9-tetrahydro-1H-beta-carboline-3-carboxylic acid	CWGHZIOJFVTEPN-UHFFFAOYSA-N
1,2-Dihexadecanoyl-3-(9Z-octadecenoyl)-sn-glycerol	YHMDGPZOSGBQRH-YYSBDVFPSA-N
1,2-Dilinoleoyl-sn-glycero-3-phosphocholine	FVXDQWZBHIXIEJ-LNDKUQBDSA-N
1,2-Dimyristoyl-sn-glycerol	JFBCSFJKETUREV-LJAQVGFWSA-N
1,2-Dioctadecanoyl-sn-glycerol	UHUSDOQQWJGJQS-QNGWXLTQSA-N
1,2-dioleoyl-sn-glycero-3-phosphatidylcholine	SNKAWJBJQDLSFF-NVKMUCNASA-N
1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine	MWRBNPKJOOWZPW-NYVOMTAGSA-N
1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine-N,N-dimethyl	XHPZRQBHFOVLEJ-UNUIOPIBSA-N
1,2-Dioleoyl-sn-glycerol	AFSHUZFNMVJNKX-LLWMBOQKSA-N
1,2-Dipentadecanoyl-sn-glycero-3-phosphocholine	LJARBVLDSOWRJT-PSXMRANNSA-N
1,3-Diolein	DRAWQKGUORNASA-CLFAGFIQSA-N
1,7-Dimethyluric acid	NOFNCLGCUJJPKU-UHFFFAOYSA-N
12-Methoxycarnosic acid	QQNSARJGBPMQDI-MZVUKIKXSA-N
13-cis-Retinol	FPIPGXGPPPQFEQ-HWCYFHEPSA-N
18-beta-Glycyrrhetinic acid	MPDGHEJMBKOTSU-YKLVYJNSSA-N
1-Arachidoyl-2-hydroxy-sn-glycero-3-phosphocholine	UATOAILWGVYRQS-HHHXNRCGSA-N
1H-Indole-3-propanoic acid	GOLXRNDWAUTYKT-UHFFFAOYSA-N
1-Methyl-3,6-(1H,2H)-pyridazinedione	UAECOHJYXUJDOF-UHFFFAOYSA-N
1-Methyluric acid	QFDRTQONISXGJA-UHFFFAOYSA-N
1-Monolinolenin	GGJRAQULURVTAJ-PDBXOOCHSA-N
1-Octadecanoyl-2-octadecenoyl-sn-glycero-3-phosphocholine	ATHVAWFAEPLPPQ-QPOMNCEOSA-N
1-O-Hexadecyl-2-O-acetyl-sn-glyceryl-3-phosphorylcholine	HVAUUPRFYPCOCA-AREMUKBSSA-N

Metabolite ID	InChIKey
1-Oleoyl-2-linoleoyl-rac-glycerol	BLZVZPYMHLXLHG-RQOIEFAZSA-N
1-Oleoyl-2-palmitoyl-rac-glycerol	DOZKMFVMCATMEH-ZCXUNETKSA-N
1-Oleoyl-sn-glycero-3-phosphoethanolamine	PYVRVRFVLRNJLY-MZMPXXGTSA-N
1-Palmitoyl-2-docosahexaenoyl-sn-glycero-3-phosphocholine	IESVDEZGAHUQJU-ZLBXKVHBSA-N
1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	YVYMBNSKXOXSKW-HXUWFJFHSA-N
1-Palmitoyl-2-linoleoyl-rac-glycerol	SVXWJFFKLMLOHO-BCTRXSSUSA-N
1-Palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine	JLPULHDHAOZNQI-ZTIMHPMXSA-N
1-Palmitoyl-2-oleoyl-3-linoleoyl-rac-glycerol	KGLAHZTWGPHKFF-FBSASISJSA-N
1-Palmitoyl-2-oleoyl-sn-glycerol	YEJYLHKQOBOSCP-OZKTZCCCSA-N
1-Palmitoyl-sn-glycero-3-phosphocholine	ASWBNKHCZGQVJV-HSZRJFAPSA-N
1-Pentadecanoyl-sn-glycero-3-phosphocholine	RJZVWDTYEWCUAR-JOCHJYFZSA-N
1-Stearoyl-2-hydroxy-sn-glycero-3-phosphocholine	IHNKQIMGVNPMTC-RUZDIDTESA-N
1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	BBYWOYAFBUOUFP-JOCHJYFZSA-N
1-Stearoyl-2-linoleoyl-sn-glycero-3-phosphoethanolamine	YDTWOEYVDRKKCR-KNERPIHHSA-N
1-Stearoyl-2-oleoyl-sn-glycero-3-phosphocholine	ATHVAWFAEPLPPQ-VRDBWYNSSA-N
2-({[3-Methyl-4-(2,2,2-trifluoroethoxy)-2-pyridinyl]methyl}sulfanyl)-1H-benzimidazole	CCHLMSUZHFPSFC-UHFFFAOYSA-N
2'-(4-Methylumbelliferyl)-.alpha.-D-N-acetylneuraminic acid	KKDWIUJBUSOPGC-GKHMPSLRSA-N
2-(Hexadecylamino)ethanol	SIFHZKKXWMJWOB-UHFFFAOYSA-N
2-(O-beta-D-Glucopyranosyl)atractyligenin	ZUDBDCBPZHGAEL-ICRASABDSA-N
2,3,4,9-Tetrahydro-1H-beta-carboline-3-carboxylic acid	FSNCEEGOMTYXKY-UHFFFAOYSA-N
2,3-Dihydroxypropyl octadecanoate	VBICKXHEKHSIBG-UHFFFAOYSA-N
2,4-Bis(1,1-dimethylethyl)phenol phosphate	GUDSEWUOWPVZPC-UHFFFAOYSA-N
2-[(2R,4aS,8S,8aS)-8-{2-[(4aS,7R,8aR)-7-(1-Carboxyvinyl)-1-hydroxy-4a-methyl-2-oxo-1,2,4a,5,6,7,8,8a-octahydro-1-naphthalenyl]ethyl}-4a-methyl-7-oxo-1,2,3,4,4a,7,8,8a-octahydro-2-naphthalenyl]acrylic acid	BSSWIRVAMSGQMG-RTTXQYISSA-N
2-[3-[3-Hydroxy-4-(4-hydroxyphenyl)-2,5-dimethoxyphenyl]-5-oxo-2H-furan-2-yl]acetic acid	CLNCCYGZTAZYLJ-UHFFFAOYSA-N
24,25-Epoxylanost-7-ene-3,23-diol	UMTABACRBSGXGK-UHFFFAOYSA-N
2'-Deoxyadenosine	OLXZPDWKRNYJJZ-RRKCRQDMSA-N
2'-Deoxyinosine	VGONTNSXDCQUGY-JXBXZBNISA-N
2-Hydroxypalmitic acid	JGHSBPIZNUXPLA-UHFFFAOYSA-N
2-Linoleoyl-1-palmitoyl-sn-glycero-3-phosphoethanolamine	HBZNVZIRJWODIB-NHCUFCNUSA-N
2-Oleoyl-1-palmitoyl-sn-glycero-3-phosphocholine	WTJKGGKOPKCXLL-VYOBOKEXSA-N
2-Palmitoyl-rac-glycerol	BBNYCLAREVXOSG-UHFFFAOYSA-N
2-Thio-acetyl MAGE	SJNRNWWBXZOALQ-OAQYLSRUSA-N
3-(1H-Imidazol-4-yl)propanoic acid	ZCKYOWGFRHAZIQ-UHFFFAOYSA-N
3-(3-Hydroxyphenyl)-2-oxopropanoic acid	PNYWALDMLUDDTA-UHFFFAOYSA-N
3-(Tetradecylamino)-1-propanol	FDZNQCUTOMWCJE-UHFFFAOYSA-N
3,7,12-Trihydroxycholan-24-oic acid (stereoisomer unknown)	BHQCQFFYRZLCQQ-UHFFFAOYSA-N
3,7-Dimethyluric acid	HMLZLHKHNBLLJD-UHFFFAOYSA-N

Metabolite ID	InChIKey
3,7-Dimethyluric acid	HMLZLHKHNBLLJD-UHFFFAOYSA-N
3-[5-(2-Methylpropyl)-3,6-dioxopiperazin-2-yl]propanoic acid	CSRXCVLPMJBQIM-UHFFFAOYSA-N
3-Hydroxyurs-12-en-23-oic acid	NBGQZFQREPIKMG-UHFFFAOYSA-N
3-Methylxanthine	GMSNIKWWOQHZGF-UHFFFAOYSA-N
4-(3,4-Dihydroxyphenyl)-6,7-dihydroxy-2-naphthalenecarboxylic acid	ZSKDVJYWOHBGNI-UHFFFAOYSA-N
4-Amino-5-methoxy-2-methylbenzenesulfonic acid	JBAVAJIXZVRJHT-UHFFFAOYSA-N
4-Cholestenone	NYOXRYYXRWJDKP-GYKMGIIDSA-N
4-Ketopimelic acid	UDDSEESQRGPVIL-UHFFFAOYSA-N
4-Oxododecanedioic acid	HHXMOTDTSDYYEI-UHFFFAOYSA-N
4-Pyridoxic acid	HXACOUQIXZGNBF-UHFFFAOYSA-N
5,6-Dihydrouridine	ZPTBLXKRQACLCR-XVFCMESISA-N
5-Hydroxyindole-3-acetic acid	DUUGKQCEGZLZNO-UHFFFAOYSA-N
6-Heptadecen-16-yne-1,2,4-triol	YPRUUOUDKJWVMW-VAWYXSNFSA-N
6-Methyl-1,2,3-oxathiazin-4(3H)-one 2,2-dioxide	YGCFIWIQZPHFLU-UHFFFAOYSA-N
8-(3-Octyl-2-oxiranyl)octanoic acid	IMYZYCNQZDBZBQ-UHFFFAOYSA-N
8-Acetamido-2-methyl-7-oxononanoic acid	UALLEVOGEQZKSX-UHFFFAOYSA-N
9,10-Dihydroxy-12Z-octadecenoic acid	XEBKSQSGNGRGDW-CJWPDFJNSA-N
9-Oxo-10(E),12(E)-octadecadienoic acid	LUZSWWYKKLTDHU-SIGMCMEVSA-N
9Z,11E,13E-Octadecatrienoic acid	CUXYLFPMQMFGPL-WPOADVJFSA-N
Adipic acid	WNLRTRBMVRJNCN-UHFFFAOYSA-N
Allyl 4-amino-1-piperidinecarboxylate	XKLGRJGBMKOQDK-UHFFFAOYSA-N
alpha,alpha-Dilaurin	KUVAEMGNHJQSMH-UHFFFAOYSA-N
alpha-Hyodeoxycholic acid methyl ester	BWDRDVHYVJQWBO-QWXHOCAMSA-N
alpha-Tocopheryl acetate	ZAKOWWREFLAJOT-PDNQPUDYSA-N
Anacardic acid	ADFWQBGTDJIESE-UHFFFAOYSA-N
Anacardic acid diene	KAOMOVYHGLSFHQ-UTOQUPLUSA-N
Asn-Pro	GADKFYNESXNRLC-WDSKDSINSA-N
Asp-Pro	UKGGPJNBONZZCM-WDSKDSINSA-N
Asp-Val	CPMKYMGGYUFOHS-CAHLUQPWSA-N
Azaspiracid-8	VTXJIDBQMAXDIY-CQQSUDGLSA-N
Bacilysocin	DIGHTWUQPWHBPG-UHFFFAOYSA-N
Benzenesulfonic acid, 4-(acetylamino)-	ZQPVMSLLKQTRMG-UHFFFAOYSA-N
Benzyldimethyldodecylammonium cation	CYDRXTMLKJDRQH-UHFFFAOYSA-N
Benzyldimethylstearylammonium cation	FWLORMQUOWCQPO-UHFFFAOYSA-N
beta-Cryptoxanthin	DMASLKHVQRHNES-FKKUPVFPSA-N
beta-Hyodeoxycholic acid	DGABKXLVXPYZII-MMTMODRTSA-N
beta-Sitosterol	KZJWDPNRJALLNS-VJSFXXLFSA-N
Bilirubin	BPYKTIZUTYGOLE-IFADSCNNSA-N
Bis-(3,4-dimethyldibenzylidenesorbitol)	YWEWWNPYDDHZDI-JJKKTNRVSA-N
Chenodeoxycholic acid	RUDATBOHQWOJDD-BSWAIDMHSA-N

Metabolite ID	InChIKey
Cholan-24-oic acid, 12-hydroxy-3-oxo-	WMUMZOAFCDOTRW-UHFFFAOYSA-N
Cholan-24-oic acid, 12-hydroxy-3-oxo-, methyl ester	LOJPGRHPJBGMMF-JJNSWQHQSA-N
Cholestan-3-ol, (3 alpha,5 beta)-	QYIXCDOBOSTCEI-VZNZRSCWSA-N
Cholesterol	HVYWMOMLDIMFJA-DPAQBDIFSA-N
Cholesterol 3-sulfate	BHYOQNUELFTYRT-DPAQBDIFSA-N
Cholic acid	BHQCQFFYRZLCQQ-OELDTZBJSA-N
cis,cis-9,12-Octadecadien-1-ol	JXNPEDYJTDQORS-HZJYTTRNSA-N
cis-Vaccenic acid	UWHZIFQPPBDJPM-FPLPWBNLSA-N
Citrulline	RHGKLRLOHDJJDR-UHFFFAOYSA-N
Cocamidopropyl betaine	MRUAUOIMASANKQ-UHFFFAOYSA-N
D-alpha-Tocopherol succinate	IELOKBJPULMYRW-NJQVLOCASA-N
Dehydropipernonaline	KAYVDASZRFLFRZ-PQECNABGSA-N
Deoxycholic acid	KXGVEGMKQFWNSR-LLQZFEROSA-N
D-erythro-N-Stearoylsphingosine	WWUZIQQURGPMPG-KRWOKUGFSA-N
D-erythro-Sphinganine	OTKJDMGTUTTYMP-ZWKOTPCHSA-N
D-Glucosyl-beta-1,1-N-palmitoyl-D-erythro-sphingosine	VJLLLMIZEJJZTE-NNTBDIJYSA-N
Didymin	RMCRQBAILCLJGU-HIBKWJPLSA-N
Dihydrodaidzein	JHYXBPPMXZIHKG-UHFFFAOYSA-N
Dimethyl azelate	DRUKNYVQGHETPO-UHFFFAOYSA-N
Di-tert-butyl L-glutamate	NTUGPDFKMVHCCJ-VIFPVBQESA-N
DL-beta-Hydroxypalmitic acid	CBWALJHXHCJYTE-UHFFFAOYSA-N
DL-Indole-3-lactic acid	XGILAAMKEQUXLS-UHFFFAOYSA-N
DL-Serine, succinate (ester)	ZAHSBRLHJRVFAU-UHFFFAOYSA-N
Docosanoyl ethanolamide	XHFWUECSNJWBJU-UHFFFAOYSA-N
Dodecanedioic acid	TVIDDXQYHWJXFK-UHFFFAOYSA-N
Ergocalciferol	MECHNRXZTMCUDQ-ANGGWTPUSA-N
Erucamide	UAUDZVJPLUQNMU-KTKRTIGZSA-N
Formamide, N-[(4-amino-2-methyl-5-pyrimidinyl)methyl]-N-[4-hydroxy-1-methyl-2-(propyldithio)-1-buten-1-yl]-	UDCIYVVYDCXLSX-SDNWHVSQSA-N
Gabapentin	UGJMXCAKCUNAIE-UHFFFAOYSA-N
Gamithromycin	VWAMTBXLZPEDQO-UZSBJOJWSA-N
gamma-Glu-Phe	XHHOHZPNYFQJKL-QWRGUYRKSA-N
Genkwanin	JPMYFOBNRRGFNO-UHFFFAOYSA-N
Ginkgolic acid I	YXHVCZZLWZYHSA-FPLPWBNLSA-N
Gln-Val	MRVYVEQPNDSWLH-XPUUQOCRSA-N
Glu-Glu	KOSRFJWDECSPRO-UHFFFAOYSA-N
Glu-Leu	YBAFDPFAUTYYRW-SFYZADRCSA-N
Glycerol monooleate	RZRNAYUHWVFMIP-KTKRTIGZSA-N
Glycerol trilaurate	VMPHSYLJUKZBJJ-UHFFFAOYSA-N
Glycerol trioleate	PHYFQTYBJUILEZ-IUPFWZBJSA-N
Glycodeoxycholic acid	WVULKSPCQVQLCU-UHFFFAOYSA-N

Metabolite ID	InChIKey
Gly-Pro	KZNQNBZMBZJQJO-YFKPBYRVSA-N
Gly-Val	STKYPAFSDFAEPH-LURJTMIESA-N
Guanine	UYTPUPDQBNUYGX-UHFFFAOYSA-N
Guanosine	NYHBQMYGNKIUIF-UUOKFMHZSA-N
Guanosine 3',5'-cyclic monophosphate	ZOOGRGPOEVQQDX-BDXYJKHTSA-N
Guineensine	FPMPOFBEYSSYDQ-AUVZEZIHSA-N
Heptadecasphinganine	KFQUQCFJDMSIJF-DLBZAZTESA-N
His-Pro	LNCFUHAPNTYMJB-IUCAKERBSA-N
Histidine	HNDVDQJCIGZPNO-UHFFFAOYSA-N
Ile-Glu	KTGFOCFYOZQVRJ-ACLDMZEESA-N
Ile-Phe	WMDZARSFSMZOQO-UHTWSYAYSA-N
Ile-Thr	DRCKHKZYDLJYFQ-YWIQKCBGSA-N
Ile-Val	BCXBIONYYJCSDF-CIUDSAMLSA-N
Inositol hexanicotinate	MFZCIDXOLLEMOO-GYSGTQPESA-N
Lactosyl-Ceramide(d18:1/22:0)	QYWVASPEUXEHSY-DYGYHFHFSA-N
Laurylsulfuric acid	MOTZDAYCYVMXPC-UHFFFAOYSA-N
Leu-Leu	LCPYQJIKPJDLLB-ZJUUUORDSA-N
Leu-Pro	VTJUNIYRYIAIHF-IUCAKERBSA-N
Linoleic acid-biotin	TXYWCIYPXRNWJU-MJDXKYDGSA-N
Linoleoyl ethanolamide	KQXDGUVSAAQARU-HZJYTTRNSA-N
Lithocholic acid	SMEROWZSTRWXGI-HVATVPOCSA-N
Lumichrome	ZJTJUVIJVLLGSP-UHFFFAOYSA-N
Lysine	KDXKERNSBIXSRK-UHFFFAOYSA-N
Lys-Leu	ATIPDCIQTUXABX-UWVGGRQHSA-N
Lys-Pro	AIXUQKMMBQJZCU-IUCAKERBSA-N
Methionine sulfoxide	QEFRNWWLZKMPFJ-YGVKFDHGSA-N
Methyl linolenate	DVWSXZIHSUZZKJ-YSTUJMKBSA-N
Met-Pro	DZMGFGQBRYWJOR-YUMQZZPRSA-N
Met-Val	BJFJQOMZCSHBMY-SFYZADRCSA-N
Monobehenin	OKMWKBLSFKFYGZ-UHFFFAOYSA-N
Myristyl sulfate	URLJMZWTXZTZRR-UHFFFAOYSA-N
N-(14-Methylpentadecanoyl)phenylalanine	WICAPTTWPBEVPW-UHFFFAOYSA-N
N-(2-Amino-4-methylpentanoyl)phenylalanine	KFKWRHQBZQICHA-UHFFFAOYSA-N
N,N-Dimethylsphingosine	YRXOQXUDKDCXME-YIVRLKKSSA-N
N2-Isobutyryl-2'-deoxyguanosine	SIDXEQFMTMICKG-UHFFFAOYSA-N
N6-Palmitoyl-L-lysine	IWKZTTDNVPAHNP-FQEVSTJZSA-N
N-Acetyldihydrosphingosine	CRJGESKKUOMBCT-VQTJNVASSA-N
N-Acetyl-L-alanine	KTHDTJVBEPMMGL-UHFFFAOYSA-N
N-Acetyl-L-aspartic acid	OTCCIMWXFLJLIA-BYPYZUCNSA-N
N-Acetyl-L-glutamic acid	RFMMMVDNIPUKGG-YFKPBYRVSA-N

Metabolite ID	InChIKey
N-Acetyl-L-tyrosine	CAHKINHBCWCHCF-JTQLQIEISA-N
N-Acetylputrescine	KLZGKIDSEJWEDW-UHFFFAOYSA-N
N-Acetyltyramine	ATDWJOOPFDQZNK-UHFFFAOYSA-N
Nadolol	VWPOSFSPZNDTMJ-UCWKZMIHSA-N
N-alpha-Acetyl-L-lysine	VEYYWZRYIYDQJM-ZETCQYMHSA-N
N-Benzyl-N,N-dimethyl-1-tetradecanaminium cation	WNBGYVXHFTYOBY-UHFFFAOYSA-N
N-Desmethylloperamide	ZMOPTLXEYOVARP-UHFFFAOYSA-N
N-Docosanoyl-4-sphingenyl-1-O-phosphorylcholine	FJJANLYCZUNFSE-TWKUQIQBSA-N
N-Lauroyl-D-erythro-sphinganine	UHWYQXNZIBLESO-URLMMPGGSA-N
N-L-gamma-Glutamyl-L-leucine	MYFMARDICOWMQP-UHFFFAOYSA-N
N-Methyl-L-phenylalanine	SCIFESDRCALIIM-VIFPVBQESA-N
N-Oleoyl-4-sphingenine	OBFSLMQLPNKVRW-RHPAUOISSA-N
N-Oleoyl-D-erythro-sphinganine	MJQIARGPQMNBGT-WWUCIAQXSA-N
N-Palmitoyl-D-sphingosine	YDNKGFDKKRUKPY-TURZORIXSA-N
N-Palmitoyltryptamine	YZWBXGMMZOEBAN-UHFFFAOYSA-N
N-Stearoylsphinganine	KZTJQXAANJHSCE-OIDHKYIRSA-N
N-Tetracosenoyl-4-sphingenine	VJSBNBBOSZJDKB-KPEYJIHVSA-N
Oleanolic acid	MIJYXULNPSFWEK-GTOFXWBISA-N
Oleic acid	ZQPPMHVWECSIRJ-KTKRTIGZSA-N
Oleoyl ethanolamide	BOWVQLFMWHZBEF-KTKRTIGZSA-N
Oleoyl ethylamide	JZJYYCFYGXPUMF-QXMHVHEDSA-N
Oleyl sulfate	ZUBJEHHGZYTRPH-KTKRTIGZSA-N
Palmitoyl ethanolamide	HXYVTAGFYLMHSO-UHFFFAOYSA-N
Palmitoyl sphingomyelin	RWKUXQNLWDTSLO-GWQJGLRPSA-N
Palmitoyleicosapentaenoyl phosphatidylcholine	KLTHQSWIRFFBRI-KOQZQRJKSA-N
Pantothenic acid	GHOKWGTUZJEAQD-UHFFFAOYSA-N
PEG15	ILLKMACMBHTSHP-UHFFFAOYSA-N
PEG16	OWTQQPNDSWCHOV-UHFFFAOYSA-N
Pentacosanoic acid	MWMPEAHGUXCSMY-UHFFFAOYSA-N
Pheophorbide a	NSFSLUUZQIAOOX-QEWKCGBTSA-N
Phe-Val	IEHDJWSAXBGJIP-RYUDHWBXSA-N
Phytosphingosine 1-phosphate	AYGOSKULTISFCW-KSZLIROESA-N
Piperine	MXXWOMGUGJBKIW-YPCIICBESA-N
Pregnanetriol	SCPADBBISMMJAW-UHHUKTEYSA-N
Pregnenolone sulfate	DIJBBUIOWGGQOP-QGVNFLHTSA-N
Pro-Val	AWJGUZSYVIVZGP-YUMQZZPRSA-N
Riboflavin	AUNGANRZJHBGPY-SCRDCRAPSA-N
Ricinoleic acid	WBHHMMIMDMUBKC-QJWNTBNXSA-N
Saccharin	CVHZOJJKTDOEJC-UHFFFAOYSA-N
Secalciferol	FCKJYANJHNLEEP-XRWYNYHCSA-N

Metabolite ID	InChIKey
Ser-Pro	WBAXJMCUFIXCNI-WDSKDSINSA-N
Sertraline	VGKDLMBJGBXTGI-YVEFUNNKSA-N
Stearamide	LYRFLYHAGKPMFH-UHFFFAOYSA-N
Stearoyl ethanolamide	OTGQIQQTPXJQRG-UHFFFAOYSA-N
Stercobilin	TYOWQSLRVAUSMI-WKULSOCRSA-N
Stigmastadienone	MKGZDUKUQPPHFM-DPZQRLQDSA-N
Suberic acid	TYFQFVWCELRYAO-UHFFFAOYSA-N
Sucralose	BAQAVOSOZGMPRM-QBMZZYIRSA-N
Termitomycamide E	JTSNNDBVDQFVEG-HZJYTTRNSA-N
Terrestribisamide	CHEMZHJQHCVLFI-MKICQXMISA-N
Tetradecanedioic acid	HQHCYKULIHKCEB-UHFFFAOYSA-N
Tetradecanedioic acid, 5,6-dihydroxy-	JCOCIXAHAFALHI-UHFFFAOYSA-N
Tetrahydroharman-3-carboxylic acid	ZUPHXNBLQCSEIA-UHFFFAOYSA-N
Thiamine cation	JZRWCGZRTZMZEH-UHFFFAOYSA-N
Threonine	AYFVYJQAPQTCCC-UHFFFAOYSA-N
Thr-Leu	BQBCIBCLXBKYHW-BIIVOSGPSA-N
Thr-Val	CKHWEVXPLJBEOZ-QYNIQEEDSA-N
Tomatidine	XYNPYHXGMWJBLV-KLPIGKMYSA-N
trans-Petroselinic Acid	CNVZJPUDSLNTQU-OUKQBFOZSA-N
trans-Vaccenic acid	UWHZIFQPPBDJPM-BQYQJAHWSA-N
Tridemorph	YTOPFCCWCSOHFV-UHFFFAOYSA-N
Trihydroxycholestanoic acid	CNWPIIOQKZNXBB-VCVMUKOKSA-N
Trilinolein	HBOQXIRUPVQLKX-BBWANDEASA-N
Tyr-Pro	VNYDHJARLHNEGA-RYUDHWBXSA-N
Tyr-Val	OYOQKMOWUDVWCR-RYUDHWBXSA-N
Ubiquinone 9	UUGXJSBPSRROMU-WJNLUYJISA-N
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N
Uric acid	LEHOTFFKMJEONL-UHFFFAOYSA-N
Urobilin	KDCCOOGTVSRCHX-UYMYUHGCSA-N
Urocanic acid	LOIYMIARKYCTBW-OWOJBTEDSA-N
Val-Glu	UPJONISHZRADBH-POYBYMJQSA-N
Val-Pro	GIAZPLMMQOERPN-YUMQZZPRSA-N
Val-Val	KRNYOVHEKOBTEF-UHFFFAOYSA-N
Valylphenylalanine	GJNDXQBALKCYSZ-UHFFFAOYSA-N
Victoria Pure Blue BO	ALMMLSQBERCMRB-UHFFFAOYSA-O
Xanthine	LRFVTYWOQMYALW-UHFFFAOYSA-N

Appendix G. Detailed Methods and Results for NMR Metabolomics

G.1. Bligh-Dyer Biphasic Solvent Extraction of Polar Metabolites - 700 MHz

Extraction solvents (methanol, chloroform, and deionized water) were pre-chilled in a -80 °C freezer and then kept on ice during use. Each sample was slightly thawed under cold running water (~30 s) to facilitate subsequent transfer of the frozen stool sample into a pre-labelled 10 mL conical glass tube kept on ice. The sample was then quickly returned to a rack in the -80 °C freezer. These pre-extraction steps were repeated for all omnivore and vegetarian samples. Once samples were transferred into 10 mL conical glass tubes, they were placed on ice and a modified Bligh-Dyer solvent extraction protocol was used for polar metabolite extraction. Specifically, a 2:2:1.8 volume ratio solvent system of methanol:chloroform:water was added to each sample. The specific solvent volumes to use in the extraction were determined based on the moisture determination conducted on RM 8048 (average water content = 97 %). The volumes were added in a two-step method according to Wu et al. (2008). First, 1100 µL (2 x 550 µL) of cold methanol were added to each sample. Water was not added to the fecal samples as the moisture content was taken into account in the solvent volume calculations. Extraction blanks required 440 µL of cold Milli-Q water (minimum resistivity 18.2 MΩ-cm) to induce solvent partitioning. Samples were vortexed for 60 s. Then, 1100 µL (2 x 550 µL) of cold chloroform were added to each sample. Water was not added to the fecal samples while extraction blanks required 550 µL of cold Milli-Q water. Samples were quickly vortexed after addition of chloroform to allow mixing of different solvents and returned to ice. When addition of non-polar solvents was completed for all samples in a batch, each sample was vortexed for 60 s (2 to 4 samples at a time) and incubated on ice for 10 min, followed by centrifugation (Eppendorf 5810R centrifuge, Eppendorf AG, Hamburg, Germany) at 2000 x *g* at 4 °C for 5 min to promote phase separation. Subsequently, the polar fraction (top layer in each sample) was separated using a glass pipet, carefully avoiding disturbance of the macromolecule layer at the phase partition interface between polar and non-polar fractions. The polar fraction was then transferred into preweighed, labelled 1.5 mL microcentrifuge tubes (Eppendorf). Due to the large solvent volume, the polar fraction was split and transferred into two sets of 1.5 mL microcentrifuge tubes. The first set of preweighed tubes were quickly spun (≈ 10 s) using a benchtop centrifuge and dried in a Vacufuge Concentrator 5301 (Eppendorf AG, Hamburg, Germany) at room temperature using the Alcohol setting (V-AL) for ≈ 2 h, before switching to the Water setting (V-AQ). When sample volumes were reduced to ≈ 250 µL, the remaining polar fraction (which had been put into the second set of microcentrifuge tubes and was maintained on ice) was added, and samples were dried to completion overnight. The dried metabolites were weighed using an analytical balance and were rehydrated in 600 µL of 100 mmol/L sodium phosphate NMR buffer in D_2O (pH = 7.3, measured) containing 1.0 mmol/L sodium 3-Trimethylsilyl 2,2,3,3-d_4 propionate (TMSP, CAS: 24493-21-8) as the chemical shift reference standard. The rehydrated samples were vortexed for ≈ 30 s and centrifuged at 14,000 rcf at 4 °C for 5 min to pellet any particulates. Samples (550 µL) were then transferred into 5 mm NMR tubes (Sample Vault Series™, Norell®, Morganton, NC), spun with a manual centrifuge, and randomized for subsequent NMR spectroscopy analysis.

NMR Analysis and Data Processing. One dimensional (1D) ^{1}H NMR spectra were acquired using the pulse program '*noesygppr1d*', a 60 ms mixing period for solvent suppression, and an acquisition time of 2.34 s for a total repetition time (D1 + AQ) of 5.34 s, with a relaxation delay of 3.0 s. The 1D experiments were executed with eight dummy scans and 120 acquisition scans with a spectral width of 20.0 ppm for a total of 65,536 data points. The resulting spectra were processed by zero-filling to 65,536 complex points and by multiplying the free induction decay (FID) by an exponential line broadening function of 0.3 Hz prior to Fourier transformation. The spectra were referenced to the chemical shift reference, TMSP, phased, and baseline-corrected with further manual corrections when necessary using TopSpin3.5 software.

Two-dimensional edited ^{1}H,^{13}C-HSQC spectra with adiabatic ^{13}C decoupling (*hsqcedetgpsisp2.2*) were collected on selected samples to aid metabolite identification. In general, 2048 data points with 256 scans and 512 increments were acquired with spectral widths of 11 ppm in the F2 dimension (^{1}H) and 180 ppm in F1 (^{13}C). A relaxation delay equal to 1.5 s was used between acquisitions, and a refocusing delay corresponding to a 145 Hz $1J_{C\text{-}H}$ coupling was used. The FIDs were weighted using a shifted sine-square function in both dimensions. Manual two-dimensional phasing was applied; all spectra were referenced to the TMSP internal standard at 0.00 ppm for ^{1}H and ^{13}C. Selected samples were also used for 2D Total Correlation Spectroscopy (TOCSY) analysis to elaborate the ^{1}H-^{1}H coupling networks offering more confidence in compound identification. A phase-sensitive TOCSY with a DIPSI2 sequence (90 ms mixing time) and water suppression via pre-saturation and excitation sculpting (*dipsi2esgpph*) was used. In general, 48 scans and 2048 data points with 256 increments were acquired with spectral widths of 10 ppm in the F1 and F2 dimensions. A relaxation delay equal to 1.5 s was used between acquisitions. The FIDs were weighted using a shifted sine-squared function in both dimensions, and the data were Fourier transformed to 4096 by 1024 complex points by zero filling. Manual two-dimensional phasing was applied; all spectra were referenced to the TMSP internal standard at 0.00 ppm for ^{1}H. Two-dimensional phase-sensitive gradient-enhanced ^{1}H, ^{13}C - HSQC-TOCSY (*hsqcdietgpsisp.2*) was acquired with 2048 data points with 312 scans and 512 increments were acquired with spectral widths of 11 ppm in F2 and 180 ppm in F1 (^{13}C). The relaxation delay was 1.5 s, and a refocusing delay corresponding to a 145 Hz $1J_{C\text{-}H}$ coupling was used.

Coefficient of Disagreement (COD) for homogeneity and stability assessment. Given the peak presence/absence for each identified bin from the spectra of two different samples, we are interested in quantifying how similar the two distributions of peak presence are to each other. The greater the agreement, the more homogeneous (or stable) the population of vials are. The greater the agreement between present peaks from samples measured at different times, the more homogeneous (or across time points, stable) the population of vials are. With this in mind, we define a metric that quantifies the level of discrepancy or disagreement between spectra from any pair of samples. We call this metric the coefficient of disagreement (COD). The procedure for calculating a COD between any pair of samples is as follows:

1. We first identify the bins for each group using the procedure described in the previous section.

2. For each sample in each bin, we must define what we consider as an identified "peak." If we consider z_{i1} to be the minimum boundary and z_{i2} to be the maximum boundary of bin i for $i = 1, \ldots, N$ (where N is the number of bins), then we are searching for local maxima in (z_{i1}, z_{i2}). Let $f(z)$ represent the intensity function within bin i, where $z \in (z_{i1}, z_{i2})$. We are searching for local maxima $z_{\text{peak}} \in (z_{i1}, z_{i2})$ such that:

   (a) The intensity at the peak satisfies:

   $$f(z_{\text{peak}}) > \mu_{\text{noise}} + 3\sigma_{\text{noise}}$$

   where μ_{noise} is the average intensity in the noise region (above 9.0 ppm), and σ_{noise} is the standard deviation of the intensity in the same region.

   (b) There exists a valley z_{valley} on either side of z_{peak}, with:

   $$f(z_{\text{peak}}) \geq 1.025 \cdot f(z_{\text{valley}})$$

   ensuring at least a 2.5 % increase from the valley to the peak.

   Each bin i can contain at most one such peak. The boundaries z_{i1} and z_{i2} cannot be identified as peaks.

3. Let P be a $K \times N$ matrix, where K is the number of samples and N is the number of bins. The matrix entry $P_{k,l}$ for sample k and bin i is defined as:

   We can then calculate the Jaccard distance for each pair of samples across all bins to quantify the dissimilarity of samples based on the presence or absence of identified peaks. We can calculate the values as

   where $P_{k_1} \cap P_{k_2}$ represents the bins where both samples k_1 and k_2 have a peak (intersection), $P_{k_1} \cup P_{k_2}$ represents the bins where at least one of the samples k_1 or k_2 has a peak (union), and $|\cdot|$ denotes the cardinality.

The COD measures the disagreement in peak identification between a pair of samples, based on potential peaks within bins determined by NMRProcFlow.

Metabolite Identification. Metabolites were assigned using the 1D ^{1}H, and 2D ^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY spectra and were putatively identified at a Level 2 where

possible[19]. Metabolite assignments were based on chemical shift comparisons using reference spectra from the Human Metabolome Database (HMDB)[20], the Biological Magnetic Resonance Bank (BMRB)[21], the spectral library in Chenomx NMR Analysis software (Chenomx Inc. version 8.5), as well as an in-house database. Specifically, in Chenomx Profiler module, the 1D ^{1}H Bruker pre-processed spectra (1r) were imported, the pH was calibrated to 7.3 ± 0.5 and the TMSP peak was fitted to the Chemical Shape Indicator (CSI) for 1 mmol/L. Once processed, the spectra were profiled against both the Chenomx 700 MHz spectral reference library and the HMDB 700 MHz spectral library to identify metabolites by chemical shift, peak shape and signal multiplicity (splitting). One dimensional ^{1}H NMR spectra can have significant metabolite peak overlap, which can lead to false positives during metabolite identification. For improved confidence in metabolite annotations, ^{1}H and ^{13}C chemical shift values were obtained from 2D ^{1}H,^{13}C-HSQC spectra acquired on representative samples for both omnivore and vegetarian donor pools, which allows enhanced peak resolution and signal dispersion resulting in reduced peak overlap. Additionally, the COLMARm web tool (https://spin.ccic.osu.edu/index.php/colmarm)[22] was used for automated metabolite IDs using the 2D NMR spectra including ^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY. Input files for the COLMARm tool were generated using MestReNova v.14.3.3 and included NMR data files for each spectrum (^{1}H,^{13}C-HSQC, ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC-TOCSY).

[19]Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics 3(3):211-221. https://doi.org/10.1007/s11306-007-0082-2

[20] Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y, Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M, Manach C, Scalbert A (2017) HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Research 46(D1):D608-D617. https://doi.org/10.1093/nar/gkx1089

[21] Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2007) BioMagResBank. Nucleic Acids Research 36(suppl_1):D402-D408. https://doi.org/10.1093/nar/gkm957

[22] Bingol K, Li D-W, Zhang B, Brüschweiler R (2016) Comprehensive Metabolite Identification Strategy Using Multiple Two-Dimensional NMR Spectra of a Complex Mixture Implemented in the COLMARm Web Server. Analytical Chemistry 88(24):12411-12418. https://doi.org/10.1021/acs.analchem.6b03724

G.1.1. Table - Omnivore Polar Metabolite Characterization List by NMR

Final NMR polar metabolite list for RM 8048 omnivore material (Bligh-Dyer extraction). ^{1}H and ^{13}C chemical shifts of metabolites identified by 1D and 2D NMR. ^{13}C shifts follow the same order as ^{1}H shifts in the preceding column. Corresponding values in the two columns are chemical shifts of $^{1}H/^{13}C$ pairs from HSQC experiments. Values in bold in the $\delta^{13}C$ column correspond to peaks detected in the HSQC; those not in bold were low intensity peaks. Signals marked "-" were too weak for detection by HSQC experiments. Multiplicity key: s, singlet; d, doublet, dd, doublet of doublets; ddd, doublet of doublet of doublets; t, triplet; q, quartet; m, multiplet.

Metabolite ID	InChIKey	$\delta^{1}H$ (ppm) and multiplicity	$\delta^{13}C$ (ppm)
(3-carboxypropyl)trimethylammonium	JHPNVNIEXXLNTR-UHFFFAOYSA-O	2.27(t), 3.14(s)	**36.41**, **55.72**
1,7-Dimethyluric acid	NOFNCLGCUJJPKU-UHFFFAOYSA-N	3.21(s), 3.31(s)	**30.0**, **30.53**
2-Aminobutyrate	QWCKQJZIFLGMSD-UHFFFAOYSA-N	0.99(t), 3.72(dd)	11.41, 58.70
2'-Deoxyuridine	MXHRCPNRJAMMIM-UHFFFAOYSA-N	2.40(m), 4.05(m), 5.90(d)	**41.48**, **89.60**, **105.14**
2-Methylbutyrate	WLAMNBDJUVNPJU-BYPYZUCNSA-N	0.86(t), 1.05(d), 1.39(m), 1.49(m), 2.21(m)	**14.38**, -, **30.09**, 30.18, **47.56**
3-Hydroxyphenylacetate	FVMDYYGIDFPZAX-UHFFFAOYSA-N	3.48(s)	**47.21**
3-Methyl-2-oxovalerate	JVQYSWDUAOAHFM-UHFFFAOYSA-M	1.10(d)	**16.66**
3-Phenylpropionate	XMIIGOLPHOKFCH-UHFFFAOYSA-N	2.49(t), 2.89(t), 7.27(t), 7.32(d), 7.37(t)	**42.11**, **34.74**, **129.10**, **131.20**, **131.61**
4-Hydroxyphenylacetate	XQXPVVBIMDBYFF-UHFFFAOYSA-M	3.45(s), 6.85(m), 7.18(m)	**46.42**, 118.39, 133.66
4-Hydroxyphenylpropionate	ZHMMPVANGNPCBW-UHFFFAOYSA-M	2.44(t), 2.81(t), 6.85(m), 7.18(m)	**42.38**, **33.96**, **118.39**, **132.52**
5-Aminopentanoate	JJMDCOVWQOJGCB-UHFFFAOYSA-M	1.64(m), 1.68(m), 2.24(t), 3.02(t)	**25.46**, **29.41**, **39.59**, **42.03**
Acetate	QTBSBXVTEAMEQO-UHFFFAOYSA-M	1.92(s)	**26.14**
Alanine	QNAYBMKLOCPYGJ-UHFFFAOYSA-N	1.48(d), 3.78(q)	**18.94**, **53.35**
Alpha-ketoisovaleric acid	QHKABHOOEWYVLI-UHFFFAOYSA-N	1.13(d)	**19.17**
Aminocaproic acid	SLXKOJJOQWFEFD-UHFFFAOYSA-N	1.60(m)	**27.90**
Arabinose	SRBFZHDQGSBBOR-HWQSCIPKSA-N	4.51(d)	**99.65**
Aspartate	CKLJMWTZIZZHCS-UHFFFAOYSA-L	2.68(dd), 2.82(dd), 3.90(dd)	**39.44**, **39.44**, **55.11**
Butyrate	FERIUCNNQQJTOY-UHFFFAOYSA-M	0.90(t), 1.56(m), 2.16(t)	**16.05**, **22.19**, **42.47**
Caprylate	WWZKQHOCKIZLMA-UHFFFAOYSA-N	0.87(t), 1.27(m), 1.29(m), 1.30(m), 1.55(m),	**16.13**, **33.82**, **31.41**, **24.76**, **28.60**,
		2.18(m)	**40.36**

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
Choline	OEYIOHPDSNJKLS-UHFFFAOYSA-N	3.21(s), 3.53(m)	**56.78**, **70.21**
Citrulline	RHGKLRLOHDJJDR-UHFFFAOYSA-N	1.86(m), 1.88(m), 3.14(m)	**30.44**, **30.53**, **42.03**
Deoxycholic acid	KXGVEGMKQFWNSR-LLQZFEROSA-N	0.71(s), 0.94(s), 0.96(s)	**14.91**, **25.09**, **19.47**
Ethanolamine	HZAXFHJVJLSVMW-UHFFFAOYSA-N	3.14(t), 3.83(t)	**44.05**, **60.46**
Formate	BDAGIHXWWSANSR-UHFFFAOYSA-M	8.46(s)	
Fructose	BJHIKXHVCXFQLS-UHFFFAOYSA-N	3.71(m), 3.83(m), 3.90(dd), 4.12(m)	**66.31**, **83.77**, **72.42**, **77.74**
Fucose	PNNNRSAQSRJVSB-UHFFFAOYSA-N	1.21(d), 1.25(d), 3.44(dd), 4.55(d), 5.21(d)	**18.51**, **18.42**, **74.81**, **99.08**, **95.07**
Fumarate	VZCYOOQTPOCHFL-OWOJBTEDSA-L	6.52(s)	**138.23**
Galactose	GZCGUPFRVQAUEE-UHFFFAOYSA-N	3.50(dd), 3.67(dd), 3.71(m), 3.81(dd), 3.93(dd),	74.68, **75.52**, **78.02**, 71.26, **71.70**,
		4.09(m), 4.58(d), 5.27(d)	**73.37, 99.35, 95.19**
Galacturonic acid	IAJILQKETJEXLJ-RSJOWCBRSA-N	3.50(dd), 3.90(dd), 4.06(d), 4.28(dd), 5.30(d)	74.67, **72.22**, **78.45**, **73.51**, **94.77**
Gluconate	RGHNJXZEOKUKBD-UHFFFAOYSA-M	3.78(m), 4.11(d)	**65.38**, **76.96**
Glucose	GZCGUPFRVQAUEE-UHFFFAOYSA-N	3.25(dd), 3.41(t), 3.42(t), 3.47(m), 3.49(t),	**77.05**, **72.49**, **72.52**, **78.81**, **78.63**,
		3.54 (dd), 3.72(dd), 3.72(t), 3.77(dd), 3.84(m),	**63.62, 74.33, 75.64, 63.50, 74.33,**
		3.84(dd), 3.90(dd), 4.65(d), 5.24(d)	**63.40, 63.62, 98.82, 94.95**
Glutamate	WHUUTDBJXJRKMK-UHFFFAOYSA-N	2.09(m), 2.36(m), 3.76(dd)	**29.83**, **36.32**, **57.48**
Glutamine	ZDXPYRJPNDTMRX-UHFFFAOYSA-N	2.15(m), 2.46(m), 3.78(t)	29.11, **33.68**, 56.96
Glutarate	JFCQEDHGNNZCLN-UHFFFAOYSA-N	1.79(m), 2.19(t)	**25.79**, 40.13
Glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N	3.56(dd), 3.65(dd), 3.79(m)	**65.35**, **65.40**, **74.98**
Glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N	3.56(s)	**44.22**
Glycolate	AEMRFAOFKBGASW-UHFFFAOYSA-M	3.94(s)	**64.15**
Hexanoic acid	FUZZWVXGSFPDMH-UHFFFAOYSA-N	1.29(t)	**33.96**
Hypoxanthine	FDGQSTZJBFJUBT-UHFFFAOYSA-N	8.19(s), 8.21(s)	**148.06**, **144.46**
Isobutyrate	KQNPFQTWMSNSAP-UHFFFAOYSA-M	1.06(d), 2.39(m)	**22.28**, **39.83**
Isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N	0.94(t), 1.01(d), 1.27(m), 1.46(m), 1.98(m),	**13.94**, **17.45**, **27.37**, **27.37**, **38.78**,
		3.68(d)	**62.39**
Isovalerate	GWYFCOCPABKNJV-UHFFFAOYSA-N	0.91(d), 1.95(m), 2.05(d)	**24.83**, **29.04**, **50.10**
Lactate	JVTAAEKCZFNVCJ-UHFFFAOYSA-M	1.33(d), 4.11(q)	**22.82**, **71.33**
Leucine	ROHFNLRQFUQHCH-UHFFFAOYSA-N	0.96(d), 0.97(d), 1.70(m), 1.71(m), 3.74(m)	**23.77**, **24.83**, **42.66**, **27.02**, **56.34**
Lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N	1.45(m), 1.73(m), 1.91(m), 3.03(t), 3.76(t)	24.30, **29.21**, **32.72**, **41.85**, **57.37**

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
Malate	BJEPYKJPYRNKOW-UHFFFAOYSA-L	2.37(dd), 2.67(dd), 4.30(dd)	**45.54**, **45.54**, **73.28**
Maltose	DKXNBNKWCZZMJT-SILGPGCPSA-N	3.28(dd), 3.60(m), 5.24(d), 5.41(d)	**77.05**, **77.49**, **94.79**, **102.42**
Mannose	GZCGUPFRVQAUEE-UHFFFAOYSA-N	3.93(m), 4.90(d), 5.19(d)	**73.36**, **96.62**, **96.89**
Methanol	OKKJLVBELUTLKV-UHFFFAOYSA-N	3.36(s)	**51.86**
Methionine	FFEARJCKVFRZRR-UHFFFAOYSA-N	2.13(m), 2.14(s), 2.21(m), 2.65(t), 3.86(dd)	**32.46**, **16.75**, 32.60, **31.67**, **56.63**
Methionine sulfoxide	QEFRNWWLZKMPFJ-YGVKFDHGSA-N	2.32(m), 2.76(s), 3.00(m), 3.88(m)	**26.76**, **39.31**, 50.92, 56.27
Methylamine	BAVYZALUXZFZLV-UHFFFAOYSA-N	2.60(s)	27.37
Methylsuccinate	MUXOBHXGJLMRAB-UHFFFAOYSA-N	1.09(d), 2.13(dd), 2.53(dd), 2.63(m)	**20.0**, **45.31**, **45.31**, **43.0**
myo-Inositol	CDAISMWEOUEBRE-UHFFFAOYSA-N	3.28(t), 3.54(dd), 3.63(t), 4.06(t)	77.27, 74.04, 75.31, **75.12**
N-Acetylalanine	KTHDTJVBEPMMGL-VKHMYHEASA-N	1.34(d), 2.02(s), 4.13(m)	**20.18**, **24.87**, **53.97**
N-Acetylglucosamine	MBLBDJOUHNCFQT-UHFFFAOYSA-N	2.06(d), 3.49(t), 3.87(m), 4.72(d), 5.21(d)	**24.91**, **72.93**, **56.86**, **97.85**, **93.73**
N-acetylglutamine	KSMRODHGGIIXDV-YFKPBYRVSA-N	1.93(m), 2.04 (s), 2.12 (m), 2.33 (m), 4.19 (m)	30.51, **24.92**, 30.51, **34.34**, 57.58
Nicotinate	PVNIIMVLHYAWGP-UHFFFAOYSA-M	7.52(ddd), 8.26(dt), 8.61(dd), 8.94(dd)	**127.08**, 140.65, 153.50, **151.92**
Ornithine	AHLPHDHHMVZTML-UHFFFAOYSA-N	1.76(m), 1.84(m), 1.94(m), 3.06(t), 3.77(t)	25.60, 25.60, **30.35**, **41.59**, **57.11**
Pantothenate	GHOKWGTUZJEAQD-UHFFFAOYSA-N	0.90(s), 0.94(s), 2.42(t), 3.44(m), 3.99(s)	**22.37**, **23.60**, **39.40**, 39.05, **78.81**
Phenylacetate	IPBVNPXQWQGGJP-UHFFFAOYSA-N	3.54(s), 7.31(m), 7.31(m), 7.38(m)	**47.38**, **129.45**, **132.08**, **131.65**
Phenylalanine	COLNVLDHVKWLRT-UHFFFAOYSA-N	3.13(dd), 3.29(dd), 4.00(dd), 7.33(m), 7.38(m),	**39.22**, **39.22**, **58.88**, **132.35**, **130.59**,
		7.43(m)	**132.08**
Pipecolic acid	HXEACLLIILLPRG-UHFFFAOYSA-N	1.62(m), 1.86(m), 2.22(m), 3.02(ddd), 3.41(m),	24.57, **24.42**, **29.51**, 46.40, **46.40**,
		3.58(dd)	**61.95**
Proline	ONIBWKKTOPOVIA-UHFFFAOYSA-N	2.00(m), 2.07(m), 3.34(m), 3.42(m), 4.13(dd)	**26.49**, **31.76**, **48.88**, **48.88**, **64.06**
Propionate	XBDQKXXYIPTUBI-UHFFFAOYSA-M	1.06(t), 2.18 (q)	**13.06**, **33.60**
Pyroglutamate	ODHCTXKNWHHXJC-UHFFFAOYSA-M	2.04(m), 2.40(ddd), 2.52(m), 4.18(dd)	28.17, **32.58**, 28.21, **61.30**
Ribose	PYMYPHUHKUWMLA-UHFFFAOYSA-N	3.70(dd), 3.84(m), 3.87(m), 4.00(m), 4.11(m),	65.81, **65.94**, **70.12**, **78.19**, **71.78**,
		4.11(m), 4.21(dd), 4.87(m), 4.93(d), 5.25(d)	**73.54, 73.28, 96.45, 96.69, 103.64**
Serine	MTCFGRXMJLQNBG-UHFFFAOYSA-N	3.85(dd), 3.96(dd), 4.00(dd)	**59.23**, **63.01**, **63.01**
Stachydrine	CMUNUTVVOOHQPW-UHFFFAOYSA-N	3.30(s)	**54.93**
Succinate	KDYFGRWQOYBRFD-UHFFFAOYSA-L	2.41(s)	**37.03**
Threonine	AYFVYJQAPQTCCC-UHFFFAOYSA-N	1.33(d), 3.59(d), 4.26(m)	**22.28**, **63.27**, **68.80**
Thymidine	IQFYYKKMVGJFEH-XLPZGREQSA-N	1.90(d), 4.03(m), 7.65(d)	**14.38**, **89.51**, **140.51**

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
Thymine	RWQNBRDOKXIBIV-UHFFFAOYSA-N	1.87(d), 7.37(d)	**14.12**, **142.0**
Tryptophan	QIVBCDIJIAJPQS-UHFFFAOYSA-N	3.31(dd), 3.49(dd), 4.06(dd), 7.20(t), 7.28(t),	**29.17**, **29.28**, **57.92**, **122.43**, **125.06**,
		7.32(s), 7.54(d), 7.73(d)	**127.96, 114.92, 121.37**
Tyrosine	OUYCCCASQSFEME-UHFFFAOYSA-N	3.06(dd), 3.20(dd), 3.94(dd), 6.90(d), 7.19(d)	**38.34**, **38.34**, **58.97**, **118.83**, **133.75**
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N	5.80(d), 7.54(d)	**103.91**, **146.48**
Uridine	DRTQHJPVMGBUCF-XVFCMESISA-N	4.23(t), 4.36(t), 5.91(d), 5.92(d), 7.88(d)	**72.31**, **76.47**, **104.96**, **92.09**, **144.70**
Urocanate	LOIYMIARKYCTBW-OWOJBTEDSA-N	6.39(d), 7.31(d), 7.40(s), 7.85(s)	**124.94**, **133.74**, **124.09**, **139.98**
Valerate	NQPDZGIKBAWPEJ-UHFFFAOYSA-M	0.89(t), 1.30(m), 1.53(m), 2.18(t)	16.05, **24.83**, **30.97**, **40.36**
Valine	KZSNJWFQEVHDMF-UHFFFAOYSA-N	0.99(d), 1.05(d), 2.28(m), 3.61(d)	**19.47**, **20.70**, **31.93**, **63.27**
Beta-Alanine	UCMIRNVEIXFBKS-UHFFFAOYSA-N	2.56(t), 3.18(t)	36.45, **39.40**

G.1.2. Table - Vegetarian Polar Metabolite Characterization List by NMR

Final NMR polar metabolite list for RM 8048 vegetarian material (Bligh-Dyer extraction). ^{1}H and ^{13}C chemical shifts of metabolites identified by 1D and 2D NMR. ^{13}C shifts follow the same order as ^{1}H shifts in the preceding column. Corresponding values in the two columns are chemical shifts of $^{1}H/^{13}C$ pairs from HSQC experiments. Values in bold in the $\delta^{13}C$ column correspond to peaks detected in the HSQC; those not in bold were low intensity peaks. Signals marked "-" were too weak for detection by HSQC experiments. Multiplicity key: s, singlet; d, doublet, dd, doublet of doublets; ddd, doublet of doublet of doublets; t, triplet; q, quartet; m, multiplet.

Metabolite ID	InChIKey	$\delta^{1}H$ (ppm) and multiplicity	$\delta^{13}C$ (ppm)
(3-carboxypropyl)trimethylammonium	JHPNVNIEXXLNTR-UHFFFAOYSA-O	2.27(t), 3.14(s)	**36.41**, **55.72**
1,3-Dimethyluric acid	OTSBKHHWSQYEHK-UHFFFAOYSA-N	3.34(s), 3.45(s)	**30.44**, **31.06**
1-Methyluric acid	QFDRTQONISXGJA-UHFFFAOYSA-N	3.30(s)	**30.44**
2-Aminobutyrate	QWCKQJZIFLGMSD-UHFFFAOYSA-N	0.99(t), 3.72(dd)	11.41, 58.70
2'-Deoxyuridine	MXHRCPNRJAMMIM-UHFFFAOYSA-N	2.40(m), 5.89(d)	**41.45**, **105.14**
2-Methylbutyrate	WLAMNBDJUVNPJU-BYPYZUCNSA-N	0.86(t), 1.05(d), 1.38(m), 1.49(m), 2.20(m)	**14.38**, ¯, **30.27**, 30.18, **47.56**
3-Hydroxyphenylacetate	FVMDYYGIDFPZAX-UHFFFAOYSA-N	3.48(s)	**47.21**
3-Phenylpropionate	XMIIGOLPHOKFCH-UHFFFAOYSA-N	2.49(t), 2.89(t), 7.27(t), 7.31(d), 7.38(t)	**42.12**, **34.74**, **129.10**, **131.29**, **131.73**
4-Hydroxyphenylacetate	XQXPVVBIMDBYFF-UHFFFAOYSA-M	3.45(s), 6.85(m), 7.18(m)	**46.33**, 118.39, 133.66
5-Aminopentanoate	JJMDCOVWQOJGCB-UHFFFAOYSA-M	1.63(m), 2.23(t)	**25.53**, **39.66**
6-Hydroxynicotinate	BLHCMGRVFXRYRN-UHFFFAOYSA-M	6.62(dd), 8.09(m)	**120.32**, **146.56**
Acetate	QTBSBXVTEAMEQO-UHFFFAOYSA-M	1.92(s)	**26.14**
Alanine	QNAYBMKLOCPYGJ-UHFFFAOYSA-N	1.48(d), 3.79(q)	**18.94**, **53.35**
Aminocaproic acid	SLXKOJJOQWFEFD-UHFFFAOYSA-N	1.60(m)	**27.90**
Arabinose	SRBFZHDQGSBBOR-HWQSCIPKSA-N	4.51(d)	**99.69**
Asparagine	DCXYFEDJOCDNAF-REOHCLBHSA-N	2.86(dd), 2.96(dd), 4.00(dd)	37.48, 37.48, **54.14**
Aspartate	CKLJMWTZIZZHCS-REOHCLBHSA-N	2.68(dd), 2.82(dd), 3.90(dd)	**39.44**, **39.44**, **55.11**
Butyrate	FERIUCNNQQJTOY-UHFFFAOYSA-M	0.90(t), 1.56(m), 2.16(t)	**16.05**, **22.19**, **42.47**

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
Cadaverine	VHRGRCVQAFMJIZ-UHFFFAOYSA-N	1.47(m), 1.73(m), 3.03(t)	**25.18**, 29.11, 42.0
Caprylate	WWZKQHOCKIZLMA-UHFFFAOYSA-N	0.87(t), 1.27(m), 1.29(m), 1.30(d), 1.55(m),	**16.15**, **33.95**, **31.41**, **31.41**, **28.69**,
		2.18(t)	**40.36**
Choline	OEYIOHPDSNJKLS-UHFFFAOYSA-N	3.21(s), 3.52(m)	**56.77**, **70.38**
Citrulline	RHGKLRLOHDJJDR-UHFFFAOYSA-N	1.86(m), 1.88(m), 3.15(m)	**30.44**, **30.53**, **42.03**
Deoxycholic acid	KXGVEGMKQFWNSR-LLQZFEROSA-N	0.71(s), 0.94(s), 0.96(s)	**14.91**, **25.11**, **19.47**
Ethanolamine	HZAXFHJVJLSVMW-UHFFFAOYSA-N	3.14(t), 3.82(t)	**44.05**, **60.46**
Formate	BDAGIHXWWSANSR-UHFFFAOYSA-M	8.46(s)	
Fructose	BJHIKXHVCXFQLS-UHFFFAOYSA-N	3.70(m), 3.83(m), 3.90(dd), 4.12(m)	**66.34**, **83.72**, **72.22**, **78.10**
Fucose	PNNNRSAQSRJVSB-UHFFFAOYSA-N	1.20(d), 1.25(d), 3.46(dd), 4.55(d), 5.21(d)	**18.14**, **18.31**, **74.68**, **99.17**, **94.98**
Fumarate	VZCYOOQTPOCHFL-OWOJBTEDSA-L	6.52(s)	**138.23**
Gabapentin	UGJMXCAKCUNAIE-UHFFFAOYSA-N	1.38(m), 1.38(m), 1.49(m), 1.49(m), 1.49(m),	**28.07**, **36.24**, **23.68**, **28.07**, **36.24**,
		2.43(s), 3.01(s)	**48.79, 51.16**
Galactose	GZCGUPFRVQAUEE-UHFFFAOYSA-N	3.50(dd), 3.68(dd), 3.71(m), 3.81(dd), 3.94(dd),	74.68, **75.56**, **78.01,** 71.26, **71.70,**
		4.09(m), 4.59(d), 5.27(d)	**73.37, 99.34, 95.19**
Galacturonic acid	IAJILQKETJEXLJ-RSJOWCBRSA-N	3.50(dd), 3.90(dd), 4.06(d), 4.28(dd), 5.30(d)	74.67, **72.22**, **78.45**, **73.51**, **94.43**
Gluconate	RGHNJXZEOKUKBD-UHFFFAOYSA-M	3.78(m), 4.04(t), 4.12(d)	**65.38**, **73.89**, **77.09**
Glucose	GZCGUPFRVQAUEE-UHFFFAOYSA-N	3.25(dd), 3.41(t), 3.47(m), 3.49(t), 3.55 (dd),	**77.05**, **72.49**, **78.80**, **78.63**, **74.33**,
		3.72(dd), 3.72(t), 3.77(dd), 3.84(m), 3.84(dd),	**75.64, 63.62, 63.44, 74.33, 63.44,**
		3.90(dd), 4.64(d), 5.24(d)	**63.62, 98.82, 94.95**
Glutamate	WHUUTDBJXJRKMK-UHFFFAOYSA-N	2.09(m), 2.36(m), 3.76(dd)	**29.83**, **36.32**, **57.39**
Glutamine	ZDXPYRJPNDTMRX-UHFFFAOYSA-N	2.15(m), 2.46(m), 3.78(t)	29.11, **33.69**, 56.96
Glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N	3.56(dd), 3.65(dd), 3.79(m)	**65.38**, **65.38**, **74.94**
Glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N	3.56(s)	**44.22**
Glycolate	AEMRFAOFKBGASW-UHFFFAOYSA-M	3.94(s)	**64.15**
Hexanoic acid	FUZZWVXGSFPDMH-UHFFFAOYSA-N	1.27(t)	**33.95**
Hypoxanthine	FDGQSTZJBFJUBT-UHFFFAOYSA-N	8.19(s), 8.21(s)	**148.14**, **144.46**
Isobutyrate	KQNPFQTWMSNSAP-UHFFFAOYSA-M	1.06(d), 2.39(m)	**22.28**, **39.83**
Isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N	0.94(t), 1.01(d), 1.27(m), 1.46(m), 1.98(m),	**13.94**, **17.45**, **27.37**, **27.46**, **38.78**,

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
		3.68(d)	**62.39**
Isovalerate	GWYFCOCPABKNJV-UHFFFAOYSA-N	0.91(d), 1.95(m), 2.05(d)	**24.83**, **29.04**, **50.10**
Lactate	JVTAAEKCZFNVCJ-UHFFFAOYSA-M	1.33(d), 4.11(q)	**22.82**, **71.33**
Leucine	ROHFNLRQFUQHCH-UHFFFAOYSA-N	0.96(d), 0.97(d), 1.71(m), 1.71(m), 3.74(m)	**23.77**, **24.83**, **42.68**, **26.93**, **56.34**
Lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N	1.45(m), 1.73(m), 1.91(m), 3.03(t), 3.76(t)	24.30, **29.21**, **32.72**, **41.85**, **57.39**
Malate	BJEPYKJPYRNKOW-UHFFFAOYSA-L	2.37(dd), 2.67(dd), 4.30(dd)	**45.54**, **45.45**, **73.28**
Maltose	DKXNBNKWCZZMJT-SILGPGCPSA-N	3.28(dd), 3.61(m), 5.24(d), 5.42(d)	**76.85**, **77.40**, **94.79**, **102.23**
Mannobiose	GUBGYTABKSRVRQ-PZPXDAEZSA-N	3.57(m), 3.82(m), 3.91(m), 3.99(m), 4.00(m),	**69.59**, **79.24**, **79.59**, **71.96**, **73.19**,
		4.08(m), 4.75(m), 4.93(d), 5.19(d)	**73.54, 103.12, 96.62, 96.89**
Methanol	OKKJLVBELUTLKV-UHFFFAOYSA-N	3.36(s)	**51.77**
Methionine	FFEARJCKVFRZRR-UHFFFAOYSA-N	2.13(m), 2.14(s), 2.21(m), 2.65(t), 3.86(dd)	**32.46**, **16.75**, 32.60, **31.67**, **56.63**
Methionine sulfoxide	QEFRNWWLZKMPFJ-YGVKFDHGSA-N	2.32(m), 2.76(s), 3.00(m), 3.88(m)	**26.76**, **39.31**, 50.92, 56.27
Methylamine	BAVYZALUXZFZLV-UHFFFAOYSA-N	2.60(s)	27.37
Methylsuccinate	MUXOBHXGJLMRAB-UHFFFAOYSA-N	1.09(d), 2.13(dd), 2.52(dd), 2.63(m)	20.0, 45.20, **45.20**, 43.0
myo-Inositol	CDAISMWEOUEBRE-UHFFFAOYSA-N	3.28(t), 3.54(dd), 3.63(t), 4.06(t)	77.27, 74.04, 75.31, **75.12**
N-Acetylalanine	KTHDTJVBEPMMGL-VKHMYHEASA-N	1.34(d), 2.02(s), 4.14(m)	**20.26**, **24.94**, **53.97**
N-Acetylglucosamine	MBLBDJOUHNCFQT-UHFFFAOYSA-N	2.05(d), 3.47(t), 3.87(m), 4.72(d), 5.21(d)	**24.91**, **72.49**, **56.60**, **97.91**, **93.73**
N-acetylglutamine	KSMRODHGGIIXDV-YFKPBYRVSA-N	1.93(m), 2.04 (s), 2.12 (m), 2.33 (m), 4.19 (m)	30.51, **24.91**, 30.51, **34.41**, 57.58
Nicotinate	PVNIIMVLHYAWGP-UHFFFAOYSA-M	7.52(ddd), 8.26(dt), 8.61(dd), 8.94(dd)	**126.99**, 140.65, 153.50, **151.92**
Ornithine	AHLPHDHHMVZTML-UHFFFAOYSA-N	1.76(m), 1.84(m), 1.94(m), 3.06(t), 3.77(t)	25.60, 25.60**, 30.36**, **41.59,** 57.11
Pantothenate	GHOKWGTUZJEAQD-UHFFFAOYSA-N	0.90(s), 0.94(s), 2.42(t), 3.44(m), 3.99(s)	**22.37**, **23.60**, **39.40**, 39.05, **78.80**
Phenylacetate	IPBVNPXQWQGGJP-UHFFFAOYSA-N	3.54(s), 7.30(m), 7.31(m), 7.38(m)	**47.38**, **129.45**, **132.17**, **131.73**
Phenylalanine	COLNVLDHVKWLRT-UHFFFAOYSA-N	3.13(dd), 3.29(dd), 4.00(dd), 7.33(m), 7.38(m),	**39.22**, **39.22**, **58.88**, **132.35**, **130.68**,
		7.43(m)	**132.08**
Pipecolic acid	HXEACLLIILLPRG-UHFFFAOYSA-N	1.62(m), 1.89(m), 2.21(m), 3.02(ddd), 3.41(m),	24.57, **24.59**, **29.83**, 46.43, **46.43**,
		3.58(dd)	**61.93**
Proline	ONIBWKKTOPOVIA-UHFFFAOYSA-N	2.00(m), 2.08(m), 3.34(m), 3.42(m), 4.14(dd)	**26.49**, **31.76**, **48.96**, **48.87**, **64.06**
Propionate	XBDQKXXYIPTUBI-UHFFFAOYSA-M	1.06(t), 2.18 (q)	**13.06**, **33.60**
Pyroglutamate	ODHCTXKNWHHXJC-UHFFFAOYSA-M	2.04(m), 2.40(ddd), 2.52(m), 4.17(dd)	28.17, **32.55**, 28.21, **61.34**

Metabolite ID	InChIKey	δ^{1}H (ppm) and multiplicity	δ^{13}C (ppm)
Ribose	PYMYPHUHKUWMLA-UHFFFAOYSA-N	3.70(dd), 3.84(m), 3.89(m), 4.00(m), 4.11(m),	65.81**, 65.94**, **70.03**, **77.84**, **71.78**,
		4.11(m), 4.21(dd), 4.87(m), 4.93(d), 5.25(d)	**73.45, 73.28, 96.36, 96.57, 103.68**
Serine	MTCFGRXMJLQNBG-UHFFFAOYSA-N	3.85(dd), 3.97(dd), 4.00(dd)	**59.23**, **63.01**, **63.01**
Succinate	KDYFGRWQOYBRFD-UHFFFAOYSA-L	2.41(s)	**37.03**
Threonine	AYFVYJQAPQTCCC-UHFFFAOYSA-N	1.33(d), 3.59(d), 4.26(m)	**22.28**, **63.27**, **68.80**
Thymidine	IQFYYKKMVGJFEH-XLPZGREQSA-N	1.90(d), 4.03(m), 7.65(d)	**14.38**, **89.33**, **140.42**
Thymine	RWQNBRDOKXIBIV-UHFFFAOYSA-N	1.87(d), 7.37(d)	**14.12**, **142.0**
Tryptophan	QIVBCDIJIAJPQS-UHFFFAOYSA-N	3.31(dd), 3.49(dd), 4.06(dd), 7.20(t), 7.28(t),	**29.25**, **29.25**, **57.92**, **122.43**, **125.06**,
		7.32(s), 7.54(d), 7.73(d)	**127.96, 114.88, 121.37**
Tyrosine	OUYCCCASQSFEME-UHFFFAOYSA-N	3.06(dd), 3.20(dd), 3.94(dd), 6.90(d), 7.19(d)	**38.34**, **38.34**, **58.97**, **118.83**, **133.75**
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N	5.80(d), 7.54(d)	**103.91**, **146.48**
Uridine	DRTQHJPVMGBUCF-XVFCMESISA-N	4.23(t), 4.36(t), 5.91(d), 5.92(d), 7.88(d)	**72.29**, **76.47**, **105.12**, **92.09**, **144.70**
Urocanate	LOIYMIARKYCTBW-OWOJBTEDSA-N	6.39(d), 7.31(d), 7.40(s), 7.85(s)	**124.89**, **133.84**, **124.09**, **140.24**
Valerate	NQPDZGIKBAWPEJ-UHFFFAOYSA-M	0.89(t), 1.30(m), 1.53(m), 2.18(t)	16.05, **24.83**, **30.97**, **40.36**
Valine	KZSNJWFQEVHDMF-UHFFFAOYSA-N	0.99(d), 1.05(d), 2.28(m), 3.61(d)	**19.47**, **20.70**, **31.93**, **63.27**
Beta-Alanine	UCMIRNVEIXFBKS-UHFFFAOYSA-N	2.56(t), 3.18(t)	36.45, **39.48**

G.2. Aqueous Extraction (Fecal Water) - Polar Metabolites - 600 MHz

G.2.1. Mass of the reagents and extract used in the analysis of omnivore samples for each target volume.

		Buffered extract[1]			NMR sample preparation	
Omnivore		Aqueous Extract	Phos. buffer	NaN_3	Buffered extract	D_2O/DSS
Date	Diet/box-vial	500 µL (g)	50 µL (g)	50 µL (g)	500 µL (g)	50 µL (g)
8/15/2023	o25-1	0.5012	0.0581	0.0517	0.4999	0.0558
8/15/2023	o6-1	0.4942	0.0580	0.0518	0.4953	0.0557
8/15/2023	o36-1	0.4969	0.0565	0.0513		0.0555
8/15/2023	o15-1		0.0569	0.0511	0.4983	0.0554
11/1/2023	o25-2		0.0577	0.0524	0.5007	0.0563
11/1/2023	o6-2	0.4938	0.0572	0.054		0.0560
11/2/2023	o36-2	0.5068		0.0521		0.0567
11/2/2023	o15-2	0.5043	0.0564	0.0514	0.5112	0.0568
11/2/2023	c2[2]		0.0564	0.0512	0.5107	0.0579
11/28/2023	o25-3	0.4935	0.056	0.0504	0.5063	0.0535
11/28/2023	o6-3	0.4932	0.0547	0.0501	0.4992	0.0547
11/28/2023	o36-3	0.4929	0.0554	0.0506		0.0546
11/28/2023	o15-3	0.5020	0.0547	0.0506	0.5059	0.0549
11/28/2023	c-5	0.5017	0.0551	0.0500		0.0543
1/23/2024	c-8	0.5057	0.0579		0.5101	0.0557
1/23/2024	o6-4	0.5036	0.0568	0.0468		0.0564
1/23/2024	o36-4	0.5050	0.0576	0.0523	0.5110	0.0566
1/23/2024	o15-4	0.4972	0.0561	0.0507		0.056
1/23/2024	o25-4	0.5048	0.0564	0.0515		0.0564
5/14/2024	o25-5	0.4951	0.0569		0.5084	0.0549
5/14/2024	o1-4	0.5021	0.0567	0.0510	0.5104	0.0565
5/14/2024	o44-3	0.5044	0.0602	0.0522	0.5106	0.0566
5/14/2024	o1-3	0.5004	0.0564	0.0525	0.5081	0.0563
5/14/2024	o44-4	0.5015	0.06	0.0531	0.5073	0.057
5/14/2024	o15-5	0.5025	0.0589	0.0526	0.5046	
Mean		0.5001	0.0570	0.0514	0.5058	0.0559
SD		0.0046	0.0014	0.0014	0.0052	0.0010
RSD %		0.93	2.51	2.73	1.02	1.78
N		22	24	23	18	24

[1]In this step, the extract is mixed with the phosphate buffer and sodium azide. This step is prior to the NMR sample preparation.
[2]Control samples were labeled to c1, c2, etc.

G.2.2. Mass of the reagents and extract used in the analysis of vegetarian samples for each target volume.

Vegetarian		Buffered extract[1]			NMR sample preparation	
		Aqueous Extract	Phos. buffer	NaN_3	Buffered extract	D_2O/DSS
Date	Diet/box-vial	500 µL (g)	50 µL (g)	50 µL (g)	500 µL (g)	50 µL (g)
8/16/2023	v36-1	0.4984	0.0567	0.0521	0.4975	0.0549
8/16/2023	v6-2	0.4952	0.0579	0.0519	0.4965	0.0553
8/16/2023	v25-2	0.4988	0.0565	0.0513	0.5005	0.0553
8/16/2023	v15-2	0.4984	0.0561	0.051	0.5004	0.0558
11/2/2023	v6-2	0.4938	0.0558	0.052	0.4956	
11/2/2023	v36-2	0.4906	0.0557	0.0519	0.4948	0.0558
11/2/2023	v25-2	0.4905	0.0566	0.0519	0.4951	0.0556
11/2/2023	v15-2	0.4927	0.0542	0.052	0.4958	0.0557
11/2/2023	c-4[2]	0.5021	0.0572	0.0516	0.5101	0.0556
1/22/2024	v6-4	0.4952	0.0562	0.0521	0.4971	0.0546
1/22/2024	v36-4	0.4939	0.0557	0.0511	0.4962	0.0555
1/22/2024	v25-4	0.4921	0.0581	0.0532	0.4971	0.0554
1/22/2024	v15-4	0.492	0.0576	0.0514		0.0551
1/22/2024	c-7	0.502	0.0576	0.0519		
5/24/2024	v25-5	0.486	0.0577	0.0529	0.4923	0.0543
5/24/2024	v1-4	0.4909	0.0587	0.0521	0.4965	0.0567
5/24/2024	v44-3	0.491	0.0564	0.0522	0.4966	0.0573
5/24/2024	v1-3	0.4941	0.0582	0.054	0.4951	0.0568
5/24/2024	v44-4	0.4931	0.0574	0.052	0.4933	0.0569
5/24/2024	v15-5	0.4898	0.0588	0.0536	0.4927	0.057
Mean		0.4940	0.0570	0.0521	0.4968	0.0558
SD		0.0042	0.0012	0.0008	0.0041	0.0009
RSD %		0.84	2.05	1.49	0.82	1.53
N		23	23	23	20	20

[1]In this step, the extract is mixed with the phosphate buffer and sodium azide. This step is prior to the NMR sample preparation.
[2]Control samples were labeled to c1, c2, etc.

G.2.3. Coefficient of disagreement (COD) values among omnivore vials.

	O25.1	O6.1	O36.1	O15.1	O25.2	O6.2	O36.2	O15.2	O25.3	O6.3	O36.3	O15.3	O6.4	O36.4	O15.4	O25.4	O25.5	O1.4	O44.3	O1.3	O44.4	O15.5
O25.1	0	0.08	0.01	0.02	0.01	0.09	0.02	0.01	0.01	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.04	0.02
O6.1	0.08	0	0.08	0.08	0.08	0.01	0.07	0.08	0.07	0.01	0.08	0.05	0.02	0.07	0.08	0.06	0.05	0.02	0.1	0.02	0.11	0.07
O36.1	0.01	0.08	0	0.02	0.01	0.09	0.02	0.01	0.01	0.07	0.01	0.04	0.07	0.01	0.02	0.01	0.03	0.07	0.02	0.09	0.04	0.02
O15.1	0.02	0.08	0.02	0	0.02	0.09	0.01	0.02	0.03	0.07	0.02	0.04	0.07	0.02	0.01	0.03	0.04	0.08	0.02	0.09	0.04	0.02
O25.2	0.01	0.08	0.01	0.02	0	0.09	0.01	0.01	0.01	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.03	0.02
O6.2	0.09	0.01	0.09	0.09	0.09	0	0.08	0.09	0.08	0.02	0.09	0.06	0.02	0.08	0.09	0.07	0.06	0.02	0.11	0.01	0.12	0.08
O36.2	0.02	0.07	0.02	0.01	0.01	0.08	0	0.02	0.02	0.06	0.02	0.03	0.06	0.02	0.01	0.02	0.03	0.07	0.03	0.08	0.04	0.01
O15.2	0.01	0.08	0.01	0.02	0.01	0.09	0.02	0	0.02	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.04	0.02
O25.3	0.01	0.07	0.01	0.03	0.01	0.08	0.02	0.02	0	0.06	0.02	0.03	0.06	0.01	0.03	0.01	0.02	0.07	0.03	0.08	0.04	0.02
O6.3	0.07	0.01	0.07	0.07	0.07	0.02	0.06	0.07	0.06	0	0.07	0.03	0.01	0.06	0.07	0.05	0.04	0.02	0.08	0.03	0.1	0.06
O36.3	0.01	0.08	0.01	0.02	0.01	0.09	0.02	0.01	0.02	0.07	0	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.1	0.03	0.02
O15.3	0.04	0.05	0.04	0.04	0.04	0.06	0.03	0.04	0.03	0.03	0.04	0	0.03	0.03	0.04	0.02	0.02	0.04	0.05	0.06	0.07	0.03
O6.4	0.07	0.02	0.07	0.07	0.07	0.02	0.06	0.07	0.06	0.01	0.07	0.03	0	0.06	0.07	0.05	0.04	0.02	0.08	0.03	0.1	0.06
O36.4	0.01	0.07	0.01	0.02	0.01	0.08	0.02	0.01	0.01	0.06	0.01	0.03	0.06	0	0.03	0.01	0.02	0.07	0.03	0.09	0.04	0.02
O15.4	0.02	0.08	0.02	0.01	0.02	0.09	0.01	0.02	0.03	0.07	0.02	0.04	0.07	0.03	0	0.03	0.04	0.08	0.02	0.09	0.04	0.02
O25.4	0.02	0.06	0.01	0.03	0.02	0.07	0.02	0.02	0.01	0.05	0.02	0.02	0.05	0.01	0.03	0	0.02	0.06	0.04	0.08	0.05	0.02
O25.5	0.03	0.05	0.03	0.04	0.03	0.06	0.03	0.03	0.02	0.04	0.03	0.02	0.04	0.02	0.04	0.02	0	0.05	0.04	0.06	0.06	0.03
O1.4	0.08	0.02	0.07	0.08	0.08	0.02	0.07	0.08	0.07	0.02	0.08	0.04	0.02	0.07	0.08	0.06	0.05	0	0.09	0.02	0.11	0.07
O44.3	0.02	0.1	0.02	0.02	0.02	0.11	0.03	0.02	0.03	0.08	0.02	0.05	0.08	0.03	0.02	0.04	0.04	0.09	0	0.11	0.02	0.03
O1.3	0.09	0.02	0.09	0.09	0.09	0.01	0.08	0.09	0.08	0.03	0.1	0.06	0.03	0.09	0.09	0.08	0.06	0.02	0.11	0	0.13	0.08
O44.4	0.04	0.11	0.04	0.04	0.03	0.12	0.04	0.04	0.04	0.1	0.03	0.07	0.1	0.04	0.04	0.05	0.06	0.11	0.02	0.13	0	0.04
O15.5	0.02	0.07	0.02	0.02	0.02	0.08	0.01	0.02	0.02	0.06	0.02	0.03	0.06	0.02	0.02	0.02	0.03	0.07	0.03	0.08	0.04	0

G.2.4. Coefficient of Disagreement (COD) among vegetarian vials.

0	V36.1	V6.1	V25.1	V15.1	V6.2	V36.2	V25.2	V15.2	V6.3	V36.3	V25.3	V15.3	V6.4	V36.4	V25.4	V15.4	V25.5	V1.4	V43.3	V1.3	V44.4	V15.5
V36.1	0	0.04	0.02	0.04	0.04	0.03	0.04	0.06	0.05	0.02	0.03	0.07	0.07	0.02	0.05	0.06	0.04	0.1	0.02	0.14	0.02	0.07
V6.1	0.04	0	0.03	0.04	0.02	0.07	0.02	0.04	0.02	0.05	0.02	0.04	0.04	0.03	0.04	0.04	0.03	0.06	0.05	0.1	0.04	0.03
V25.1	0.02	0.03	0	0.03	0.03	0.04	0.02	0.05	0.04	0.03	0.02	0.06	0.05	0.02	0.05	0.05	0.03	0.09	0.02	0.12	0.02	0.05
V15.1	0.04	0.04	0.03	0	0.02	0.06	0.02	0.03	0.03	0.03	0.01	0.04	0.04	0.02	0.03	0.03	0.02	0.07	0.04	0.11	0.03	0.03
V6.2	0.04	0.02	0.03	0.02	0	0.07	0.02	0.03	0.02	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.07	0.04	0.1	0.04	0.03
V36.2	0.03	0.07	0.04	0.06	0.07	0	0.06	0.09	0.08	0.03	0.05	0.1	0.09	0.04	0.07	0.08	0.07	0.13	0.03	0.16	0.04	0.09
V25.2	0.04	0.02	0.02	0.02	0.02	0.06	0	0.03	0.02	0.04	0.01	0.04	0.03	0.02	0.03	0.03	0.01	0.07	0.04	0.1	0.03	0.03
V15.2	0.06	0.04	0.05	0.03	0.03	0.09	0.03	0	0.02	0.06	0.04	0.01	0.01	0.05	0.03	0.01	0.03	0.04	0.07	0.07	0.06	0.03
V6.3	0.05	0.02	0.04	0.03	0.02	0.08	0.02	0.02	0	0.05	0.03	0.03	0.02	0.04	0.03	0.03	0.02	0.05	0.06	0.08	0.05	0.03
V36.3	0.02	0.05	0.03	0.03	0.04	0.03	0.04	0.06	0.05	0	0.03	0.07	0.06	0.02	0.04	0.06	0.04	0.11	0.02	0.14	0.03	0.06
V25.3	0.03	0.02	0.02	0.01	0.02	0.05	0.01	0.04	0.03	0.03	0	0.05	0.04	0.01	0.02	0.04	0.02	0.08	0.03	0.11	0.02	0.04
V15.3	0.07	0.04	0.06	0.04	0.03	0.1	0.04	0.01	0.03	0.07	0.05	0	0.02	0.06	0.04	0.02	0.03	0.03	0.08	0.06	0.07	0.03
V6.4	0.07	0.04	0.05	0.04	0.03	0.09	0.03	0.01	0.02	0.06	0.04	0.02	0	0.05	0.03	0.02	0.03	0.04	0.07	0.07	0.06	0.04
V36.4	0.02	0.03	0.02	0.02	0.03	0.04	0.02	0.05	0.04	0.02	0.01	0.06	0.05	0	0.03	0.05	0.03	0.09	0.02	0.12	0.02	0.05
V25.4	0.05	0.04	0.05	0.03	0.02	0.07	0.03	0.03	0.03	0.04	0.02	0.04	0.03	0.03	0	0.03	0.02	0.07	0.04	0.1	0.04	0.03
V15.4	0.06	0.04	0.05	0.03	0.02	0.08	0.03	0.01	0.03	0.06	0.04	0.02	0.02	0.05	0.03	0	0.02	0.05	0.07	0.08	0.06	0.03
V25.5	0.04	0.03	0.03	0.02	0.02	0.07	0.01	0.03	0.02	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0	0.06	0.05	0.1	0.04	0.02
V1.4	0.1	0.06	0.09	0.07	0.07	0.13	0.07	0.04	0.05	0.11	0.08	0.03	0.04	0.09	0.07	0.05	0.06	0	0.11	0.03	0.1	0.04
V43.3	0.02	0.05	0.02	0.04	0.04	0.03	0.04	0.07	0.06	0.02	0.03	0.08	0.07	0.02	0.04	0.07	0.05	0.11	0	0.14	0.01	0.07
V1.3	0.14	0.1	0.12	0.11	0.1	0.16	0.1	0.07	0.08	0.14	0.11	0.06	0.07	0.12	0.1	0.08	0.1	0.03	0.14	0	0.13	0.08
V44.4	0.02	0.04	0.02	0.03	0.04	0.04	0.03	0.06	0.05	0.03	0.02	0.07	0.06	0.02	0.04	0.06	0.04	0.1	0.01	0.13	0	0.06
V15.5	0.07	0.03	0.05	0.03	0.03	0.09	0.03	0.03	0.03	0.06	0.04	0.03	0.04	0.05	0.03	0.03	0.02	0.04	0.07	0.08	0.06	0

G.2.5. Table - Vegetarian polar metabolites characterized using an aqueous extraction protocol. The samples were analyzed by ^{1}H NMR and ^{1}H, ^{13}C HSQC, and the metabolites were annotated using Chenomx and COLMAR respectively.

Metabolite ID	InChIKey
Acetate	QTBSBXVTEAMEQO-UHFFFAOYSA-N
Hypoxanthine	FDGQSTZJBFJUBT-UHFFFAOYSA-N
Fumarate	VZCYOOQTPOCHFL-OWOJBTEDSA-N
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N
Formate	BDAGIHXWWSANSR-UHFFFAOYSA-N
Methanol	OKKJLVBELUTLKV-UHFFFAOYSA-N
Glucose	WQZGKKKJIJFFOK-VFUOTHLCSA-N
Phenylalanine	COLNVLDHVKWLRT-QMMMGPOBSA-N
Tyrosine	OUYCCCASQSFEME-QMMMGPOBSA-N
Gabapentin	UGJMXCAKCUNAIE-UHFFFAOYSA-N
Xylose	SRBFZHDQGSBBOR-IOVATXLUSA-N
Nicotinate	PVNIIMVLHYAWGP-UHFFFAOYSA-N
Aspartate	CKLJMWTZIZZHCS-REOHCLBHSA-N
Succinate	KDYFGRWQOYBRFD-UHFFFAOYSA-N
Isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N
Arabinose	SRBFZHDQGSBBOR-HWQSCIPKSA-N
L-Malic acid	BJEPYKJPYRNKOW-REOHCLBHSA-N
Valine	KZSNJWFQEVHDMF-BYPYZUCNSA-N
Threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N
Tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N
Glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N
Ribose	IHMFHBZSHGGEWLO-SOOFDHNKSA-N
Sarcosine	FSYKKLYZXJSNPZ-UHFFFAOYSA-N
Propionate	XBDQKXXYIPTUBI-UHFFFAOYSA-N
Methionine	FFEARJCKVFRZRR-BYPYZUCNSA-N
p-Cresol	IWDCLRJOBJJRNH-UHFFFAOYSA-N

Metabolite ID	InChIKey
Urocanate	LOIYMIARKYCTBW-OWOJBTEDSA-N
Galactose	WQZGKKKJIJFFOK-SVZMEOIVSA-N
Phenylacetate	WLJVXDMOQOGPHL-UHFFFAOYSA-N
Butyrate	FERIUCNNQQJTOY-UHFFFAOYSA-N
Xanthine	LRFVTYWOQMYALW-UHFFFAOYSA-N
Uridine	DRTQHJPVMGBUCF-XVFCMESISA-N
Leucine	ROHFNLRQFUQHCH-YFKPBYRVSA-N
Lactate	JVTAAEKCZFNVCJ-REOHCLBHSA-N
Choline	OEYIOHPDSNJKLS-UHFFFAOYSA-N
Citrulline	RHGKLRLOHDJJDR-BYPYZUCNSA-N
Glutamate	WHUUTDBJXJRKMK-VKHMYHEASA-N
Methylsuccinate	WXUAQHNMJWJLTG-GSVOUGTGSA-N
Cadaverine	VHRGRCVQAFMJIZ-UHFFFAOYSA-N
3-Phenylpropionate	XMIIGOLPHOKFCH-UHFFFAOYSA-N
Proline	ONIBWKKTOPOVIA-BYPYZUCNSA-N
Serine	MTCFGRXMJLQNBG-REOHCLBHSA-N
Alanine	QNAYBMKLOCPYGJ-REOHCLBHSA-N
Lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N
Glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N
Caprylate	WWZKQHOCKIZLMA-UHFFFAOYSA-N
Glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N
Isovalerate	GWYFCOCPABKNJV-UHFFFAOYSA-N
Valerate	NQPDZGIKBAWPEJ-UHFFFAOYSA-N

G.2.6. Table - Omnivore polar metabolite characterized using an aqueous extraction protocol. The samples were analyzed by ^{1}H NMR and ^{1}H, ^{13}C HSQC, and the metabolites were annotated using Chenomx and COLMAR respectively.

Metabolite ID	InChI
Acetate	QTBSBXVTEAMEQO-UHFFFAOYSA-N
Hypoxanthine	FDGQSTZJBFJUBT-UHFFFAOYSA-N
1,3-Dihydroxyacetone	RXKJFZQQPQGTFL-UHFFFAOYSA-N
Uracil	ISAKRJDGNUQOIC-UHFFFAOYSA-N
Glucose	WQZGKKKJIJFFOK-VFUOTHLCSA-N
Phenylalanine	COLNVLDHVKWLRT-QMMMGPOBSA-N
Methanol	OKKJLVBELUTLKV-UHFFFAOYSA-N
Valine	KZSNJWFQEVHDMF-BYPYZUCNSA-N
Succinate	KDYFGRWQOYBRFD-UHFFFAOYSA-N
Xylose	SRBFZHDQGSBBOR-IOVATXLUSA-N
Isoleucine	AGPKZVBTJJNPAG-WHFBIAKZSA-N
Tyrosine	OUYCCCASQSFEME-QMMMGPOBSA-N
Choline	OEYIOHPDSNJKLS-UHFFFAOYSA-N
Ribose	HMFHBZSHGGEWLO-SOOFDHNKSA-N
Alanine	QNAYBMKLOCPYGJ-REOHCLBHSA-N
Threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N
p-Cresol	IWDCLRJOBJJRNH-UHFFFAOYSA-N
Methionine	FFEARJCKVFRZRR-BYPYZUCNSA-N
Nicotinate	PVNIIMVLHYAWGP-UHFFFAOYSA-N
3-(1H-imidazol-5-yl)prop-2-enoate	LOIYMIARKYCTBW-OWOJBTEDSA-N
Phenylacetate	WLJVXDMOQOGPHL-UHFFFAOYSA-N
Galactose	WQZGKKKJIJFFOK-SVZMEOIVSA-N
L-Malic acid	BJEPYKJPYRNKOW-REOHCLBHSA-N
Propionate	XBDQKXXYIPTUBI-UHFFFAOYSA-N
Aspartate	CKLJMWTZIZZHCS-REOHCLBHSA-N
Butyrate	FERIUCNNQQJTOY-UHFFFAOYSA-N
Tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N

Metabolite ID	InChI
Glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N
Leucine	ROHFNLRQFUQHCH-YFKPBYRVSA-N
Lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N
Xanthine	LRFVTYWOQMYALW-UHFFFAOYSA-N
Desaminotyrosine	NMHMNPHRMNGLLB-UHFFFAOYSA-N
Arabinose	SRBFZHDQGSBBOR-HWQSCIPKSA-N
Caprylate	WWZKQHOCKIZLMA-UHFFFAOYSA-N
Glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N
Proline	ONIBWKKTOPOVIA-BYPYZUCNSA-N
3-Phenylpropionate	XMIIGOLPHOKFCH-UHFFFAOYSA-N
Glutamate	ZDXPYRJPNDTMRX-VKHMYHEASA-N
Methylsuccinate	WXUAQHNMJWJLTG-GSVOUGTGSA-N
Citrulline	RHGKLRLOHDJJDR-BYPYZUCNSA-N
Serine	MTCFGRXMJLQNBG-REOHCLBHSA-N
Glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N
Valerate	NQPDZGIKBAWPEJ-UHFFFAOYSA-N
Isovalerate	GWYFCOCPABKNJV-UHFFFAOYSA-N

Appendix H. Detailed Methods and Results for Flow Cytometry

H.1. Detailed Methods for Flow Cytometry

Fluorescence Standardization

A MoFlo Astrios EQ cell sorter (Beckman Coulter) was used. In all cases, MFI corresponds to Median Fluorescent Intensity. The MoFlo was started, and QC was performed according to the manufacturer's protocol [Sphero UltraRainbow Fluorescent Particles, 3.1 µm, Cat# URFP01-30-2k, Lot# AN04]. Rainbow Calibration Particles [8peakRCP - SpheroTech, RCP-30-5A, Lot AN04] were used to standardize the fluorescent intensities across experiments by setting the detector voltages. Intensity values from the eight peaks were recorded on the first day of experimentation. For all subsequent experiments the voltages were adjusted until the median intensity of the highest intensity peak was aligned with the value from the first day.

File Acquisition

All *.FCS raw data files are available at (https://doi.org/10.18434/mds2-3633). Samples were run on the instrument in semi-random order (the first set of technical replicates from all samples were acquired in random order before moving on to the second set of technical replicates). The file naming convention was "Date_DonorPool_Box#_Sample#" where Date was expressed as YYYYMMDD. Data from each sample were acquired for a total of two min or ≈ 490,000 events depending on which occurred first (500,000 is the maximum number of events recorded by the instrument in a single file).

Data Analysis Overview:

All *.R and *.RMD data analysis scripts are available at (https://doi.org/10.18434/mds2-3633). All files were imported into RStudio using the 'flowCore' package[23]. Each set of RM 8048 files in an experiment was processed as a batch.

The overall workflow was:

(1) For each date, use the RCP file to produce a fluorescent intensity calibration

(2) For each date, apply the calibration to the raw data and perform quality control (QC) using the 'PeacoQC' package[24]

(3) Concatenate the homogeneity data and generate a self-organizing map (SOM) using the 'FlowSOM' package[25]

(4) Map each calibrated file onto the SOM, enumerate cells and CBP beads, calculate normalized cell counts

[23] Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M, Finak G (2025). flowCore: flowCore: Basic structures for flow cytometry data. doi:10.18129/B9.bioc.flowCore, R package version 2.20.0, https://bioconductor.org/packages/flowCore.

[24] Emmaneel A (2025). *PeacoQC: Peak-based selection of high quality cytometry data*. doi:10.18129/B9.bioc.PeacoQC, R package version 1.18.0, https://bioconductor.org/packages/PeacoQC

[25] Van Gassen S, Callebaut B, Van Helden M, Lambrecht B, Demeester P, Dhaene T, Saeys Y (2015). "FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data." *Cytometry Part A*, **87**(7), 636-645. https://onlinelibrary.wiley.com/doi/full/10.1002/cyto.a.22625.

(5) Export calibrated intensity distributions and normalized count values for statistical analysis.

Fluorescence Calibration:

Gates were generated using the 'gate_flowclust_2d' function from 'flowClust' package[26]. The primary bead population was gated in FSC1 and SSC1, six separate clusters corresponding to the intensity subpopulations were identified in FL25 and FL20, and the MFI in the three fluorescent channels of interest was calculated.

The manufacturer provides lot-specific values for the equivalent intensity of the beads using standardized units (which are named according to common fluorophores for that channel) at https://www.spherotech.com/templates.html. These intensities span orders of magnitude, so a weighted linear regression was calculated for each primary fluorescent channel. A linear fit was selected because the relationship between fluorophore concentration and fluorescent emission is expected to be linear. The slope and intercept of the fitted lines were exported for use in the next script.

Intensity Quality Control:

Samples were preprocessed before QC: the calibration from the previous script was applied to all files, events on detection margins were removed, and a logicle transform estimated from the RCP bead file was applied. Quality control was performed using the 'PeacoQC' package to remove bins of data with shifted fluorescent intensities.

SOM (self-organizing map):

Cell and bead classifications were determined using the 'FlowSOM' package. Calibrated and QC'ed files for the two homogeneity experiments (2023.08.02, 2023.08.04) were concatenated to build the SOM. Note that no compensation was applied due to minimal spillover between the channels and fluorophores selected. The resulting cluster and metacluster assignments were evaluated by overlaying the cluster assignments on raw data and manually classified as beads, cells, or debris. After the clusters and metaclusters had been defined, all files were mapped onto the SOM to enumerate the number of events in each classification. A normalized cell count for each cluster and metacluster was calculated from the cell count and the CBP bead count.

H.2. Detailed Results for Flow Cytometry Intensities

Intensity distributions for additional channels (Figure 7-7) are shown here.

[26] Lo K, Hahne F, Brinkman R, Gottardo R (2009). "FlowClust: a Bioconductor package for automated gating of flow cytometry data." *BMC Bioinformatics*, **10**. R package version 4.5.0.

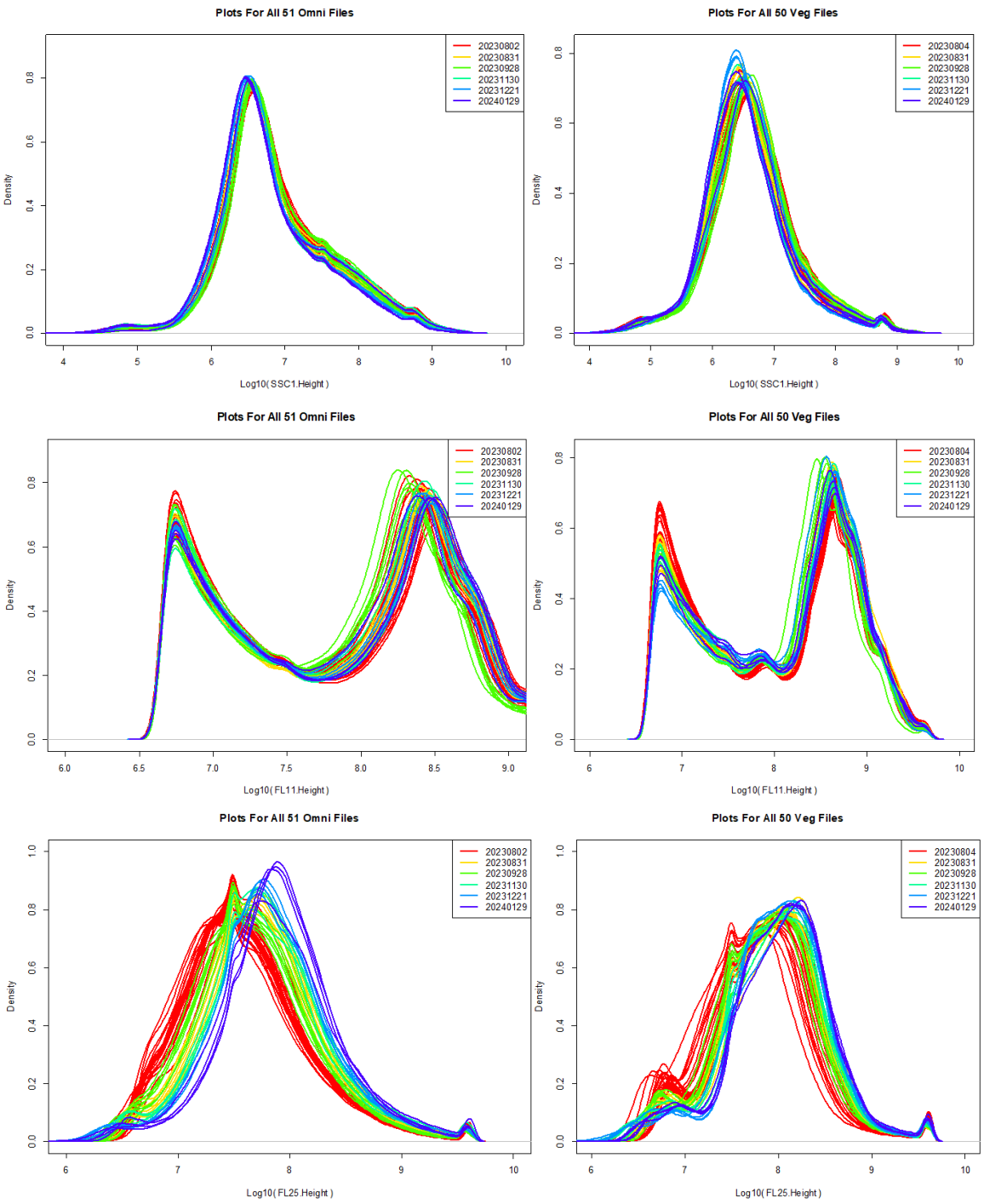


Figure H-1: Density plots of intensity distributions across channels and cohorts via FCM. Each row corresponds to a different channel of the cytometer (from top: SSC, FL11, FL25) and the columns correspond to donor pools (left – omnivore, right – vegetarian). Each plot shows the log10-transformed intensity of the channel on the *x*-axis and the density on the *y*-axis with color-coding corresponding to experiment date. *X*-axis scales vary across plots.

Identified peak positions in intensity distributions for additional channels (Figure 7-8) are shown here for the homogeneity data.

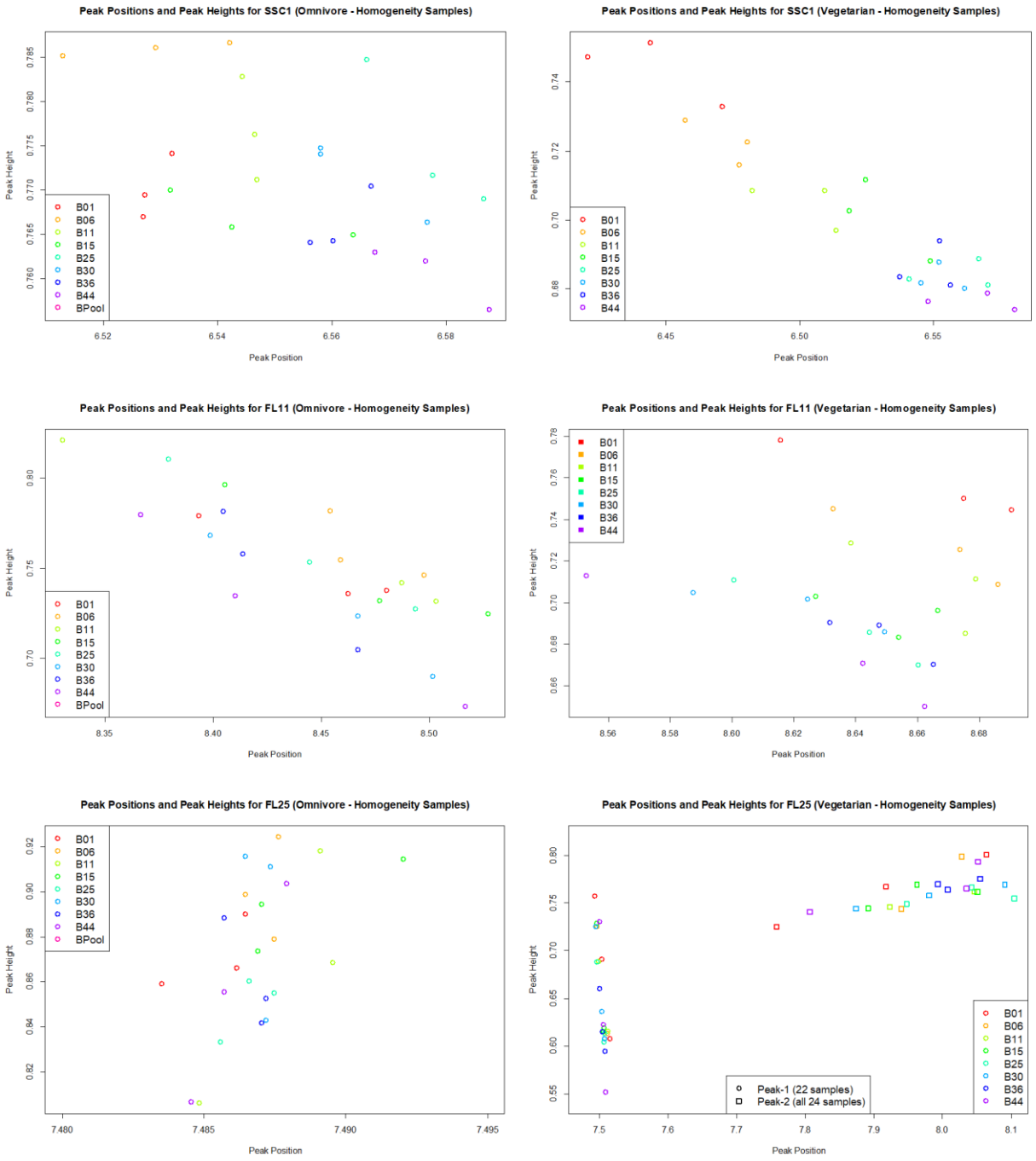


Figure H-2: Intensity peak position and height across cohort and channel via FCM. Each row corresponds to a different channel of the cytometer (from top: SSC, FL11, FL25) and the columns correspond to donor pools (left – omnivore, right – vegetarian). Each plot shows location of the identified peak in terms of the log10-transformed intensity (*x*-axis) and the density (*y*-axis) with color-coding corresponding to box number. *X*-axis and *y*-axis scales vary across plots.

Values supporting Figure 7-9 are shown in Table H-1: .

Table H-1: Coverage and Limits Across Peak Position, Peak Height, and Cohort.
The estimated 95 % confidence interval lower and upper limit for the log10-transformed intensities and peak height are given for each channel and donor pool.

Cohort	Channel	Peak	Axis	Lower Limit	Upper Limit	Confidence
Omni	FSC1	1	Position	7.285579	7.313782	0.95
Omni	FSC1	1	Height	0.535374	0.576286	0.95
Omni	FSC1	2	Position	7.448022	7.491804	0.95
Omni	FSC1	2	Height	0.508059	0.593826	0.95
Omni	SSC1	1	Position	6.50139	6.608523	0.95
Omni	SSC1	1	Height	0.751568	0.792796	0.95
Omni	FL11	1	Position	8.40256	8.490701	0.95
Omni	FL11	1	Height	0.71393	0.784683	0.95
Omni	FL20	1	Position	4.647133	4.651852	0.95
Omni	FL20	1	Height	1.604866	1.630049	0.95
Omni	FL25	1	Position	7.484801	7.488943	0.95
Omni	FL25	1	Height	0.835856	0.914701	0.95
Veg	FSC1	1	Position	7.214038	7.388805	0.95
Veg	FSC1	1	Height	0.51067	0.549287	0.95
Veg	SSC1	1	Position	6.403016	6.647247	0.95
Veg	SSC1	1	Height	0.636948	0.768843	0.95
Veg	FL11	1	Position	8.60064	8.684853	0.95
Veg	FL11	1	Height	0.638679	0.776974	0.95
Veg	FL20	1	Position	4.645421	4.652987	0.95
Veg	FL20	1	Height	1.617547	1.679138	0.95
Veg	FL25	1	Position	7.497992	7.509337	0.95
Veg	FL25	1	Height	0.596211	0.711335	0.95
Veg	FL25	2	Position	7.902855	8.106188	0.95
Veg	FL25	2	Height	0.746825	0.780223	0.95

The stability plots (similar to Figure 7-10) for the remaining channels of interest are shown in Figure I-3.

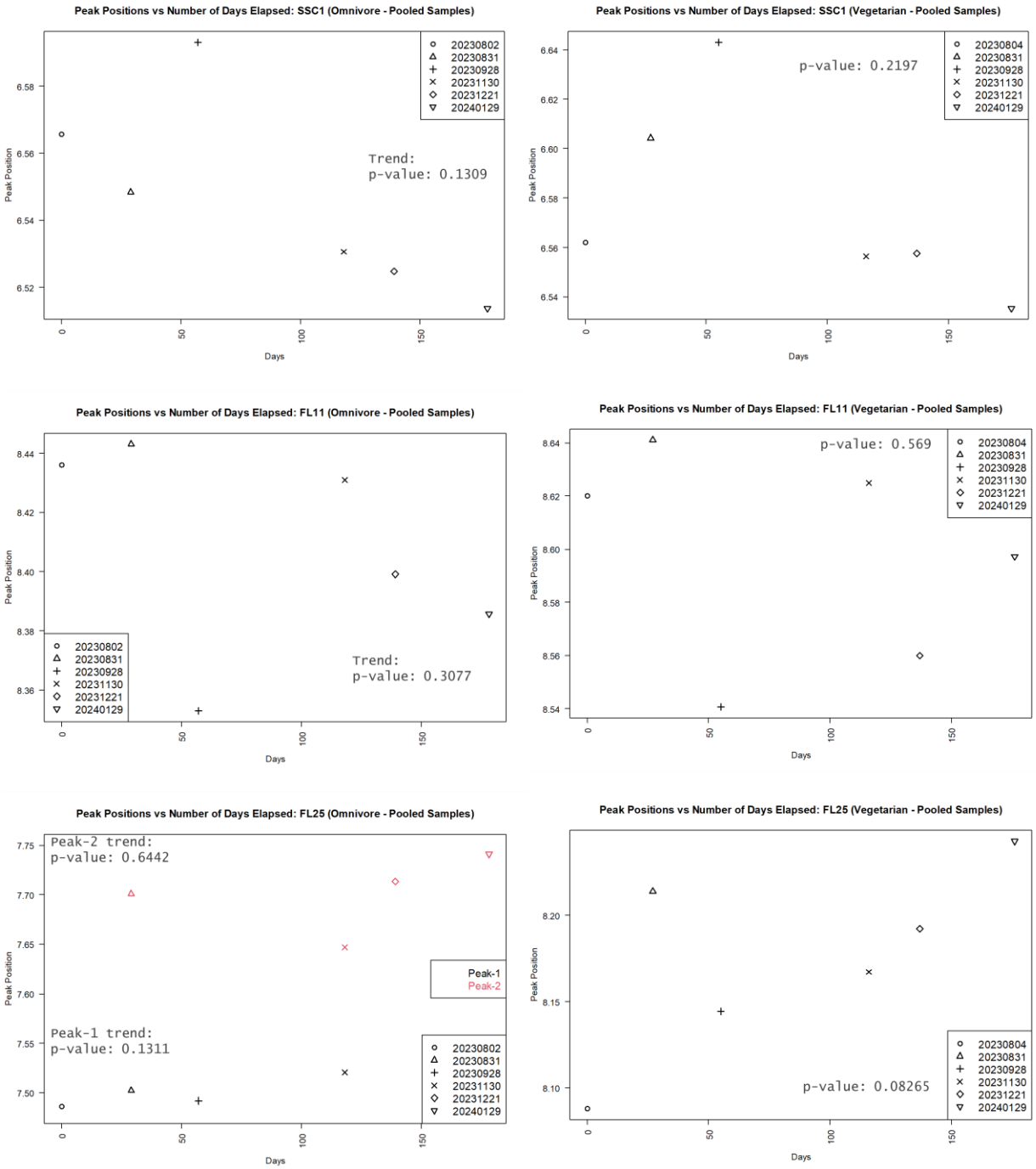


Figure I-3: Stability plots for intensities in additional channels. The peak intensity values for each of the pooled samples in were evaluated for trends as stability data. The *x*-axis gives the elapsed time since the pooled samples were generated. The *y*-axis represents the peak position. (left – omnivore, right – vegetarian samples).

Appendix I. **Report of Statistical Analysis**

I.1. Introduction

NIST defines a Reference Material (RM)[27] as follows:

> Reference Material (RM) - Material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process.

1. RM is a generic term.
2. Properties can be quantitative or qualitative, e.g. identity of substances or species.
3. Uses may include the calibration of a measurement system, assessment of a measurement procedure, assigning values to other materials, and quality control.
4. A single RM cannot be used for both calibration and validation of results in the same measurement procedure.
5. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms has an analogous definition (VIM - ISO/IEC Guide 99:2007, 5.13), but restricts the term "measurement" to apply to quantitative values and not to qualitative properties. However, Note 3 of ISO/IEC Guide 99:2007, 5.13, specifically includes the concept of qualitative attributes, called "nominal properties."

> (ISO Guide 30:1992/Amd 1:2008)

The phrase "sufficiently homogeneous and stable" is not unambiguously defined but is tied to the idea of "fitness for purpose." Therefore it seems more appropriate to characterize the degree of heterogeneity than to report a binary conclusion such as homogeneous or heterogeneous.

Often a reference material is characterized in terms of several attributes. For instance, the Nuclear Magnetic Resonance (NMR) measurements of the human fecal reference material (RM 8048) result in information on several metabolites. One could discuss the degree of heterogeneity of each metabolite across vials of RM 8048. Such information is tabulated in supplementary material section when appropriate. For a "fitness of purpose" judgment for RM 8048 as a whole it seemed reasonable to make a statement such as what follows:

If you select any two vials of RM 8048 from the current batch of human fecal reference material and measure K different attributes, we can say with $100 \times (1 - \alpha)$ percent confidence that $100 \times (1 - \beta)$ percent of the corresponding pairs of measured values of the attributes will differ from one another by at most $100 \times p$ percent. Smaller values of correspond to greater homogeneity. The scientist or the user of RM can make a judgment regarding fitness for purpose. The values for α and β are chosen by the subject matter expert and the value of p is

[27] Beauchamp, CR et al. (2021) Metrological Tools for the Reference Materials and Reference Instruments of the NIST Material Measurement Laboratory (U.S. Department of Commerce, Washington, D.C.) NIST Special Publication (SP) 260-132 2021 Edition. https://doi.org/10.6028/NIST.SP.260-136-2021

calculated using any appropriate statistical inference method. We now discuss this more formally.

I.2. Coefficient of Disagreement (COD)

Let there be N items (vials) of the reference material that are being characterized. Data are available on a (random) sample of η items. Let the population items be denoted by $\nu_1, \nu_2, \ldots, \nu_{\mathrm{N}}$. For each ν_i we (could) measure K different attributes (e.g. metabolites), either in terms of a 'peak height' (representing abundance) or relative abundance. For the purposes of explaining the measurement, we have included figure I-1 as representative data (Figure I-1). Let us denote these by $y_{i1}, y_{i2}, \ldots, y_{iK}$ $(1 \leq i \leq \mathrm{N})$. The y_{ij} are assumed to be nonnegative. This is true in all different types of measurements made on RM 8048. For any pair of items $\{\nu_i, \nu_j\}$ we denote the 'discrepancy' between their k^{th}component measurements by $PPD_{i,j}(k)$ (PPD stands for 'Peak to Peak discrepancy', the name chosen based on the initial scenario (NMR) we considered, but can simply be thought of as any interpretable measure of discrepancy between y_{ik} and y_{jk}). One such measure, used when analyzing NMR spectra is

$$PPD_{i,j}(k) = \frac{max\{y_{ik}, y_{jk}\}}{min\{y_{ik}, y_{jk}\}} - 1$$

One can say that the measurements (peak heights) for attribute k for ν_i and ν_j are within $100 \times PPD_{i,j}(k)$ percent of each other.

We then define a 'coefficient of disagreement' between and as

$$COD_{\beta(\nu_i,\nu_j)} = 100 \times (1 - \beta) \text{ percentile of the collection of values } \{PPD_{i,j}(\kappa) | \ 1 \leq \kappa \leq K\} \quad (1)$$

We will omit the subscript β from COD when the value of β used is clear from the context. If β is chosen to be zero then $COD_{\beta(\nu_i,\nu_j)}$ is the maximum value among $PPD_{i,j}(\kappa) | \ 1 \leq \kappa \leq K$. If it turns out that a $y_{i,k}$ is equal to zero then $PPD_{ij}(k)$ can be infinity. Therefore, we typically choose β to be a small value greater than 0 (and less than 1).

Another measure, sometimes considered when analyzing 'agreement' between two measurements is

$$PPD_{i,j}(k) = \frac{|y_{ik}, y_{jk}|}{mean(y_{ik}, y_{jk})}.$$

With this definition of PPD, one can say that the absolute difference between the measurements (peak heights) for attribute κ for ν_i and ν_j are within $100 \times PPD_{i,j}(k)$ percent of their mean.

$PPD_{i,j}(k)$ could be defined in other suitable ways depending on the application context. In any case, the definition of COD is the same as in Equation (1).

For information regarding homogeneity, one asks the following question:

> If two items (vials) are randomly chosen from the population, what is a percent upper confidence bound for their coefficient of disagreement (COD)?

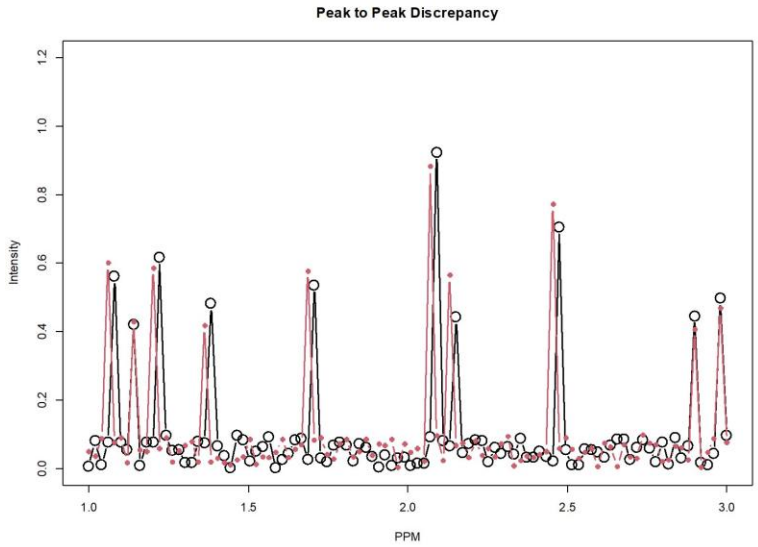


Figure I-1: Two 600 MHz NMR spectra from 1.0 ppm to 3.0 ppm. The black line and black circles (peaks) denote the intensities of spectrum 1 while the red line and red dots (peaks) denote the intensities of spectrum 2. Each peak in spectrum 1 is matched with the closest peak in spectrum 2, and then each peak has a PPD comparing peak heights. This creates a set of PPD values equal to the number of peaks in the spectra.

To answer the above question, we consider the collection of all pairwise COD values for the items in the sample of η items. This gives us $\eta(\eta - 1)/2$ pairwise COD values. If we consider these as a random sample (they are not; more on that later) from the $N(N - 1)/2$ pairs of COD values for the population of items, then we seek a bound p such that, with percent confidence, the COD value for the chosen pair will be at most . (As indicated above, this value of is easily interpretable based on the way it is defined.)

To compute such a confidence bound we use the quantCI function from the R package QuantileNPCI. This function can be used to calculate nonparametric confidence intervals for quantiles using fractional order statistics, based on Hutson 1999[28].

The confidence bound obtained in this manner is only an approximate confidence bound since we are treating a non i.i.d sample as if it is an i.i.d sample. A confidence bound approach that acknowledges the dependence among the $COD(v_i, v_j)$ values for items in the sample is under investigation. However, even when doing so, the confidence bound is not expected to be exact but perhaps it would yield a 'better' approximation. This will be investigated via simulation studies (future work) or asymptotic considerations. It is true that a bootstrap confidence interval could be carried out in this case that takes into account the inter-dependence of the

[28] Hutson, Alan D. 1999. "Calculating Nonparametric Confidence Intervals for Quantiles Using Fractional Order Statistics." Journal of Applied Statistics 26 (3): 343-53. https://doi.org/10.1080/02664769922458.

COD values. However, in all cases of this particular set of measurements, the number of COD values is quite low - at most we may have 40 samples. This means we can have at most 780 different combinations of pairs of samples to compare and calculate PPD and COD values from. When we resample, say, 1000 times for 1000 bootstrap replicates we then have encompassed the entirety of the available sample and will get values that are almost identical (to 4 decimal places in our checks) to the approximate interval we describe above.

I.3. Methods: COD Calculation

As we described in the previous Section, a different PPD can be defined for each application as appropriate. In particular, it is important that the PPD chosen is intuitive and has an easy interpretation for scientists and users. Given that the goal of each application is different, we will define the PPD for each. We will describe more details about each application and their respective data structures and analysis methods in appendices.

The applications will be discussed as follows: 600 MHz NMR will be discussed in Section I.3.1, where we are concerned with how far apart the peaks of any two spectra are; 700 MHz NMR will be discussed in Section I.3.2, where we are concerned with the presence / absence of peaks between spectra only; Metagenomics will be discussed in Section I.3.3, where we are concerned with how much the relative abundances of two samples differ; LC-MS metabolomics will be discussed in Section I.3.4, where we are interested in quantifying how well the two sets of metabolites agree with each other. Although we cannot use the PPD for Flow Cytometry, it is mentioned in Section I.3.5 and full methods are explored in Section I.8.

I.3.1. NMR: 600 MHz

Given NMR spectra from two different samples (these may be two separate vials or subsamples from a single vial) we are interested in quantifying how well the two spectra agree with one another. Greater agreement indicates more homogeneity (or stability). In this case, "agreement" is how far apart the K corresponding peaks of the spectra are from each other. We must first define how to identify corresponding peaks between samples:

1. We first set a threshold value for peak heights and consider only peaks whose heights exceed this threshold value. We have chosen this value to be 10^{-5}.

2. We identify, in each spectrum, peaks whose heights exceed 10^{-5} and note the corresponding peak position in terms of the ppm value. In the spectra, the ppm values are recorded in discrete steps of 0.0002 units.

3. From experience we know that a peak corresponding to a particular metabolite in one spectrum may be slightly offset from the corresponding peak in the second spectrum. We associate a peak from one spectrum with a peak in the second spectrum if they are within 0.0002 units of each other (one ppm step length).

4. Once pairs of corresponding peaks have been identified in the two spectra, we record the corresponding peak heights as well.

5. Suppose, for a particular peak, the heights in the two spectra are and . The discrepancy between the two peak heights is quantified as follows:

$$PPD_{i,j}(k) = \frac{max\{y_{ik}, y_{jk}\}}{min\{y_{ik}, y_{jk}\}} - 1$$

   which is as described in the previous section.

In order to exclude only some occasional outliers that may occur, but which do not indicate a deviation from homogeneity and stability that is extreme enough to rule out that the material is fit for purpose, we choose $\beta = 0.01$. Then the COD can be found by following Equation (1), and with $\alpha = 0.05$ an approximate 95 % upper confidence limit can be found using the nonparametric approach for quantiles implemented in quantCI. See Section I.4 for further details and analysis.

I.3.2. NMR: 700 MHz

Similarly to the NMR measurement at 600 MHz, we are interested in the level of agreement between any two spectra. However, since methodology and spectrum pre-processing steps differed, we will not consider the distance between K corresponding peaks. Instead, we have a PPD based on peak presence:

1. The spectra were processed and separated into buckets with an intelligent binning procedure using NMRProcFlow v1.4.
2. For each sample in each bin, we must define what we consider as an identified "peak." If we consider $z_{min,\kappa}$ to be the minimum boundary and $z_{max,\kappa}$ to be the maximum boundary of bin κ, then we are searching for local maxima in $(z_{min,\kappa}, z_{max,\kappa})$ Let $f(z)$ represent the intensity function within bin κ, where $y \in (z_{min,\kappa}, z_{max,\kappa})$. We are searching for local maxima $z_{peak} \in (z_{min,\kappa}, z_{max,\kappa})$ such that:
   a. The intensity at the peak satisfies:

$$f(z_{peak}) > (\mu_{noise} + 3\sigma_{noise})$$

      where μ_{noise} is the average intensity in the noise region (above 9.0 ppm), and $3\sigma_{noise}$ is the standard deviation of the intensity in the same region.

   b. There exists a valley on either side of , with:

$$f(z_{peak}) \geq 1.025 \cdot f(z_{valley})$$

      ensuring at least a 2.5 % increase from the valley to the peak.

   Each bin κ can contain at most one such peak. We know from experience that the boundaries $z_{min,\kappa}$ and $z_{max,\kappa}$ cannot be identified as peaks.

3. Let P be a $N \times K$ matrix, where N is the number of samples and K is the number of bins. Let y_{ik} be the matrix entry $P_{i,\kappa}$ for sample i and bin κ, which we define as:

$$y_{ik} = \begin{cases} 1 & if\ a\ peak\ is\ identified\ in\ bin\ \kappa\ for\ sample\ i, \\ 0 & if\ a\ peak\ is\ identified\ in\ bin\ \kappa\ for\ sample\ i. \end{cases}$$

We can then calculate the Jaccard distance for each pair of samples across all bins to quantify the dissimilarity of samples based on the presence or absence of identified peaks. We can calculate the discrepancy as

$$PPD = 1 - \frac{|y_{ik} \cap y_{jk}|}{|y_{ik} \cup y_{jk}|}$$

where $y_{ik} \cap y_{jk}$ represents the bins where both samples have a peak (intersection), $y_{ik} \cup y_{jk}$ represents the bins where at least one of the samples has a peak (union), and $|\cdot|$ denotes the cardinality.

In this case, we have K binary responses based on the presence or absence of peaks in each of the bins. We find the level of agreement between the two samples by summarizing the difference between these responses with the Jaccard distance. Given the binary response, we can define $\beta = 0.01$ in this case and $\alpha = 0.05$ as before. We find the approximate upper 95 % confidence bound using the same procedure. See Section I.5 for details and further analysis.

I.3.3. Metagenomics

The data for metagenomics is not a spectrum, but relative abundances for each of the K identified 16S DNA sequences. Given the relative abundances of any two samples, we are interested in quantifying how similar the two distributions of abundances agree with one another:

1. We first set a threshold for the number of unique sequences considered per group. We have chosen the sequences, with relative abundances greater than 0.0008.

2. If we have y_{ik} as a the relative abundance of one a particular sequence in sample i and y_{jk} as the matching relative abundance in sample j, we can find the discrepancy between the two relative abundances as:

$$PPD = \frac{\log(y_{ik}) - \log(y_{jk})}{\log(y_{ik}) + \log(y_{jk})}$$

If $y_{ik} = y_{jk} = 0$, we define PPD = 0. If $y_{ik} = y_{jk} = 1$, we also define PPD = 0. If either y_{ik} or y_{jk} are 0 while the other is not, we define $PPD = \infty$. This metric has several qualities which make it desirable for our purpose: it is symmetric, bounded between -1 and 1, and sensitive to relative changes. We perform a log-transform in order to account for the fact that we have many "rare" sequences with low relative abundance, and would like to minimize impacts of potential extreme values.

3. Since we are not specifically interested in the direction of discrepancy, we can take the absolute value of the PPD so that we have a metric between 0 and 1.

Note that this PPD value is how far apart the two relative abundances are in reference to their midpoint, and is closely related to the second example PPD described in Section I.2. With the threshold of 0.0008 for the relative abundances at most 0.3 % of PPD values will be infinite, and thus be excluded. This could be avoided by increasing the threshold value, but we would like to include as much information as possible.

In this application, we set $\beta = 0.01$ and $\alpha = 0.05$ and can, again, find an approximate upper 95 % confidence limit for the COD from Equation (1) using the methods described in Section I.2.

Additional analysis was performed in order to determine homogeneity at the 16S sequence level. See section I.6 for details and results.

I.3.4. LC-MS Metabolomics

Given LC-MS data from two different samples, we are interested in quantifying how well the two sets of measurements agree with one another. With this in mind, we define a PPD that quantifies how well the relative abundance of K identified metabolites from the samples agree. We can find a discrepancy measure as follows:

1. Suppose for a particular chromatogram feature, the response for the samples are y_{ik} and y_{jk} .

2. We can perform a log-transform in order to more easily handle the fact that the data spans multiple orders of magnitude.

3. We can then define the discrepancy measure as:

$$PPD = \frac{\left|\log(y_{ik}) - \log(y_{jk})\right|}{mean\left(\log(y_{ik}), \log(y_{jk})\right)}.$$

In this application, we set $\beta = 0.10$ and $\alpha = 0.05$ and can, again, find approximate 95 % confidence limits for the COD from Equation (1) using the methods described in Section I.2. See section I.7 for full analysis and details.

I.3.5. Flow Cytometry

No COD was used for Flow Cytometry since only univariate features were used to evaluate homogeneity and stability (as opposed to entire spectra or an entire vector of relative abundances). Consequently, nonparametric confidence intervals for the features of interest could be directly calculated using the R function quantCI from the package QuantileNPCI. See Section I.8 for analysis details and results.

Further details and analyses can be found in the following sections: Section I.4 includes the full report of analysis for 600 MHz NMR, Section I.5 includes the full analysis for 700 MHz NMR,

Section I.6 has all details for the Metagenomics analysis, and Section I.7 has all details for LC-MS metabolomics. All these analyses include COD calculations and characterizations as described in the previous sections along with additional exploratory analyses or other appropriate modeling approaches. Section I.8 has the full report of analysis for Flow Cytometry. As already mentioned, the measurement for Flow Cytometry did not lend itself to be applicable to the COD approach, and so the Appendix includes an alternate homogeneity and stability characterization.

I.4. Metabolomics: Nuclear Magnetic Resonance 600 MHz

Twenty two Omnivore samples and Twenty two Vegetarian samples were analyzed following the 'unfiltered' protocol. Statistical analyses were conducted on the resulting NMR spectra to assess homogeneity and stability of the RM8048 samples. The regions of ppm values that were included in the analysis were ppm = 0.5 to ppm = 4.67 and ppm = 4.9 to ppm = 9.5. The vial labels and time points at which the samples were analyzed are shown in Table I-1.

Time.Point	Days.Elapsed	Vial
1	0	omni_25-1
1	0	omni_6-1
1	0	omni_36-1
1	0	omni_15-1
2	78	omni_25-2
2	78	omni_6-2
2	78	omni_36-2
2	78	omni_15-2
3	105	omni-25-3
3	105	omni-6-3
3	105	omni-36-3
3	105	omni-15-3
4	160	omni-6-4
4	160	omni-36-4
4	160	omni-15-4
4	160	omni-25-4
5	274	omni-25-5
5	274	omni-1-4
5	274	omni-44-3
5	274	omni-1-3
5	274	omni-44-4
5	274	omni-15-5

Time.Point	Days.Elapsed	Vial
1	0	veg_36-1
1	0	veg_6-1
1	0	veg_25-1
1	0	veg_15-1
2	79	veg_6-2
2	79	veg_36-2
2	79	veg_25-2
2	79	veg_15-2
3	105	veg-6-3
3	105	veg-36-3
3	105	veg-25-3
3	105	veg-15-3
4	160	veg-6-4
4	160	veg-36-4
4	160	veg-25-4
4	160	veg-15-4
5	280	veg-25-5
5	280	veg-1-4
5	280	veg-44-3
5	280	veg-1-3
5	280	veg-44-4
5	280	veg-15-5

Table I-1 Vials and timepoints sampled for Nuclear Magnetic Resonance 600 MHz

I.4.1. Comparison of Omnivore Samples with Vegetarian Samples

There are 44 samples (vials) in all. For every pair of samples from this collection we compute the corresponding COD. This results in a 44 by 44 matrix which we denote by **A**. The diagonal elements of this matrix are 0 because the disagreement between a sample and itself is zero.

The 44 by 44 matrix A may be visualized using a multidimensional scaling plot shown in Figure I-2.

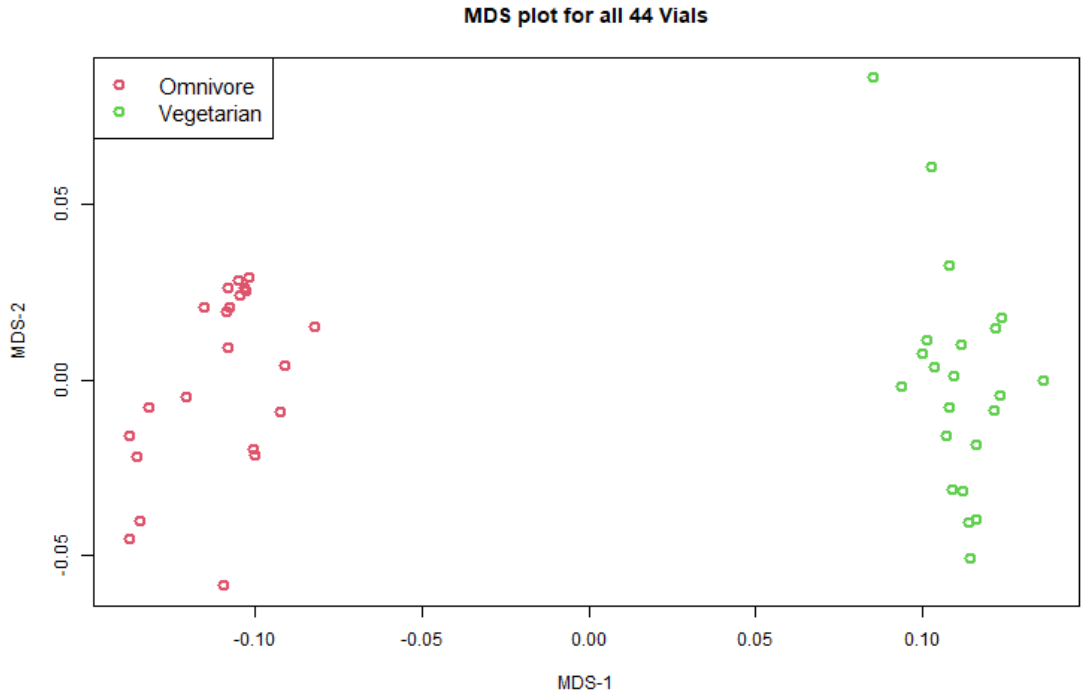


Figure I-2 Multidimensional Scaling Plot for visualizing the matrix **A** of Coefficients of disagreement between all pairs of samples, both Omnivore and Vegetarian. This plot makes it clear that omnivore samples are demonstrably different from vegetarian samples with respect to their NMR spectra.

The entries in the matrix A can also be viewed as a point plot shown in Figure I-3. For each sample, identified by its label along the -axis, its coefficients of disagreement with all other samples are shown along the vertical axis.

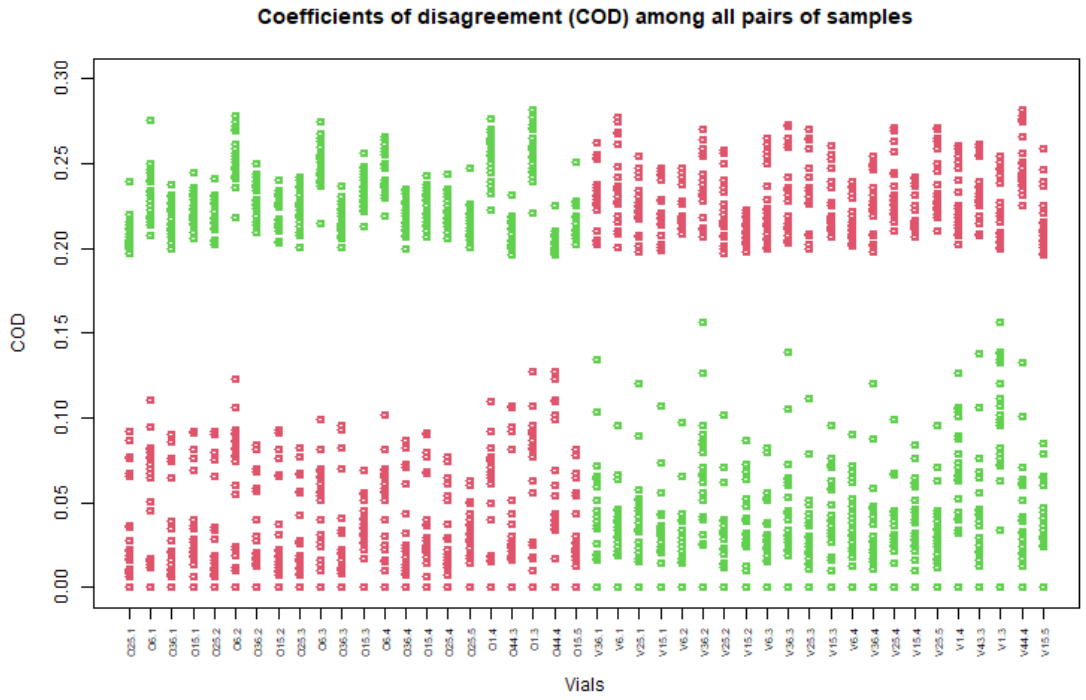


Figure I-3 Coefficients of disagreement for each sample when compared to all other samples. Red points correspond to omnivore samples and green points correspond to vegetarian samples. The larger COD values come from a comparison of an omnivore sample with a vegetarian samples. For example, the very first sample (O25.1) along the horizontal axis is an omnivore sample. When it is compared to other omnivore samples (red points along the vertical corresponding to O25.1) the values remain below 0.15 on the vertical scale but COD values when compared to vegetarian samples are all above 0.15 on the vertical scale).

As in the case of the MDS plot in Figure I-2, we again see that the Omnivore spectra differ greatly from the Vegetarian spectra. This is seen in the plot shown in Figure I-3 as a big gap between the lower set of points (-axis values below 0.15) and the upper set of points (-axis values above 0.15). This difference is also seen from the heat map shown in Figure I-4.

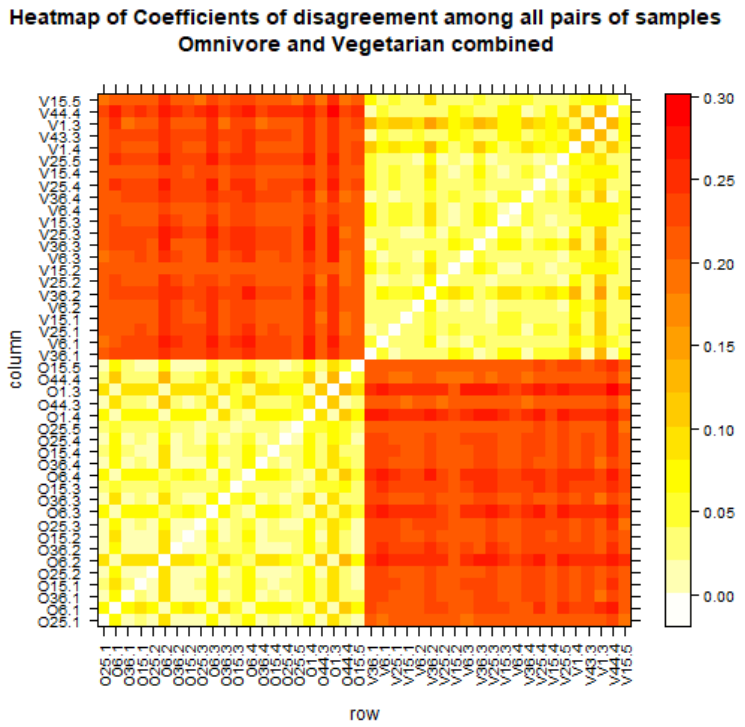


Figure I-4 Heatmap of pairwise COD values for all samples. Lighter colors indicate less disagreement and darker (more red) colors indicate more disagreement. It is clearly seen that the omnivore samples agree well among each other and vegetarian samples also agree well with each other but omnivore samples noticeably disagree with vegetarian samples.

I.4.2. Homogeneity and Stability of Omnivore Samples

In this subsection we will investigate the COD values among the 22 omnivore samples only. These 22 samples represent the entire collection of omnivore samples over approximately 9 months time window. The 22 by 22 matrix of COD values among these samples, represented by the matrix B is shown in Table I-2

	O25.1	O6.1	O36.1	O15.1	O25.2	O6.2	O36.2	O15.2	O25.3	O6.3	O36.3	O15.3	O6.4	O36.4	O15.4	O25.4	O25.5	O1.4	O44.3	O1.3	O44.4	O15.5
O25.1	0	0.08	0.01	0.02	0.01	0.09	0.02	0.01	0.01	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.04	0.02
O6.1	0.08	0	0.08	0.08	0.08	0.01	0.07	0.08	0.07	0.01	0.08	0.05	0.02	0.07	0.08	0.06	0.05	0.02	0.09	0.02	0.11	0.07
O36.1	0.01	0.08	0	0.02	0.01	0.09	0.02	0.01	0.01	0.06	0.01	0.03	0.07	0.01	0.02	0.01	0.03	0.07	0.02	0.09	0.04	0.02
O15.1	0.02	0.08	0.02	0	0.02	0.09	0.01	0.02	0.03	0.07	0.02	0.04	0.07	0.02	0.01	0.03	0.04	0.08	0.02	0.09	0.04	0.02
O25.2	0.01	0.08	0.01	0.02	0	0.09	0.01	0.01	0.01	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.03	0.02
O6.2	0.09	0.01	0.09	0.09	0.09	0	0.08	0.09	0.08	0.02	0.09	0.06	0.02	0.08	0.09	0.07	0.06	0.02	0.11	0.01	0.12	0.08
O36.2	0.02	0.07	0.02	0.01	0.01	0.08	0	0.02	0.02	0.06	0.02	0.03	0.06	0.02	0.01	0.02	0.03	0.07	0.03	0.08	0.04	0.01
O15.2	0.01	0.08	0.01	0.02	0.01	0.09	0.02	0	0.02	0.07	0.01	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.09	0.04	0.02
O25.3	0.01	0.07	0.01	0.03	0.01	0.08	0.02	0.02	0	0.06	0.02	0.03	0.06	0.01	0.03	0.01	0.02	0.07	0.03	0.08	0.04	0.02
O6.3	0.07	0.01	0.06	0.07	0.07	0.02	0.06	0.07	0.06	0	0.07	0.03	0.01	0.06	0.07	0.05	0.04	0.02	0.08	0.03	0.1	0.05
O36.3	0.01	0.08	0.01	0.02	0.01	0.09	0.02	0.01	0.02	0.07	0	0.04	0.07	0.01	0.02	0.02	0.03	0.08	0.02	0.1	0.03	0.02
O15.3	0.04	0.05	0.03	0.04	0.04	0.06	0.03	0.04	0.03	0.03	0.04	0	0.03	0.03	0.04	0.02	0.02	0.04	0.05	0.06	0.07	0.03
O6.4	0.07	0.02	0.07	0.07	0.07	0.02	0.06	0.07	0.06	0.01	0.07	0.03	0	0.06	0.07	0.05	0.04	0.02	0.08	0.03	0.1	0.06
O36.4	0.01	0.07	0.01	0.02	0.01	0.08	0.02	0.01	0.01	0.06	0.01	0.03	0.06	0	0.03	0.01	0.02	0.07	0.03	0.09	0.04	0.02
O15.4	0.02	0.08	0.02	0.01	0.02	0.09	0.01	0.02	0.03	0.07	0.02	0.04	0.07	0.03	0	0.03	0.04	0.08	0.02	0.09	0.04	0.02
O25.4	0.02	0.06	0.01	0.03	0.02	0.07	0.02	0.02	0.01	0.05	0.02	0.02	0.05	0.01	0.03	0	0.01	0.06	0.04	0.08	0.05	0.02
O25.5	0.03	0.05	0.03	0.04	0.03	0.06	0.03	0.03	0.02	0.04	0.03	0.02	0.04	0.02	0.04	0.01	0	0.05	0.04	0.06	0.06	0.03
O1.4	0.08	0.02	0.07	0.08	0.08	0.02	0.07	0.08	0.07	0.02	0.08	0.04	0.02	0.07	0.08	0.06	0.05	0	0.09	0.02	0.11	0.06
O44.3	0.02	0.09	0.02	0.02	0.02	0.11	0.03	0.02	0.03	0.08	0.02	0.05	0.08	0.03	0.02	0.04	0.04	0.09	0	0.11	0.02	0.03
O1.3	0.09	0.02	0.09	0.09	0.09	0.01	0.08	0.09	0.08	0.03	0.1	0.06	0.03	0.09	0.09	0.08	0.06	0.02	0.11	0	0.13	0.08
O44.4	0.04	0.11	0.04	0.04	0.03	0.12	0.04	0.04	0.04	0.1	0.03	0.07	0.1	0.04	0.04	0.05	0.06	0.11	0.02	0.13	0	0.04
O15.5	0.02	0.07	0.02	0.02	0.02	0.08	0.01	0.02	0.02	0.05	0.02	0.03	0.06	0.02	0.02	0.02	0.03	0.06	0.03	0.08	0.04	0

Table I-2. 22 by 22 matrix of COD values among omnivore samples over approximately 9 months time.

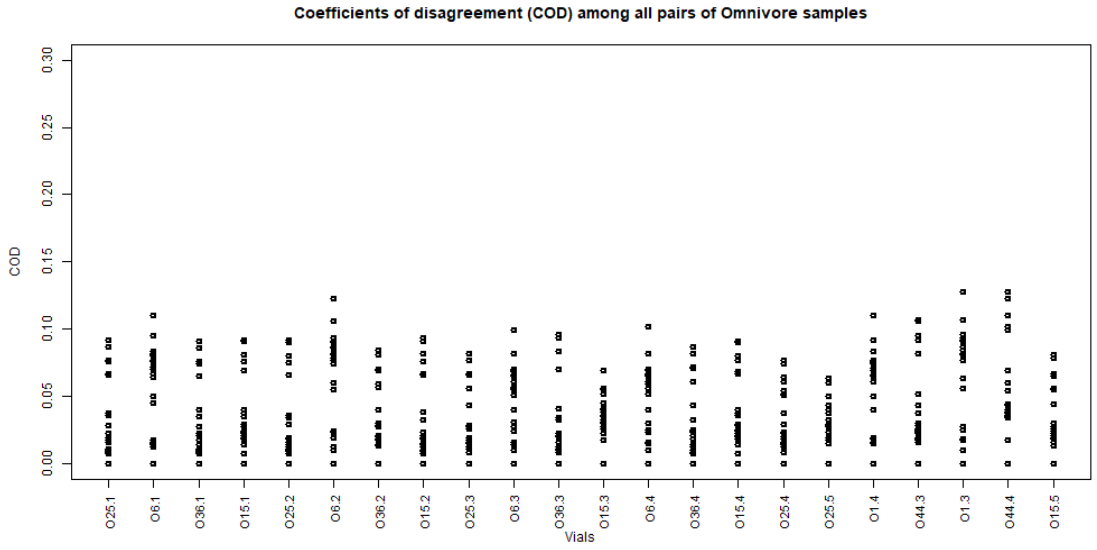


Figure I-5 Coefficients of Disagreement for each omnivore sample (as labeled along the horizontal axis) with all other omnivore samples.

The 22 by 22 matrix B may be visualized as a point plot shown in Figure I-5. For each sample, identified by its label along the -axis, its coefficients of disagreement with all other samples are shown along the vertical axis. A heat map of this matrix is shown in Figure I-6. Cells in the heat map that are more red correspond to samples that show greater disagreement with the bulk of the other samples.

An approximately 99 % upper confidence bound for the coefficient of disagreement among any pair of samples from the entire collection of vials is equal to 0.1069. This value is obtained using the R function quantCI from the package QuantileNPCI based on the COD values for the 231 pairs of omnivore samples. The formula for this confidence statement is based on the assumption that the 231 COD values are independent but, due to the manner in which the 22 vials were

selected, the 231 COD values are only approximately independent. Hence the provided value is only an approximate 95 % confidence bound.

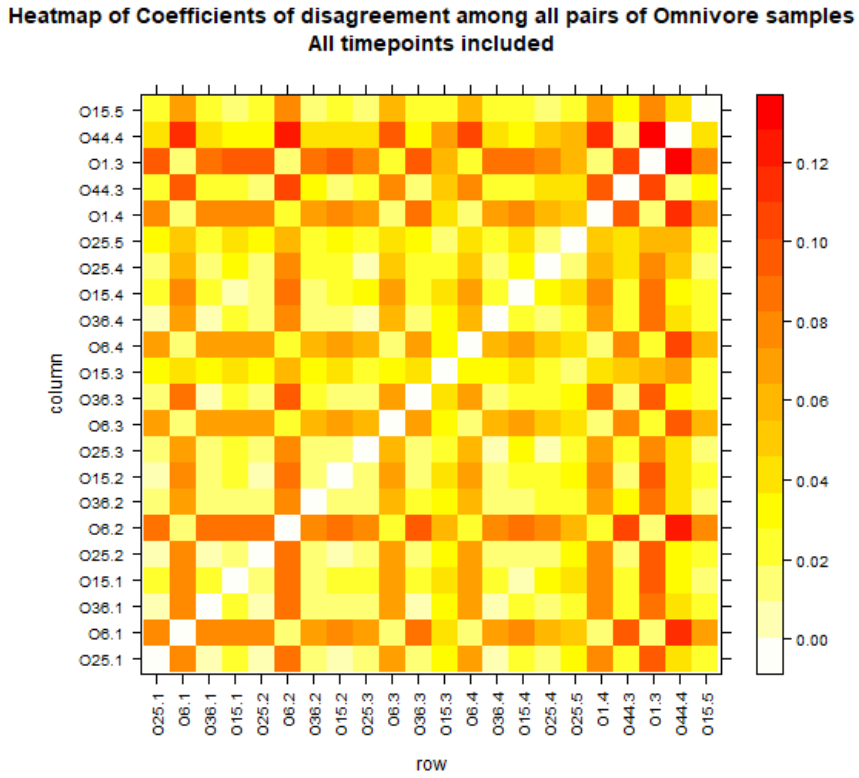


Figure I-6 Heat map showing COD values for pairs of omnivore samples. Cells in the heat map that are more red correspond to samples that show greater disagreement with the bulk of the other samples.

I.4.3. Homogeneity and Stability of Vegetarian Samples

In this subsection we will investigate the COD values among the 22 vegetarian samples only. These 22 samples represent the entire collection of vegetarian samples over approximately 9 months time window. The 22 by 22 matrix of COD values among these samples (Table I-3), denoted by C, is shown in Figure I-7

	V36.1	V6.1	V25.1	V15.1	V6.2	V36.2	V25.2	V15.2	V6.3	V36.3	V25.3	V15.3	V6.4	V36.4	V25.4	V15.4	V25.5	V1.4	V43.3	V1.3	V44.4	V15.5
V36.1	0	0.04	0.02	0.03	0.04	0.03	0.04	0.06	0.05	0.02	0.03	0.07	0.07	0.02	0.05	0.06	0.04	0.1	0.02	0.13	0.02	0.07
V6.1	0.04	0	0.03	0.04	0.02	0.07	0.02	0.04	0.02	0.05	0.02	0.04	0.04	0.03	0.04	0.04	0.03	0.06	0.05	0.1	0.04	0.03
V25.1	0.02	0.03	0	0.03	0.03	0.04	0.02	0.05	0.04	0.03	0.02	0.06	0.05	0.02	0.05	0.05	0.03	0.09	0.02	0.12	0.02	0.05
V15.1	0.03	0.04	0.03	0	0.02	0.06	0.02	0.03	0.03	0.03	0.01	0.04	0.04	0.02	0.03	0.03	0.02	0.07	0.04	0.11	0.03	0.03
V6.2	0.04	0.02	0.03	0.02	0	0.07	0.01	0.03	0.02	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.07	0.04	0.1	0.04	0.03
V36.2	0.03	0.07	0.04	0.06	0.07	0	0.06	0.09	0.08	0.03	0.05	0.1	0.09	0.04	0.07	0.08	0.07	0.13	0.02	0.16	0.04	0.09
V25.2	0.04	0.02	0.02	0.02	0.01	0.06	0	0.03	0.02	0.04	0.01	0.04	0.03	0.02	0.03	0.03	0.01	0.07	0.04	0.1	0.03	0.03
V15.2	0.06	0.04	0.05	0.03	0.03	0.09	0.03	0	0.02	0.06	0.04	0.01	0.01	0.05	0.03	0.01	0.03	0.04	0.07	0.07	0.06	0.03
V6.3	0.05	0.02	0.04	0.03	0.02	0.08	0.02	0.02	0	0.05	0.03	0.03	0.02	0.04	0.03	0.03	0.02	0.05	0.06	0.08	0.05	0.03
V36.3	0.02	0.05	0.03	0.03	0.04	0.03	0.04	0.06	0.05	0	0.03	0.07	0.06	0.02	0.04	0.06	0.04	0.11	0.02	0.14	0.03	0.06
V25.3	0.03	0.02	0.02	0.01	0.02	0.05	0.01	0.04	0.03	0.03	0	0.05	0.04	0.01	0.02	0.04	0.02	0.08	0.03	0.11	0.02	0.04
V15.3	0.07	0.04	0.06	0.04	0.03	0.1	0.04	0.01	0.03	0.07	0.05	0	0.02	0.06	0.04	0.02	0.03	0.03	0.08	0.06	0.07	0.03
V6.4	0.07	0.04	0.05	0.04	0.03	0.09	0.03	0.01	0.02	0.06	0.04	0.02	0	0.05	0.03	0.02	0.03	0.04	0.07	0.07	0.06	0.04
V36.4	0.02	0.03	0.02	0.02	0.03	0.04	0.02	0.05	0.04	0.02	0.01	0.06	0.05	0	0.03	0.05	0.03	0.09	0.02	0.12	0.02	0.05
V25.4	0.05	0.04	0.05	0.03	0.02	0.07	0.03	0.03	0.03	0.04	0.02	0.04	0.03	0.03	0	0.03	0.01	0.07	0.04	0.1	0.04	0.03
V15.4	0.06	0.04	0.05	0.03	0.02	0.08	0.03	0.01	0.03	0.06	0.04	0.02	0.02	0.05	0.03	0	0.02	0.04	0.06	0.08	0.06	0.03
V25.5	0.04	0.03	0.03	0.02	0.02	0.07	0.01	0.03	0.02	0.04	0.02	0.03	0.03	0.03	0.01	0.02	0	0.06	0.05	0.1	0.04	0.02
V1.4	0.1	0.06	0.09	0.07	0.07	0.13	0.07	0.04	0.05	0.11	0.08	0.03	0.04	0.09	0.07	0.04	0.06	0	0.11	0.03	0.1	0.04
V43.3	0.02	0.05	0.02	0.04	0.04	0.02	0.04	0.07	0.06	0.02	0.03	0.08	0.07	0.02	0.04	0.06	0.05	0.11	0	0.14	0.01	0.06
V1.3	0.13	0.1	0.12	0.11	0.1	0.16	0.1	0.07	0.08	0.14	0.11	0.06	0.07	0.12	0.1	0.08	0.1	0.03	0.14	0	0.13	0.08
V44.4	0.02	0.04	0.02	0.03	0.04	0.04	0.03	0.06	0.05	0.03	0.02	0.07	0.06	0.02	0.04	0.06	0.04	0.1	0.01	0.13	0	0.06
V15.5	0.07	0.03	0.05	0.03	0.03	0.09	0.03	0.03	0.03	0.06	0.04	0.03	0.04	0.05	0.03	0.03	0.02	0.04	0.06	0.08	0.06	0

Table I-3. 22 by 22 matrix of COD values among vegetarian samples over approximately 9 months time

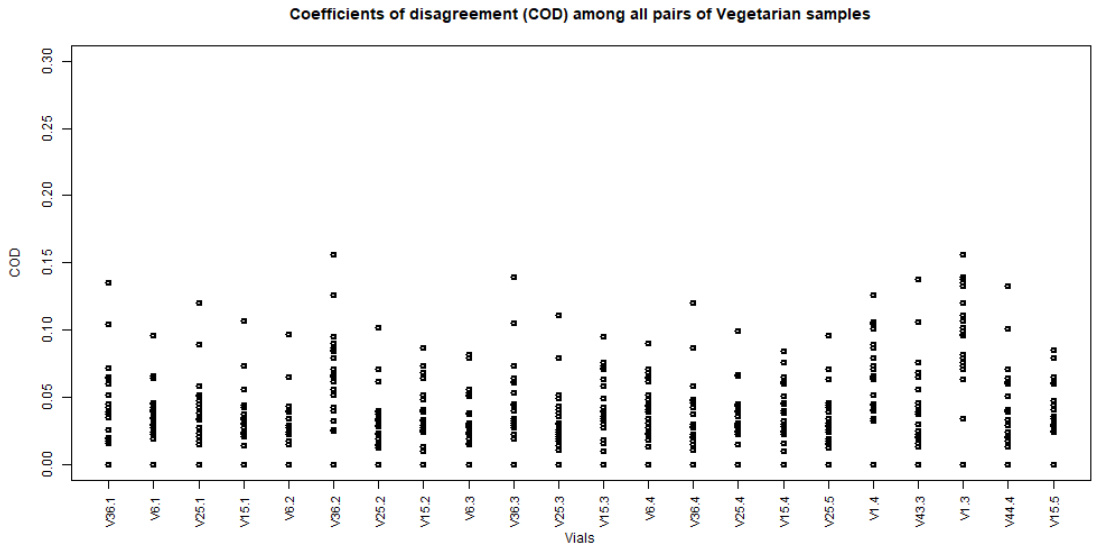


Figure I-7 Coefficients of Disagreement for each vegetarian sample (as labeled along the horizontal axis) with all other vegetarian samples.

The 22 by 22 matrix may be visualized as a point plot shown in Figure I-7. For each sample, identified by its label along the -axis, its coefficients of disagreement with all other samples are shown along the vertical axis. A heat map of this matrix is shown in Figure I-8.

An approximately 95 % upper confidence bound for the coefficient of disagreement among any pair of samples from the entire collection of vials is equal to 0.1323. This value is obtained using the R function quantCI from the package QuantileNPCI based on the COD values for the 231 pairs of Vegetarian samples. The formula for this confidence statement is based on the assumption that the 231 COD values are independent but, due to the manner in which the 16 vials were

selected, the 231 COD values are only approximately independent. Hence the provided value is only an approximate 95 % confidence bound.

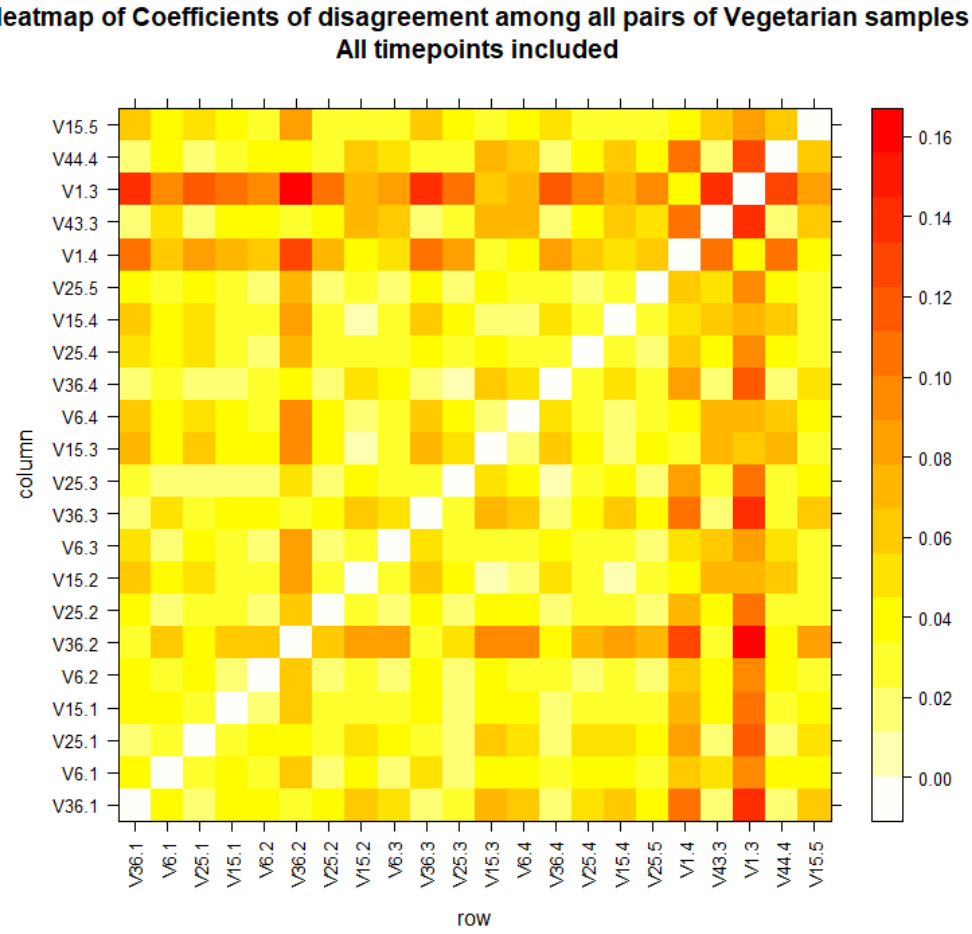


Figure I-8 Heat map showing COD values for pairs of vegetarian samples. Cells in the heat map that are more red correspond to samples that show greater disagreement with the bulk of the other samples.

I.5. Metabolomics: Nuclear Magnetic Resonance 700 MHz

Polar metabolites were either extracted in 2 separate batches (one for Omnivore and one for Vegetarian) of 10 vials each extracted at the same time (for homogeneity testing), or in a single batch containing both Omnivore and Vegetarian samples (for each stability time point). Each extraction batch also included 3 aliquots of a NIST standard reference material (SRM 1946 – Lake Superior Fish Tissue, 100 mg +/- 3 mg per aliquot), in addition to 1 procedural blank for both homogeneity and stability testing.

Time Point	Box
T0	1, 6, 11, 15, 20, 25, 30, 36, 44, 50
T1 (+2mo)	1, 15, 30, 50

Time Point	Box
T2 (+4mo)	1, 15, 30, 50
T3 (+6mo)	1, 15, 30, 50
T4 (+8mo)	6, 20, 25, 36
T5 (+10mo)	6, 20, 25, 36
T6 (+12mo)	6, 20, 25, 36

Table I-4 Time Points and Corresponding Boxes Nuclear Magnetic Resonance 700 MHz Assessment

For homogeneity, 10 vials were analyzed from Omnivores and 10 from Vegetarians using the sampling scheme noted in Table I-4. In each subsequent time point for stability measurements, 4 vials were sampled. This leads to a total of 34 samples each for Omnivores and Vegetarians. One dimensional (1D) 1H NMR spectra were acquired, and all acquired spectra were referenced to the chemical shift reference TMSP, phased, and baseline-corrected with further manual corrections when necessary prior to any statistical analysis.

For experimental sample analysis, intelligent binning (bucketing) was performed using NMRProcFlow v1.4. The spectra were binned from 0.20 ppm to 10.0 ppm with a resolution factor of 0.5 and a signal-to-noise ratio threshold of 3. The water region and contaminants identified in the blank sample were removed from the spectra creating 537 bins for analysis. All data were normalized using normalization by sum of intensities, and centered and scaled using Pareto scaling. This procedure resulted in 695 bins for Omnivores and 652 bins for Vegetarians.

One effective approach for exploratory assessment of homogeneity and stability is a distance metric. An intuitive measure of distance between spectra is to use the correlation between spectra. Specifically, we use 1 - correlation in order to ensure that all values are positive. From these distances, we can create a multi-dimensional scaling (MDS) plot to view the similarity between the samples. If we include all time points, we see in Figure I-9 that there is a separation over time. Specifically, the homogeneity time points (labeled as “t0”) marked in red are slightly separate from the other time points. There is no clear trend over time - the time point 5 data, however, is clearly separated from the other time points in both groups. This could be due to variability that can be expected in the measurement. It is not unexpected that the stability time points might have more variation overall than the homogeneity time points.

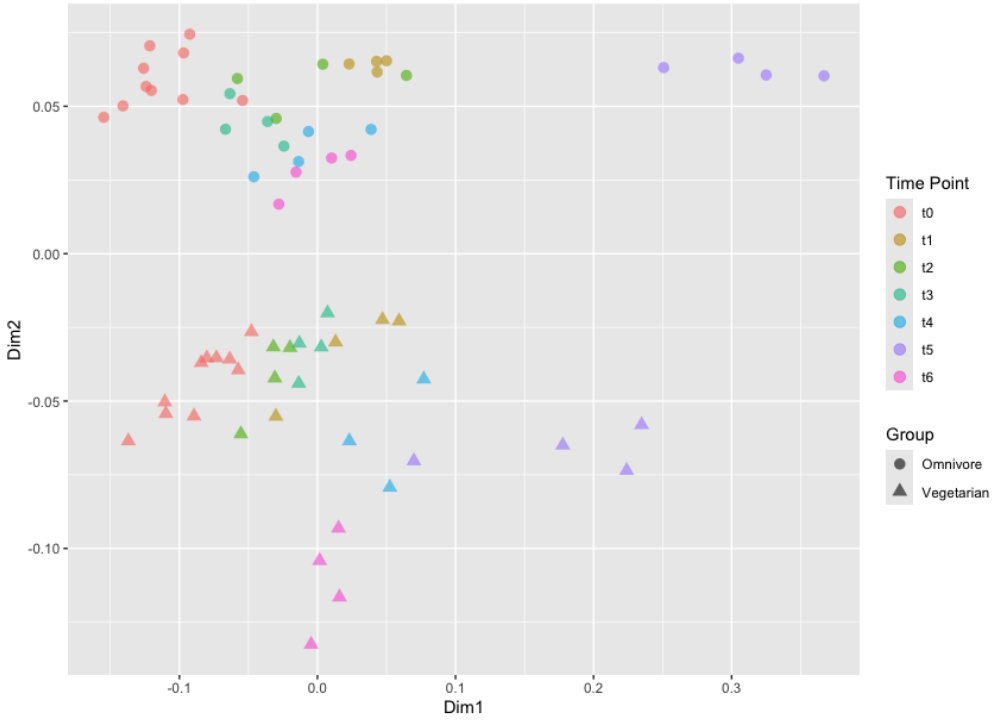


Figure I-9 MDS plot for stability of NMR spectra using 1 - correlation as the dissimilarity matrix. Shapes are the two groups and colors are the time points. Time point 5 for both groups is further away from the other time points, and the homogeneity time point is more similar to itself than the stability time points.

I.5.1. Homogeneity and Stability of Omnivore Samples

In this subsection we will investigate the COD values among the Omnivore samples only. For homogeneity, we have ten samples and so have a 10 by 10 matrix of pairwise COD values as shown in Table 1-5. The diagonal of this matrix is 0, as the COD between any sample and itself is 0.

	B01	B06	B11	B15	B20	B25	B30	B36	B44	B50
B01	0.00000000	0.06327160	0.05657492	0.05222734	0.07515337	0.08049536	0.06466165	0.08179012	0.07001522	0.07035928
B06	0.06327160	0.00000000	0.06811146	0.06074766	0.07198748	0.08346457	0.07317073	0.08176101	0.07275542	0.08472012
B11	0.05657492	0.06811146	0.00000000	0.05701079	0.06211180	0.07943925	0.05471125	0.07476636	0.06000000	0.06927711
B15	0.05222734	0.06074766	0.05701079	0.00000000	0.06376361	0.08411215	0.06221548	0.07644306	0.05864198	0.07088989
B20	0.07515337	0.07198748	0.06211180	0.06376361	0.00000000	0.08044164	0.07022901	0.06962025	0.05468750	0.08761329
B25	0.08049536	0.08346457	0.07943925	0.08411215	0.08044164	0.00000000	0.09893455	0.07507987	0.08709176	0.10454545
B30	0.06466165	0.07317073	0.05471125	0.06221548	0.07022901	0.09893455	0.00000000	0.09146341	0.05927052	0.05105105
B36	0.08179012	0.08176101	0.07476636	0.07644306	0.06962025	0.07507987	0.09146341	0.00000000	0.08242613	0.10574018
B44	0.07001522	0.07275542	0.06000000	0.05864198	0.05468750	0.08709176	0.05927052	0.08242613	0.00000000	0.07669173
B50	0.07035928	0.08472012	0.06927711	0.07088989	0.08761329	0.10454545	0.05105105	0.10574018	0.07669173	0.00000000

Table I-5 10 by 10 matrix of COD values among omnivore homogeneity samples

We can visualize the COD matrix as a point plot as shown in Figure I-10. For each sample, identified by its label along the -axis, its coefficients of disagreement with all other samples are shown along the vertical axis. In the plot, the points are colored by the paired comparison. For example, all red dots are comparisons with the sample on the -axis with the sample from box 1, and so on. Box 25 has among the highest dissimilarity with the other homogeneity samples. Overall, the samples have CODs less than 0.11. We see in Table I-5 that the largest COD value is 0.1057402 between the sample from box 50 and the sample from box 36. We calculate an upper confidence bound for these values using quantCI. As discussed, this will be an approximate confidence limit due to the fact that the CODs are not i.i.d. We estimate the upper 95 % confidence limit as 0.1025.

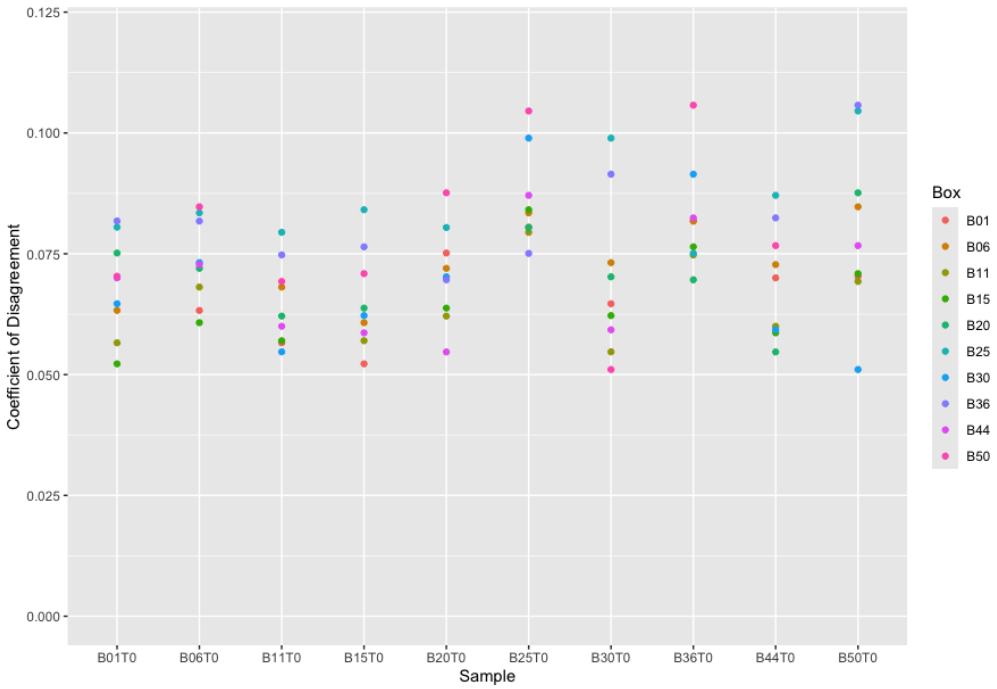


Figure I-10 COD point plot of homogeneity data for Omnivores. Points are colored by the paired box.

For stability, we have 34 samples and have a 34 by 34 matrix of COD values, as shown in Table I-6. We can visualize this again with a point plot. This time, points are colored by time point to elucidate any trend in time. We can clearly see from both Figure I-11 and Table I-6 that time point 5 has the most dissimilarity when compared to all other samples.

Using a percentile bootstrapping procedure, the upper 95 % confidence limit for the COD values for Omnivores (with the time point 5 data included) is 0.2288. Since the data from time point 5 is somewhat irregular, it is possible that there were some experimental variations that caused a bias. If we do not include time point 5, the COD values are much more consistent with one another. While this alone is not sufficient justification for removing time point 5 as the discrepancy in identified peaks for time point 5 could be within the expected variation of results, we will report the upper 95 % confidence limit for the data not including time point 5 for informative purposes.

Without time point 5 included, the approximate 95 % upper confidence limit for the COD values for the Omnivore stability data is 0.1435. This is much lower than the confidence limit when time point 5 is included.

	B01T0	B06T0	B11T0	B15T0	B20T0	B25T0	B30T0	B36T0	B44T0	B50T0	B01T1	B15T1	B30T1	B50T1	B01T2	B15T2	B30T2	B50T2	B01T3	B15T3	B30T3	B50T3	B06T4	B20T4	B25T4	B36T4	B06T5	B20T5	B25T5	B36T5	B06T6	B20T6	B25T6	B36T6
B01T0	0.000	0.075	0.059	0.056	0.075	0.100	0.065	0.067	0.088	0.067	0.116	0.121	0.131	0.110	0.090	0.075	0.128	0.109	0.092	0.101	0.091	0.104	0.109	0.114	0.106	0.104	0.169	0.147	0.212	0.177	0.118	0.110	0.114	0.114
B06T0	0.075	0.000	0.083	0.059	0.078	0.080	0.083	0.076	0.088	0.090	0.121	0.123	0.127	0.128	0.085	0.054	0.115	0.087	0.101	0.114	0.094	0.102	0.121	0.118	0.105	0.108	0.196	0.168	0.250	0.218	0.122	0.134	0.138	0.138
B11T0	0.059	0.083	0.000	0.064	0.064	0.104	0.051	0.072	0.074	0.062	0.134	0.136	0.137	0.105	0.103	0.082	0.129	0.105	0.084	0.097	0.089	0.105	0.113	0.104	0.116	0.102	0.168	0.142	0.210	0.180	0.113	0.103	0.121	0.109
B15T0	0.056	0.059	0.064	0.000	0.059	0.085	0.055	0.057	0.057	0.069	0.124	0.120	0.124	0.112	0.092	0.056	0.125	0.088	0.091	0.095	0.084	0.088	0.113	0.104	0.103	0.097	0.180	0.149	0.225	0.195	0.105	0.098	0.113	0.110
B20T0	0.075	0.078	0.064	0.059	0.000	0.085	0.061	0.070	0.067	0.084	0.137	0.127	0.131	0.124	0.102	0.071	0.129	0.089	0.098	0.099	0.076	0.098	0.123	0.108	0.104	0.112	0.181	0.153	0.232	0.202	0.115	0.113	0.126	0.120
B25T0	0.100	0.080	0.104	0.085	0.085	0.000	0.107	0.084	0.093	0.115	0.140	0.142	0.149	0.155	0.107	0.094	0.144	0.106	0.115	0.116	0.102	0.106	0.129	0.137	0.121	0.133	0.229	0.187	0.282	0.245	0.141	0.147	0.152	0.134
B30T0	0.065	0.083	0.051	0.055	0.061	0.107	0.000	0.063	0.060	0.062	0.140	0.125	0.129	0.111	0.106	0.076	0.135	0.102	0.090	0.097	0.083	0.108	0.124	0.101	0.110	0.102	0.170	0.151	0.208	0.191	0.107	0.100	0.115	0.104
B36T0	0.067	0.076	0.072	0.057	0.070	0.084	0.063	0.000	0.065	0.085	0.135	0.123	0.127	0.114	0.103	0.076	0.136	0.105	0.105	0.103	0.101	0.102	0.116	0.121	0.111	0.102	0.199	0.168	0.238	0.212	0.122	0.120	0.130	0.121
B44T0	0.088	0.088	0.074	0.057	0.067	0.093	0.060	0.065	0.000	0.085	0.146	0.137	0.144	0.128	0.120	0.070	0.139	0.102	0.102	0.103	0.101	0.102	0.130	0.118	0.122	0.111	0.196	0.171	0.231	0.209	0.133	0.117	0.135	0.135
B50T0	0.067	0.090	0.062	0.069	0.084	0.115	0.062	0.085	0.085	0.000	0.113	0.113	0.125	0.101	0.096	0.072	0.125	0.106	0.086	0.093	0.094	0.109	0.106	0.103	0.112	0.107	0.155	0.138	0.200	0.176	0.106	0.096	0.117	0.105
B01T1	0.116	0.121	0.134	0.124	0.137	0.140	0.140	0.135	0.146	0.113	0.000	0.100	0.097	0.100	0.117	0.119	0.133	0.128	0.128	0.126	0.135	0.146	0.125	0.127	0.137	0.137	0.202	0.166	0.245	0.199	0.142	0.140	0.144	0.133
B15T1	0.121	0.123	0.136	0.120	0.127	0.142	0.125	0.123	0.137	0.113	0.100	0.000	0.094	0.087	0.102	0.113	0.109	0.116	0.121	0.116	0.111	0.136	0.141	0.123	0.130	0.127	0.196	0.162	0.233	0.195	0.130	0.122	0.137	0.123
B30T1	0.131	0.127	0.137	0.124	0.131	0.149	0.129	0.127	0.144	0.125	0.097	0.094	0.000	0.100	0.118	0.114	0.119	0.114	0.111	0.129	0.109	0.134	0.119	0.127	0.128	0.120	0.205	0.166	0.240	0.200	0.122	0.115	0.116	0.104
B50T1	0.110	0.128	0.105	0.112	0.124	0.155	0.111	0.114	0.128	0.101	0.100	0.087	0.100	0.000	0.119	0.118	0.129	0.121	0.098	0.105	0.117	0.124	0.118	0.109	0.124	0.129	0.156	0.139	0.213	0.169	0.112	0.108	0.120	0.117
B01T2	0.090	0.085	0.103	0.092	0.102	0.107	0.106	0.103	0.120	0.096	0.117	0.102	0.118	0.119	0.000	0.102	0.094	0.089	0.107	0.114	0.094	0.122	0.121	0.126	0.107	0.116	0.196	0.162	0.252	0.204	0.116	0.134	0.144	0.127
B15T2	0.075	0.054	0.082	0.056	0.071	0.094	0.076	0.076	0.070	0.072	0.119	0.113	0.114	0.118	0.102	0.000	0.117	0.086	0.089	0.102	0.090	0.092	0.120	0.114	0.109	0.104	0.192	0.153	0.229	0.197	0.123	0.110	0.126	0.126
B30T2	0.128	0.115	0.129	0.125	0.129	0.144	0.135	0.136	0.139	0.125	0.133	0.109	0.119	0.129	0.094	0.117	0.000	0.111	0.131	0.135	0.113	0.138	0.148	0.147	0.132	0.126	0.203	0.175	0.246	0.208	0.148	0.152	0.159	0.159
B50T2	0.109	0.087	0.105	0.088	0.089	0.106	0.102	0.105	0.102	0.106	0.128	0.116	0.114	0.121	0.089	0.086	0.111	0.000	0.100	0.101	0.090	0.098	0.109	0.128	0.112	0.101	0.212	0.154	0.254	0.209	0.121	0.127	0.137	0.131
B01T3	0.092	0.101	0.084	0.091	0.098	0.115	0.090	0.105	0.102	0.086	0.128	0.121	0.111	0.098	0.107	0.089	0.131	0.100	0.000	0.063	0.070	0.095	0.088	0.093	0.097	0.095	0.194	0.142	0.229	0.180	0.089	0.090	0.102	0.099
B15T3	0.101	0.114	0.097	0.095	0.099	0.116	0.097	0.103	0.103	0.093	0.126	0.116	0.129	0.105	0.114	0.102	0.135	0.101	0.063	0.000	0.059	0.099	0.101	0.089	0.096	0.093	0.181	0.131	0.224	0.184	0.101	0.102	0.106	0.100
B30T3	0.091	0.094	0.089	0.084	0.076	0.102	0.083	0.101	0.101	0.094	0.135	0.111	0.109	0.117	0.094	0.090	0.113	0.090	0.070	0.059	0.000	0.087	0.107	0.101	0.078	0.082	0.182	0.157	0.233	0.193	0.105	0.109	0.110	0.095
B50T3	0.104	0.102	0.105	0.088	0.098	0.106	0.108	0.102	0.102	0.109	0.146	0.136	0.134	0.124	0.122	0.092	0.138	0.098	0.095	0.099	0.087	0.000	0.126	0.117	0.107	0.107	0.201	0.168	0.255	0.201	0.132	0.131	0.129	0.123
B06T4	0.109	0.121	0.113	0.113	0.123	0.129	0.124	0.116	0.130	0.106	0.125	0.141	0.119	0.118	0.121	0.120	0.148	0.109	0.088	0.101	0.107	0.126	0.000	0.096	0.109	0.109	0.211	0.161	0.240	0.194	0.091	0.101	0.105	0.090
B20T4	0.114	0.118	0.104	0.104	0.108	0.137	0.101	0.121	0.118	0.103	0.127	0.123	0.127	0.109	0.126	0.114	0.147	0.128	0.093	0.089	0.101	0.117	0.096	0.000	0.102	0.105	0.188	0.141	0.226	0.188	0.091	0.083	0.093	0.090
B25T4	0.106	0.105	0.116	0.103	0.104	0.121	0.110	0.111	0.122	0.112	0.137	0.130	0.128	0.124	0.107	0.109	0.132	0.112	0.097	0.096	0.078	0.107	0.109	0.102	0.000	0.089	0.201	0.173	0.244	0.201	0.109	0.127	0.137	0.111
B36T4	0.104	0.108	0.102	0.097	0.112	0.133	0.102	0.102	0.111	0.107	0.137	0.127	0.120	0.129	0.116	0.104	0.126	0.101	0.095	0.093	0.082	0.107	0.109	0.105	0.089	0.000	0.197	0.150	0.237	0.192	0.092	0.110	0.117	0.114
B06T5	0.169	0.196	0.168	0.180	0.181	0.229	0.170	0.199	0.196	0.155	0.202	0.196	0.205	0.156	0.196	0.192	0.203	0.212	0.194	0.181	0.182	0.201	0.211	0.188	0.201	0.197	0.000	0.127	0.141	0.150	0.181	0.190	0.194	0.202
B20T5	0.147	0.168	0.142	0.149	0.153	0.187	0.151	0.168	0.171	0.138	0.166	0.162	0.166	0.139	0.162	0.153	0.175	0.154	0.142	0.131	0.157	0.168	0.161	0.141	0.173	0.150	0.127	0.000	0.161	0.132	0.131	0.143	0.147	0.160
B25T5	0.212	0.250	0.210	0.225	0.232	0.282	0.208	0.238	0.231	0.200	0.245	0.233	0.240	0.213	0.252	0.229	0.246	0.254	0.229	0.224	0.233	0.255	0.240	0.226	0.244	0.237	0.141	0.161	0.000	0.116	0.226	0.224	0.234	0.232
B36T5	0.177	0.218	0.180	0.195	0.202	0.245	0.191	0.212	0.209	0.176	0.199	0.195	0.200	0.169	0.204	0.197	0.208	0.209	0.180	0.184	0.193	0.201	0.194	0.188	0.201	0.192	0.150	0.132	0.116	0.000	0.191	0.192	0.202	0.204
B06T6	0.118	0.122	0.113	0.105	0.115	0.141	0.107	0.122	0.133	0.106	0.142	0.130	0.122	0.112	0.116	0.123	0.148	0.121	0.089	0.101	0.105	0.132	0.091	0.091	0.109	0.092	0.181	0.131	0.226	0.191	0.000	0.081	0.079	0.091
B20T6	0.110	0.134	0.103	0.098	0.113	0.147	0.100	0.120	0.117	0.096	0.140	0.122	0.115	0.108	0.134	0.110	0.152	0.127	0.090	0.102	0.109	0.131	0.101	0.083	0.127	0.110	0.190	0.143	0.224	0.192	0.081	0.000	0.059	0.059
B25T6	0.114	0.138	0.121	0.113	0.126	0.152	0.115	0.130	0.135	0.117	0.144	0.137	0.116	0.120	0.144	0.126	0.159	0.137	0.102	0.106	0.110	0.129	0.105	0.093	0.137	0.117	0.194	0.147	0.234	0.202	0.079	0.059	0.000	0.063
B36T6	0.114	0.138	0.109	0.110	0.120	0.134	0.104	0.121	0.135	0.105	0.133	0.123	0.104	0.117	0.127	0.126	0.159	0.131	0.099	0.100	0.095	0.123	0.090	0.090	0.111	0.114	0.202	0.160	0.232	0.204	0.091	0.059	0.063	0.000

Table I-6 34 by 34 matrix of COD values among omnivore stability samples

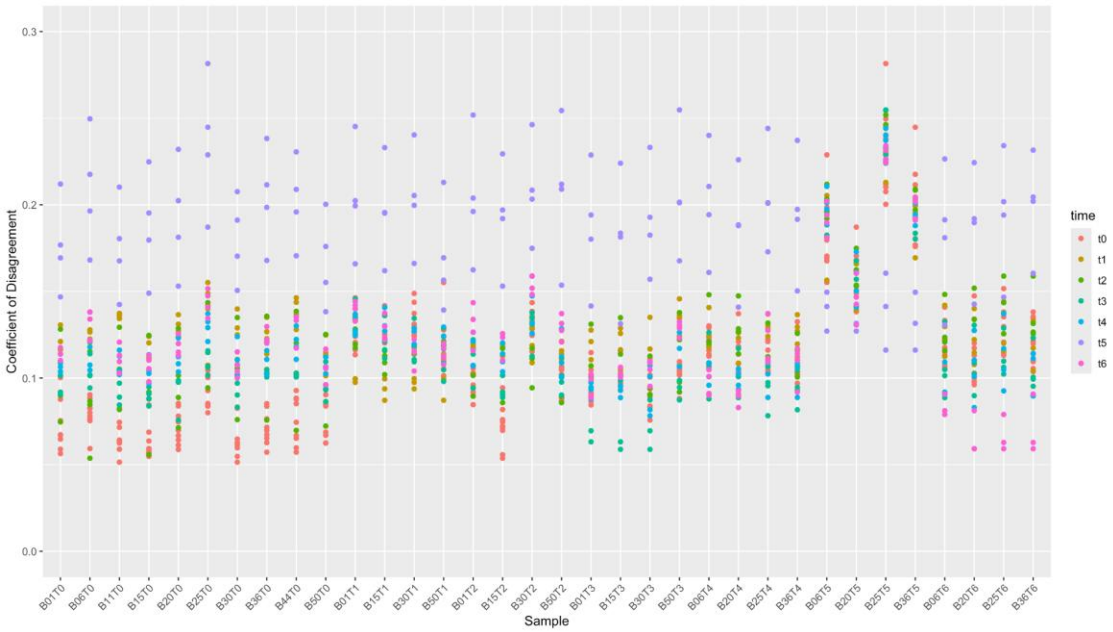


Figure I-11 COD point plot of stability data for Omnivores. Points are colored by time point.

I.5.2. Homogeneity and Stability of Vegetarian Samples

In this subsection we will investigate the COD values among the Vegetarian samples only. For homogeneity, we have ten samples and so have a 10 by 10 matrix of pairwise COD values as shown in Table I-7. The diagonal of this matrix is 0, as the COD between any sample and itself is 0.

	B01	B06	B11	B15	B20	B25	B30	B36	B44	B50
B01	0.00000000	0.05400982	0.06451613	0.06622517	0.06331169	0.07592892	0.07802548	0.06851550	0.07096774	0.06178862
B06	0.05400982	0.00000000	0.04952077	0.06666667	0.04508857	0.04830918	0.05079365	0.05331179	0.04025765	0.04983923
B11	0.06451613	0.04952077	0.00000000	0.06763285	0.03054662	0.04006410	0.03645008	0.05440000	0.03205128	0.04160000
B15	0.06622517	0.06666667	0.06763285	0.00000000	0.04746318	0.06655844	0.06880000	0.06219313	0.06472492	0.05863192
B20	0.06331169	0.04508857	0.03054662	0.04746318	0.00000000	0.04186795	0.04444444	0.04684976	0.03697749	0.03392569
B25	0.07592892	0.04830918	0.04006410	0.06655844	0.04186795	0.00000000	0.04140127	0.04376013	0.03069467	0.04662379
B30	0.07802548	0.05079365	0.03645008	0.06880000	0.04444444	0.04140127	0.00000000	0.05254777	0.03968254	0.04603175
B36	0.06851550	0.05331179	0.05440000	0.06219313	0.04684976	0.04376013	0.05254777	0.00000000	0.02926829	0.03257329
B44	0.07096774	0.04025765	0.03205128	0.06472492	0.03697749	0.03069467	0.03968254	0.02926829	0.00000000	0.03225806
B50	0.06178862	0.04983923	0.04160000	0.05863192	0.03392569	0.04662379	0.04603175	0.03257329	0.03225806	0.00000000

Table I-7 10 by 10 matrix of COD values among vegetarian homogeneity samples

We can visualize the COD matrix as a point plot as shown in Figure I-12. For each sample, identified by its label along the -axis, its coefficients of disagreement with all other samples are shown along the vertical axis. In the plot, the points are colored by the paired comparison. For example, all red dots are comparisons with the sample on the -axis with the sample from box 1, and so on. Box 25 has among the highest dissimilarity with the other homogeneity samples. Overall, the samples have CODs less than 0.08. We find that the approximate upper 95 % confidence limit is 0.0682.

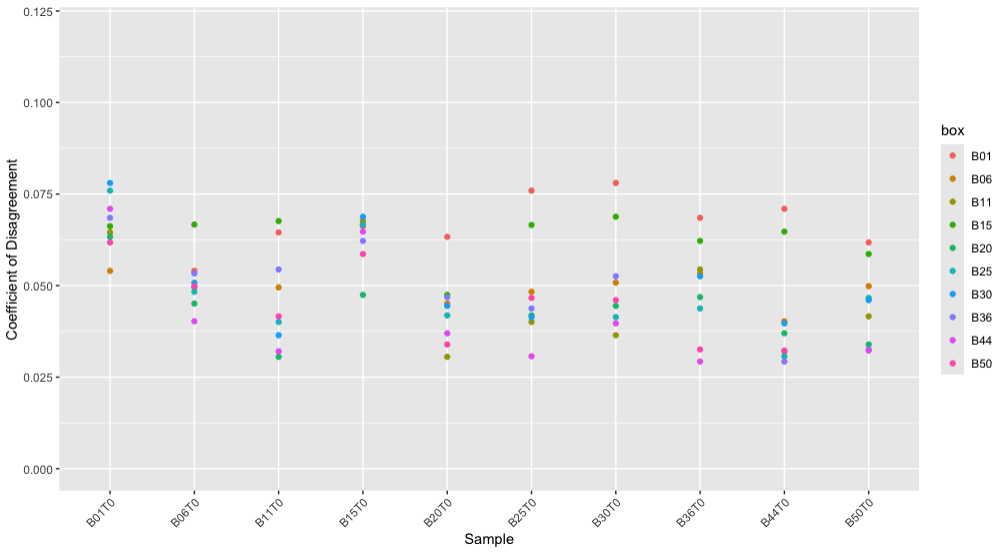


Figure I-12 COD point plot of homogeneity data for Vegetarians. Points are colored by the paired box.

For stability, we have 34 samples and have a 34 by 34 matrix of COD values, as shown in Table I-8. We can visualize this again with a point plot with points again colored by time point to elucidate any trend in time. We can clearly see from both Figure I-13 and Table I-8 that for Vegetarians, the time point 5 does not differ from the other time points in the way that it does for Omnivores for this metric. However, the sample from Box 1 from time point 1 has a larger discrepancy in identified peak presence / absence using the definitions for this analysis described previously.

	B01T0	B06T0	B11T0	B15T0	B20T0	B25T0	B30T0	B36T0	B44T0	B50T0	B01T1	B15T1	B30T1	B50T1	B01T2	B15T2	B30T2	B50T2	B01T3	B15T3	B30T3	B50T3	B06T4	B20T4	B36T4	B06T5	B20T5	B25T5	B36T5	B06T6	B20T6	B25T6	B36T6
B01T0	0.000	0.053	0.060	0.064	0.057	0.065	0.066	0.062	0.065	0.051	0.153	0.095	0.080	0.080	0.053	0.064	0.072	0.064	0.082	0.085	0.104	0.090	0.107	0.086	0.111	0.116	0.118	0.095	0.104	0.128	0.114	0.115	0.113
B06T0	0.053	0.000	0.049	0.047	0.043	0.045	0.049	0.035	0.048	0.049	0.157	0.081	0.082	0.084	0.067	0.062	0.064	0.059	0.074	0.071	0.096	0.079	0.083	0.084	0.100	0.099	0.104	0.084	0.090	0.108	0.091	0.095	0.090
B11T0	0.060	0.049	0.000	0.060	0.038	0.036	0.043	0.052	0.039	0.041	0.174	0.085	0.070	0.102	0.049	0.047	0.065	0.050	0.071	0.074	0.087	0.074	0.099	0.081	0.100	0.084	0.098	0.088	0.078	0.105	0.089	0.092	0.081
B15T0	0.064	0.047	0.060	0.000	0.042	0.056	0.060	0.056	0.059	0.051	0.162	0.098	0.086	0.095	0.062	0.061	0.084	0.076	0.057	0.085	0.095	0.093	0.100	0.092	0.120	0.116	0.118	0.104	0.104	0.131	0.108	0.118	0.113
B20T0	0.057	0.043	0.038	0.042	0.000	0.040	0.050	0.040	0.040	0.029	0.160	0.094	0.074	0.097	0.056	0.051	0.071	0.057	0.062	0.081	0.094	0.089	0.102	0.085	0.104	0.097	0.107	0.091	0.088	0.115	0.086	0.099	0.091
B25T0	0.065	0.045	0.036	0.056	0.040	0.000	0.039	0.045	0.035	0.040	0.176	0.092	0.081	0.098	0.066	0.055	0.076	0.058	0.070	0.064	0.086	0.078	0.098	0.080	0.105	0.095	0.106	0.090	0.086	0.110	0.091	0.103	0.092
B30T0	0.066	0.049	0.043	0.060	0.050	0.039	0.000	0.055	0.033	0.041	0.175	0.081	0.064	0.096	0.061	0.062	0.067	0.053	0.083	0.071	0.098	0.073	0.086	0.072	0.090	0.078	0.085	0.075	0.075	0.095	0.073	0.086	0.072
B36T0	0.062	0.035	0.052	0.056	0.040	0.045	0.055	0.000	0.038	0.037	0.166	0.096	0.085	0.090	0.076	0.059	0.067	0.062	0.074	0.080	0.086	0.079	0.104	0.087	0.103	0.102	0.113	0.090	0.090	0.120	0.100	0.098	0.099
B44T0	0.065	0.048	0.039	0.059	0.040	0.035	0.033	0.038	0.000	0.021	0.170	0.092	0.072	0.092	0.069	0.055	0.066	0.055	0.070	0.064	0.083	0.072	0.091	0.074	0.096	0.092	0.103	0.077	0.089	0.104	0.085	0.091	0.083
B50T0	0.051	0.049	0.041	0.051	0.029	0.040	0.041	0.037	0.021	0.000	0.157	0.094	0.070	0.091	0.062	0.054	0.062	0.054	0.065	0.078	0.084	0.083	0.105	0.076	0.103	0.097	0.107	0.088	0.093	0.114	0.089	0.099	0.094
B01T1	0.153	0.157	0.174	0.162	0.160	0.176	0.175	0.166	0.170	0.157	0.000	0.138	0.147	0.114	0.154	0.163	0.142	0.152	0.162	0.183	0.169	0.178	0.186	0.177	0.193	0.188	0.170	0.166	0.170	0.190	0.184	0.188	0.194
B15T1	0.095	0.081	0.085	0.098	0.094	0.092	0.081	0.096	0.092	0.094	0.138	0.000	0.062	0.070	0.072	0.091	0.066	0.061	0.113	0.110	0.110	0.103	0.110	0.110	0.132	0.113	0.091	0.090	0.092	0.116	0.123	0.118	0.110
B30T1	0.080	0.082	0.070	0.086	0.074	0.081	0.064	0.085	0.072	0.070	0.147	0.062	0.000	0.071	0.060	0.071	0.064	0.052	0.092	0.098	0.107	0.091	0.110	0.092	0.105	0.089	0.082	0.093	0.080	0.116	0.109	0.115	0.101
B50T1	0.080	0.084	0.102	0.095	0.097	0.098	0.096	0.090	0.092	0.091	0.114	0.070	0.071	0.000	0.081	0.088	0.065	0.070	0.103	0.097	0.110	0.111	0.122	0.107	0.132	0.119	0.124	0.102	0.104	0.125	0.127	0.121	0.128
B01T2	0.053	0.067	0.049	0.062	0.056	0.066	0.061	0.076	0.069	0.062	0.154	0.072	0.060	0.081	0.000	0.062	0.061	0.056	0.080	0.086	0.108	0.088	0.101	0.087	0.115	0.099	0.107	0.090	0.084	0.111	0.103	0.107	0.099
B15T2	0.064	0.062	0.047	0.061	0.051	0.055	0.062	0.059	0.055	0.054	0.163	0.091	0.071	0.088	0.062	0.000	0.065	0.066	0.059	0.081	0.075	0.071	0.102	0.082	0.107	0.097	0.110	0.097	0.085	0.123	0.101	0.105	0.094
B30T2	0.072	0.064	0.065	0.084	0.071	0.076	0.067	0.067	0.066	0.062	0.142	0.066	0.064	0.065	0.061	0.065	0.000	0.050	0.099	0.096	0.099	0.095	0.111	0.111	0.115	0.111	0.113	0.094	0.102	0.111	0.107	0.095	0.102
B50T2	0.064	0.059	0.050	0.076	0.057	0.058	0.053	0.062	0.055	0.054	0.152	0.061	0.052	0.070	0.056	0.066	0.050	0.000	0.087	0.081	0.097	0.077	0.094	0.091	0.107	0.100	0.099	0.067	0.082	0.103	0.093	0.096	0.088
B01T3	0.082	0.074	0.071	0.057	0.062	0.070	0.083	0.074	0.070	0.065	0.162	0.113	0.092	0.103	0.080	0.059	0.099	0.087	0.000	0.078	0.066	0.080	0.097	0.082	0.107	0.115	0.111	0.104	0.109	0.127	0.111	0.114	0.109
B15T3	0.085	0.071	0.074	0.085	0.081	0.064	0.071	0.080	0.064	0.078	0.183	0.110	0.098	0.097	0.086	0.081	0.096	0.081	0.078	0.000	0.081	0.052	0.069	0.066	0.076	0.106	0.111	0.073	0.100	0.103	0.093	0.090	0.082
B30T3	0.104	0.096	0.087	0.095	0.094	0.086	0.098	0.086	0.083	0.084	0.169	0.110	0.107	0.110	0.108	0.075	0.099	0.097	0.066	0.081	0.000	0.074	0.100	0.088	0.111	0.107	0.117	0.095	0.103	0.122	0.114	0.106	0.110
B50T3	0.090	0.079	0.074	0.093	0.089	0.078	0.073	0.079	0.072	0.083	0.178	0.103	0.091	0.111	0.088	0.071	0.095	0.077	0.080	0.052	0.074	0.000	0.071	0.066	0.069	0.096	0.095	0.072	0.087	0.111	0.089	0.095	0.081
B06T4	0.107	0.083	0.099	0.100	0.102	0.098	0.086	0.104	0.091	0.105	0.186	0.110	0.110	0.122	0.101	0.102	0.111	0.094	0.097	0.069	0.100	0.071	0.000	0.073	0.080	0.116	0.105	0.067	0.100	0.103	0.093	0.093	0.088
B20T4	0.086	0.084	0.081	0.092	0.085	0.080	0.072	0.087	0.074	0.076	0.177	0.110	0.092	0.107	0.087	0.082	0.111	0.091	0.082	0.066	0.088	0.066	0.073	0.000	0.074	0.086	0.090	0.065	0.085	0.107	0.087	0.091	0.098
B36T4	0.111	0.100	0.100	0.120	0.104	0.105	0.090	0.103	0.096	0.103	0.193	0.132	0.105	0.132	0.115	0.107	0.115	0.107	0.107	0.076	0.111	0.069	0.080	0.074	0.000	0.105	0.091	0.084	0.095	0.126	0.103	0.107	0.105
B06T5	0.116	0.099	0.084	0.116	0.097	0.095	0.078	0.102	0.092	0.097	0.188	0.113	0.089	0.119	0.099	0.097	0.111	0.100	0.115	0.106	0.107	0.096	0.116	0.086	0.105	0.000	0.072	0.098	0.058	0.128	0.108	0.118	0.110
B20T5	0.118	0.104	0.098	0.118	0.107	0.106	0.085	0.113	0.103	0.107	0.170	0.091	0.082	0.124	0.107	0.110	0.113	0.099	0.111	0.111	0.117	0.095	0.105	0.090	0.091	0.072	0.000	0.097	0.069	0.121	0.107	0.114	0.109
B25T5	0.095	0.084	0.088	0.104	0.091	0.090	0.075	0.090	0.077	0.088	0.166	0.090	0.093	0.102	0.090	0.097	0.094	0.067	0.104	0.073	0.095	0.072	0.067	0.065	0.084	0.098	0.097	0.000	0.083	0.083	0.085	0.076	0.089
B36T5	0.104	0.090	0.078	0.104	0.088	0.086	0.075	0.090	0.089	0.093	0.170	0.092	0.080	0.104	0.084	0.085	0.102	0.082	0.109	0.100	0.103	0.087	0.100	0.085	0.095	0.058	0.069	0.083	0.000	0.124	0.090	0.109	0.095
B06T6	0.128	0.108	0.105	0.131	0.115	0.110	0.095	0.120	0.104	0.114	0.190	0.116	0.116	0.125	0.111	0.123	0.111	0.103	0.127	0.103	0.122	0.111	0.103	0.107	0.126	0.128	0.121	0.083	0.124	0.000	0.068	0.062	0.057
B20T6	0.114	0.091	0.089	0.108	0.086	0.091	0.073	0.100	0.085	0.089	0.184	0.123	0.109	0.127	0.103	0.101	0.107	0.093	0.111	0.093	0.114	0.089	0.093	0.087	0.103	0.108	0.107	0.085	0.090	0.068	0.000	0.052	0.044
B25T6	0.115	0.095	0.092	0.118	0.099	0.103	0.086	0.098	0.091	0.099	0.188	0.118	0.115	0.121	0.107	0.105	0.095	0.096	0.114	0.090	0.106	0.095	0.093	0.091	0.107	0.118	0.114	0.076	0.109	0.062	0.052	0.000	0.047
B36T6	0.113	0.090	0.081	0.113	0.091	0.092	0.072	0.099	0.083	0.094	0.194	0.110	0.101	0.128	0.099	0.094	0.102	0.088	0.109	0.082	0.110	0.081	0.088	0.098	0.105	0.110	0.109	0.089	0.095	0.057	0.044	0.047	0.000

Table I-8 34 by 34 matrix of COD values among vegetarian stability samples

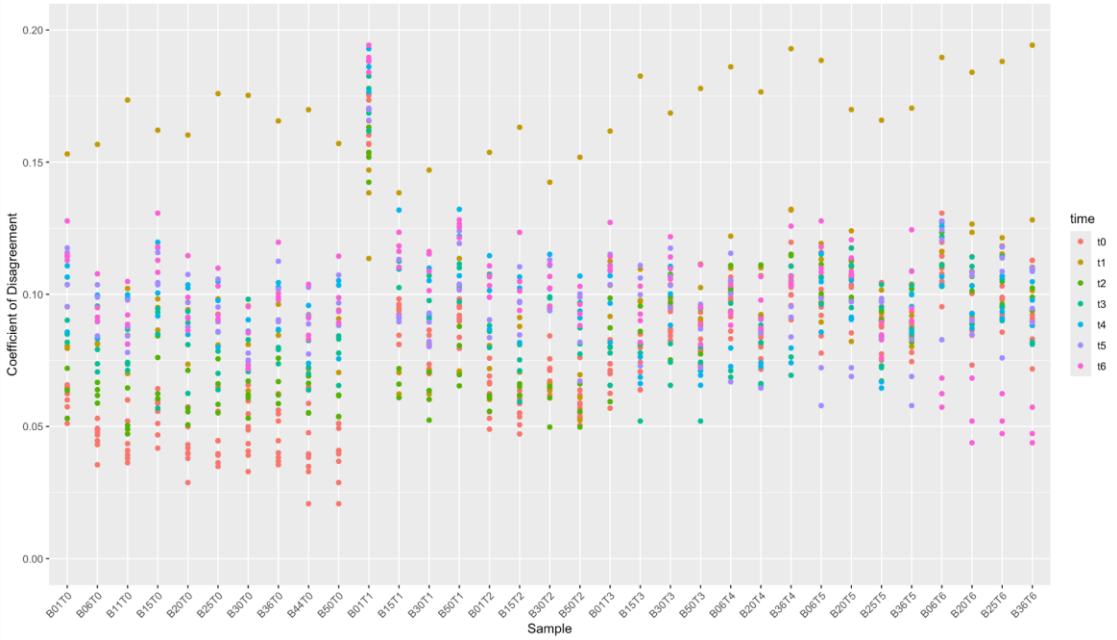


Figure I-13 COD point plot of stability data for Vegetarians. Points are colored by time point.

Using quantCI, the approximate upper 95 % confidence limit for the COD values for Vegetarians (with the time point 5 data included) is 0.1632. Since the data from time point 5 is not irregular for Vegetarians as it was for Omnivores, there will not be as distinct of a difference if time point 5 is removed. If we were to remove time point 5, we can see in Figure I-13 that the range of COD values would be unchanged if time point 5 were removed. In fact, with time point 5 removed for Vegetarians the upper confidence limit may be larger due to the fact that there are fewer COD values and those with the largest discrepancies are still included.

I.6. Metagenomics

The data for the homogeneity analysis is made up of measurements taken from 10 omnivore and 10 vegetarian boxes. For each box, 1 vial was taken and separated into 4 aliquots as replicate measurements. Two of the vegetarian samples (box 20, replicate 3 and box 36, replicate 3) had an experimental error that resulted in a lack of data with less than 2000 reads, so the samples were not considered. As a result, there are 78 (38 vegetarian and 40 omnivore) rather than 80 total samples. Water and ZymoBIOMICS Fecal Reference with TruMatrix™ Technology were included as negative and positive controls, respectively. Additionally, the samples were spiked-in with several known bacteria - Deinococcus, Delftia, and Dickeya (Brenneria). All replicates were thawed the same length of time and were run together to minimize confounding errors. These controls were included to help resolve if any observed aberrations were the result experimental error; to date, additional analysis has not been conducted beyond that.

The resulting homogeneity data consists of a total of 1066 unique identified DNA sequences and their relative abundances for each of the 78 measurements. We have compositional data for each sample, with the sum of the relative abundances being 1. In the case where a particular sequence is not identified in a sample, it is assigned a relative abundance of 0. However, this does not automatically imply that the sequence is not present. Rather, it indicates that the sequence was not identified up to the machine measurement precision.

Day	Omnivore Box	Vegetarian Box
10	1, 25	1, 25
16	20, 50	20, 50
31	6, 30	6, 30
59	15, 44	15, 44
87	1, 6, 11, 25, 30, 36	11, 36
112	20, 50	1, 25
140	15, 44	20, 50
168	25, 36	11, 36

Table I-9 Time points and Corresponding Boxes for Omnivores and Vegetarians Metagenomic Analysis

As we can see in Table I-9, the stability data has a slightly more complex structure. While all 10 boxes were measured for the homogeneity data (considered time point 0), the additional measurements taken on days 10, 16, 31, 59, 87, 112, 140, and 168 typically included only 2 boxes per time point. For omnivores, boxes 1, 6, 11, and 25 had repeat measures at the fifth time point. Since the 4 replicate vials at each box can be considered repeated measurements for the stability assessment, we can choose to aggregate to the box level. There are a total of 2353 unique sequences across all the time points (including time point 0), and a total of 30 omnivore samples and 26 vegetarian samples.

Omnivore and vegetarian sequences were analyzed separately, and only the sequences with a relative abundance above 0.0008 for each group were included. This results is the top 122 sequences in omnivores and the top 100 in vegetarians. As the relative abundance of the DNA sequence decreases, so does the reliability of the measurement. For both omnivores and vegetarians, the sequences with relative abundance greater than 0.0008 constitute over 98 % of the relative abundance. The relative abundances in omnivore and vegetarian samples are different from one another - there may be sequences found in the omnivore samples that are not found in the vegetarian samples and vice versa.

I.6.1. Homogeneity: Bayesian Modeling Approach

We have an unbalanced design, for which some traditional likelihood approaches are computationally difficult or intractable. There are some likelihood methods such as maximum likelihood and its restricted counterpart that are more computationally efficient. However, Bayesian models more naturally incorporate multiple sources of uncertainty including uncertainty in model parameters and the underlying data generation process, and give the ability to consider distributional results rather than point estimates.

We consider a hierarchical Bayesian modeling approach with priors that are generally noncommittal. However, it should be kept in mind that even weakly informative priors can be influential, especially if there is a small number of measured values. The distributions selected for the Bayesian model are the following for each group:

1. μ has a prior half-Guassian distribution with mean 0 and a small standard deviation (10^{-5}). This proper prior is more informative and includes specific information regarding the measurand that we must adhere to - relative abundances are greater than 0 and generally have small values.

2. σ_A and σ_E have half Student's t-distributions with 4 degrees of freedom. This distribution is weakly informative and has been suggested as an alternative to a half-Cauchy.

3. $\theta_i|\sigma_A \sim \mathrm{N}(0, \sigma_A)$.

4. $y_{ij}|\mu, \sigma_E, \theta_i \sim \mathrm{N}(\mu + \theta_i, \sigma_E)$.

We have a model that follows from a standard ANOVA table:

$$y_{ij} = \mu + \theta_i + \epsilon_{ij}.$$

We can define each identified 16S sequence to be "sufficiently" homogeneous if the biological variability (σ_A^2) is less than or equal to the machine variability (σ_E^2), or equivalently, if $\sigma_A^2 / \sigma_E^2 \leq 1$. Using the model described above, we can find the 95 % credible interval for this ratio. For the top 300 most abundant sequences in omnivores and vegetarians, we can see these results in Figure I-14. We can also use this model to get credible intervals for the coefficient of variation. We can see from Figure I-15 that as the sequences become less abundant, the width of the credible intervals increases and the overall intervals are above or include 1. This indicates, unsurprisingly, that as the sequences become more rare, the measurement becomes more unreliable. This helps inform a cut-off of 0.0008 for relative abundance in the COD analyses which results in the top 122 sequences in omnivores and the top 100 sequences in vegetarians.

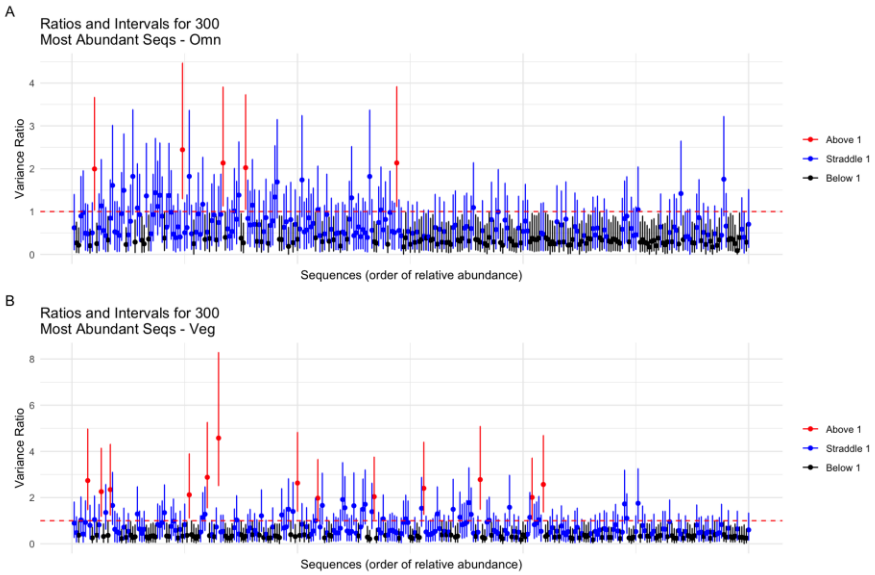


Figure I-14 Credible intervals for top 300 most abundant sequences in descending order of abundance for omnivores (panel A) and vegetarians (panel B). Red lines indicate an interval fully above 1, a blue line indicates that the interval straddles 1, and a black line indicates that the interval is fully below 1.

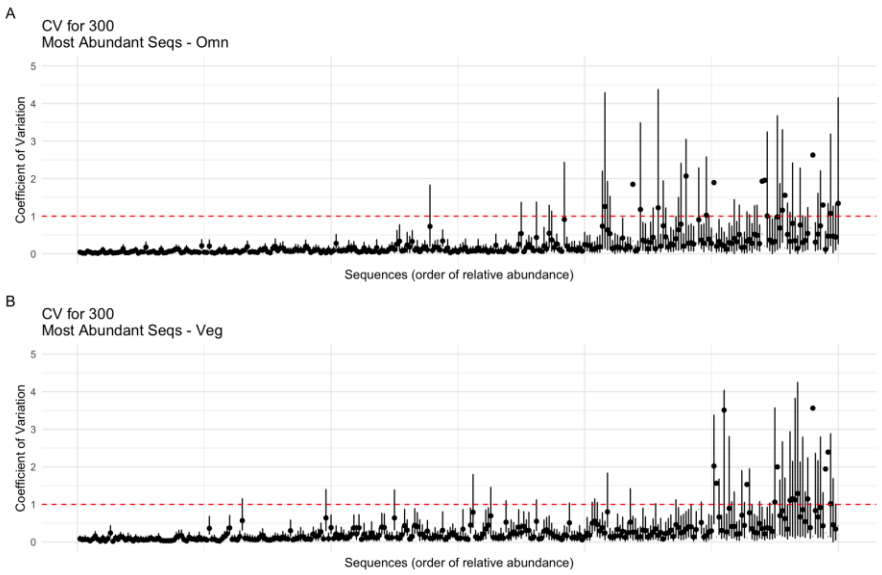


Figure I-15 Credible intervals for top 300 most abundant sequences in descending order of abundance for omnivores (panel A) and vegetarians (panel B). The red dotted line indicates a coefficient of variation of 1 or 100 %.

I.6.2. Homogeneity and Stability of Omnivore Samples

When considering the omnivore material for homogeneity, we have 40 samples (replicates) in all and we calculate the corresponding COD for each pair of samples. We use the relative abundance threshold of 0.0008 in this analysis, which results in 122 unique sequences to be compared for the omnivore samples. This results in a 40 by 40 symmetric matrix which has diagonal elements 0 because the disagreement between a sample and itself is 0. For the pairwise omnivore samples, before taking the 99th percentile, 0.3 % of RADs were infinite.

In Figure I-16 we see the 40 CODs for homogeneity visualized in panel A. Each sample is identified by its label on the -axis (vial 1, replicate 1 is labeled “V1R1” and so on until vial 10, replicate 4 as “V10R4”). Along the -axis for each sample are the COD for each of its pairwise comparisons. So for “V1R1,” for example, there are 39 points along the vertical axis corresponding to a COD for each of the 39 other samples. The points are colored by vial that each pairwise comparison is made from (so a red circle for “V1R1” signifies a comparison between one of the other 3 replicates from vial 1, while an orange triangle would signify a comparison between one of the replicates from vial 2).

Figure I-16 COD plots of homogeneity for omnivore (panel A) and vegetarians (panel B).

Using the quantCI function available in the QuantileNPCI package in R, we can find an approximate 95 % confidence interval for the COD values. Using this method, we get an approximate upper confidence bound of 0.0727. As this non-parametric approach assumes independence of values, and we know the COD values are correlated, this is truly only an approximate confidence interval.

We can assess the level of stability of the material using the same COD procedure. We aggregate the replicates of the vials at each time point, since these can be considered as repeated measurements. Then, each sample is an aggregate mean of a vial at a time point. We include all measurements, including those made at time point 0 (or day 1). For the Omnivore stability samples, 0.2 % of RADs were infinite and were not included in the COD. When we calculate the CODs for each of the pairs, we get a 30 by 30 matrix with diagonal elements of 0.

In Figure I-17 we have visualized the CODs for the disagreement matrix described above in panel A. The -axis is labeled as follows: the COD values for the measurements from time point 0 (homogeneity time point), vial 1 are labeled “T0V1” until the measurements from time point 8, vial 2 values which are labeled “T8V2.” Note again that every vial is not measured at every time point - vial 1 at time point 0 is not vial 1 from time point 1, and so on. Each sample along the - axis has 29 points corresponding to the comparisons with the other samples. The points are colored by time point, so that a red point for time point 0 signifies a comparison to another box measured on time point 0, while an orange triangle signifies a comparison to the vials measured at time point 1. We can see that the time point 7 and 8 measurements are similar to

each other, but most dissimilar from the other time points. Using quantCI, we estimate an approximate upper 95 % confidence limit of 0.1140.

Figure I-17 COD plot for stability of omnivore samples (panel A) and vegetarian samples (panel B).

I.6.3. Homogeneity and Stability of Vegetarian Samples

For the homogeneity assessment, we have two removed samples for the Vegetarian material due to low read counts, and have 38 samples (replicates) in all. This results in a 38 by 38 matrix with diagonal elements 0. For the pairwise Vegetarian samples, before taking the 99th percentile, 0.1 % of RADs were infinite. In Figure I-16 we see the 38 CODs visualized for homogeneity in panel B. The plot is labeled and colored similarly to the plot for the Omnivore samples, with the exception that since we have 38 CODs, each sample on the -axis has 37 points that correspond to the 37 other samples. The approximate upper limit for a confidence interval using quantCI is 0.1057.

For the Vegetarian stability samples, 0.3 % of the RADs were infinite and were not included in the COD. There are 26 stability samples for Vegetarians in all, which results in a 26 by 26 matrix. In panel B of Figure I-17 we visualize these CODs. The approximate 95 % upper confidence bound using quantCI is 0.1504. Generally, the Vegetarian material displayed more variability than the Omnivore material, which could be attributed to several factors. The difference in diet means that the material may inherently have a higher level of diversity, or it can cause differences in sample preparation.

I.7. Metabolomics: LC-MS

The data analyzed are contained in 8 Excel files with the final lists for polar and non-polar compounds for each diet material (Omnivore and Vegetarian) corresponding to the 8 measurement scenarios shown in Table I-10. Ten vials from each of the omnivore and vegetarian donor pools (20 vials total) were characterized for metabolites and assessed for homogeneity.

Measurement Scenarios

Experiment
Non-Polar_Homogeneity_Omni_NEG_Lumos
Non-Polar_Homogeneity_Omni_POS_LumosFSC1
Non-Polar_Homogeneity_Veg_NEG_Lumos
Non-Polar_Homogeneity_Veg_POS_Lumos
Polar_Homogeneity_Omni_NEG_Lumos
Polar_Homogeneity_Omni_POS_LumosFSC1
Polar_Homogeneity_Veg_NEG_Lumos
Polar_Homogeneity_Veg_POS_Lumos

Table I-10 Measurement Scenarios for LC-MS Analysis

For annotation of polar metabolites, samples were run on two different instruments. For annotation of non-polar metabolites, samples were run on a single intrument. Resulting raw files were processed and searched with Compound Discoverer (CD, version 3.3.2) using both the local NIST23 library and the online database mzCloud (https://www.mzcloud.org/). Note that we only have homogeneity results for the LC-MS application. The measurement protocol chosen does not lend itself easily to stability checks.

I.7.4. Homogeneity Results for All Samples

Table I-11 gives the uncertainty intervals for the Coefficients of Disagreement for the 8 measurement scenarios.

Measurement Scenarios

Experiment	COD.Lower.Limit	COD.Upper.Limit	Confidence
Non-Polar_Homogeneity_Omni_NEG_Lumos	0.087	0.094	95%
Non-Polar_Homogeneity_Omni_POS_LumosFSC1	0.117	0.129	95%
Non-Polar_Homogeneity_Veg_NEG_Lumos	0.112	0.149	95%
Non-Polar_Homogeneity_Veg_POS_Lumos	0.141	0.151	95%
Polar_Homogeneity_Omni_NEG_Lumos	0.113	0.123	95%
Polar_Homogeneity_Omni_POS_LumosFSC1	0.143	0.151	95%
Polar_Homogeneity_Veg_NEG_Lumos	0.130	0.161	95%
Polar_Homogeneity_Veg_POS_Lumos	0.091	0.119	95%

Table I-11 Uncertainty intervals for the Coefficients of Disagreement for the 8 measurement scenarios

Consider the Non-Polar_Homogeneity_Omni_NEG_Lumos measurements. The results can be interpreted as follows. Here the COD lower limit is 0.087 (8.7 %) and the upper limit is 20.094 (9.4 %). We can say, with 95 % confidence, the following: Select any two samples from the population of vials and compare the chromatogram response for any feature. In 90 % of the comparisons the discrepancy will be between 8.7 % and 9.4 %.

I.8. Flow Cytometry

Fifteen (15) vials were selected from 9 different boxes as shown in the following table. In addition, a pooled sample from 8 boxes (B01, B06, B11, B15, B25, B30, B36, B44) were prepared from which 6 aliquots were made and analyzed on 6 different dates as shown in the Table I-12.

	B01	B06	B11	B15	B20	B25	B30	B36	B44	BPool
20230802	3	3	3	3	0	3	3	3	3	1
20230831	3	0	0	0	0	3	0	0	0	1
20230928	0	3	0	0	0	0	3	0	0	1
20231130	0	0	3	0	0	0	0	0	0	1
20231221	0	0	0	3	0	0	0	0	0	1
20240129	0	0	0	0	3	0	0	0	0	1

Table I-12 Box Numbers corresponding to the omnivore vials measured by Flow cytometry on different dates. Three aliquots were measured on each vial except for the pooled sample (Bpool).

Unpooled samples from 2023/08/02 for omnivore and from 2023/08/04 for vegetarian are referred to as “homogeneity samples”. For each sample analyzed, the following measurements were made: (1) Normalized counts and (2) Intensity measurements for the following five channels: FSC1, SSC1, FL11, FL20, FL25. For each measurement, degree of homogeneity and stability were examined. We now report the types of analyses conducted and the results.

I.8.1. Homogeneity analysis based on normalized cell counts

A hierarchical Gaussian model was used to account for variability across vials and variability among technical replicates (aliquots) within each vial. The following model was assumed.

$$Y_{ij} = \mu + v_i + r_{ij}$$

Where

μ = mean normalized count for the population of vials.

v_i = deviation of vial normalized count from μ

r_{ij} = deviation of the count for aliquot j of vial i from the normalized count for vial i

v_i, r_{ij} are all assumed to be mutually independent and normally distributed with $v_i \sim N(0, \sigma_B^2)$ and $r_{ij} \sim N(0, \sigma_R^2)$.

We calculate the posterior predictive distribution for the normalized count of a randomly selected vial from the population of vials.

Normalized cell counts for the 8 omnivore vials (each from a different box) are shown in Figure I-18.

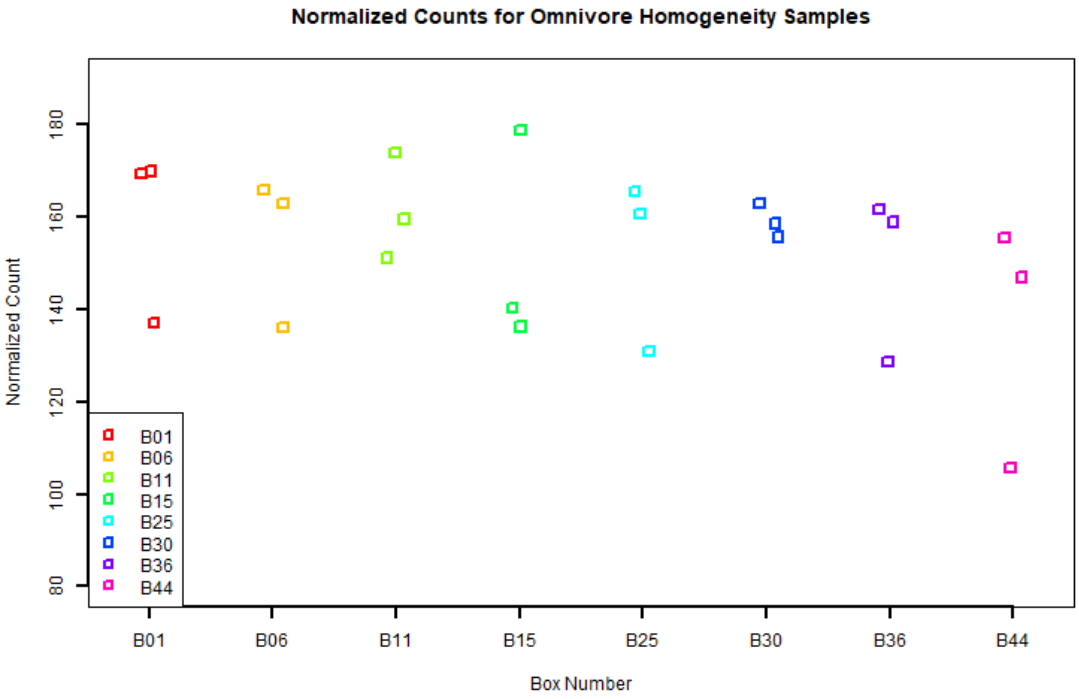


Figure I-18 Normalized cell counts for omnivore homogeneity samples (pooled samples not included).

A 95 % two-sided posterior predictive interval for the normalized count value of a randomly chosen omnivore vial from the population is (136.54, 171.55).

Normalized cell counts for the 8 vegetarian vials (each from a different box) are shown in Figure I-19.

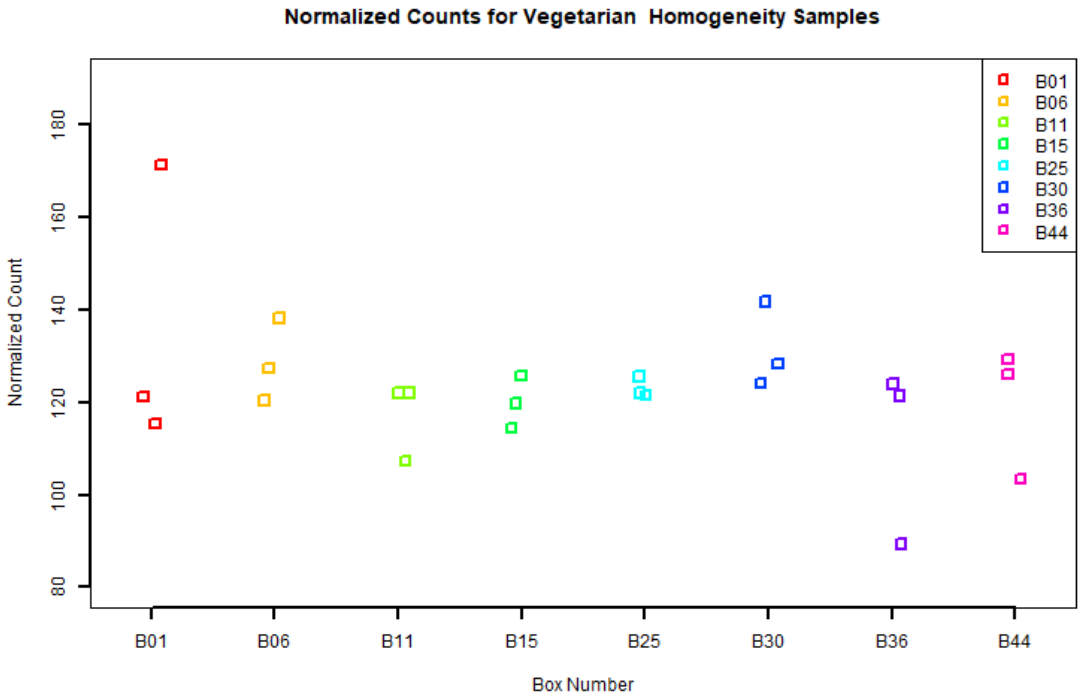


Figure I-19 Normalized cell counts for vegetarian homogeneity samples (pooled samples not included).

A 95 % two-sided posterior predictive interval for the normalized count value of a randomly chosen vegetarian vial from the population is (105.3775, 141.4082).

Stability analysis for normalized counts was conducted using only the pooled samples measured at 6 different time points. For omnivore samples, measurements were conducted on day 0 (2023/08/02), day 29, day 57, day 118, day 139, and day 178. For vegetarian samples, measurements were conducted on day 0 (2023/08/04), day 27, day 55, day 116, day 137, and day 176. Plots of normalized cell counts versus days elapsed are shown in Figure I-20 and Figure I-21, for omnivore pooled samples and vegetarian pooled samples respectively. The R-package zyp was used to compute 90 % confidence intervals for the slope of the trend using the function zyp.sen. For omnivores, a 90 % confidence interval for the Sen slope is (-0.179, 0.090). For vegetarians, a 90 % confidence interval for the Sen slope is (-0.722, 0.480). The confidence intervals include 0 for the omnivore samples as well as the vegetarian samples implying that there is no statistically significant trend in either case (in both cases).

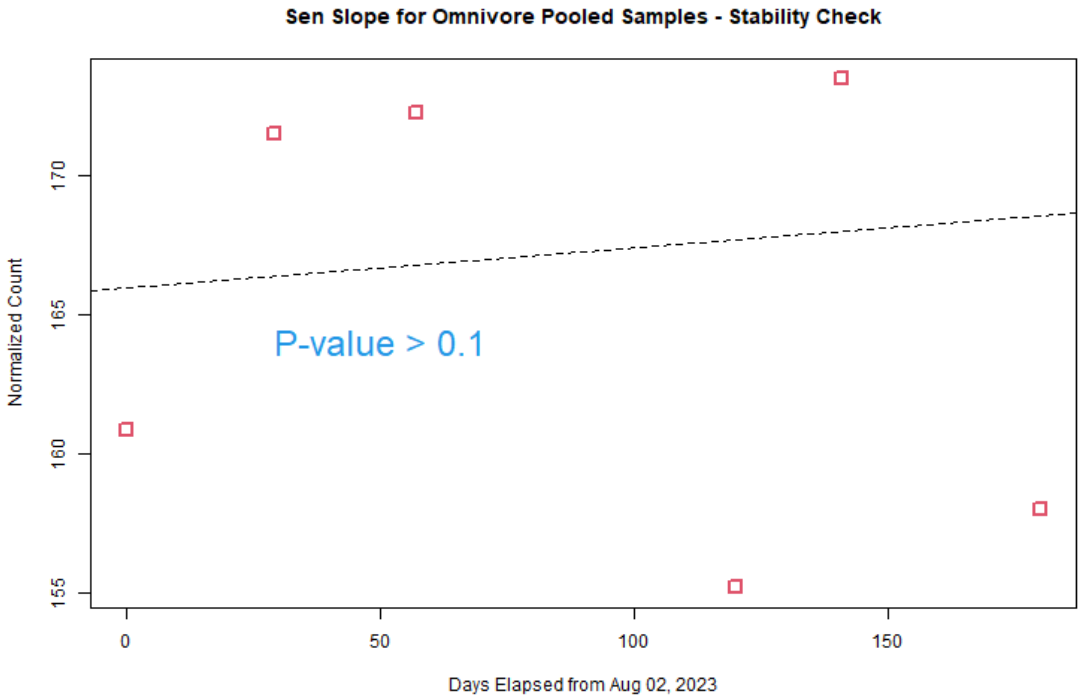


Figure I-20 Normalized cell counts for Omnivore Pooled samples as a function of days elapsed.

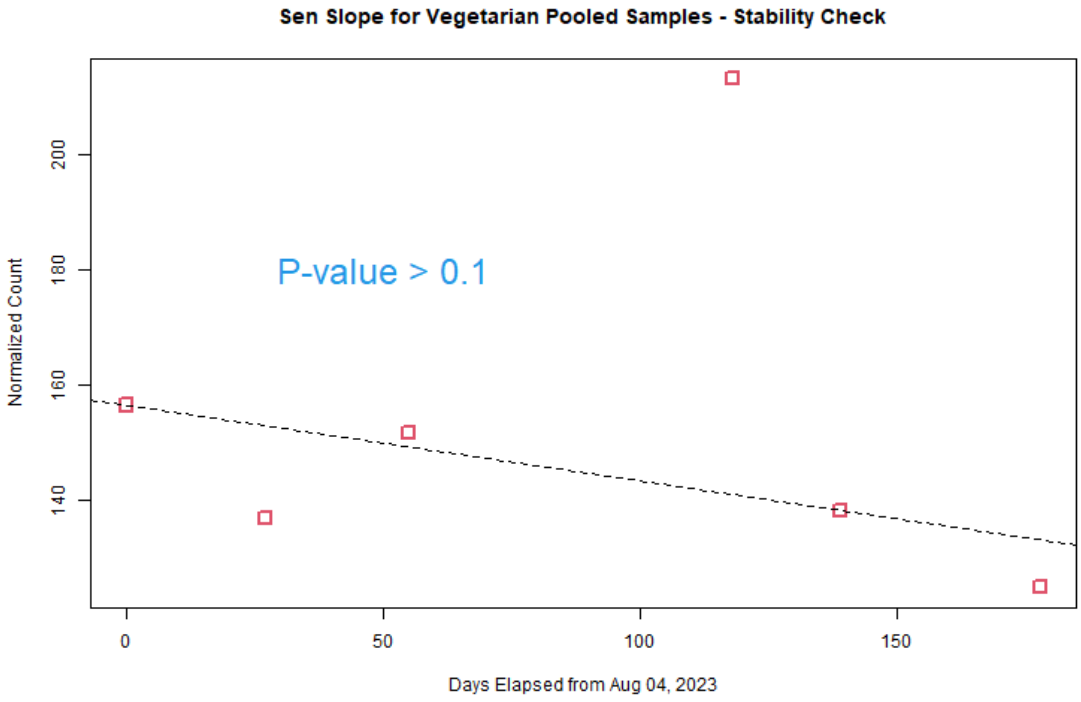


Figure I-21 Normalized cell counts for Vegetarian Pooled samples as a function of days elapsed.

I.8.2. Homogeneity and Stability based on Fluorescence Intensity Data

For each of the 51 omnivore samples (including pooled samples) and each of the vegetarian samples (including pooled samples) fluorescence intensity distributions from 5 channels (FSC1, SSC1, FL11, FL20, FL25) were examined. Homogeneity assessment was made using the 24 omnivore samples from 08/02/2023 and the 24 vegetarian samples from 04/04/2023. Kernel density estimates of log10(intensity) were examined for each sample and each channel was analyzed to assess homogeneity and stability of the SRM 8084 samples. Omnivore samples and Vegetarian samples were analyzed separately.

I.8.2.1. Homogeneity: Omnivore Samples

Figure I-22 through Figure I-26 show the kernel densities for log10(intensity) for the 5 channels (Figure I-22 Channel FSC1, range 7.1 and 7.6; I-23 Channel SSC1, range 6.4 and 6.8; I-24 Channel FL11, range 8.2 and 8.7; I-25 Channel FL20, range 4.4-4.7; I-26 Channel FL25, range 7.4-7.6). The density function from R was used with default bandwidth. The R script is provided in the appendix. These density estimates show certain modes (peaks), and the peak positions are expected to be at the same intensity value for all homogeneity samples. In some channels there may be two modes (peaks) of interest. We first find the peak positions for the log10(intensity) densities for each channel and then evaluate the variability in the peak positions between vials.

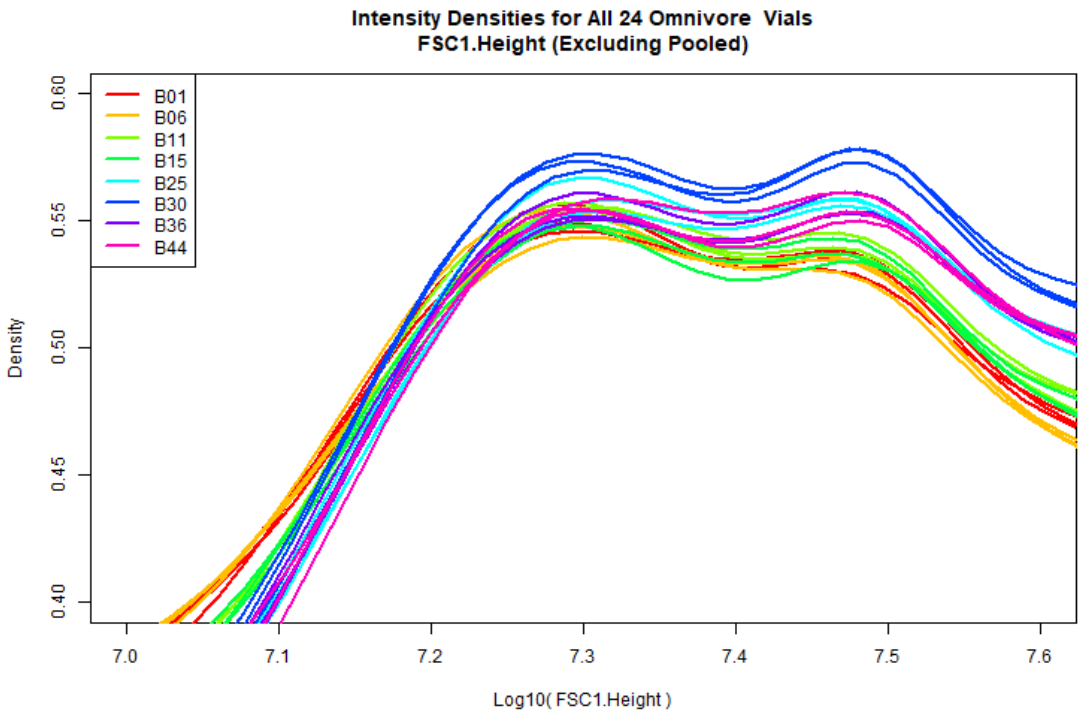


Figure I-22 Kernel densities for Channel FSC1.Height log10(Intensities). The density shows two modes in the interval for log10(Intensity) between 7.1 and 7.6. These modes are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

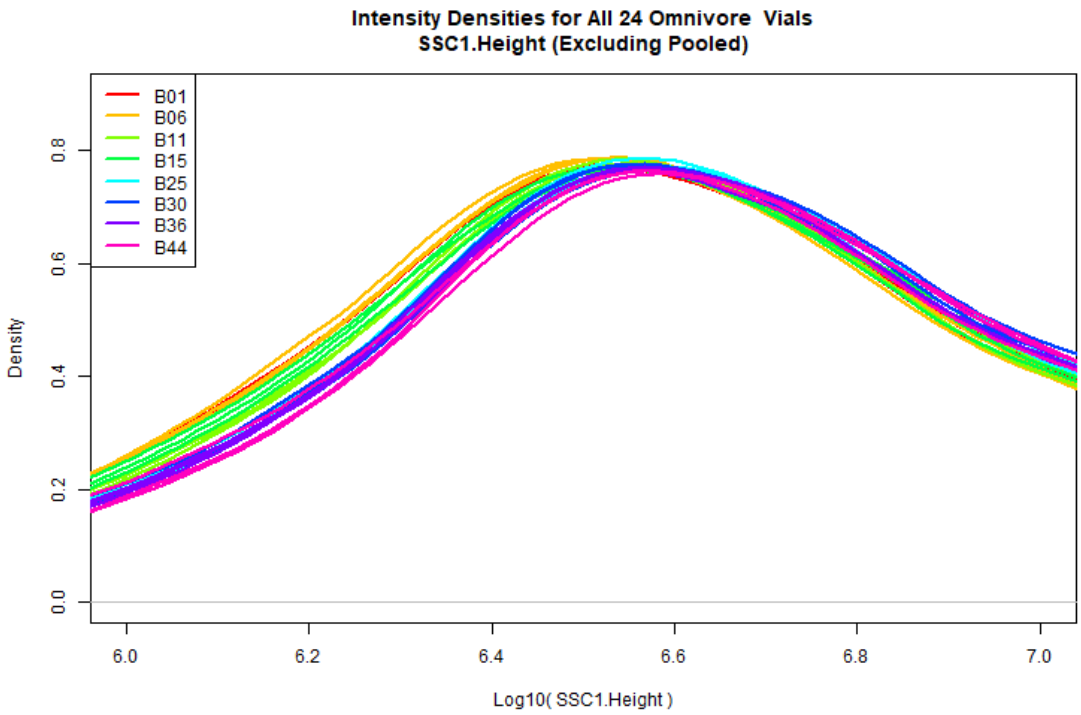


Figure I-23 Kernel densities for Channel SSC1.Height log10(Intensities). The modes in the interval for log10(Intensity) between 6.4 and 6.8 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

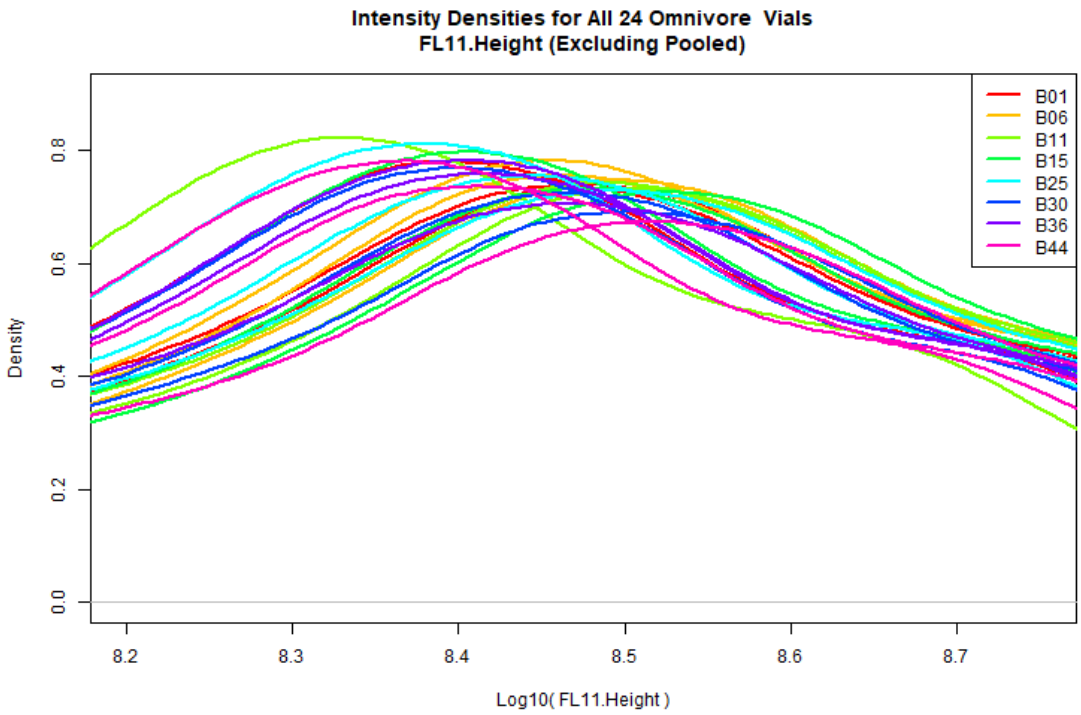


Figure I-24 Kernel densities for Channel FL11.Height log10(Intensities). The modes in the interval for log10(Intensity) between 8.2 and 8.7 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

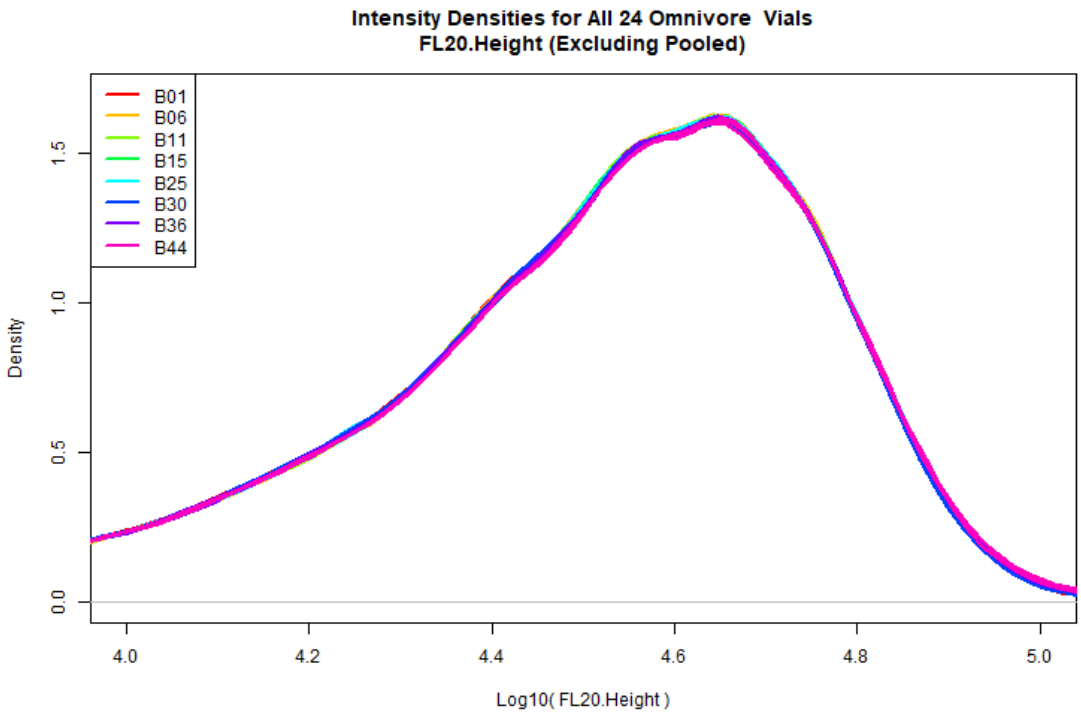


Figure I-25 Kernel densities for Channel FL20.Height log10(Intensities). The modes in the interval for log10(Intensity) between 4.4 and 4.7 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

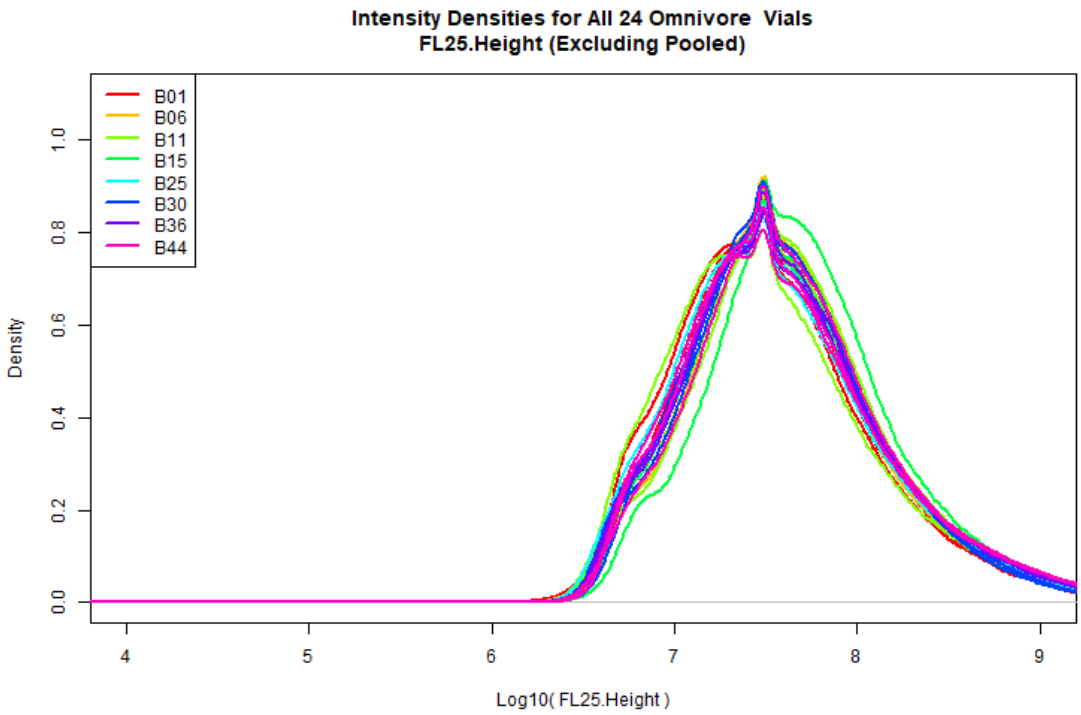


Figure I-26 Kernel densities for Channel FL25.Height log10(Intensities). The modes in the interval for log10(Intensity) between 7.4 and 7.6 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

I.8.2.2. Peak Positions: Omnivore samples

Peak positions were determined by looking for the local maxima of the density values in the specified interval of interest. In most cases, only a single peak was found although two peaks were found in the channel FSC1.Height for omnivores. The peak positions and peak heights were recorded. A hierarchical Bayesian model (discussed earlier) was used to account for within vial variability and between vials variability. Table I-13 shows the coverage intervals for peak positions and peak heights, for the different channels, obtained from the respective posterior predictive distributions.

	Channel	Peak	LowerLimit	UpperLimit	Confidence
1	FSC1	Peak-1 Position	7.285	7.313	95%
2	FSC1	Peak-1 Height	0.535	0.576	95%
3	FSC1	Peak-2 Position	7.448	7.493	95%
4	FSC1	Peak-2 Height	0.508	0.593	95%
5	SSC1	Peak-1 Position	6.503	6.609	95%
6	SSC1	Peak-1 Height	0.750	0.792	95%
7	FL11	Peak-1 Position	8.402	8.495	95%
8	FL11	Peak-1 Height	0.716	0.782	95%
9	FL20	Peak-1 Position	4.647	4.652	95%
10	FL20	Peak-1 Height	1.606	1.630	95%
11	FL25	Peak-1 Position	7.485	7.489	95%
12	FL25	Peak-1 Height	0.832	0.917	95%

Table I-13 95 % coverage intervals for peak positions and peak heights for the different channels. FSC1 has two peaks in the region of interest. The rest of the channels show only a single peak.

I.8.2.3. Homogeneity: Vegetarian Samples

Figure I-27 through Figure I-31 show the kernel densities for log10(intensity) for the 5 channels (Figure I-27 Channel FSC1, range 7.0 and 7.5; I-28 Channel SSC1, range 6.2 and 6.6; I-29 Channel FL11, range 8.4 and 8.7; I-30 Channel FL20, range 4.4-4.7; I-31 Channel FL25, range 7.2-8.5). The density function from R was used with default bandwidth. The R script is provided in the appendix. These density estimates show certain modes (peaks), and the peak positions are expected to be at the same intensity value for all homogeneity samples. In some channels there may be two modes (peaks) of interest. We first find the peak positions for the log10(intensity) densities for each channel and then evaluate the variability in the peak positions between vials.

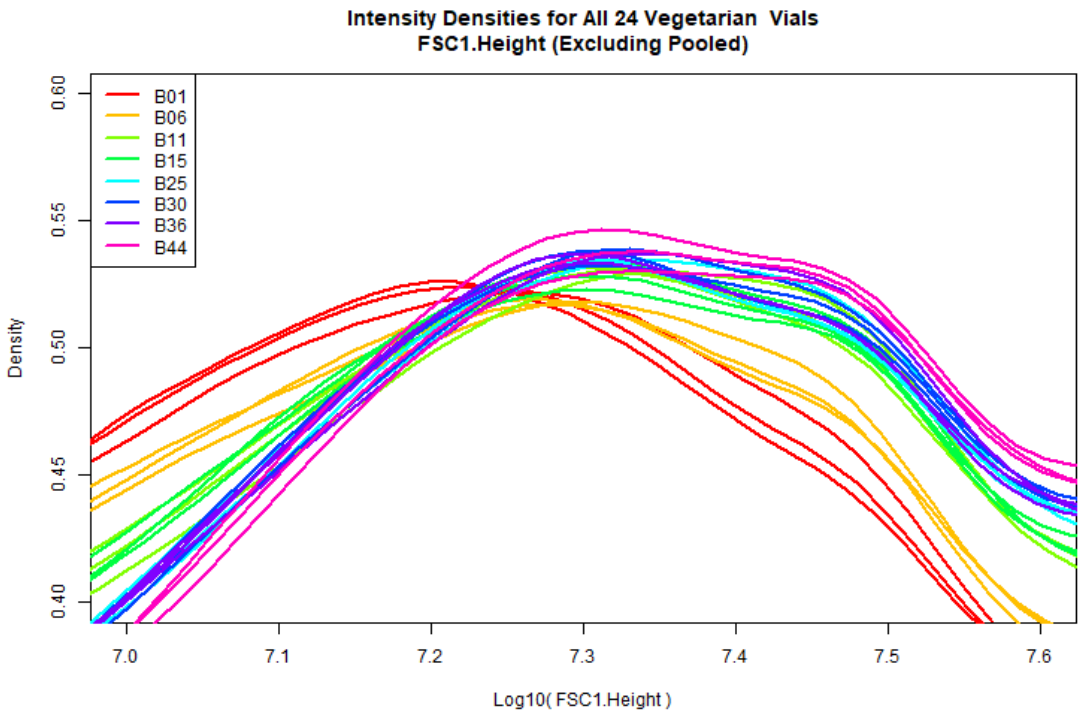


Figure I-27 Kernel densities for Channel FSC1.Height log10(Intensities). The density shows a single peak in the interval for log10(Intensity) between 7.0 and 7.5. Homogeneity assessment will be made by evaluating the variability in the mode positions and heights.

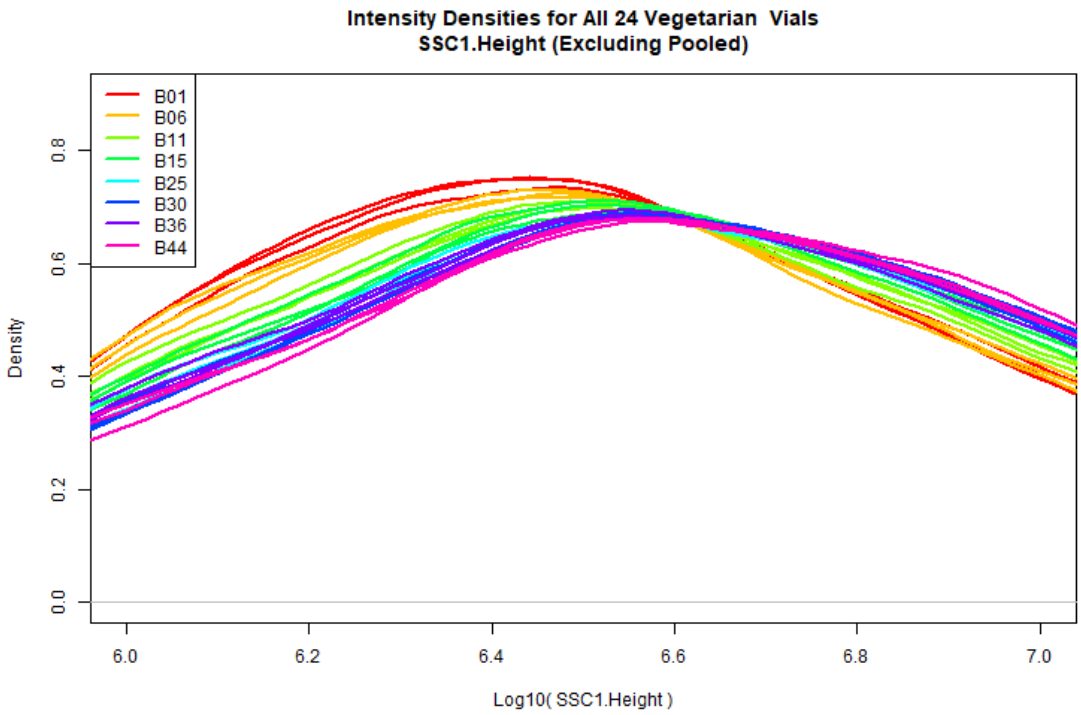


Figure I-28 Kernel densities for Channel SSC1.Height log10(Intensities). The modes in the interval for log10(Intensity) between 6.2 and 6.6 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions and heights.

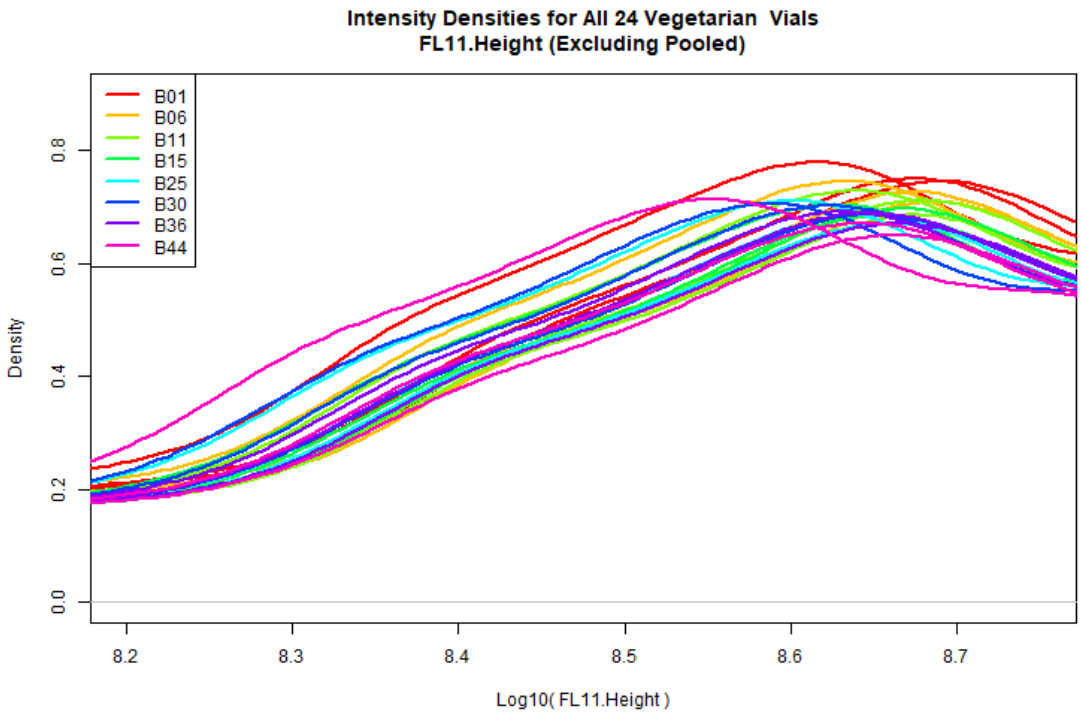


Figure I-29 Kernel densities for Channel FL11.Height log10(Intensities). The modes in the interval for log10(Intensity) between 8.4 and 8.7 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

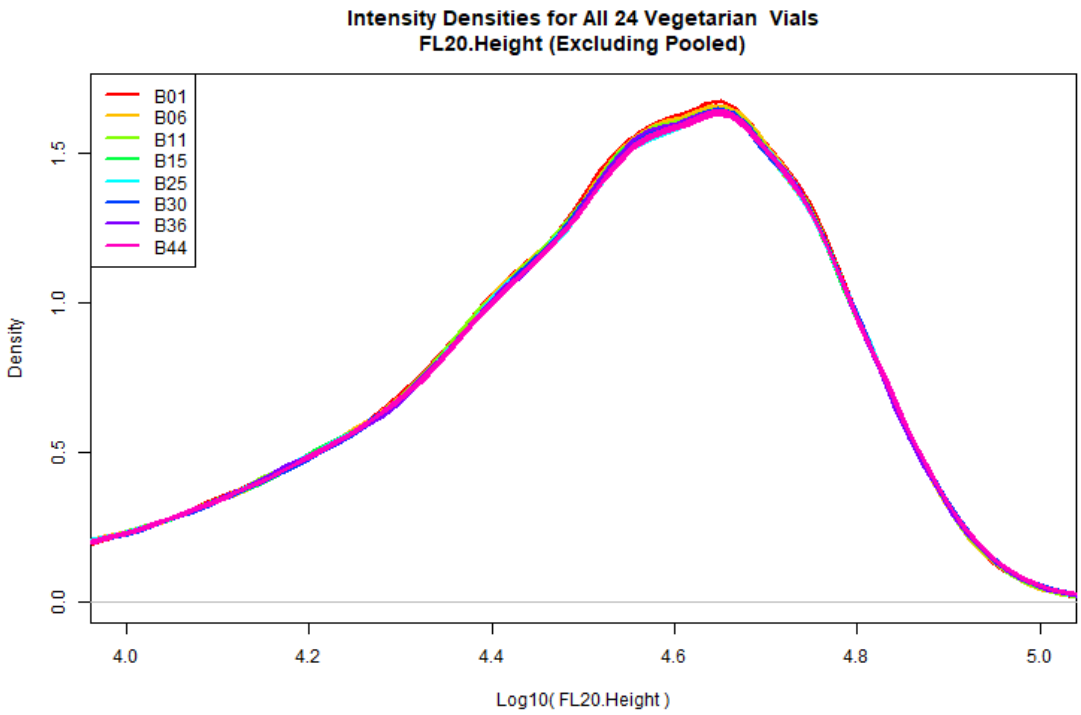


Figure I-30 Kernel densities for Channel FL20.Height log10(Intensities). The modes in the interval for log10(Intensity) between 4.4 and 4.7 are of interest. Homogeneity assessment will be made by evaluating the variability in these mode positions.

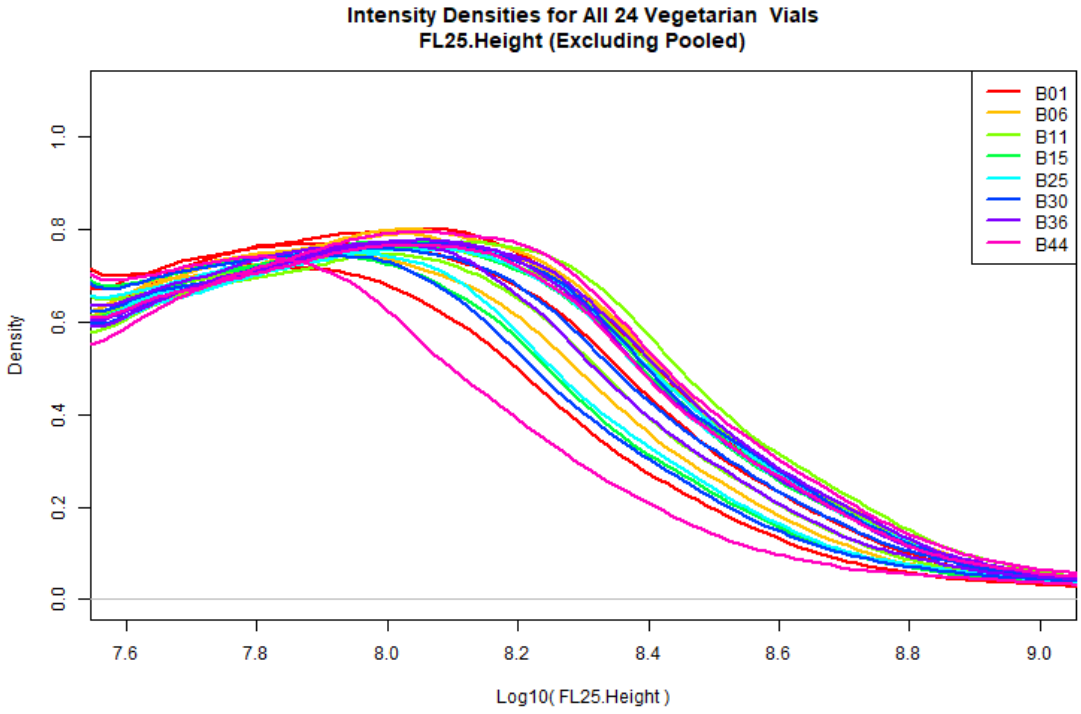


Figure I-31 Kernel densities for Channel FL25.Height log10(Intensities). The modes in the interval for log10(Intensity) between 7.2 and 8.5 are of interest. There are two modes present, one around 7.5 and the other

around 8.2. Homogeneity assessment will be made by evaluating the variability in these mode positions and their heights.

I.8.2.4. Peak Positions: Vegetarian samples

Peak positions were determined by looking for the local maxima of the density values in the specified interval of interest. In most cases, only a single peak was found although two peaks were found in the channel FL25.Height for vegetarians. The peak positions and peak heights were recorded. A hierarchical Bayesian model (discussed earlier) was used to account for within vial variability and between vials variability. Table I-14 shows the coverage intervals for peak positions and peak heights, for the different channels, obtained from the respective posterior predictive distributions.

	Channel	Peak	LowerLimit	UpperLimit	Confidence
1	FSC1	Peak-1 Position	7.218	7.389	95%
2	FSC1	Peak-1 Height	0.513	0.550	95%
3	SSC1	Peak-1 Position	6.401	6.648	95%
4	SSC1	Peak-1 Height	0.637	0.768	95%
5	FL11	Peak-1 Position	8.608	8.684	95%
6	FL11	Peak-1 Height	0.635	0.775	95%
7	FL20	Peak-1 Position	4.645	4.653	95%
8	FL20	Peak-1 Height	1.618	1.679	95%
9	FL25	Peak-1 Position	7.372	7.740	95%
10	FL25	Peak-1 Height	0.596	0.735	95%
11	FL25	Peak-2 Position	7.903	8.102	95%
12	FL25	Peak-2 Height	0.746	0.782	95%

Table I-14 95 % coverage intervals for peak positions and peak heights for the different channels. FL25 has two peaks in the region of interest. The rest of the channels show only a single peak.

I.8.3. Comparing Omnivore Samples with Vegetarian Samples

For the 24 omnivore homogeneity samples and 24 vegetarian homogeneity samples we compared the peak positions and peak heights for each of the channels FSC1 (Figure I-32), SSC1 (Figure I-33), FL11 (Figure I-34), FL20 (Figure I-35), and FL25 (Figure I-36) . We see that peak positions and peak heights differentiate the omnivore samples from vegetarian samples very well. When a particular channel had 2 peaks for one donor group and only one peak for the

other donor group we associated the single peak position in the second group with the closer of the two peak positions of the other group.

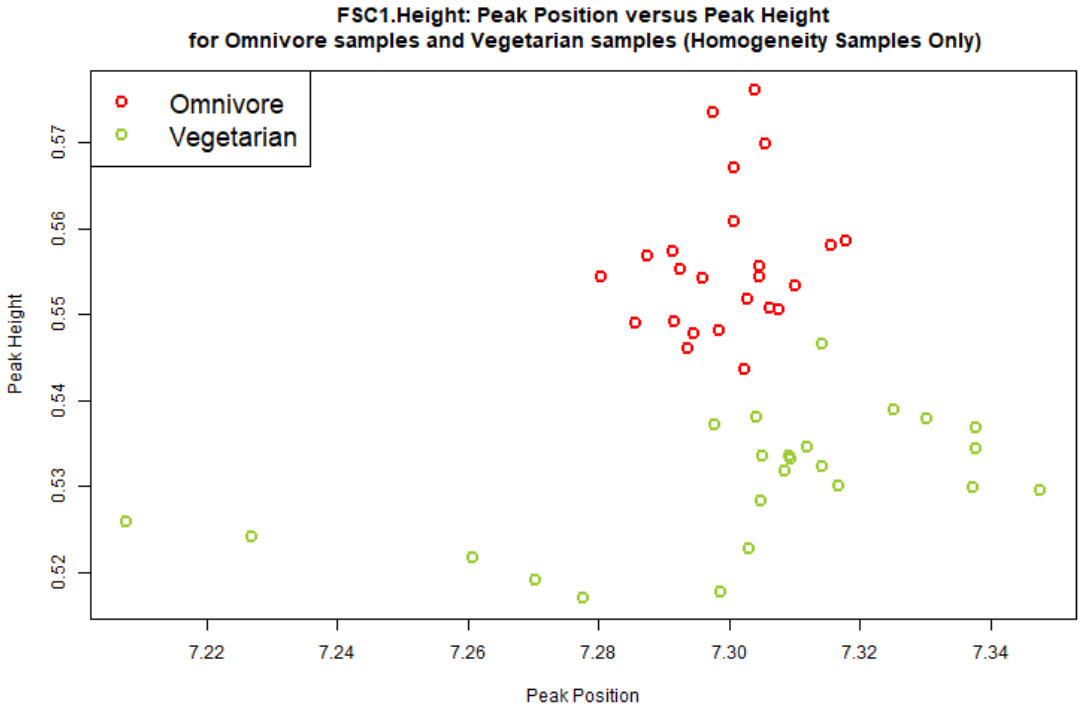


Figure I-32 Peak Positions and Peak Heights of the log10(intensity) density for omnivore homogeneity samples (red) and vegetarian homogeneity samples (green) for the channel FSC1.Height.

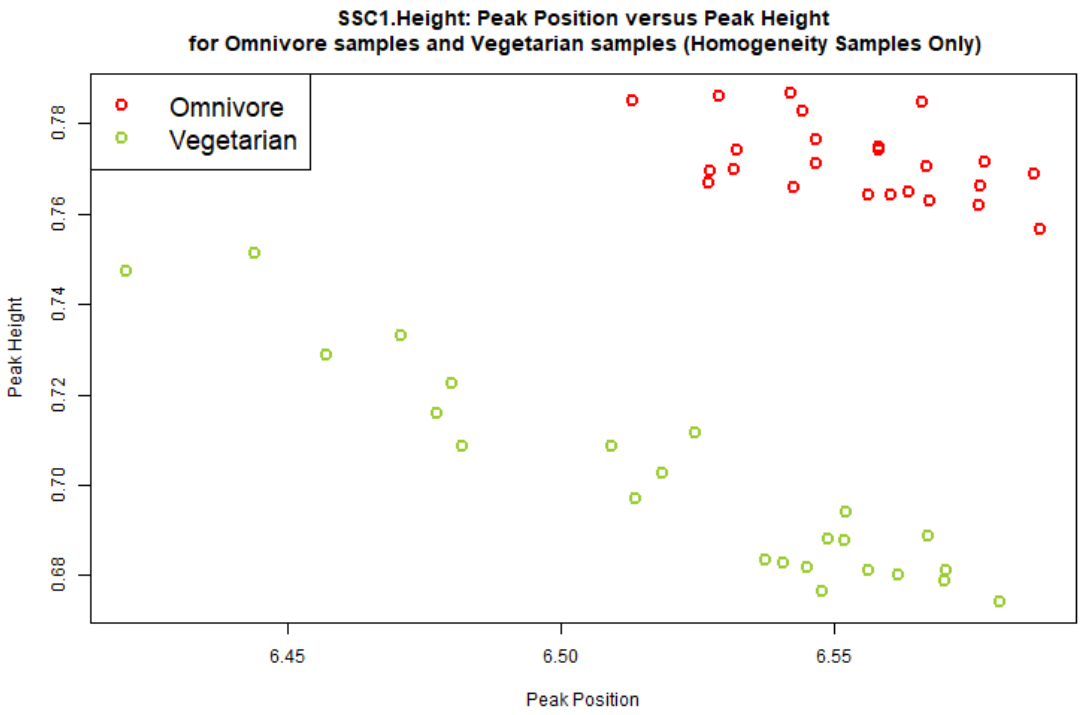


Figure I-33 Peak Positions and Peak Heights of the log10(intensity) density for omnivore homogeneity samples (red) and vegetarian homogeneity samples (green) for the channel SSC1.Height.

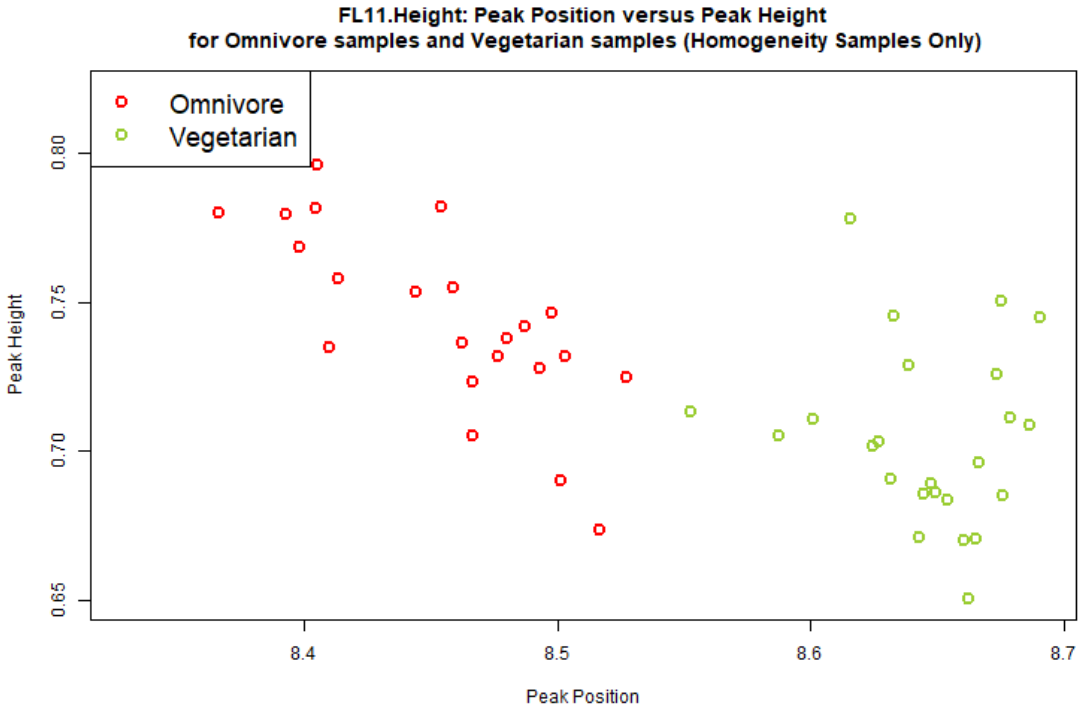


Figure I-34 Peak Positions and Peak Heights of the log10(intensity) density for omnivore homogeneity samples (red) and vegetarian homogeneity samples (green) for the channel FL11.Height.

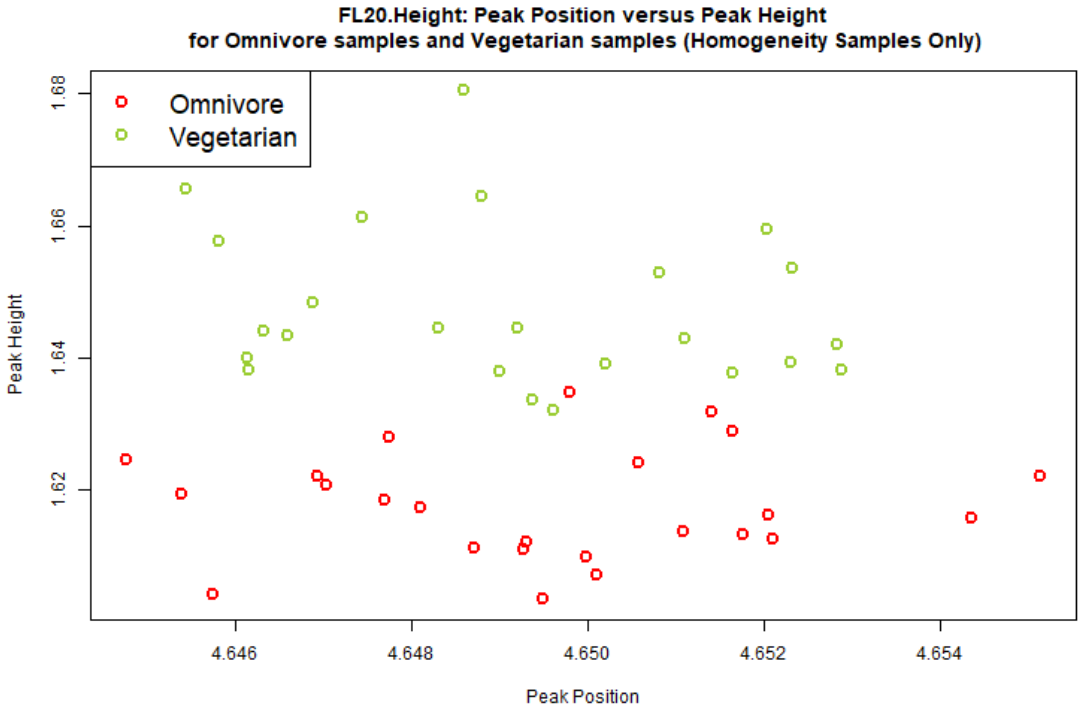


Figure I-35 Peak Positions and Peak Heights of the log10(intensity) density for omnivore homogeneity samples (red) and vegetarian homogeneity samples (green) for the channel FL20.Height.

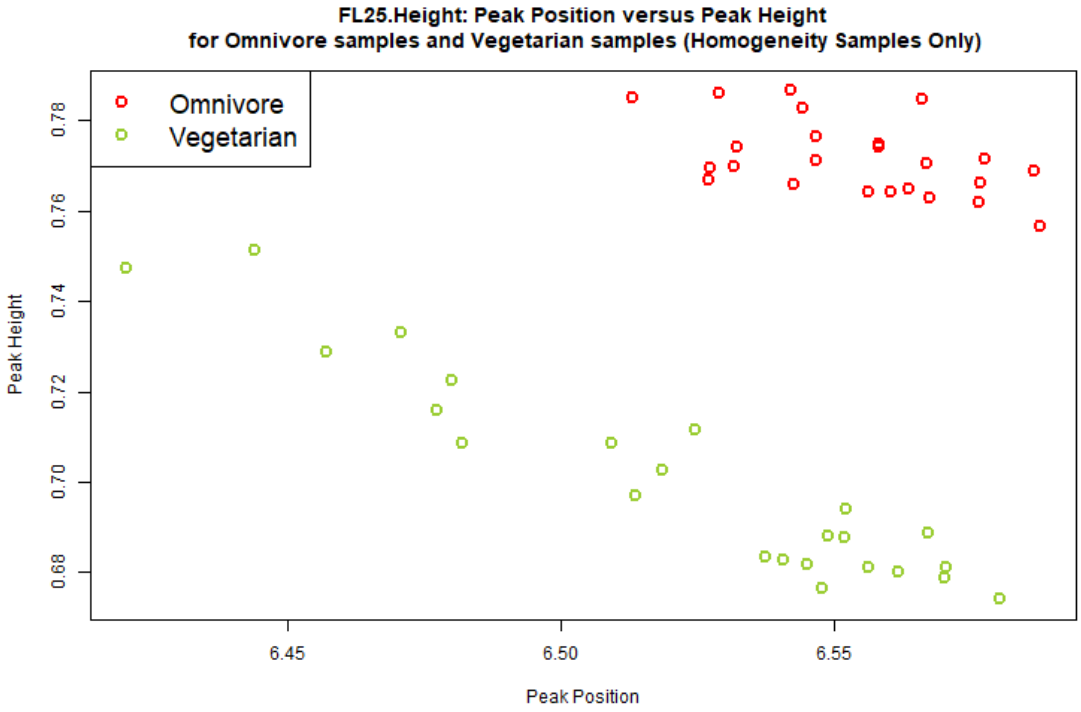


Figure I-36 Peak Positions and Peak Heights of the log10(intensity) density for omnivore homogeneity samples (red) and vegetarian homogeneity samples (green) for the channel FL25.Height.

We also formed a 48 by 10 data matrix consisting of the 5 pairs of peak positions and peak heights (hence 10 columns) for the 48 samples (the 48 rows). We applied a multidimensional scaling transformation to this data matrix using the function **cmdscale** from **R**. Figure I-37 shows the corresponding multidimensional plot. It is very clear from this plot that the omnivore samples and the vegetarian samples have very different intensity distributions for the 5 channels considered.

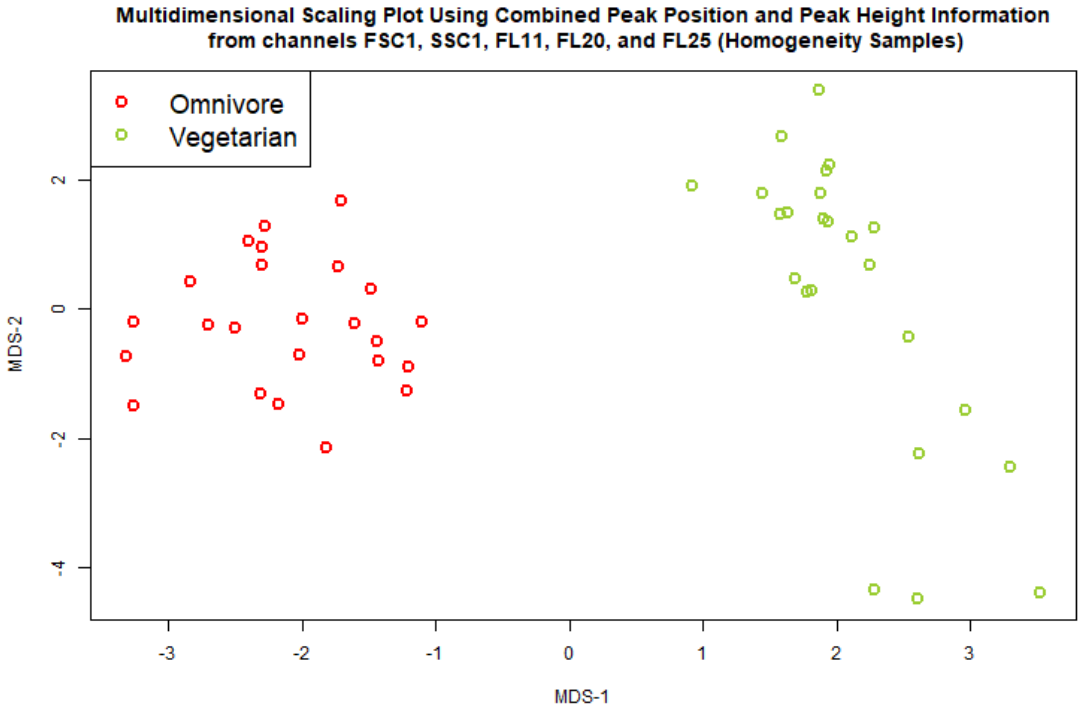


Figure I-37 Plot of the first two multidimensional scaling components based on the peak positions and peak heights for the log10(Intensity) distributions in the five channels (FSC1, SSC1, FL11, FL20, and FL25) in the homogeneity samples.