Master Thesis

The Effect of Homogenization Techniques on the Microheterogeneity of Soil Microbiome Samples

An application of techniques from Incremental Sampling Methodology

Submitted by

Laura Dietrich

For the academic degree of

Master of Science in Biobanking

(MSc)

at the

Medical University of Graz

Executed at

Biobank Graz

Under the supervision of

Prof. Berthold Huppertz, Ph.D.

Submitted 2018

Baltimore, Maryland, 09 July, 2018

Statutory Declaration

I declare on my honor that I have written this dissertation independently and without assistance, that no sources other than those cited were used and that the sources used verbatim or in substance have been marked as such.

Graz, …… Signature (handwritten)

Acknowledgements

The author would like to give thanks to

Prof. Berthold Huppertz, Ph.D., my supervisor, for his selfless donation of time, advice, editing skills, and especially sanity when I had lost mine.

The Sweet Briar College Department of Biology

Especially

Dr. John Morrissey, Ph.D., department chair, for not only allowing me to come back and sample soil, but for plotting with me one last time as well.

Dr. Michael R. Davis, Jr., Ph.D., for advice, planning support, and even printer support when the situation called for it.

Mona Browning, for patience, logistical support, supplies, and a pep talk when it was sorely needed.

Mag. Gabriele Hartl, for kindly translating my abstract, helping me with MEDOnline, and keeping me on track with the paperwork.

The staff of Microbiome Inc, for laboratory and bioinformatics services and for suggesting the addition of fungi to this study.

Dr. Christine Moissl-Eichinger for technical advice on decontamination

The advice and support of my family and Oliver Allen, Matthew McGee-Crane, Komlanvi Gbete, Ashley Carpenter, Megan Hurst, Alison Herman, Ray Rogers, Serena Flood, Melissa Swaby, and Mona Patel

Zusammenfassung

Der Effekt von Homogenisierungsverfahren auf die Mikroheterogenität von Bodenmikrobiomproben

Anwendung von Techniken zur inkrementellen Probensammlung

Der Boden ist eine der heterogensten Umgebungen, die es gibt, vielleicht vor allem in der Verteilungsheterogenität der verschiedenen Bodenpartikel. Das Bodenmikrobiom teilt diese charakteristische Heterogenität, da die Bodenorganismen eine sehr heterogene Verteilung im Boden aufweisen. Dies macht es extrem schwierig, reproduzierbare Ergebnisse aus Bodenproben für die Mikrobiomforschung zu erhalten. Dies ist für ein breites Spektrum an Fachgebieten wie Landwirtschaft, Forstwirtschaft, Medizin und Biobanking von Bedeutung. Daher zielte diese Studie darauf ab, die Variabilität des Mikrobioms in einer Probe nach der Anwendung verschiedener Homogenisierungsverfahren zu vergleichen.

Es wurden Bodenproben von fünf verschiedenen Feldern gesammelt und in drei Gruppen eingeteilt. Die Proben der Gruppe 1 wurden einfach direkt aliquotiert und als Kontrolle verwendet. Die Proben der Gruppe 2 wurden nach der Homogenisierung mit einem Mixer zum Mischen und Disaggregieren der Bodenaggregate aliquotiert. Die Proben der Gruppe 3 wurden mit einem Mixer homogenisiert, um die Bodenaggregate zu mischen und zu disaggregieren, gefolgt von einer zweidimensionalen japanischen Slabcake-Technik, um die Auswirkung der Schichtung während der Probengrößenreduzierung vorweg zu reduzieren, und dann aliquotiert. Die Genomanalyse wurde an drei Aliquotes jeder Gruppe und jedes Feldes durchgeführt, um Bakterien und Pilze zu identifizieren, und anschließend wurde eine bioinformatische Analyse durchgeführt.

Weniger Mikroheterogenität konnte in den Aliquots der Gruppe 3 als in den Aliquots der Gruppe 2 beobachtet werden. Aliquots aus beiden Behandlungsgruppen zeigten eine geringere Mikroheterogenität als Aliquots aus der Kontrollgruppe - Gruppe 1, die überhaupt nicht homogenisiert waren.

Die Homogenisierung von Bodenproben unter der Verwendung von Techniken zur inkrementellen Probensammlung hatte zur Folge, dass Aliquots dieser Proben weniger Variabilität in der Vielfalt an Bakterien und Pilzen in ihren Mikrobiomen aufwiesen. Die Homogenisierung von Proben unter Verwendung dieser speziellen Techniken ist besser als

die Verwendung von Bodenproben, die nicht homogenisiert sind oder mit anderen Techniken homogenisiert werden, die nicht die gleichen Verzerrungs- und Fehlerquellen berücksichtigen.

Abstract

The Effect of Homogenization Techniques on the Microheterogeneity of Soil Microbiome Samples

An application of techniques from Incremental Sampling Methodology

Soil is one of the most heterogeneous environments in existence, perhaps most notably in the distributional heterogeneity of different soil particles. Shaped by this environment, the soil microbiome shares this characteristic heterogeneity as soil organisms display a very heterogeneous distribution within the soil. This makes it extremely difficult to obtain replicable results from soil samples collected for microbiome research. This has significance for a wide range of fields including agriculture, forestry, medicine, and biobanking. Therefore this study aimed to compare microbiome intra-sample variability after the application of different homogenization procedures.

Soil samples from five different fields were collected and separated into three groups. The samples of group 1 were simply aliquoted directly and used as a control. The samples of group 2 were aliquoted after being homogenized with a blender to mix and disaggregate the soil aggregates. The samples of group 3 were homogenized with a blender to mix and disaggregate the soil aggregates followed by using two dimensional Japanese slabcake technique to reduce the effects of stratification during sample size reduction prior to analysis and then aliquoted. Genomic analysis was conducted upon three aliquots of each group and field to identify bacteria and fungi, followed by bioinformatics analysis.

Less microheterogeneity was observed between aliquots in group 3 than between aliquots in group 2. Aliquots from both of these treatment groups demonstrated less microheterogeneity than aliquots in the control group, group 1, that were not homogenized at all. Thus, homogenizing soil samples using techniques from the Incremental Sampling Methodology resulted in aliquots from those samples demonstrating less variability in the diversity of the bacterial and fungal taxa present in their microbiome. Homogenizing samples using these specific techniques is superior to using soil samples that are not homogenized or are homogenized using other techniques which do not take the same sources of bias and error into account.

Table of Contents

List of Figures

List of Tables

Introduction

Background

From the turn of the 20th Century, it has been popular to define soil as "more or less the loose and friable material in which, by means of their roots, plants may or do find a foothold and nourishment, as well as other conditions of growth" [1]. While in acknowledging the connection between soil and vegetative life, this definition certainly underpins the importance of the soil to agriculture, it fails to adequately capture the full extent of the soil's complexity. Soil is far more than just a physical or mineral entity; it is a function of a number of factors (Figure 1). These chiefly include climate, organisms, topography, parent material, and time [2]. Organisms are stated here more broadly to include both macro and microflora as well as macro and micro fauna. As soil microorganisms act upon the soil, so too are they acted upon by the rest of the factors as well as other properties of the soil.

$$s = f\,(cl, o, r, p, t, \ldots)$$

Figure 1. Fundamental Equation of Soil-Forming Factors [2]. This equation defines soil (s) as function (f) of a number of factors including climate (cl), organisms (o), topography (r), parent material (p), and time (t) as well as an infinite number of additional factors or properties (…).

Soil bacteria (and also fungal spores) are transported vast distances by the wind which brings about "a continuous reinoculation of soils," and therefore most all soil microorganisms are introduced to most all soils [2]. It is the "constellations" of factors and soil properties which determine which organisms take hold in a given location [2]. This relationship between soil microorganisms and their environment occurs down to a microscopic level as "soil organisms live in the microscale environments within and between soil particles. Differences over short distances in pH, moisture, pore size, and the types of food available create a broad range of habitats" [3]. Indeed, the range of habitats created by conditions in the soil is nearly limitless, providing both spatial diversity and genetic isolation which in turn makes an extreme level of diversity possible—a handful of

soil may contain a diversity of prokaryotes similar to that of the macrofauna and flora in the Amazon basin [4].

Importance

Representative samples and aliquots, regardless of sample type, are a key component of successful biobanking and research. Reproducibility is an important part of the scientific method and homogeneous aliquots allow researchers to reproduce experiments. This homogeneity of aliquots is also what enables biobanks to maintain retain specimens for later comparison. Any heterogeneity of aliquots of any kind leads to irreproducibility and such aliquots are useless to science.

In biobanking of human and animal samples, samples of a tissue or organ should be representative of the whole tissue or organ (or piece of it) of interest. In order to enable repetition of experiments and to allow reproducibility of data, today samples are divided into smaller pieces/parts [5]. These aliquots again should represent the whole sample and should be homogeneous insomuch as possible.

With all of its different substructures such as the parenchyma, blood vessels, muscle, nerves, connective tissues etc., a piece of tissue cannot be divided into homogeneous aliquots. Therefore, each subsample of a larger tissue sample is again a sample in its own right. Homogeneous aliquots of a tissue sample can be obtained only by homogenization of the tissue and subsequent aliquoting of the homogenate. This allows the generation of homogeneous aliquots; however, it prohibits drawing any morphological relations within the tissue. Therefore, homogenous aliquots are most commonly generated from fluids such as full blood, plasma, serum and urine.

Looking at the soil, traditionally, the exploration of the soil microbiome has been confined to the organisms which can be successfully cultured in the lab—a small number of species which comprise less than 0.1% of the species present in the soil [4]. It is now possible to get a much clearer picture of the whole microbiome through advancements in genomic sequencing and computing technology [4]. Göhler et al. [6] performed a direct comparison between detection of a specific bacterial strain *Burkholderia pseudomallei* by culture and detection by comparative genomic analysis using qPCR (quantitative PCR). This study revealed that culture detected the bacterial strain in 94 out of 200 samples

(47%), while the combination of culture and qPCR resulted in the detection of the strain in 180 out of 200 samples (90%).

As previously stated, soil is a function of a number of factors. Chiefly among these are climate, organisms, topography, parent material, and time [2]. This latter factor is one characteristic which provides possibilities for biobanking. If a method which yields consistent, homogenous aliquots can be established, it would be possible to establish a longitudinal biobank or biobanks of soils that will provide valuable insight into how the soil microbiome changes over time in one place and how this impacts the soil itself. This would be especially useful for the study of contaminated areas and areas undergoing succession after contamination events and remediation [7].

State of the Art

There is little readily-accessible information on homogenization techniques for soil microbiome research [8]. By comparison, there are very well documented and detailed protocols and general information available on soil homogenization for environmental and mineral soil sampling operations. One of the most recent and most extensive of these is the Incremental Sampling Methodology published by the Interstate Technology and Regulatory Council (ITRC) in 2012 [9]. This largely builds off of the earlier work of Pierre Gy and the United States Environmental Protection Agency [9]. Originally developed for sampling environmental contaminants in the soil, the Incremental Sampling Methodology has been slow to catch on in biological applications. In one instance, the Incremental Sampling methodology was utilized by a team of researchers looking for *Burkeholderia psuedomallei*, a strain of bacteria which causes severe illness in humans and animals [10]. No studies were found that specifically addressed broader microbiome sampling with this methodology and none explored whether the methodology is effective for any type of biological sampling.

As pertains to soil, there are three spatial scales in which heterogeneity occurs (Figure 2). These are: large scale heterogeneity, short scale heterogeneity, and microheterogeneity [9]:

Large scale heterogeneity is the largest of these and is concerned with samples that are a considerable distance apart. An example would be two samples from opposite ends

of a field (Figure 2A, red line). Variations due to large scale heterogeneity have been nicely depicted by Heil and Schmidhalter [11] who used an electromagnetic device to analyze soil parameters such as salinity, soil texture and water content.

Short scale heterogeneity is concerned with collocated samples: samples taken so close together that they can be said to have been taken from the same place. Typically, this refers to a scale of roughly a meter apart (Figure 2A, yellow line). Even over such a short distance, massive variations of soil parameters may occur. For soil samples in a forest, Baldrian [12] identified general parameters influencing soil composition such as soil composition, litter, foliage and deadwood. All of these parameters had a direct impact on the microbial mass as well as the fungi/bacteria ratio. They also influenced the dominant taxa and groups within the microbiome [12].

Microheterogeneity is the smallest scale of heterogeneity and occurs within a single soil sample (Figure 2B) [9]. In order to address the larger scales of heterogeneity, which are relevant to making decisions and coming to conclusions about larger areas and ecosystems, it is first necessary to overcome microheterogeneity through appropriate homogenization techniques. Microheterogeneity within a given soil sample again depends on a variety of factors. Additionally, there is a strong horizontal stratification within the soil. There are mostly organic horizons upon different mineral horizons [8]. The organic horizon at the top of soil is the layer with the highest rate of fluctuations and the highest concentrations of microbes. This is the reason why most studies on the soil microbiome use the uppermost 15cm of this top soil horizon [8].

Figure 2. The three scales of spatial heterogeneity in soil. **A**: Large scale heterogeneity (red "L") is between samples taken a considerable distance apart. In this example, one is in the hay field in the foreground and the other is in a switchgrass field in the background. Short scale heterogeneity (yellow "S") is between two collocated samples; typically within about a meter apart, as in this example. **B**: Microheterogeneity is the heterogeneity occurring within a single sample: in this example, a discreet soil core.

As proposed by Pierre Gy, only Grouping and Segregation Error (GSE) can be specifically considered in regards to homogenization as homogenization only impacts GSE [13]. GSE is the result of the Distributional Heterogeneity of soil. This is perhaps best described by first looking at the soil as a whole in situ: soil in nature is anisotropic which is to say, it has horizontal layers which give it a profile [2]. Ultimately, it is the Distributional Heterogeneity: the effect of gravity acting upon soil particles and aggregates of different sizes which causes this segregation, or stratification, to occur (Figure 3).

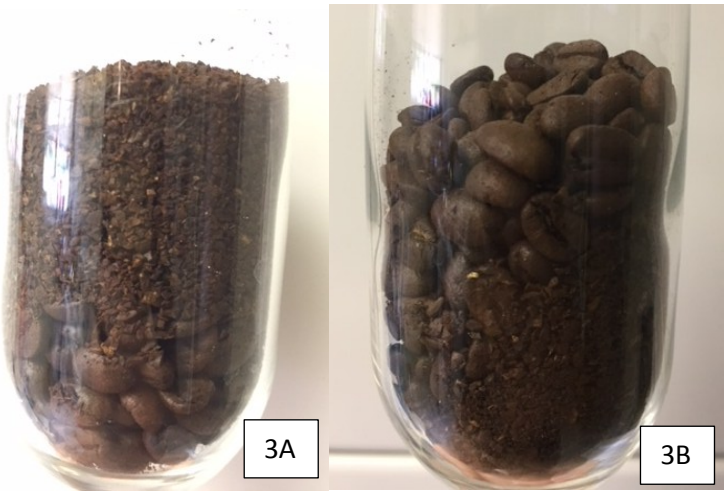

Figure 3. The Distributional Heterogeneity of particulates. **A**: Fine particles, represented by coffee grounds, are poured in a layer on top of a layer of larger particles, represented by whole coffee beans. **B**: After the vessel is shaken up and down five times, the fine particles have filtered down through the large particles, which are in the process of being pushed up to the top of the mixture. In nature this process occurs to soil particulates over time as gravity acts upon them, but when soil samples are extracted from the soil, it is accelerated by conditions during shipping and processing including some homogenization techniques.

Simple mixing by such means as stirring, shaking, or using a blender may seem to be the obvious answer to the question of how to homogenize soil samples. However, this may actually increase Grouping and Segregation Error, or GSE, within a sample [9]. Similarly, conditions such as vibration and jostling during transport of a sample between the collection site and the laboratory where it is to be processed can likewise exacerbate GSE. It can also occur during storage. GSE cannot be prevented during transport or storage, so the benefit of homogenization is short-lived and must typically be carried out right before analysis [13].

It follows that additional means to overcome GSE need to be implemented in order to obtain representative aliquots of a given sample. It is impractical from a biobanking standpoint to store the full volume of a typical Incremental Sampling Methodology composite soil sample especially when there can be many such composite samples from a single collection site. It is also impractical for either a biobank or a researcher to have to homogenize such a large volume of soil each time a test must be performed and

furthermore, it would be extremely expensive to run genomic analysis on such a large sample. The ideal solution, would be to take representative aliquots of the composite sample and discard the remnants. At the same time, as mixing will not overcome the stratification or segregation of the soil aggregates, another technique will have to be used to ensure that all of the layers which are created will be sampled. Fortunately, there is a technique in the Incremental Sampling Methodology which can address both sample size reduction (aliquotting) and Distributional Heterogeneity.

In a 1D, or one-dimensional, Japanese slabcake, the soil is spread in a linear pile on a steel plate and then a scoop is used to take a narrow subsample, or aliquot, all the way across and through the sample [14]. In 2D, or two dimensional, Japanese slabcake the soil is spread out comparatively thinly across a walled metal sheet and then small cores are taken through the entire thickness: as many cores are taken as are determined necessary in the specific situation, but generally no fewer than thirty [9]. 2D Japanese slabcake is in essence, a scaled down version of field sampling. Just as taking cores and creating a composite sample from them can overcome GSE in the field, it can also achieve this on a smaller scale in the laboratory. 2D Japanese slabcake can be used in conjunction with a blender for disaggregation and some simple mixing, to create representative and nearly homogeneous aliquots.

Hypothesis

Homogenizing soil samples using techniques from the Incremental Soil Methodology results in aliquots from those samples demonstrating less variability in the diversity of the bacterial and fungi present in their microbiome than is demonstrated by aliquots from samples which have not been homogenized. Homogenizing soil samples using these techniques is thus superior to not homogenizing soil samples or using a method of homogenization which does not take the same sources of bias and error into account.

Materials and Methods

The methodology of this study can be divided into five main parts: collection, initial processing, shipping, extraction and processing of DNA, and analysis of data.

Collection:

During collection, the principle equipment used included a steel step probe (Figure 4) and an Ohaus® spring scale. Consumable supplies used included Kimwipes™, nitrile gloves, and Ziploc® bags. 70% ethanol was used for general cleaning.

Figure 4. Step probe.

The fields selected for soil collection were under switchgrass cultivation since 2014 [15]. Previously, they were used in the production of hay for livestock.

Prior to the commencement of collecting activities, a laboratory marker was used to mark the step probe for a depth of 10 cm. The ground at the site of each coring was

prepared by gently scraping or brushing away any moss or grass to make a clear spot just big enough for the cutting end of the probe. Cores were taken by inserting the step probe straight down into the ground to the 10 cm mark (Figure 5). Accuracy was within approximately 1 cm with a preference for slightly over, as opposed to under 10 cm.

The probe was removed from the ground by pulling upwards with a slight, gentle twisting motion to ensure clean and easy removal. Each core was removed from the probe using gloved fingers. Care was taken that cores were taken up, down, and across slopes with efforts made to obtain cores from areas of varying moisture content and plant composition throughout each field. Coring sites on top of and immediately adjacent to or below animal droppings on slopes were avoided.

Figure 5. Step probe with core. The taper at the cutting end of the probe assists both with pushing into the soil and with retaining the core inside the probe as the probe is withdrawn.

For each composite sample, enough cores were taken and added to a Ziploc® bag to equal or to just exceed 500 g weight on the spring scale (Figure 6). Three composite samples were taken per field and these were collected at the same time for efficiency; cores for all three field samples were taken within approximately 1 m of each other. The probe was cleaned with 70% ethanol on a Kimwipe™ between each field.

Figure 6. Weighing soil cores with spring scale.

Due to inclement weather, it was necessary to stop collecting after the first three fields and then finish collecting in the second two fields the next day. The composite samples that had been collected the first day were placed in a -80°C freezer overnight.

Initial Processing:

After the final composite samples were collected, the composite samples of the first group were removed from the -80°C freezer and allowed to thaw at ambient temperature.

The composite samples were divided into three treatment groups with each treatment group receiving five composite samples: one from each field. The three treatment groups included one control for which no homogenization would take place, a treatment group which would be disaggregated and mixed in a blade blender, and one treatment group which would be disaggregated and mixed in the blender, but also be treated with a technique called 2D Japanese Slabcake intended to reduce the effects of distributional heterogeneity. These were known as Treatment Group 1, Treatment Group 2, and Treatment Group 3, respectively. All samples including those in the control group

were gently massaged in the Ziploc® bags to break up the cores prior to any further treatment.

At this stage, the composite samples in the control group (Treatment Group 1) were aliquoted. The process for aliquoting a single composite sample was as follows: a plastic disposable weigh boat was placed upon a Mettler™ laboratory scale accurate to 0.01 g and the scale was then tared. It is important to note that as it was necessary to reuse weigh boats several times, a line was drawn with a laboratory marker dividing each weigh boat into two halves. Care was taken to keep each aliquot on its own side of the line and between composite samples, each weigh boat was washed with diluted dish soap and then treated with Argos Technologies DR NAse Free™ reagent. A half-round metal Scoopula™ was used to scoop small quantities of soil from the Ziploc® bag into the weigh boat. Soil was added to the weigh boat until a weight of 0.6 g was reached. This was then scooped into a labeled 2 ml cryotube. A total of six aliquots were taken per field sample: three were to be sent to Microbiome Inc. for analysis and three were to be kept as retain samples in case anything should happen to the first group of samples during shipping.

Before aliquoting, field samples in Treatment Group 2 and Treatment Group 3 needed to be run through a blender. For this purpose, an Oster® Osterizer blender was used. Prior to blending, any rocks and large roots which might damage the blade assembly were removed. As placing all of the soil from a given composite sample into the blender all at once would result in ineffective disaggregation and longer processing time, only enough soil to just cover the blades was blended at any given time (Figure 7). Blending was achieved by using the pulse setting and only pulsing briefly one to three times before using a lab spoon to loosen and remove the soil from the bottom of the blender and around the blades. The soil was then turned back onto the blades and the process was repeated until the soil aggregates were approximately 2 mm to 3 mm in size. Blending beyond this point did not seem to yield consistently finer aggregates and it is very important to note that overworking the soil either overall or during any of the pulsing stages resulted in the soil aggregating again into a single mass. Aside from defeating the purpose of the disaggregation step, this risked damage to the blender.

Figure 7. Soil in the blender, demonstrating proper coverage of blade assembly. The soil is depicted prior to any pulsing; the aggregates are irregular in size.

Treatment Group 1 and Treatment Group 2 were completed on day one of processing. Treatment Group 3 was completed on day two of processing. Overnight, all samples were left in a refrigerator at 6°C.

Composite samples in Treatment Group 3 were run through the blender as samples in Treatment Group 2 were, but were also treated using a method called 2D Japanese slabcake (Figure 8). Each blended composite sample was spread out across a 25.4 cm by 43.2 cm walled metal sheet to an approximate depth of 0.5 cm. Thirty cores were taken from the composite sample using an inverted infant dosing syringe from which the plunger had been removed (Figure 9).

Figure 8. Composite sample prepared for 2D-Japanese slabcake, before cores are taken. Though there is still some degree of irregularity in size, the largest aggregates are smaller than those depicted in Fig. 7.

Figure 9. Coring implements: an infant dosing syringe and pipette tips. Soil tended to get stuck in the pipette tip used to take the cores (center), so a second pipette tip (right) was used to remove soil from the coring tip.

It was then necessary to do further sample size reduction through a second, smaller 2D Japanese slabcake step (Figure 10). The soil remaining in the metal sheet was scraped to one side and the soil which was collected via coring during the first 2D Japanese slabcake step was spread out in the empty portion of the metal sheet to a size of about 10 cm by 7 cm. It was difficult to spread the soil to a more uniform size each time as the clay content of the soil made it quite sticky in consistency.

Figure 10. Secondary 2D Japanese slabcake after thirty cores have been taken. The plastic weigh boat contains the soil from the cores.

An inverted filtered pipette tip was used to take thirty cores as before. These were combined on a clean weigh boat and aliquots were then taken from this as for Treatment Group 1 and Treatment Group 2. It should be noted that it was especially crucial to complete this step quickly as these samples seemed to dry out faster than samples from other treatment groups.

Shipping:

It was not possible to ship the aliquots immediately after initial processing: instead, the aliquots were transported to Baltimore and held for a week before being shipped to Microbiome Inc. for the analytical phase of testing. All aliquots were transported at ambient temperature to the nearest available location for the procurement of water ice; a process which took approximately forty five minutes. Water ice was used instead of dry ice for safety and legal reasons. Once packed in the ice and in an insulated cooler bag, the aliquots remained so stored for the seven hour journey to Baltimore. Upon arrival, they were removed from the ice and placed into a consumer freezer for approximately forty hours until dry ice could be procured.

Dry ice was used to store the aliquots for five days until February 26, 2018, when the aliquots were packed on a combination of gel packs (approximately -20°C) and a very small quantity of dry ice (0.46 kg) and turned over to the shipper. They arrived in British Columbia ten hours later. They were then held up in Canadian customs awaiting clearance until a broker was able to negotiate their release to the shipper on the 28th of February, who then placed the whole package into cold storage with dry ice. They were delivered to the laboratory on March 1, 2018.

Processing by Microbiome Inc.:

The work in this section was completed by Microbiome, Inc. in Vancouver, British Columbia, Canada.

In order to look at both the bacterial and fungal communities in the aliquots, they were sequenced for both the 16S (V4 region) bacterial genes and ITS (region 2) fungal genes. For both applications, the aliquots were placed in a MoBio PowerMag Soil DNA Isolation Bead Plate and their DNA was extracted using a KingFisher robot according to instructions provided by MoBio. As per the protocol of Kozich et al. (2013) [16], 16S rRNA (bacterial) genes were PCR-amplified with dual-barcoded primers targeting the V4 region and fungi were PCR-amplified using dual-barcoded primers targeting the ITS2 region. An Illumina MiSeq was used to sequence the amplicons with the 250-bp paired-end kit (version 2). Sequences were then denoised. For both fungi and bacteria, the Mothur software package (version 1.39.5) [17] was used to cluster sequences into 97%-similarity operational taxonomic units (OTUs).

For the bacteria, 1.063709×10^6 high-quality reads were obtained. There were 66,041 OTUs, including those occurring once with a count of 1, and a read range of 1.6259×10^4 and 3.302×10^4 in the final dataset. These bacterial sequences were classified taxonomically using the Greengenes (version 13_8) reference database.

For the fungi, 2.197994×10^6 high-quality reads were obtained. There were 12,132 OTUs, again including those with a singular occurrence, and a read range of 7772 and 8.8561×10^4 in the final dataset. Fungal sequences were taxonomically classified using the Unite reference database (version 7.1).

In order to address the potential for contamination, the DNA amplified from specimens was co-sequenced with that of four template-free controls and extraction kit reagents which were processed as the specimens. Cloned SUP05 DNA (number of copies = $2*10^6$) was used for two positive controls which were also included. Any OTUs with a mean abundance in the controls which met or exceeded 25% of their mean abundance in the specimens were considered "putative contaminants" and were removed.

Analysis of Data:

The following analyses were computed in the R environment by Microbiome Inc.

After filtering out contaminants, the Shannon-Weiner Diversity Index was calculated on the raw OTU data to estimate alpha diversity, intrinsic to each aliquot. An ANOVA or, analysis of variance, was used to test the significance of differences in diversity.

In estimating beta diversity between aliquots, any OTUs which occurred in fewer than 10% of the aliquots with a count of less than three were excluded. The remaining OTU abundances were converted into pairwise dissimilarities using the Bray-Curtis Index. NMDS, non-metric multidimensional, ordination was used to visualize this with an emphasis on differences between aliquots.

PERMANOVA, or permutational multivariate analysis of variance, was used to assess variations in community structure. Treatment group was used as the primary fixed factor and 4,999 permutations were used for significance testing.

The analyses above were performed on both the fungal and bacterial data.

All analyses beyond this point were conducted by the author with help from Dr. Berthold Huppertz.

Species richness, or the total number of species present, was counted for each aliquot and then the mean species richness was calculated of the three aliquots from each composite sample so that it would be possible to compare each treatment group within each field. Standard Deviation was calculated and a t-test was performed for each of the mean species richness values in order to determine formal significance.

In addition, the specific taxa shared by all three aliquots from each composite sample were counted to determine the shared species richness. This was then compared to the mean species richness for each composite sample as a proportion.

As with the analyses performed by Microbiome, Inc., these analyses were performed for both the fungi and the bacteria.

Results

The aim of this study was to demonstrate that there is greater similarity in the fungal and bacterial microbiome of aliquots from composite soil samples which have been homogenized as opposed to aliquots of composite soil samples which have not been homogenized. Though there are a number of similarities between the bacterial and fungal communities in the soil aliquots, it is ultimately more functional to split the results by bacteria and fungi to allow for an examination of their differences. Therefore, the bacterial results will be described first, followed by the fungal results.

In interpreting the naming conventions for each of the aliquots it should be noted that the prefix, "LED" is merely the initials of the author and was included on the vial labels to denote stewardship during their storage in shared fridges and freezers in the lab. Following the prefix, the middle three digits indicate which composite sample the aliquot is from. The final letter denotes the aliquot. For example, "LED001A" would be aliquot "A" of composite sample "001." The following composite samples were in group 1 (control): 001, 006, 008, 011, and 013. The following composite samples were in group 2 (blender only): 002, 004, 009, 010, and 015. The following composite samples were in group 3 (blender and slabcake): 003, 005, 007, 012, and 014. Field 1 consisted of composite samples 001, 002, and 003; field 2 consisted of composite samples 004, 005, and 006; field 3 consisted of composite samples 007, 008, and 009; field 4 consisted of composite samples 010, 011, and 012; and field 5 consisted of composite samples 013, 014, and 015.

Bacteria:

When the beta diversity, the diversity between all of the aliquots, is plotted by field (Figure 11), clear differentiation and grouping of bacterial microbiome aliquots by field can be observed. This is especially pronounced for fields 4 and 5. The bacteria from these two fields appear vastly distinct in relation to one another.

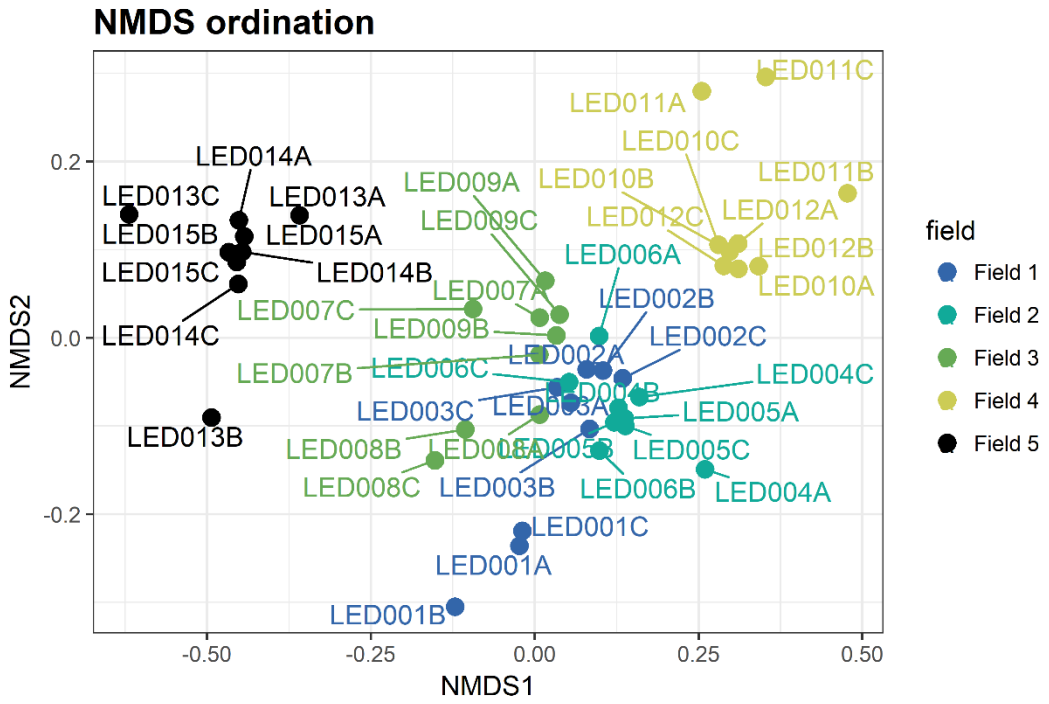

Figure 11. Beta diversity by field (Microbiome, Inc.). This plot is a visualization of the similarity in the bacterial microbiome of each of the aliquots sorted by field through the non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis Dissimilarity Index values. Both axes represent Bray-Curtis Index values which provide a scale of relative distance. The closer two aliquots are within the plot, the more similar they are to one another in terms of diversity. Conversely, the farther apart the aliquots are, the more dissimilar they are. The following composite samples were in group 1 (control): 001, 006, 008, 011, and 013. The following composite samples were in group 2 (blender only): 002, 004, 009, 010, and 015. The following composite samples were in group 3 (blender and slabcake): 003, 005, 007, 012, and 014.

When the beta diversity of the bacteria is plotted by treatment group (Figure 12), it becomes apparent that the aliquots from the samples which were homogenized are more tightly clustered than the control aliquots from group 1, control. In the clusters of the homogenized groups from each field, there is little differentiation between groups 2 and 3, blender and blender + slabcake, respectively.

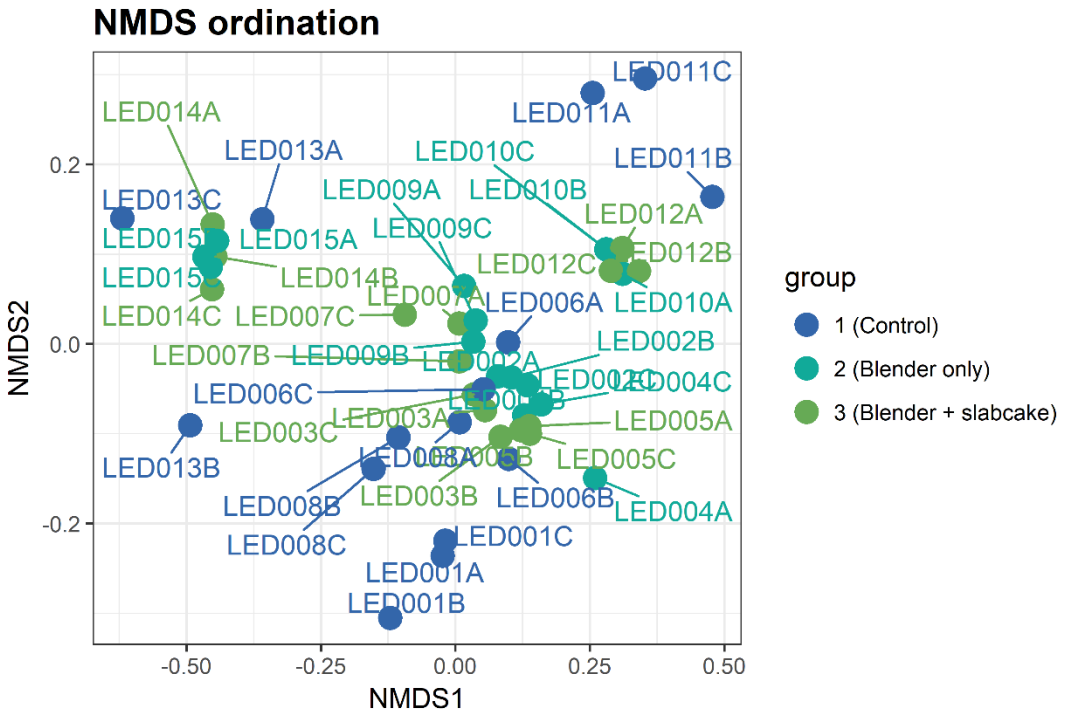

Figure 12. Beta diversity of bacteria by treatment group (Microbiome, Inc.). This plot is a visualization of the similarity in the bacterial microbiome of each of the aliquots sorted by treatment group through the non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis Dissimilarity Index values. Both axes represent Bray-Curtis Index values which provide a scale of relative distance. The closer two aliquots are within the plot, the more similar they are to one another in terms of their diversity. Conversely, the farther apart the aliquots are, the more dissimilar they are. Field 1 consisted of composite samples 001, 002, and 003; field 2 consisted of composite samples 004, 005, and 006; field 3 consisted of composite samples 007, 008, and 009; field 4 consisted of composite samples 010, 011, and 012; and field 5 consisted of composite samples 013, 014, and 015.

This trend is supported by the alpha diversity, when plotted by testing group (Figure 13). In this case the alpha diversity is the specific level of diversity intrinsic to each aliquot according to the Shannon-Weiner Diversity Index. The error bars for the samples in group 1 are longer than those of the other two groups. The two homogenized groups are much more similar to each other in terms of both error bars and interquartile ranges as well as the overall trend towards higher diversity within the single aliquots.

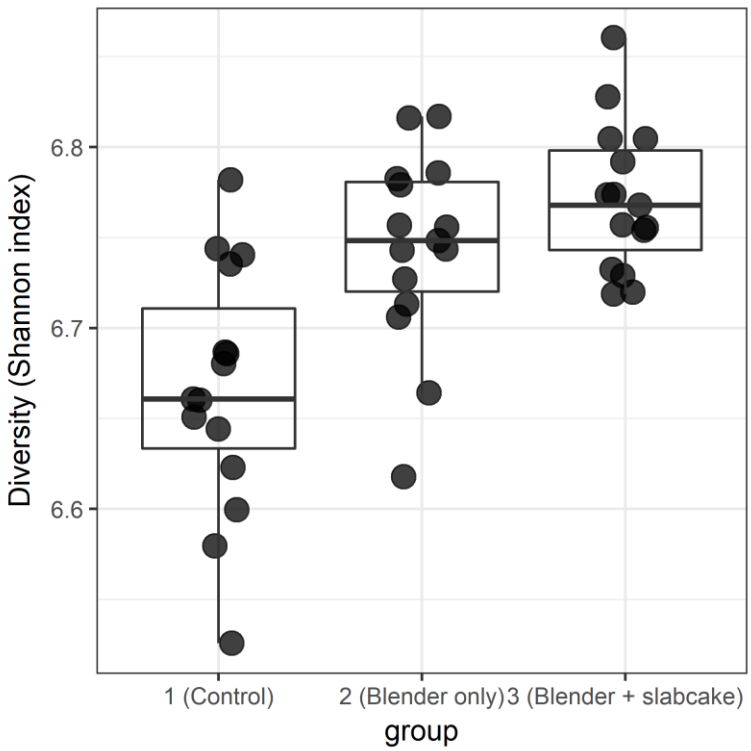

Figure 13. Alpha diversity of bacteria by treatment group (Microbiome, Inc.). This is a box and whisker plot of the specific Shannon-Diversity Index values calculated for each aliquot and grouped by treatment group. The y-axis represents a scale of values on the Shannon-Weiner Index. The higher a point (representing an aliquot) is on this scale, the higher the aliquot's level of diversity is according to the Shannon-Weiner Index.

Overall, the relative abundances of the bacterial microbiome within the aliquots appear relatively uniform when plotted by treatment group (Figure 14), though there are some aliquots particularly in group 1, the control, which are less so. It is interesting to note that though it is most pronounced in group 1, the control, and least pronounced in group 3, blender and slabcake, there is a wave-like appearance to the abundances in each treatment group overall (see especially the iii1.15_unclassified bar segments). The reason for this can be explained by plotting the data differently (Figure 15).

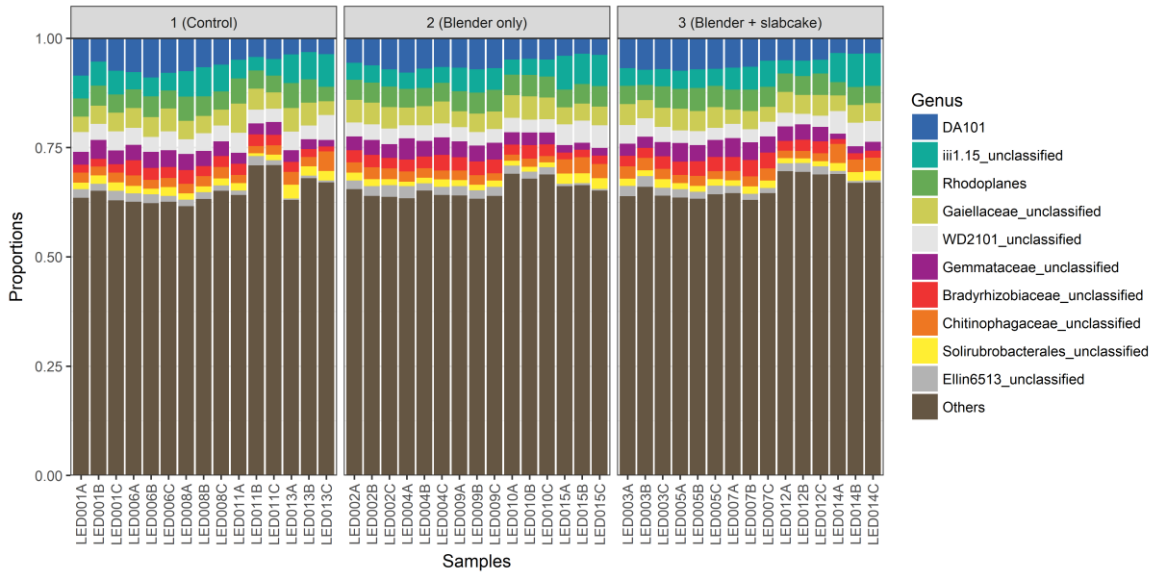

Figure 14. Barplot of relative abundances of the most abundant bacterial taxa plotted by treatment group (Microbiome, Inc.). This stacked barplot details the differences in community composition through the abundances of each of the most abundant taxa in each aliquot, grouped by treatment group. The y-axis is a proportional scale and the aliquots are grouped across the x-axis according to treatment group. Every three aliquots sharing the same number, e.g. "001" represent one composite sample. Field 1 consisted of composite samples 001, 002, and 003; field 2 consisted of composite samples 004, 005, and 006; field 3 consisted of composite samples 007, 008, and 009; field 4 consisted of composite samples 010, 011, and 012; and field 5 consisted of composite samples 013, 014, and 015.

When the relative bacterial abundances are plotted by field (Figure 15), trends become more apparent. In general, aliquots taken from field 5, are higher in *Chitinophagaceae*_unclassified, *Solirubrobacterales*_unclassified, *WD2102_unclassified*, and *iii1.15_unclassified* than are the aliquots from other fields. At the same time the aliquots from field 5 are lower in *DA101*, *Gemmataceae*_unclassified, and *Bradyrhizobiaceae*_unclassified. *Elin6513_unclassified* is virtually non-existent in field 5, whereas it is relatively uniform in the other fields, though field 1 has the highest abundance. Across all fields, the abundance of *Rhodoplanes* appears nearly uniform. Field 4 appears to have the least *Chitinophagaceae*_unclassified and the least *iii1.15_unclassified* of all the fields.

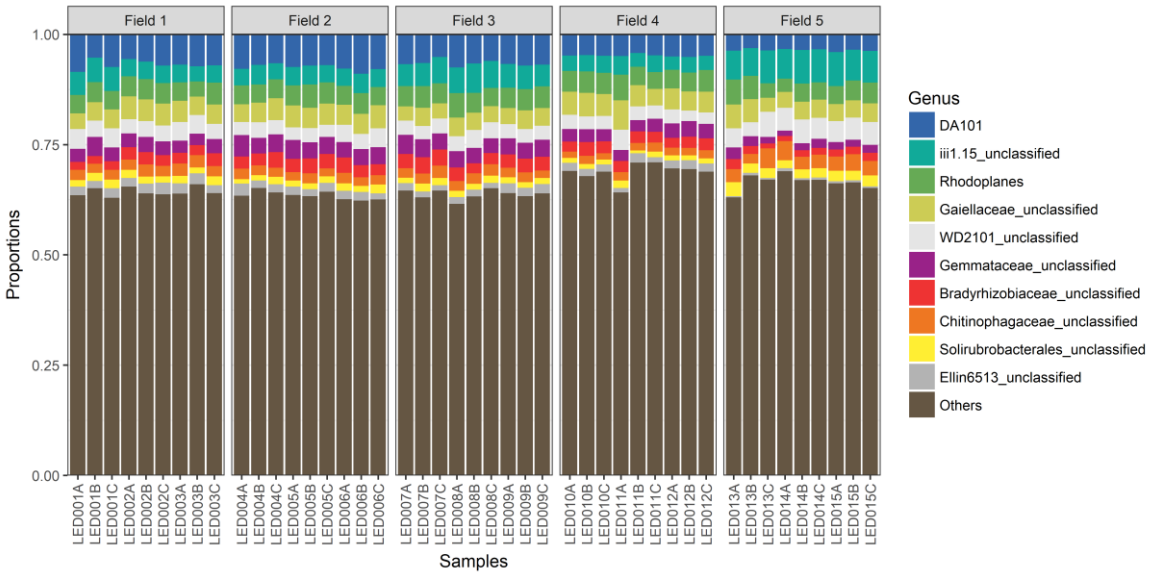

Figure 15. Barplot of relative abundances of the most abundant bacterial taxa plotted by field (Microbiome, Inc.). This stacked barplot details the differences in community composition through the abundances of each of the most abundant taxa in each aliquot, grouped by field. The y-axis is a proportional scale and the aliquots are grouped across the x-axis according to field. Every three aliquots sharing the same number, e.g. "001" represent one composite sample. The following composite samples were in group 1 (control): 001, 006, 008, 011, and 013. The following composite samples were in group 2 (blender only): 002, 004, 009, 010, and 015. The following composite samples were in group 3 (blender and slabcake): 003, 005, 007, 012, and 014.

For the most part, both of the two homogenized treatment groups tend to have higher counts of the top twelve differentially abundant bacteria. Out of the twelve, only two of the bacteria tend to be more abundant in the control group (Figure 16).

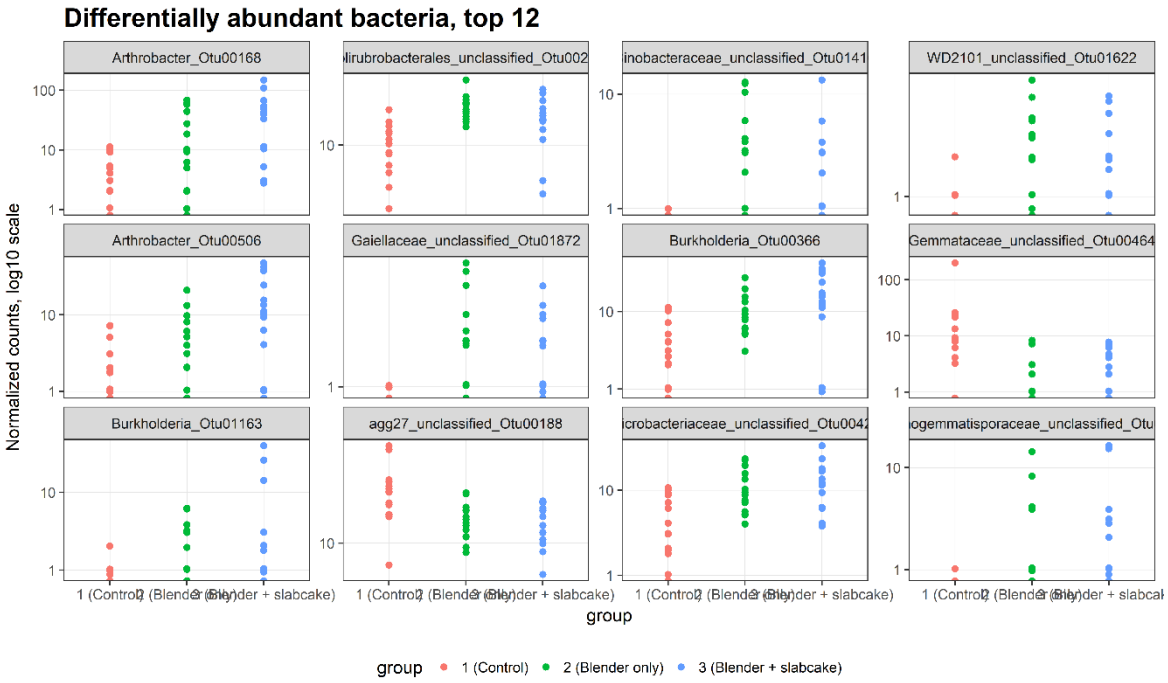

Figure 16. Differential abundance of the 12 most common bacterial genera plotted by treatment group (Microbiome, Inc.). These plots represent the counts of the 12 most differentially abundant bacteria between the treatment groups. Each plot represents one OTU and demonstrates how its abundance varies across all three treatment groups.

In order to better visualize the effectiveness of the homogenization, the species richness, or number of species present, in each aliquot was counted. Then, the specific species present in each aliquot of each composite sample were compared and the species which were present in all three aliquots were counted to come up with shared species richness (Figure 17). In two fields (fields 1 and 5), group 2, blender, (orange) had the highest shared richness. In three fields (fields 2, 3, and 4), group 3, blender and slabcake, (gray) had the highest shared richness. In all of the fields group 1, the control, (blue) had the lowest shared richness.

Figure 17. The shared species richness of the bacterial microbiome plotted by treatment group and field. This bar graph depicts the counts of species shared by all three aliquots from a given composite sample. The y-axis depicts the count of species present and the y-axis is the field from which the composite sample was collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Normalization of the values of shared species richness of the control group (group 1) enabled a direct comparison between the groups. Figure 18 shows the differences in shared species richness between the three groups with significantly higher values in groups 2 and 3 compared to the control, group 1.

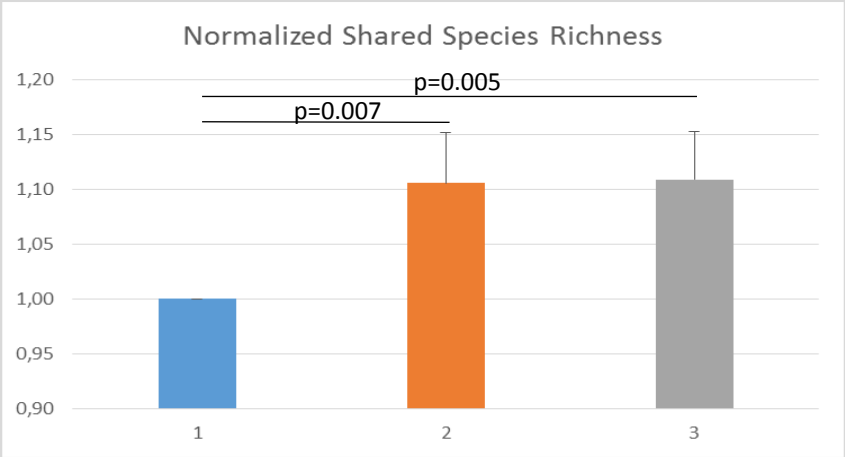

Figure 18. The normalized shared species richness of the bacterial microbiome plotted by treatment group and field. This bar graph depicts the counts of species shared by all three aliquots from a given composite sample. All values of all fields were grouped, normalized to one in the control group and calculated for the two other groups in relation to group 1. The y-axis depicts the normalized count of species present and the x-axis represents the three groups. The blue bar is the mean of composite samples in the control group, group 1; the orange bar is the mean of the blender only group, group 2; and the gray bar is the mean of the blender and slabcake group, group 3. P-values indicate significantly higher shared species richness in groups 2 and 3 compared to the control group 1.

Next, the mean bacterial species richness for each composite sample was calculated using the initial raw species richness counts for each aliquot from each composite sample (Figure 19). Though the specific values changed, the general trends remained the same as for shared species richness (Figure 17).

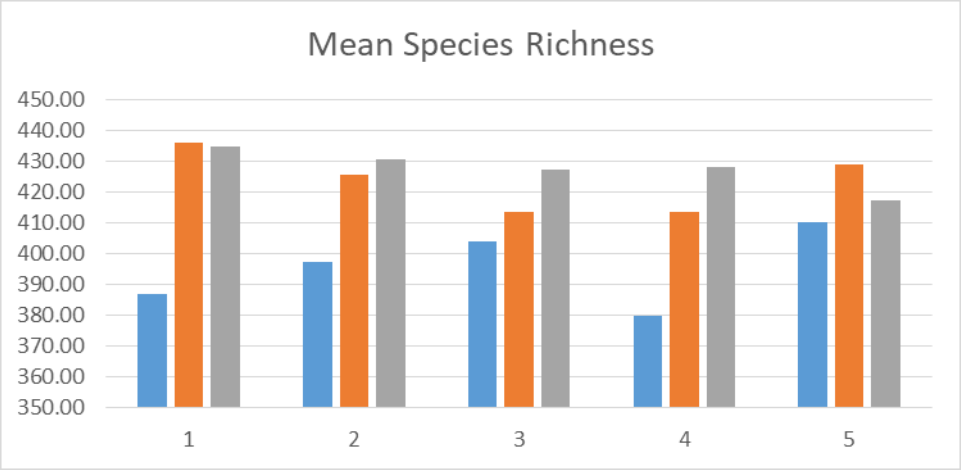

Figure 19. The mean species richness of the bacterial microbiome plotted by treatment group and field. This bar graph depicts the mean of the species richness counts of the three aliquots from each composite sample, plotted by treatment group and field. The y-axis depicts the mean of the counts of species present and the y-axis is the field from which the composite sample was collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Finally, the shared species richness was compared to the mean species richness by dividing the shared species richness by the mean species richness and multiplying by 100 to come up with a percentage, or proportional species richness (Figure 20). While generally, group 1 (blue), the control, has the lowest proportional species richness, it should be noted that this is not true for field 2. Furthermore, in all fields save for field 2, group 2 (orange), blender only, has the highest proportional richness. In field 2, group 3 (gray), blender and slabcake, has the highest proportional richness.

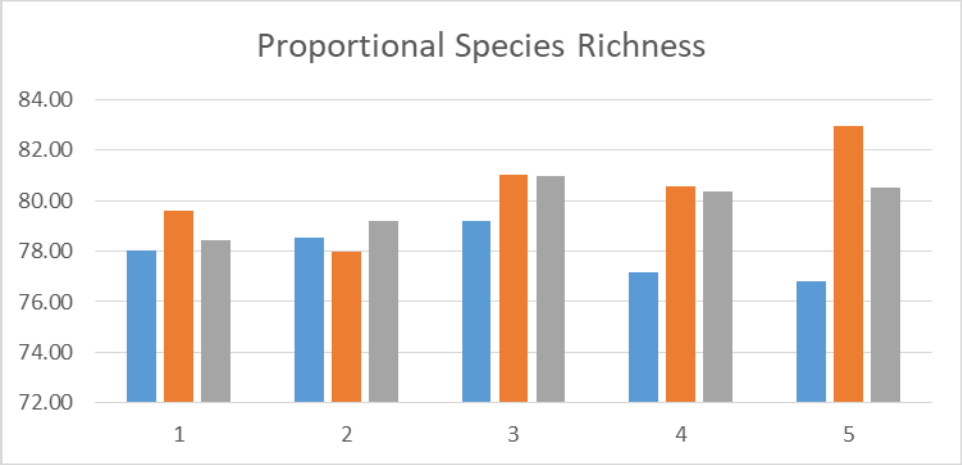

Figure 20. The proportional species richness of the bacterial microbiome plotted by treatment group and field. This bar graph depicts the proportional species richness obtained by dividing the shared species richness for each composite sample by the mean species richness of the same and multiplying this by 100 so that it is expressed as a percentage and plotted by treatment group and field. The y-axis depicts the proportional species richness by percentage and the y-axis is the field from which each of the composite samples were collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Fungi:

The aliquot, 0015A did not yield enough high quality reads during analysis of the fungi and it is for this reason excluded from all figures and discussion beyond this point.

As with the bacteria, the fungi in the aliquots do demonstrate differentiation by field when they are plotted by field. Unlike the bacteria however, the relative distances between the fungal communities are more pronounced, so clustering is not as tight. As with the bacteria, fields 4 and 5 demonstrated more distinct differentiation than did the other fields (Figure 21).

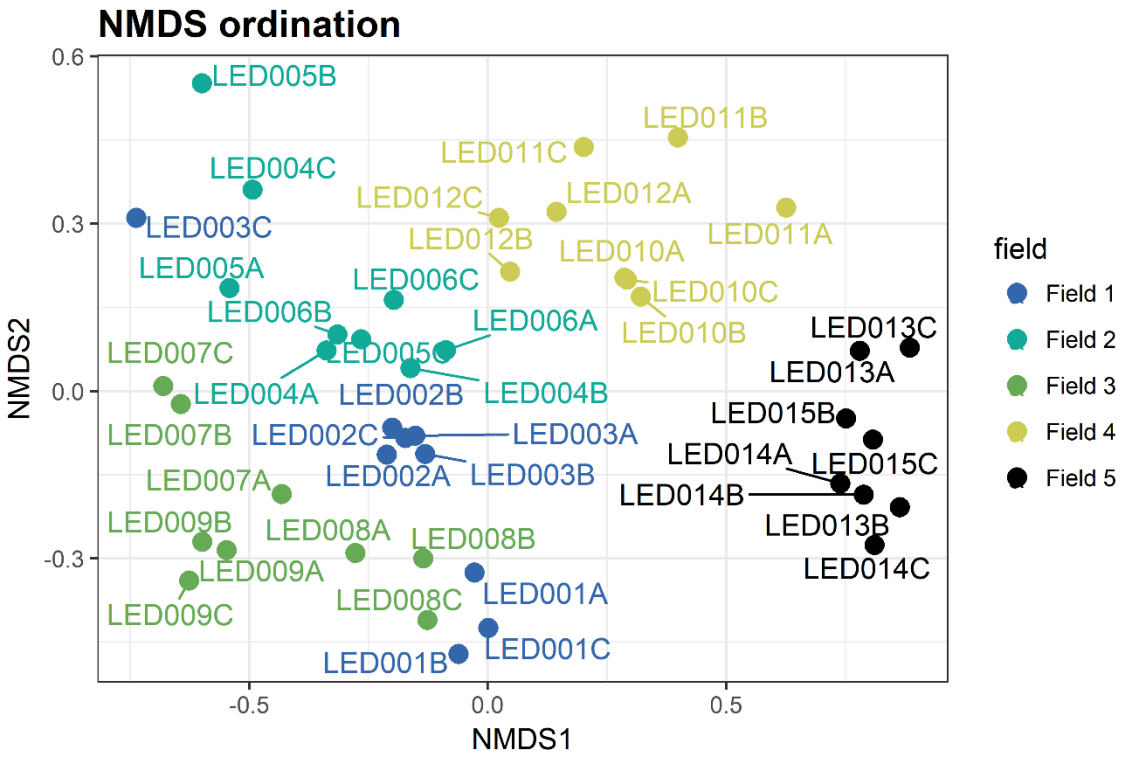

Figure 21. Beta diversity of fungi by field (Microbiome, Inc.). This plot is a visualization of the similarity in the fungal microbiome of each of the aliquots sorted by field through the non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis Dissimilarity Index values. Both axes represent Bray-Curtis Index values which provide a scale of relative distance. The closer two aliquots are within the plot, the more similar they are to one another in terms of diversity. Conversely, the farther apart the aliquots are, the more dissimilar they are. The following composite samples were in group 1 (control): 001, 006, 008, 011, and 013. The following composite samples were in group 2 (blender only): 002, 004, 009, 010, and 015. The following composite samples were in group 3 (blender and slabcake): 003, 005, 007, 012, and 014.

Unlike with the bacteria, most of the differentiation or clustering disappears when the beta diversity is plotted by treatment group. The aliquots in all treatment groups appear widely spread in general. There are only a few small, tight clusters and these only occur within the homogenized treatment groups (Figure 22).

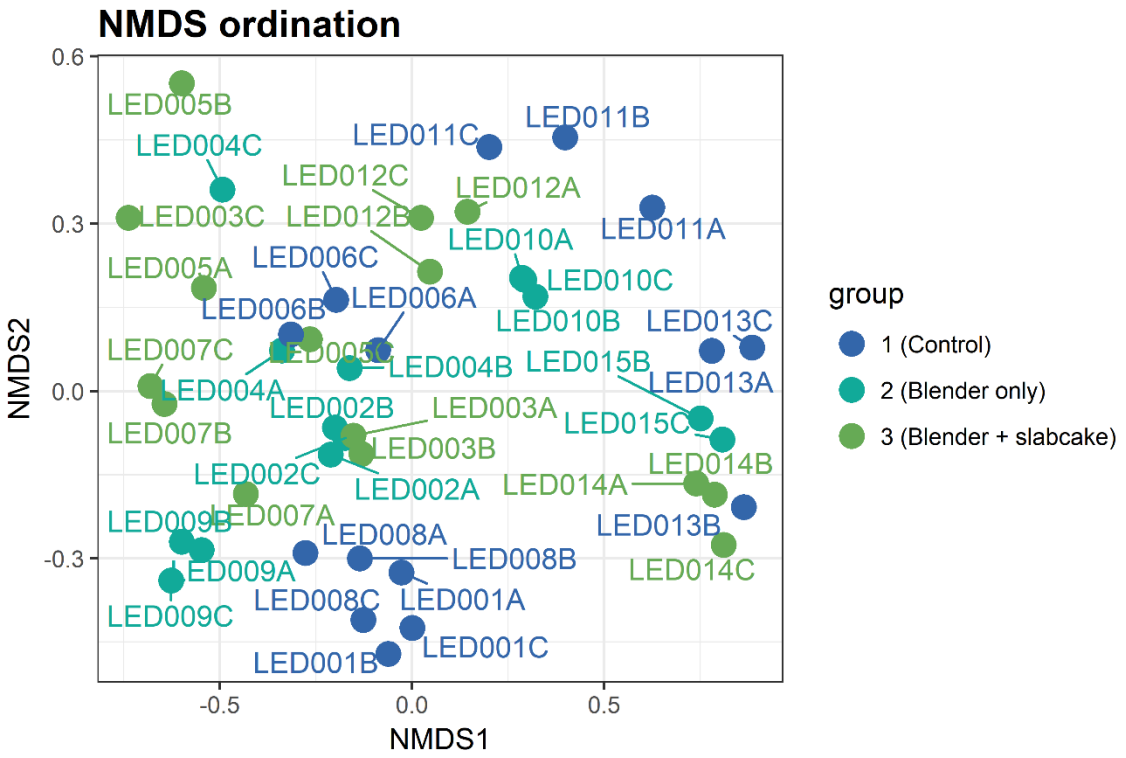

Figure 22. Beta diversity of fungi by treatment group (Microbiome, Inc.). This plot is a visualization of the similarity in the fungal microbiome of each of the aliquots sorted by testing group through the non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis Dissimilarity Index values. Both axes represent Bray-Curtis Index values which provide a scale of relative distance. The closer two aliquots are within the plot, the more similar they are to one another in terms of their diversity. Conversely, the farther apart the aliquots are, the more dissimilar they are. Field 1 consisted of composite samples 001, 002, and 003; field 2 consisted of composite samples 004, 005, and 006; field 3 consisted of composite samples 007, 008, and 009; field 4 consisted of composite samples 010, 011, and 012; and field 5 consisted of composite samples 013, 014, and 015.

Like with the bacteria, when the alpha diversity intrinsic to each aliquot is plotted, the interquartile ranges of the homogenized groups trend higher than that of the control group, however the error bars and outliers of the homogenized groups are comparable to or greater than those of the control (Figure 23).

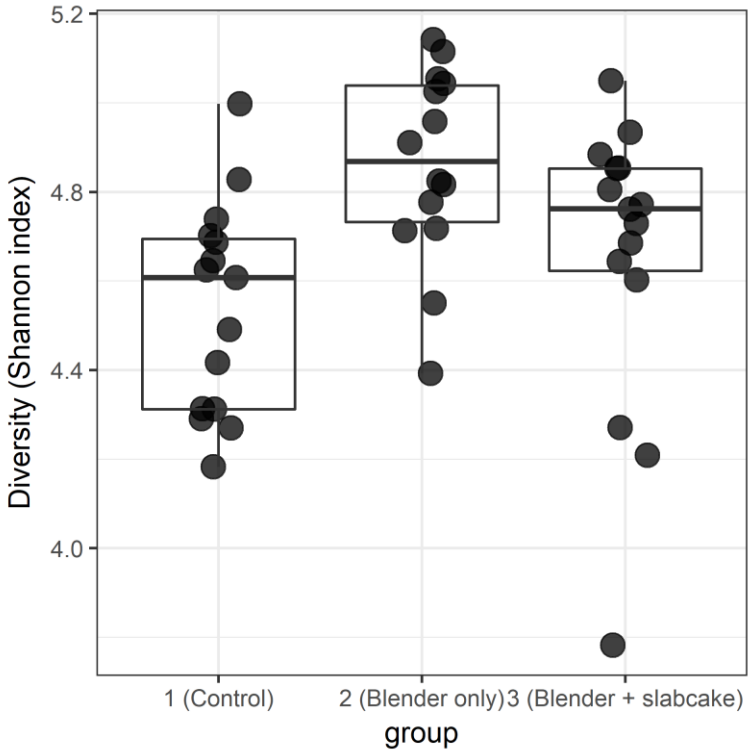

Figure 23. Alpha diversity of fungi plotted by treatment group (Microbiome,Inc.). This is a box and whisker plot of the specific Shannon-Weiner Diversity Index values calculated for each aliquot and grouped by treatment group. The y-axis represents a scale of values on the Shannon-Weiner Index. The higher a point (representing an aliquot) is on this scale, the higher the aliquot's level of diversity is according to the Shannon-Weiner Index.

As before with the bacteria, patterns are less distinct when the relative abundances of the fungi are plotted by treatment group (Figure 24). Generally, the smallest overall abundances of *Humicola*, *Mortierella*, and *Agaricomycetes*_unclassified occur in group 3 and the highest abundances of "others" occur in this group as well. The control, group 1, has the highest overall abundances of *Fusarium*, *Mortierella*, and *Humicola*. Group 2 has the lowest overall abundances of *Basidiomycota_unclassified*.

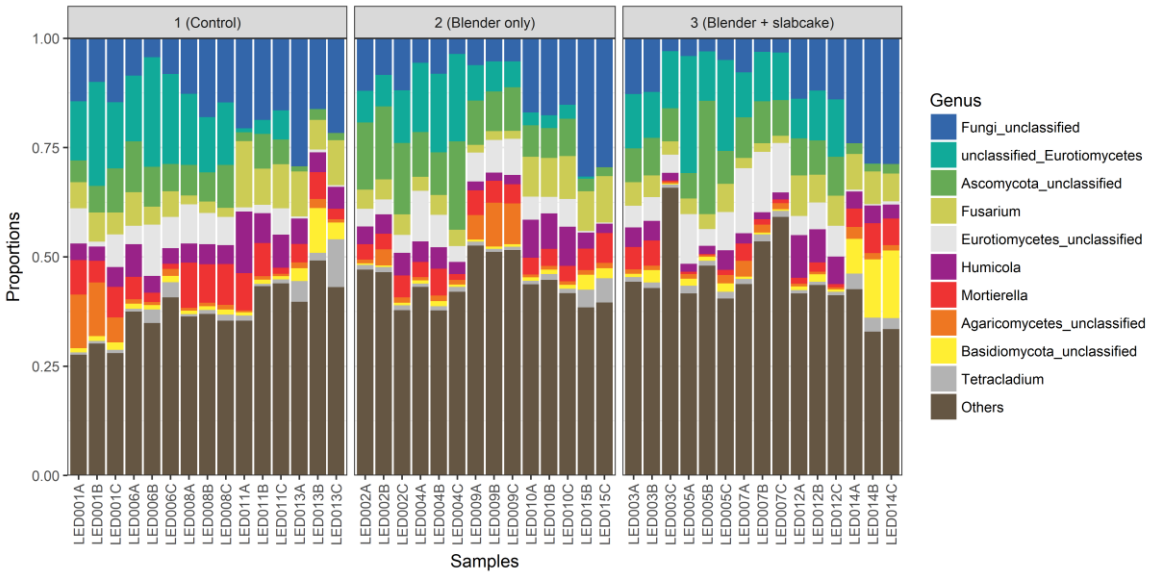

Figure 24. Barplot of relative abundances of the most abundant fungi plotted by treatment group (Microbiome, Inc.). This stacked barplot details the differences in community composition through the abundances of each of the most abundant taxa in each aliquot, grouped by treatment group. The y-axis is a proportional scale and the aliquots are grouped across the x-axis according to treatment group. Every three aliquots sharing the same number, e.g. "001" represent one composite sample. Field 1 consisted of composite samples 001, 002, and 003; field 2 consisted of composite samples 004, 005, and 006; field 3 consisted of composite samples 007, 008, and 009; field 4 consisted of composite samples 010, 011, and 012; and field 5 consisted of composite samples 013, 014, and 015.

When plotted by field (Figure 25), trends in the relative abundances of the most abundant fungi become more apparent. Field 5 has the most *Basidiomycota*, *Tetracladium*, and *Fungi_unclassified*, but also the least *Ascomycota*_unclassified and *Eurotiomycetes_unclassified*. By contrast, fields 1, 2, and 3 are characterized by relatively low levels of *Fungi_unclassified*, *Fusarium*, *Tetracladium*, and *Basidiomycota*_unclassified with relatively high abundances of *unclassified_Eurotiomycetes*, *Ascomycota*_unclassified, and *Eurotiomycetes_unclassified*. Field 4 is somewhat intermediate between fields 1-3 and field 5, having both *unclassified_Eurotiomycetes* and *Eurotiomycetes_unclassified* with very low abundances of *Tetracladium* and *Basidiomycota*; features of fields 1-3. At the same time, field 4 displays high levels of *Fungi_unclassified* and *Fusarium* which are features of field 5. Of all the fields, field 4 has the highest levels of *Humicola*.

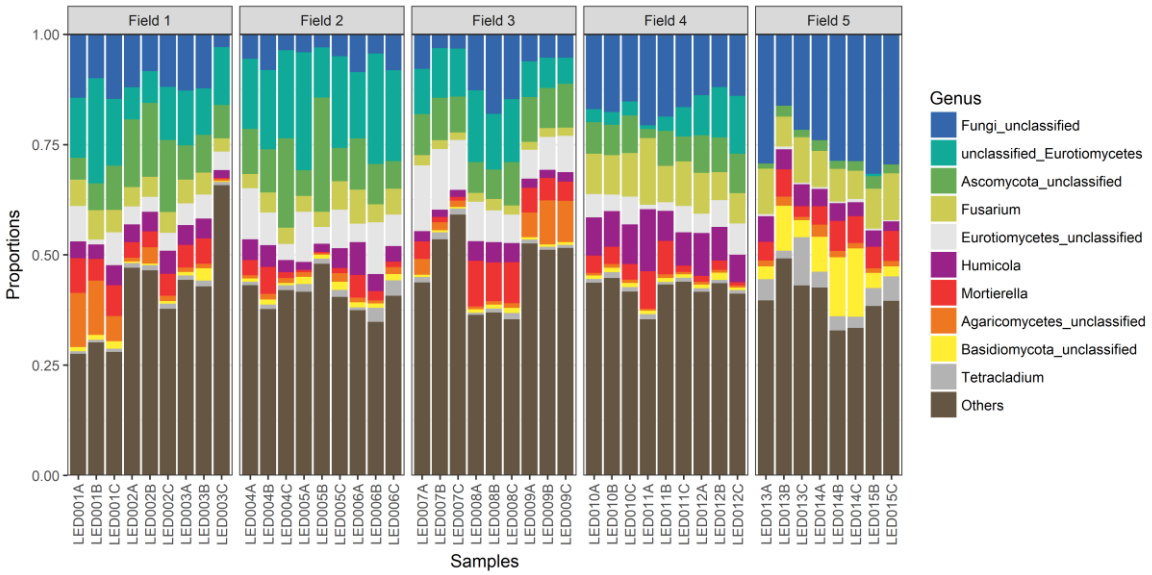

Figure 25. Barplot of relative abundances of the most abundant fungi plotted by field (Microbiome, Inc.). This stacked barplot details the differences in community composition through the abundances of each of the most abundant taxa in each aliquot, grouped by field. The y-axis is a proportional scale and the aliquots are grouped across the x-axis according to field. Every three aliquots sharing the same number, e.g. "001" represent one composite sample. The following composite samples were in group 1 (control): 001, 006, 008, 011, and 013. The following composite samples were in group 2 (blender only): 002, 004, 009, 010, and 015. The following composite samples were in group 3 (blender and slabcake): 003, 005, 007, 012, and 014.

Again as with the bacteria, the aliquots in the homogenized treatment groups tend to have higher counts of the most differentially abundant fungi (Figure 26). Three, as opposed to two in the bacteria, of the twelve fungi tend to be more abundant in the control group.

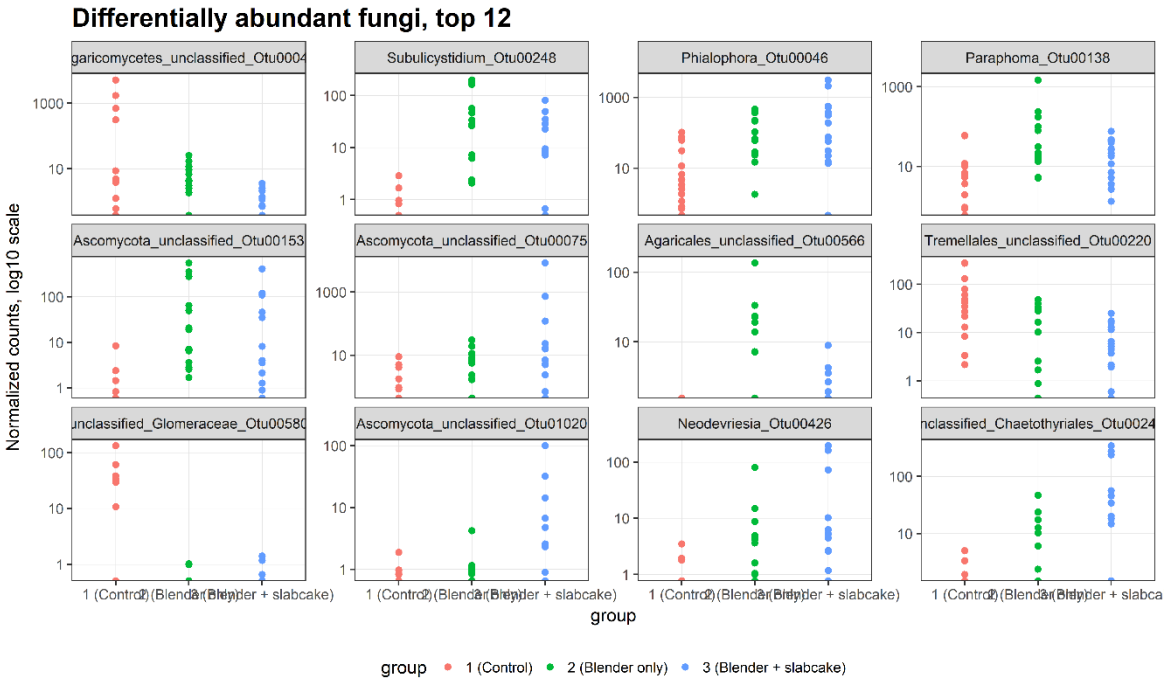

Figure 26. Abundance of the 12 most differentially abundant fungal genera plotted by treatment group (Microbiome, Inc.). These plots represent the counts of the 12 most differentially abundant fungi between the treatment groups. Each plot represents one OTU and demonstrates how its abundance varies across all three treatment groups.

As for the bacteria, to better visualize the effectiveness of the homogenization, the species richness of fungi, or number of fungal species present, in each aliquot was counted. Then, the specific species present in each aliquot of each composite sample were compared and the species which were present in all three aliquots were counted to come up with shared species richness (Figure 27). In all fields, group 2, blender, (orange) had the highest shared richness, followed by group 3, blender and slabcake, (gray) except in field 2. In all other fields group 1, the control, (blue) had the lowest shared richness.

Figure 27. The shared species richness of the fungal microbiome plotted by treatment group and field. This bar graph depicts the counts of species shared by all three aliquots from a given composite sample. The y-axis depicts the count of species present and the y-axis is the field from which the composite sample was collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Normalization of the values of shared fungal species richness of the control group (group 1) enabled a direct comparison between the groups. Figure 28 shows the differences in shared species richness between the three groups. The shared species richness in group 2 was significantly higher compared to group 1 and group 3. There was no significant difference between group 1 and group 3.

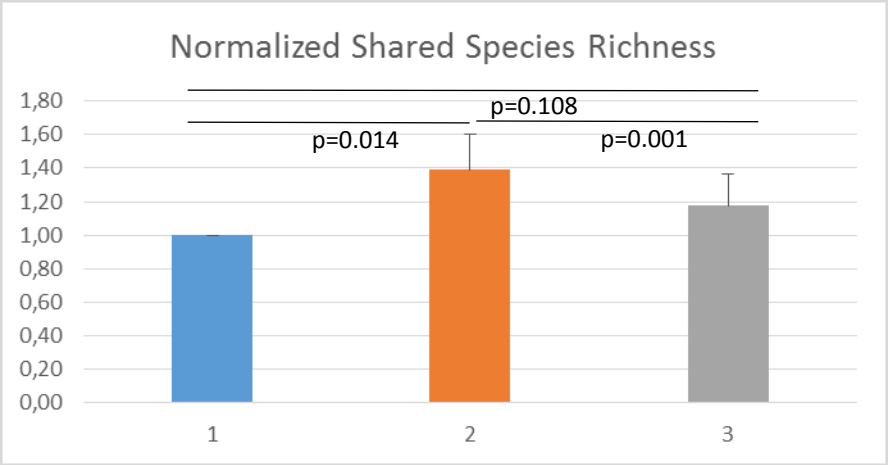

Figure 28. The normalized shared species richness of the fungal microbiome plotted by treatment group and field. This bar graph depicts the counts of species shared by all three aliquots from a given composite sample. All values of all fields were grouped, normalized to one in the control group and calculated for the two other groups in relation to group 1. The y-axis depicts the normalized count of species present and the y-axis represents the three groups. The blue bar is the mean of composite samples in the control group, group 1; the orange bar is the mean of the blender only group, group 2; and the gray bar is the mean of the blender and slabcake group, group 3. P-values indicate significantly higher shared species richness in group 2 compared to the control group 1 and treatment group 3.

Next, the mean fungal species richness for each composite sample was calculated using the initial raw species richness counts for each aliquot from each composite sample (Figure 29). Though the specific values changed, the general trends remained the same as for shared species richness (Figure 27).

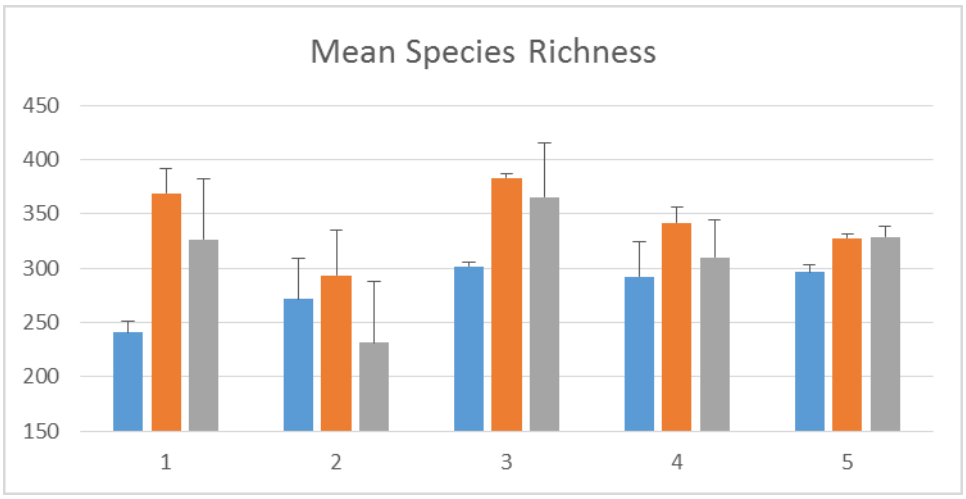

Figure 29. The mean species richness of the fungal microbiome plotted by treatment group and field. This bar graph depicts the mean of the species richness counts of the three aliquots from each composite sample, plotted by treatment group and field. The y-axis depicts the mean of the counts of species present and the y-axis is the field from which the composite sample was collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Finally, the shared fungal species richness was compared to the mean species richness by dividing the shared species richness by the mean species richness and multiplying by 100 to come up with a percentage, or proportional species richness (Figure 30). While generally, group 1 (blue), the control, has the lowest proportional species richness, it should be noted that this is not true for field 3. Furthermore, in all fields group 2 (orange), blender only, has the highest proportional richness.

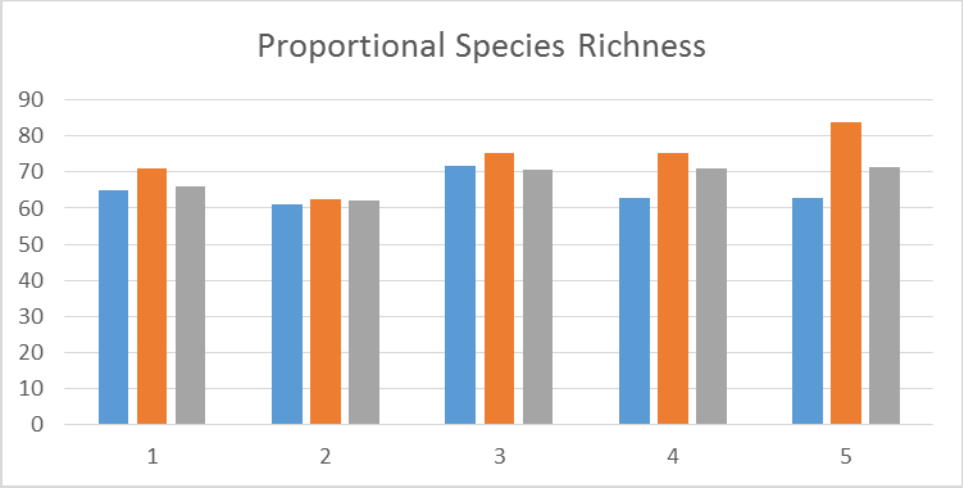

Figure 30. The proportional species richness of the fungal microbiome plotted by treatment group and field. This bar graph depicts the proportional species richness obtained by dividing the shared species richness for each composite sample by the mean species richness of the same and multiplying this by 100 so that it is expressed as a percentage and plotted by treatment group and field. The y-axis depicts the proportional species richness by percentage and the y-axis is the field from which each of the composite samples were collected. The blue bars are the composite samples in the control group, group 1; the orange bars are the blender only group, group 2; and the gray bars are the blender and slabcake group, group 3.

Discussion

The intention of this study was to demonstrate through the use of homogenization techniques, that it is possible to overcome microheterogeneity within composite soil samples and thus to achieve greater homogeneity of the microbiome between aliquots of these composite samples. This thereby increases their representativeness when compared to those from a control group not treated with any method of homogenization. The community structure and diversity of two types of soil organism were examined through the use of genomic analysis: bacteria and fungi.

Overall, the soil bacteria within the aliquots demonstrated closer community structure and there were higher levels of diversity within a given aliquot, but lower diversity between aliquots when they were homogenized using a blender (group 2) and both a blender and 2D Japanese slabcake (group 3) when compared to a control group (group 1), which received no homogenization whatsoever. Though homogenization was more effective in reducing microheterogeneity, neither method of homogenization was significantly more effective on the soil bacteria than the other.

Like the soil bacteria, the soil fungi demonstrated closer community structure and higher levels of diversity within a given aliquot but lower diversity between aliquots when homogenized when compared to a control group (group 1), but unlike with the bacteria where neither treatment with a blender only (group 2) nor treatment with a combination of a blender and 2D Japanese slabcake (group 3) was more effective than the other, with the soil fungi it was more effective to use a blender alone (group 2).

Limitations

There are a number of areas where the methods used in this study can be improved. Using the wrong shape of tool can be a source of error as it biases the sample collection towards specific portions of the sample [9]. Instead of using pipette tips to perform the coring of the secondary 2D Japanese slabcake, it might be beneficial to use small syringes. Ideally, if the closed end of such a syringe can be cleanly cut away, the plunger can then be used to push out the soil core. An implement, such as a syringe, which has straight sides would reduce the bias that pipette tips, which are tapered, have in favor of the soil aggregates and particles towards the bottom layer of the 2D Japanese slabcake [9].

It was difficult to get aggregates of uniform size using a blender. One suggestion for improving this would be to use water to wet the soil until it has the consistency of a paste prior to blending as this would make it easier to work with, but slower to separate or stratify [9].

When it comes to measuring the abundance of fungi in an area of interest, the genomic methods used in this study might not be sufficient on their own in every circumstance. With the exception of yeasts which are single-celled, most fungi are multi-cellular and cover considerable distances with their fine, filamentous hyphae which thread their way through void space around the soil particles [3]. In using DNA molecular analysis to examine the fungal hyphae present in a 20-hectare forest, scientists determined that they belonged to a single 1,500 year old organism which weighed in excess of 10,000 kg [4]. For this reason, it may be worth combining the methods of this study with others such as measures of fungal biomass [4]. Also, extraction methods to isolate DNA from environmental samples show a major impact on the taxa of bacteria and fungi isolated from different soil samples [18].

General aspects

The fields used in this study (Figure 31) are comprised of soils with a high clay content: roughly 76.4% of the combined area of the fields is made up of clay-loam soils [19] (Table 1). Clay is the constituent of soil which lends it plasticity and adhesiveness when wet [1]. The addition of too much water however, will actually bring about the separation of the soil [9]. This is because the clay diffuses and remains suspended for some time in water [1]. In order to get the most benefit out of the addition of water for ease of mixing or blending, it may be necessary to first establish the ideal water content that will provide the right consistency for use on the soil to be sampled.

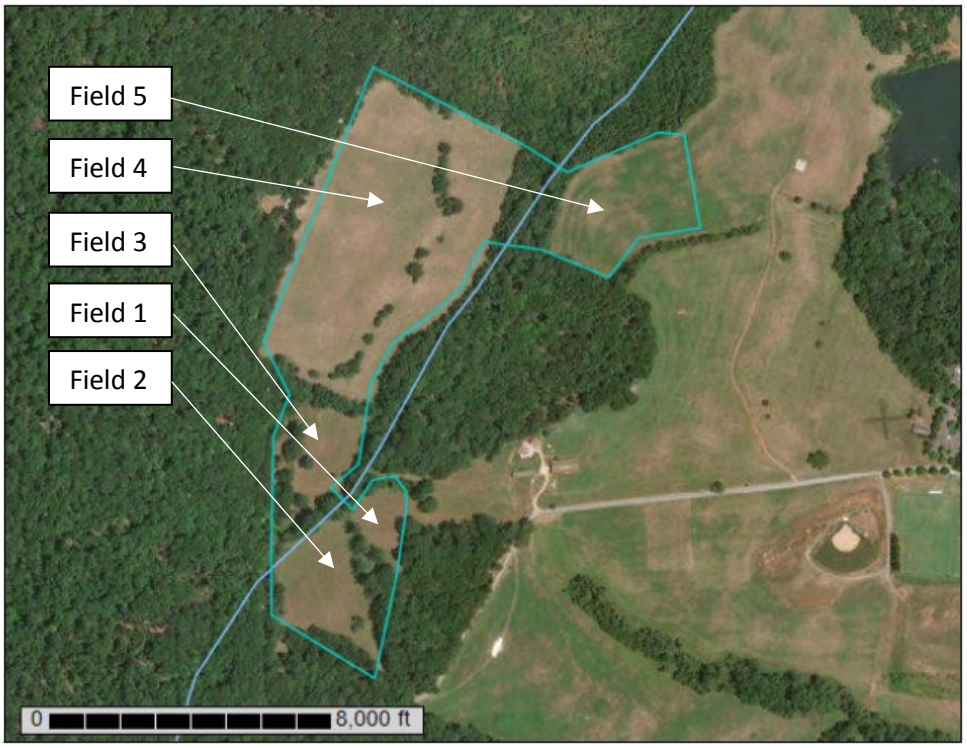

Figure 31. Fields and area of interest (modified) [19]. This satellite image depicts field 1, field 2, field 3, field 4, and field 5 at Sweet Briar College in Amherst, Virginia from which soil samples were collected for this study. The college uses these fields to grow switchgrass for the production of biofuel and they have been under cultivation since 2014 [13].

Soil	Percent of Area of Interest
Clifford clay loam, 7% to 15% slopes, severely eroded	39.30%
Clifford clay loam, 15% to 25% slopes, severely eroded	21.90%
Clifford loam 7% to 15% slopes	2.70%
Clifford loam 15% to 25% slopes	8.90%

Delanco-Elsinboro complex, 2% to 7% slopes, rarely eroded	12%
Wintergreen clay loam, 7% to 15% slopes, severely eroded	15.20%

Table 1. Summary of soils present by percentage in the area of interest from which samples were collected [19]. This table illustrates the percentage of each type of soil present in the total area of interest comprising portions of five switchgrass fields on the campus of Sweet Briar College in Amherst, Virginia. Four of the six soils present are phases of the Clifford series; one of the soils is a phase of the Wintergreen series; and one of the soils is a complex of a phase from the Delanco series and Elsinboro series. Soil series are made up of soils that have profiles that are extremely similar and phases are divisions of a series that vary in the texture of the surface horizon of the profile or other characteristics affecting usage. A complex is made up of two or more soils interwoven in such a way that they cannot be shown separately on a map [19].

As stated previously, the majority (~76.4%) of the soil sampled for this study is from deposits of clay-loam. The remaining 23.6% of the soil in the fields is loam [19]. In terms of their texture, which is to say the proportions of clay, silt, and sand particles in the soil [20], these soils are quite similar. Future studies should focus on examining the microbiome of soils with greater degrees of heterogeneity between particles in their texture.

The composition as well as the three dimensional organization of soil samples seems to have a high impact on the microbial diversity. Soil with a high clay proportion have been found to form aggregates with a prismatic or spherical shape [21]. The spherical aggregates were found to be derived from the activity of earthworms, while the prismatic do not seem to have a specific developmental process but are rather formed by a mixture of water percolation, freeze-thaw cycles and movement of soil particles [21]. Interestingly, both aggregate types further differ as the spherical type contains much more organic

matter than the prismatic type. It would be interesting to test whether the differences in the five fields tested in this study are related to differences in the type of soil aggregates.

Not only aggregate shape but also aggregate size impacts the habitats for microorganisms and defines the type of organisms in a given aggregate [22]. Also the horizontal stratification of soil impacts the diversity and composition of the soil microbiome. Hansel et al. [23] have shown that collection of a continuous profile of soil results in the discrimination of four different soil horizons with distinct differences in the bacterial and archaeal community structure between the four horizons [23]. These authors only used one continuous soil profile that was separated into the four horizons and homogenized for DNA extraction. The method by which the soil was homogenized and how the results may differ from a profile taken in the direct vicinity to define microheterogeneity as well as short-scale heterogeneity are not mentioned.

Conclusions

This study may provide a jumping off point for further exploration and experimentation in the area of soil homogenization techniques for microbiome analysis and perhaps might even provide some basis for eventually establishing standards and guidelines. Such measures are necessary as there is little emphasis placed upon collection and processing protocols currently [23]. In a recent review of literature on *Burkholderia pseudomallei* published between 1912 and 2010, it was discovered that the number of samples taken per site ranged from 2 to 100, the sample weight could range from 2 g to 1,000 g, and the sampling depth of these studies ranged from 0 cm to 90 cm [24].

There is very real danger that by using inappropriate sampling methodology, samples will fail to yield representative results leading to false conclusions and decisions. In looking at the species richness including the shared species richness and proportional species richness for both bacteria and fungi in this study, it is clear that more species may be detected when the samples are appropriately homogenized, at least in this particular instance. In the control samples for which this was not done, many species are simply not found. In the review of literature on *Burkholderia pseudomallei* [24], consensus guidelines were reached that in order to be representative, it is necessary to use 100 samples for an area of soil measuring 50 m by 50 m as in soil there can be “hotspots” of high abundance and also areas of low abundance where a given species of bacteria might

not be found at all. As this also exists in the soil on the micro level, it is imperative to come up with guidelines for exactly how to overcome the microheterogeneity of the microbiome for a wide variety of applications over a wide variety of soils.

Reference List

[1] Hilgard, EW. Soils: their Formation, Properties, Composition, and Relations to Climate and Plant Growth in the Humid and Arid Regions. New York: Macmillan. 1906

[2] Jenny, H. Factors of Soil Formation: a System of Quantitative Pedology. New York: Dover. 1994

[3] Soil Biology Primer [online]. Available: http://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/health/biology/ [10OCT2017].

[4] Brady NC, Weil RR. The Nature and Properties of Soils. 14th ed. Uttar Pradesh: Pearson. 2008.

[5] Fransson MN, Rial-Sebbag E, Brochhausen M, Litton JE. Toward a common language for biobanking. European Journal of Human Genetics. 2015; 23:22-28.

[6] Göhler A, Trung TT, Hopf V, Kohler C, Hartleib J, Wuthiekanun V, Peacock SJ, Limmathurotsakul D, Tuanyok A, Steinmetz I. Multitarget quantitative PCR improves detection and improves cultivability of the pathogen Burkholderia pseudomallei. Applied and Environmental Microbiology. 2017; 83:03212-16.

[7] Hou D, O'Conner D, Nathanial P, Tian L, Ma Y. Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. Environmental Pollution, 2017; 231:1188-1200.

[8] Keiblinger KM, Fuchs F, Zechmeister-Boltenstern S, Riedel K. Soil and leaf litter metaproteomics—a brief guideline from sampling to understanding. FEMS Microbiology Ecology. 2016; 92:1-18.

[9] ITRC (Interstate Technology & Regulatory Council). 2012. Incremental Sampling Methodology. ISM-1. Washington, D.C.: Interstate Technology & Regulatory Council, Incremental Sampling Methodology Team. www.itrcweb.org.

[10] Knappik M, Dance DAB, Rattanavong S, Pierret A, Ribolzi O, Davong V, Silisouk J, Vongsouvath M, Newton PN, Dittrich S. Evaluation of molecular methods to improve the detection of Burkholderia pseudomallei in soil and water samples from Laos. Applied and Environmental Microbiology. 2015; 81:3722–3727.[11] Heil K and Schmidhalter U. The Application of EM38: Determination of Soil Parameters, Selection of Soil Sampling Points and Use in Agriculture and Archaeology. Sensors. 2017; 17:1-44.

[12] Baldrian P. Forest microbiome: diversity, complexity and dynamics. FEMS Microbiology Reviews, 2017; 41:109–130.

[13] Mason BJ. Preparation of soil sampling protocols: sampling techniques and strategies. Las Vegas (NV): Environmental Research Center. 1992 July; EPA/600/R-92/128.

[14] Gy PM. Sampling of Dynamic and Heterogeneous Materials: theories of heterogeneity, sampling and homogenizing. Amsterdam: Elsevier. 1992.

[15] Gore S. Sweet Briar biofuel project aims to draw pollinatiors. Amherst New Era Progress. Amherst (VA) 2014 June 15.

[16] Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied Environmental and Microbiology. 2013; 79:5112-20.

[17] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing Mothur: open-source platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 2009; 75: 7537-7541.

[18] Hermans SM, Buckley HL, Lear G. Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. Molecular Ecology Resources. 2018; 18:557-569.

[19] United States Department of Agriculture, Natural Resources Conservation Service (2018) Custom Soil Resource Report for Amherst County, Virginia.

[20] Schoeneberger, PJ, Wysocki DA, Benham EC, and Soil Survey Staff. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE, 2012.

[21] Frouz J, Krištůfek V, Livečková M, van Loo D, Jacobs P, Van Hoorebeke L. Microbial properties of soil aggregates created by earthworms and other factors: spherical and prismatic soil aggregates from unreclaimed post-mining sites. Folia Microbiologica. 2011; 56:36–43.

[22] Liao H, Zhang Y, Zuo Q, Du B, Chen W, Wei D, Huang Q. Contrasting responses of bacterial and fungal communities to aggregate-size fractions and long-term fertilizations in soils of northeastern China. Science of the Total Environment 2018; 635:784–792.

[23] Hansel CM, Fendorf S, Jardine PM, Francis CA. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Applied and Environmental Microbiology. 2008; 74:1620–1633.

[24] Limmathurotsakul D, Dance DAB, Wuthiekanun V, Kaestli M, Mayo M, Warner J, Wagner DM, Tuanyok A, Wertheim H, Cheng TY, Mukhopadhyay C, Puthucheary S, Day NPJ, Steinmetz I, Currie BJ, Peacock SJ. Systematic review and consensus guidelines for environmental sampling of *Burkholderia pseudomallei*. PLOS Neglected Tropical Diseases. 2013; 7: 1-11.