MICROBIAL PROFILING OF SOIL FOR FORENSIC APPLICATIONS

By

Ethan Scott Travis Smith

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Forensic Science

2011

ABSTRACT

MICROBIAL PROFILING OF SOIL FOR FORENSIC APPLICATIONS

By

Ethan Scott Travis Smith

Soil can be of tremendous evidentiary value in forensic investigations. Historically, soil evidence has been analyzed based on physical or chemical characteristics; however, microbial analysis has recently emerged as a possible way to better characterize soil samples. Within any given soil sample there are hundreds or thousands of species of microorganisms, each differing in abundance. This variation can potentially be assayed, producing a unique and comparable microbial "fingerprint" for questioned and known samples. The aim of this research was to examine the effectiveness of real-time PCR in the analysis of forensic soil samples. This was accomplished by collecting soil from four different locations around mid-Michigan over a one year period, extracting bacterial DNA, and targeting the *16S rRNA* gene of different bacterial groups known to vary in abundance based on soil type. Several soil characteristics were examined including uniqueness among habitats, changes in bacterial communities over time, and the level of heterogeneity within a habitat. Multivariate statistical analysis was performed to determine the significance of each characteristic examined. Results showed that some habitats could be differentiated from one another using ADONIS and NMDS. Habitats had little variability at different depths; however the Agricultural Field and Marsh showed significant temporal variability. Given this, most habitats could still be distinguished from one another in a pairwise manner, which more truly reflects a forensic situation.

ACKNOWLEDGEMENTS

There are many people that need to be thanked for their contributions and help throughout the course of this project. First, I would like to express great appreciation to my advisor, Dr. David Foran. His advice and patience throughout this research, as well as help with numerous thesis revisions enabled this project to be as accurate and precise as possible. I would also like to thank Dr. Tom Schmidt and Dr. Jeff Landgraf for allowing me to use their lab's facilities and whose help in the design of primers/probes was instrumental in me completing my research. Thank you to Dr. Patrick Elia, Dr. Peter Van Berkum, Dr. George W. Sundin, and Dr. Eugene Nester who graciously provided known DNA samples for this research free of charge. Also many thanks to Brian Graff of MSU's Department of Crop And Soil Sciences, and the Fenner Nature Center for allowing me to sample from their grounds. I am also very thankful to Dr. James Hallett whose guidance on the statistical analysis allowed for the most accurate and robust examination of the data possible. Thank you to Dr. Chris Smith for agreeing to be on my committee.

I would also like to thank Dr. James Bruce and Dr. Scot Dowd. Dr. Bruce helped facilitate my interest in research as my advisor at Washington State University. Dr. Dowd furthered my knowledge in the area of DNA analysis and inspired me to pursue my interests in the area of soil forensics.

Lastly, I am so lucky to have such good friends and family. My colleagues in the forensic biology laboratory have offered advice and support on countless occasions. My wife has encouraged me throughout this long process and shown great patience in a sometimes very stressful environment. Finally, my parents and family have provided endless support and advice. I owe much of where I am now to my parents Mike and Clare Smith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES CONTINUED

INTRODUCTION

Overview

Soil can have important evidentiary value in forensic investigations wherein questioned and known samples can be either differentiated or shown to have a common origin. The ability to associate soil found on items such as a shoe, shovel, car tire, or clothing, with suspected geographic locations could help include or exclude a suspect's involvement in a crime. There are many examples of soil evidence being used in criminal cases (reviewed by Marumo, 2002). One of the first occurred in 1904 (Cengage, 2006). Georg Popp, a German chemist, was called to the murder scene of Eva Disch, a young woman who was found in a field having been strangled with her own scarf. Preliminary evidence allowed police to narrow the search to one primary suspect, Karl Laubach. Since the murder took place in a field, Popp examined the legs of the pants Laubach was wearing the day of the murder. Multiple layers of soil were retrieved and microscopically examined for their physical appearance and mineral composition. One of the soil layers was similar to samples collected from the scene, and the mineral composition of a different layer was consistent with those of the mud leading away from the scene. When confronted with this evidence, Laubach confessed to the murder.

While soil evidence has been used in court, soil analysis, along with many other forensic techniques, has recently been called into question (National Research Council, 2009). As with all scientific evidence, soil analysis must withstand *Daubert* challenges, including being generally accepted within the appropriate scientific community and the ability to establish quantitative error rates for the techniques used. In this regard, a

technique that meets these criteria would be of great utility in cases where soil evidence plays an important role.

Traditional Forensic Soil Analysis

Historically, soil analysis has been accomplished through physical or chemical examination (reviewed by Marumo, 2002), including the color of the soil, particle size distribution, and density measurements. Color determinations traditionally employ the Munsell Color System that provides indices of hue, value, and chroma using a spectrophotometer. Croft and Pye (2004) measured additional color parameters using L*a*b indices (which provide color determinations based on their position on a 3-D color sphere), and measured the percentage of light reflected over the visible wavelength producing a reflectance graph. Particle size distribution was examined using laser granulometers, which measure the diffraction of light to determine the volume of the soil particles. Examining color and particle size distribution, the authors were successful in distinguishing several different soil types. Density gradients, wherein soil is placed in a tube containing liquids of differing density and centrifuged at high speed to separate the particles, have also been used to help differentiate soils (Nute, 1975; Dudley, 1979; Petraco and Kubic, 2000).

The components of soil, such as organics, minerals, oxides, and elemental composition have also been used to distinguish between soil types. Typically, organic matter is removed in order to analyze particle size and color, but scientists have taken advantage of its removal as another way to identify soil, by measuring loss of organic matter after ignition in a furnace or after decomposition with hydrogen peroxide

(Wanogho *et al.*, 1987). Mineral composition can be analyzed using X-ray diffraction (Ruffell and Wiltshire, 2004; Rawlins *et al.*, 2006), which is non-destructive, allowing for additional testing if sample is limited. A scanning electron microscope (SEM) coupled with an energy dispersive X-ray spectrometer is often used to determine elemental composition (Zadora and Brozek-Mucha, 2003), while X-ray fluorescence spectrometry examines free oxides in the soil (Marumo, 1989). Alone, these techniques are not necessarily highly discriminatory, but when several are employed, accuracy of identification has been shown to increase (reviewed by Marumo, 2002).

Examination of plant material, including pollen, has also proven useful in forensic investigations. Plants not only contribute to the organic content of soil, but can be examined by light microscopy or SEM to help determine the origins of a soil (Marumo, 1991). Rawlins *et al.* (2006) tried to determine the provenance of plant material for forensic purposes by characterizing molecular changes in lignin (a chemical compound found only in certain types of plants) after chemical modification. Different vegetation groups produced unique chemical profiles. Microscopic examination of pollen granules involves observing pollen grain type as well as grain frequency (Horrocks, 1999; Horrocks, 2004). This has helped in multiple forensic investigations (Horrocks and Walsh, 2001; Bull *et al.*, 2006).

Unfortunately, there are several problems associated with traditional soil analysis in a forensic context, such as lengthy preparation, subjectivity of analysis, and sample size. For example, in color and particle size distribution tests, the soil must be cleaned of all organic matter by running it through a sieve several times, ignition in a furnace, or treating with hydrogen peroxide, and studies have shown that the use of different

techniques to prepare samples, such as dry versus wet sieving, can produce inconsistent results (reviewed by Marumo, 2002). Contributing to the complexity, transfer of soil to a suspect can be affected by grain size, which has been shown to cause some variation in color measurements (Croft and Pye, 2004). Results are often subjective, requiring extensive training and creating the potential for decreased reproducibility. Also, since many physical and chemical techniques reveal what are functionally class characteristics, soils from different areas can appear the same, preventing the individualization of a particular case sample. Finally, a large amount of soil is sometimes required to perform a physical examination, which might not be available in forensic cases.

Microbial Community Analysis

More recently, microbial analysis has emerged as a possible way to better characterize soils. Within a soil, there are thousands of species of microorganisms, with different groups having differing abundances not only in a single soil sample, but among habitats (Spain *et al.*, 2009). These differences can potentially be targeted and assayed, producing a unique microbial profile for a given soil sample.

Previous research on microbial analysis of soil has involved a variety of assays, including denaturing gradient gel electrophoresis (DGGE) (Muyzer and Smalla, 1998), amplified fragment length polymorphism (AFLP) (Franklin and Mills, 2003), and restriction fragment length polymorphism (RFLP) (Widmer *et al.*, 2001). Microbial analysis has even been employed to study the effect of decomposing animals on microbial communities in the surrounding soil (Carter *et al.*, 2007; Stokes *et al.*, 2009). DGGE separates polymerase chain reaction (PCR)-amplified DNA fragments using

chemicals that partially denature the DNA into its single-stranded components, affecting its electrophoretic mobility. The point at which the DNA denatures is directly related to its sequence, meaning that DNA molecules with different sequences will electrophorese differently. Once the PCR fragments are separated on the gel, individual bands can be excised and sequenced to determine which bacteria are contributing to the profile. This technique has been used in combination with PCR to assess the abundance of different bacteria within a microbial community (Heuer *et al.*, 1997). However, there are several problems with DGGE. Although different soil samples produce different patterns on the gel, each band must be analyzed by additional steps before the bacterium from which it originated can be determined. Furthermore, identical samples run on different gels can produce different images depending on the denaturing conditions. Finally, both co-migration of different DNA molecules that appear as one band in the gel, and micro-heterogeneity of gene targets with multiple copy number, can lead to a misrepresentation of the relative amount of bacteria present in the sample.

AFLP involves first digesting total DNA from a sample using restriction enzymes that cut the DNA at specific locations along the genome. The digested DNA is amplified using dye-labeled primer sets that target random locations in the genome. Amplified fragments are subjected to capillary electrophoresis where the fluorescence from the amplicons produces peaks on a corresponding electropherogram. However, this technique is not locus-specific, nor is it specific for a particular bacterial group; hence there is still an unknown aspect to the assay. In RFLP, soil DNA can be digested prior to amplification (although many cells are usually required for a robust examination, so performing PCR before restriction digestion often allows for better resolution of the

bacteria present), and restriction fragments are analyzed on an agarose gel. While this technique can be locus-specific—lending itself to selection of certain bacteria by targeting loci specific to certain groups—the profiles generated do not specify which bacterium contribute to the bands observed. The complexity of the banding patterns also makes them difficult to reproduce and interpret.

Soil Research at the Forensic Biology Laboratory of MSU

Extensive research into bacterial profiling of soil samples for forensic application has been conducted at the Forensic Biology Laboratory of Michigan State University. The goal of this research was to examine several important factors that need to be determined before microbial profiling can be used as a means of characterizing soil samples. These include inter-habitat variability to determine if habitats can be differentiated from each other, and intra-habitat variability to determine the extent to which bacteria in soil vary temporally and spatially.

In the initial studies, the effectiveness of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the typing of forensic soil samples was investigated. T-RFLP provides high resolution of bacteria in soil through generation of DNA fragments of variable sizes that can be separated via capillary electrophoresis and visualized as different peaks on an electropherogram. Meyers and Foran (2008) analyzed soils from multiple habitats using this technique by assaying the universal 16S ribosomal RNA (*16S rRNA*) gene and generating profiles that encompassed all bacteria present. The profiles were used to generate similarity indices (for all possible pair-wise comparisons) that were calculated by determining the number of peaks that two soils had

in common, with zero indicating that they shared no peaks and one signifying they shared all peaks. Results indicated that the month in which habitats were compared to each other made a statistically significant difference in the similarity index. Intra-habitat temporal variability had a greater affect on similarity indices in the spring, with the Agricultural Field being the only habitat to show significant *individual* changes over time. No significant difference was present when comparing similarity indices of soil collected different directions around the main collection site. This indicated that time resulted in more heterogeneity than position. However, targeting all bacteria led to extremely complicated datasets, with profiles that most likely contained too much information to repeatedly differentiate between two soils, and causing some bacterial strains to not be reproducibly assayed.

This approach was modified by Lenz and Foran (2010), who narrowed the species analyzed to those of the genus *Rhizobium*—bacteria widespread in soil that require legume plant hosts to propagate—by targeting the recombination A (*recA)* gene. *RecA* was chosen as it is one of the most highly conserved bacterial genes, being essential for DNA repair, yet still has hypervariable regions useful for differentiating specific bacterial groups. The goal was to decrease the complexity of the T-RFLP profiles, and by doing so, increase reproducibility and the ability to differentiate soils. The *recA* gene was amplified using soil DNA previously extracted by Meyers and Foran (2008), and was subsequently purified, digested, and analyzed by capillary electrophoresis. Multivariate statistical analysis showed that habitats could be distinguished from one another, especially when only two habitats were compared at a time, which more closely reflects a forensic situation where the prosecution argues that evidentiary soil originated from one

location while the defense argues it originated from another. Results showed improvement versus targeting all bacteria, likely stemming from a more robust statistical analysis method and simplification of the T-RFLP profiles targeting a smaller subset of bacteria.

Overall, the research conducted by Meyers and Foran (2008) demonstrated that targeting all bacteria produced complicated profiles, but still showed that soils from the same habitat on average were more similar to each other than those from different habitats. Lenz and Foran (2010) improved resolution of these habitats by narrowing the bacterial target to a single genus, using a more robust statistical technique, and making pairwise comparisons. However, while T-RFLP can be very informative, other techniques that can assess the relative amounts of different bacterial groups and that allow for statistical significance to be attributed to the results may prove even more useful.

The 16S rRNA Gene and ARB Software

The *16S rRNA* gene is highly conserved among bacteria, yet still contains hypervariable regions that allow for species identification. It is a roughly 1500 base pair sequence whose use for differentiating bacteria was greatly expanded in the 1980's as sequencing technologies improved (Woese *et al.*, 1985). In fact, many bacteria that were initially classified based on phenotype have been reclassified based on 16S sequence. GenBank®, the largest database of nucleotide sequences (maintained by the National Institutes of Health) with more than 100 million different sequences, contains over one million *16S rRNA* gene sequences. This allows many bacterial species to be

differentiated, but for the same reason it makes primer development very daunting, as sequences must be aligned in order to find variable regions of interest. This requires computer based analysis, and illustrates the complexity of the microbiological landscape, as seen in Meyers and Foran (2008) where universal primers targeting the *16S rRNA* gene generated very complicated profiles.

While sequence similarity searches can be done with the Basic Local Alignment Search Tool (BLAST), which searches GenBank®, sequences must still be aligned and analyzed manually to determine regions suitable for primer design. Instead, ARB (from Latin *arbor*, tree) software (Ludwig *et al.*, 2004) can be used to automatically generate 18 base-oligonucleotide primer sequences conserved among bacterial groups of interest using the SILVA (from Latin *silva*, forest) reference database (Pruesse *et al.*, 2007). This database contains quality controlled, near full length, aligned rRNA datasets from *Bacteria, Archaea,* and *Eukarya*. By scanning the entire database, ARB can also be used to exclude bacterial sequences that are not in the group of interest, increasing specificity of the primer set and confidence that cross-reactivity with other groups does not occur. However, there are some limitations in using the 16S gene for species identification, such as the inability to differentiate closely related organisms (reviewed by Clarridge (2004)). In these cases, the 16S-23S intergenic transcribed spacer (ITS) region or another locus specific to certain species is typically assayed (Daffonchio *et al.*, 2003).

Bacterial Groups Selected for Microbial Community Analysis

There are many bacterial groups present in different types of soil. One group of interest is Rhizobia (in the class α-proteobacteria), found in virtually all soils and well

characterized taxonomically. Rhizobia are essential in agriculture due to their ability to form symbiotic relationships with legumes and fix atmospheric nitrogen. They encompass roughly 73 species within 13 genera (Weir, 2010). The genus *Bradyrhizobium*, consisting of eight different species, was named for its slow growing, non-acid producing phenotype as well as additional genotypic traits (Jordan, 1982). *Bradyrhizobium japonicum* became the type species and is a well characterized symbiont with soybean.

The genus *Burkholderia* (in the class β-proteobacteria) is also commonly found in soil, as well as ground water. Some species serve as plant pathogens, others are opportunistic pathogens in cystic fibrosis patients, and still others have been shown to protect plant seeds from invasive bacterial species in the soil (reviewed by Parke and Gurian-Sherman, 2001). It is comprised of roughly 34 species (Coenye & Vandamme, 2003), but has a very complex taxonomy with many closely related species.

The phylum *Acidobacteria* has been shown to be in high abundance in soils from a number of different environments (Barns *et al.*, 1999; Gremion *et al*., 2003; Lee and Cho, 2009). Acidobacteria is broken into several subdivisions or subphyla, with groups I and IV being two of the more prominent. Eichorst *et al.* (2007) found that acidobacteria constituted up to 6% of total bacterial rRNA, with group I making up roughly 27% of acidobacterial strains isolated from soil in Michigan. Both Eichorst *et al.* (2007) and Sait *et al.* (2006) determined there was a correlation with soil pH and the abundance of *Acidobacteria* group 1. Variation in bacterial abundance can aid in the ability to differentiate habitats.

The genus *Agrobacterium* (in the class α-proteobacteria) is in the same family as rhizobia and is also found within many soils. This genus causes several plant diseases including hairy root and crown gall disease (reviewed by Escobar and Dandekar, 2003). It is widespread in soils, with different species targeting different plants.

Overall, these bacterial groups encompass a wide variety of habitats and function. Their abundance aids in the ability to examine them in different soil types. Janssen (2006) surveyed 21 *16S rRNA* gene libraries from different soil bacterial communities and found that the phyla *Proteobacteria* and *Acidobacteria* constituted the most abundant bacteria. In addition, β-proteobacteria and α-proteobacteria contributed the most sequences seen within the phylum *Proteobacteria*.

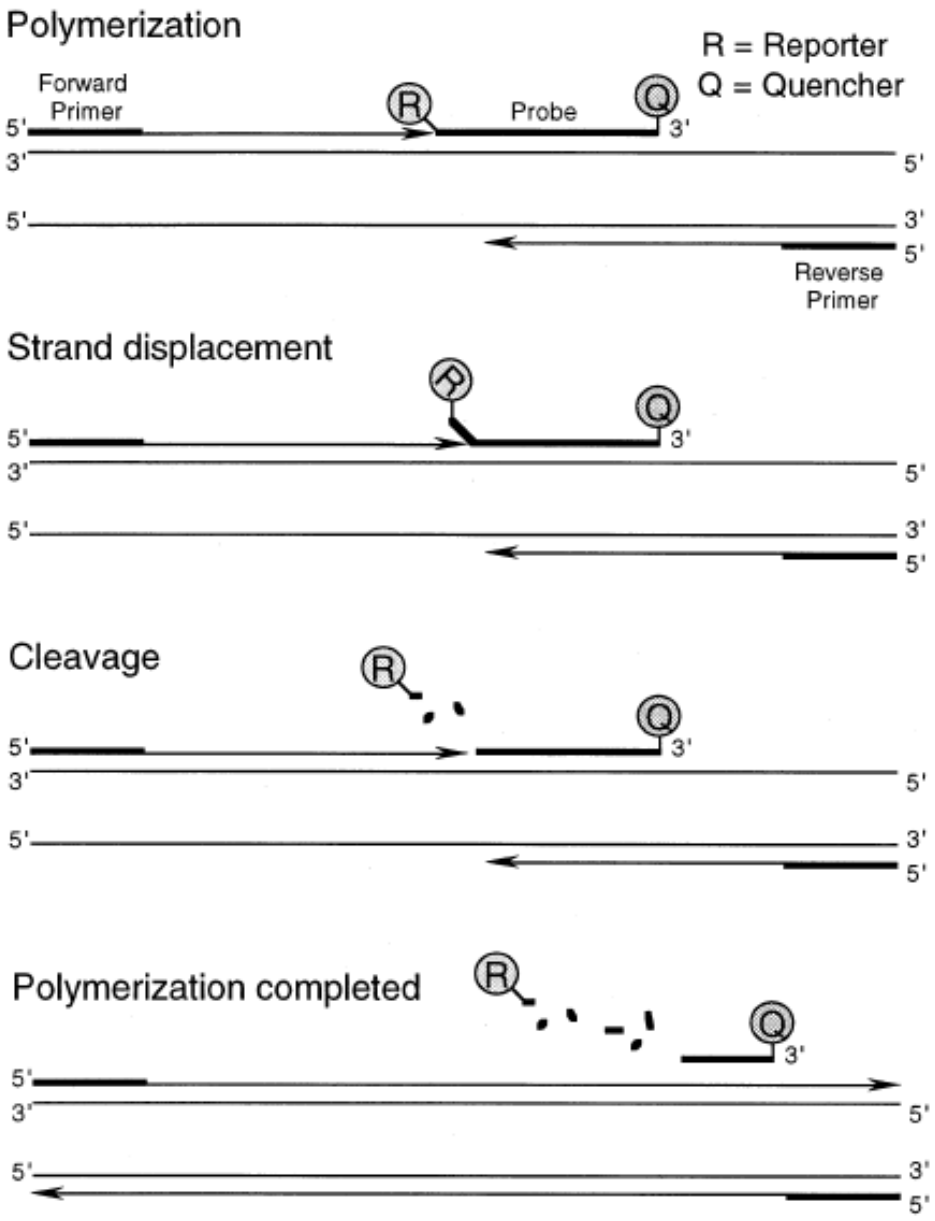
Real-time PCR Analysis

Real-time PCR was first described in the early 1990's by Higuchi *et al.* (1992). Instead of visualizing the amount of DNA on a gel at the endpoint of PCR, real-time PCR tracks the amplification of DNA throughout the PCR process by fluorescent technology. As DNA amounts increase with each cycle, so does the amount of fluorescence detected. The point at which fluorescence crosses a threshold of detection is referred to as the cycle threshold (C_T), and occurs during the exponential phase of the reaction (i.e. doubling of DNA product each cycle). The more initial DNA, the earlier the fluorescence will cross the cycle threshold, and the lower the resulting C_T value. This provides the basis for comparison to other samples, and when run with standards of known concentration can be used to determine the initial amount of DNA in an unknown sample. While regular PCR can typically only detect a 5–10-fold difference in DNA amount, real-time PCR is

sensitive enough to detect a 2-fold difference (Applied Biosystems). This allows for a much more quantitative assessment of DNA yield.

There are several fluorescent technologies available for real-time PCR. 5' nuclease assays are commonly used and take advantage of the 5' nuclease activity of *Taq* polymerase first described in relation to PCR product detection by Holland *et al.* (1991). By using dual-labeled hybridization probes (Lee *et al.*, 1993; Livak *et al.*, 1995) that have a fluorogenic dye on the 5' end whose fluorescence is quenched by a quencher dye at the 3' end, fluorescence is achieved once *Taq* degrades the probe, separating fluorophore from quencher (Figure 1). More fluorescence is detected as cycle number increases due to more DNA being available for the probe to bind (Heid *et al.* 1996). An advantage of this technique is its additional specificity for the DNA target. Primer dimers, even if they do form, are not of concern because the probe needs the specific DNA target to bind. Additionally, the amplicons size is usually no larger than 300 bp, allowing for a more robust amplification of highly degraded samples.

Figure 1. Diagram of 5' nuclease activity of *Taq* polymerase on fluorogenic probe



Once the probe hybridizes to the DNA target, the 5' nuclease activity of *Taq* polymerase separates the fluorogenic reporter dye from the quencher and fluorescence is achieved. As more DNA is made, more fluorescence is observed. (Image from Livak *et al.*, 1995)

Absolute quantification is achieved when standards of known concentration are used to determine the exact amount of initial DNA in an unknown sample; however, real-time PCR does not always have to be absolute to be informative. Relative quantification is accomplished through comparison of C_T values among samples to determine the initial amount of the DNA target of interest in relation to either an internal standard or total DNA in the system.

Considerations for the Forensic Comparison of Soils

Although real-time PCR is widely accepted in forensic and other scientific communities, and has previously been used to identify microbes in soil (Gruntzig *et al.*, 2001; Hristova *et al.*, 2001; Duodu *et al.*, 2005), comparison of soils using this technique has not been extensively tested in a forensic context. Any technique attempting to characterize soil bacteria, including real-time PCR, must reflect several factors such as differences in the microbial community between habitats (inter-habitat variability), and within a habitat over time (intra-habitat temporal variability) and over different distances and depths (intra-habitat spatial variability). If soils from all habitats are similar, there would be no way to determine if an unknown sample came from a particular location. Temporal variability is important to consider since it will most likely be several days, weeks, or even months before known or unknown soils are collected. Large changes in bacterial composition over time would make it hard to link soil from a crime scene to soil collected from a suspect or victim. Spatial variability can help determine if habitats are highly variable in a close proximity. Variability in microbial communities (both temporal and spatial) has been examined using several techniques other than real-time

PCR (Bell *et al.*, 2008; Cernohlavkova *et al*., 2009; Fuka *et al*., 2009); while, Suzuki *et al.* (2000) used real-time PCR to measure abundance of different bacteria in seawater, and found that it produced results very similar to other commonly used techniques. In general, the use of real-time PCR to measure relative abundance of bacteria in soil has not been extensively studied.

Multivariate Statistical Analysis

Using the correct statistical analysis is vital within any scientific study. A multivariate technique is required when there are multiple variables being examined. For instance, Lenz and Foran (2010) used non-metric multidimensional scaling (NMDS) to analyze T-RFLP profiles. NMDS is useful for a wide variety of data since it does not assume linearity or normal distribution of the dataset like other multivariate techniques, such as principal components analysis. In addition, NMDS provides a visual representation of multivariate patterns among observations, which could prove useful when attempting to translate the results in a courtroom. However, NMDS does not rigorously express the nature and degree of uncertainty concerning the dataset, which would be preferred in a forensic context. Non-parametric multivariate ANOVA based on dissimilarity (ADONIS), which analyzes the variance within a dataset, is one way to accomplish this. Similar to ANOVA, ADONIS is used to produce a p-value that can indicate if there are significant differences in multivariate datasets.

The Utility of Real-time PCR in Forensic Soil Analysis

The ability to differentiate soils becomes paramount when attempting to establish the origin of soil evidence. The aim of the research presented here was to determine the utility of real-time PCR in measuring the relative abundance of bacteria in order to differentiate forensic soil samples. This approach has major advantages over methods that use regular PCR, wherein a dominant species with thousands of copies can potentially give the same result as a species with 1 copy. Because relative abundance is not being assessed, the differences that help differentiate the habitats in real-time PCR are not discernible. The use of real-time PCR to measure relative abundance, combined with ADONIS and NMDS, allow for statistical significance to be attributed to the multivariate patterns observed, a feature that aids in the ability to transition the assay to a forensic setting.

MATERIALS AND METHODS

Sample Collection

Soil samples were collected at a main site from August 2009 through June 2010 from four habitats in central Michigan: an agricultural field, a marsh, a yard, and a woodlot (Figures 2 – 3). The agricultural field was located south of MSU's campus and the marsh, yard, and woodlot were located several hundred yards apart within the Fenner Nature Center, a wildlife preserve a few miles from campus. Soils were collected next to the marsh and not in the water. One scoop of soil was taken from the surface (approximately 0 to 1 inch in depth), placed in a plastic zip-style freezer bag (Kroger Co., Cincinnati, OH), and mixed thoroughly. Soils were also collected from the main site once every 3 d for a week, and once every week for a month in the fall.

In addition, every 6 mo, soils were collected 10 ft from the main collection site in each of the cardinal directions (north (N), south (S), east (E), and west (W)). The south site at the marsh could not be accessed because it was under water, and the east site at the yard could not be accessed because it extended into the woods. Also, once every six months soil was sampled from the agricultural field and the woodlot at different depths using an AMS Regular Soil Probe (AMS, Inc., American Falls, ID) that was drilled into the ground to a depth of 10 in. The core was removed and the soil cut into 2 in increments and placed in separate freezer bags. Soil samples were stored at -20ºC within an hour of collection. These were labeled based on month and year of collection, habitat, whether the soil was collected from the main site, and at what depth the soil was collected.

Figure 2. Agricultural Field and Marsh collection sites

Left: Photograph of the agricultural field located in East Lansing, MI. Soybean was planted in the field during the summer of 2009 and harvested in October. Soil was tilled, fertilized, and planted in corn at the end of May, 2010.
Right: Photograph of marsh located in Lansing, MI. This location was undisturbed by human activity during the collection period.
For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis.

Figure 3. Yard and Woodlot collection sites

Left: Photograph of the yard located in Lansing, MI. The yard was mowed on a regular basis and was used as a campground during the summer months.
Right: Photograph of woodlot located in Lansing, MI. The woodlot was populated by maples trees and was undisturbed by human activity during the collection period.

DNA Extractions

DNA extraction and purification was performed using a PowerSoil™ DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA). Each extraction required 0.25 g of soil and followed manufacturer's protocol, except for the following modifications: After the first wash with solution S5, an additional 500 µL was added to the spin filter, which was then rotated 180 degrees in the centrifuge before being spun down. DNA was eluted twice using 75 µL of TE buffer (heated to 55°C), rotating 180 degrees between elutions.

Primer/Probe Design for Real-time PCR

Primers and probes were originally designed by aligning gene sequences (retrieved from GenBank®) on BioEdit software v. 5.0.9 (Hall, 1999) targeting the recombination (*recA*) gene of the order *Rhizobiales*. *B. japonicum* and *B. elkanii* primers were designed with BioEdit software, but were specific for the 16S-23S ITS region. *S. meliloti* was also designed with BioEdit software, but targeted the *nodC* gene. Subsequently, primers and probes were designed with ARB software (Ludwig *et al.*, 2004) using the SILVA genomic database (Pruesse *et al.*, 2007) assaying the *16S rRNA* gene. Fluorogenic probes (IDT, Coralville, IA) were labeled on the 5' end with FAM™, HEX™, Cy3™, or Cy5™ reporter dyes and either Iowa Black® FQ or RQ as the quencher dye at the 3' end (Table 1).

Table 1. Primers and probes for real-time PCR

Bacterial groups are indicated in the first column, followed by the designation given each primer and probe. Universal bacterial primers 338R (Wang and Qian, 2009) and 519F (Lane, 1985) were used for *Acidobacteria* group 1 and the genus *Burkholderia*, respectively. Probes are listed with their respective 5' reporter dye and 3' quencher.

Phylogenetic group	Primer/ Probe Name	Sequence (5' – 3')	Amplicon size (bp)	Conc.
Order *Rhizobiales*	Rhi-recA F1	GCAAGGGCTCGATCATGA	217	2 µM
	Rhi-recA R2	AGATGCCGCCCTTCTTCTG		2 µM
Rhizobium leguminosarum bv. trifolii strain 2370	RhiP2370-recA	6-FAM/ ATCGAGACGATCTCGACCGGCTC /IB_FQ		125 nM
Sinorhizobium meliloti strain 1002	RhiP1002-recA	HEX/ CTCCACCGGTTCGCTCGGC /IB_FQ		125 nM
R. leguminosarum	16S-8F*	TCCAGACTTTGAT*YM*TGGCTC	210	1 µM
	R.legR3	CGGGCTCATCCTTGACC		1 µM
R. etli	R.etliF2	GTGGGAACGTACCCTTTACT	214	1 µM
	338R	CTGCTGCCTCCCGTAGG		1 µM
R. tropici	R.tropF2	GTGGGAACGTACCCTTTACT	198	1 µM
	338R	CTGCTGCCTCCCGTAGG		1 µM
S. meliloti	S.melF1	GCCGCTATCTCAATCTACGC	148	900 nM
	S.melR1	TTGAAGCTGGGGACGATAAC		900 nM
Bradyrhizobium japonicum†	B.japF3‡	ATGTAGCTCACAAGGCTGCGT	185	2 µM
	B.japR2‡	CAGAATGTTGTCTGTAAGAACTG		2 µM
	B.jap-ITS	6-FAM/ CTCGCTATCGGAACGATCTTACGAAGC /IB_FQ		250 nM
B.elkanii	B.elkF2	ATCAGCTCACGCTATCTATCGG	200	900 nM
	B.elkR2	ACAAGCCCCTAACACGAGAG		900 nM
Acidobacterium group 1†	Grp1F	GGGTCGCGGCCATTAG	107	125 nM
	338R	CTGCTGCCTCCCGTAGG		125 nM
	Grp1-16S	HEX/ CCTCTCAGGCCGGATACCGATCA /IB_FQ		125 nM

Table 1 (cont'd)

Genus *Burkholderia*†	519F	CAGCAGCCGCGGTAATAC	252	1 µM
	BurkR1	GTCAGTATTGGCCCAGGG		1 µM
	Burk-16S	Cy5/ AATTCTACCCCCCTCTGCCATACTCTAGC /IB_RQ		250 nM
Genus *Agrobacterium*†	AgroF1	AGCTCTTGACATTCGGGGT	297	2 µM
	AgroR1	GAGATTAGCTCGACATCGCTG		2 µM
	Agro-16S	Cy3/ TCCTTCAGTTAGGCTGGCCCCAG /IB_RQ		500 nM

F = forward, R = reverse
Y = C or T, M = A or C
ITS = intergenic transcribed spacer
6-FAM = 6-carboxy-fluorescein, HEX = 5-hexachloro-fluorescein, Cy = cyanine, IB = Iowa Black
*Primer taken and modified from Felske *et al.* (1997).
†Bacterial groups targeted in final assay
‡Primers were designed by Parker and Kennedy (2006) with the designations csits.f3 and csits.r2 respectively

DNA Amplification from Pure Cultures and Soil Samples

Each primer and probe combination was first screened using DNA extracted from pure bacterial cultures. Cesium gradient purified DNA extracts of *Rhizobium leguminosarum bv. trifolii* strains USDA 2370 and 2063, *R. etli* strain USDA 9032, *R. tropici* strains USDA 9030 and 9039, *S. meliloti* strain USDA 1002, *B. japonicum* strain USDA 6, and *B. elkanii* strain USDA 76, were provided by Dr. Patrick Elia (USDA-ARS National Rhizobium Germplasm Collection, Beltsville, MD). A DNA extract of *Burkholderia cepacia* strain 5SP159 was provided by Dr. George Sundin (Michigan State University). Cultures of *Acidobacterium* group 1 and group 4 were provided by Dr. Tom Schmidt (Michigan State University), and a culture of *Agrobacterium* strain 348 was provided by Dr. Eugene Nester (University of Washington, Seattle, WA). Cultures of *Acidobacterium* and *Agrobacterium* were extracted using the PowerSoil™ DNA Kit following the same protocol used for the soil samples.

Primers were tested with control DNA from all species/strains to ensure specificity. PCR was performed using 1 ng of bacterial DNA as template in 10 µL reactions containing 1.0 U Go Taq® DNA polymerase (Promega, Madison, WI), 1X Go Taq® Colorless Reaction Buffer (Promega), 0.2 mM dNTP mix (Promega), and 1µM of each primer in 0.2 mL flat-capped PCR tubes (VWR International, West Chester, PA). The temperature regime was an initial denaturation step at 95°C for 3 min, followed by 50 cycles at 95°C for 15 s, and 60°C for 1 min, in an ABI 2720 thermocycler (Applied Biosystems, Foster City, CA). Primers were tested with soil DNA using the same PCR parameters. Amplified products were viewed on a 2.0% agarose gel and compared to

either a 100bp DNA Ladder (NEB, Ipswich, MA) or a 123bp DNA Ladder (Sigma-Aldrich Co., St. Louis, MO).

Optimization of Real-Time PCR Assays

Primers that produced a single PCR product of the correct size were used in the real-time PCR assay. Specificity of the primer/probe sets was confirmed using known genomic DNA, as well as soil DNA extracts, in a 15 µL final reaction volume initially consisting of 1X iQ Supermix PCR master mix (Bio-Rad Laboratories, Hercules, CA), 900 nM forward Primer, 900 nM reverse primer, 125 nM probe, and 1 µL of DNA template. Two master mixes were made. The first, containing template DNA and 1X iQ Supermix, was dispensed into four reaction wells in 8.5 µL aliquots. A second master mix containing the primers and probe for each bacterial group was then dispensed into the corresponding wells. Reactions containing universal primers and strain specific probes targeting the *recA* gene, were additionally tested in a multiplex reaction with different proportions of control DNA, ranging from 1:10 – 1:1000.

Optimal annealing temperatures of the primers and probes were determined using the gradient feature on an iQ^{TM}5 Multicolor Real-Time PCR Detection System (Bio-Rad Laboratories). Temperatures ranged from 55°C – 65°C. The temperature that resulted in the lowest C_T value was used for all subsequent testing. Optimization of annealing temperature resulted in the following cycling conditions: 95°C for 3 min to activate the polymerase, and 50 cycles of 15 s at 95°C and 1 min at 60°C.

Primer/probe concentrations were optimized by creating a matrix of combinations of primer concentrations ranging from 125 nM to 2 µM, and probe concentrations

ranging from 125 nM to 500 nM. Concentrations that gave the lowest C_T value were used for all soil extracts (Table 1).

Amplification efficiency and limits of detection for the primers and probes were tested by generating standard curves from serial dilutions of the control DNA, and plotting the log of the starting quantity of template (or dilution factor if DNA concentration was unknown) against the C_T value. The equation for efficiency is $E = 10^{(-1/slope)}$ (Bio-Rad Laboratories), and the resultant value was converted to percent by the equation $\%E = (E-1)*100$. Standard curves that produced an R^2 value of > 0.980 and efficiencies of 90 – 105% were deemed acceptable (Bio-Rad Laboratories).

Real-time PCR was initially performed in optical domed capped PCR tubes (DOT Scientific, Burton, MI), but was subsequently adapted to a 96-well format, using unskirted 96-well optical plates (Denville Scientific Inc., Metuchen, NJ) covered with Microseal B clear adhesive seals (Bio-Rad). Resulting data were imported into Microsoft Office Excel 2007 for further analyses.

Analysis of Real-time PCR Profiles

C_T values were converted to the linear form by a 2^{-Ct} transformation, to more accurately depict abundance differences (Livak and Schmittgen, 2001). Transformed data for all bacterial groups within a soil sample were then summed and the value for an individual group divided into the summed total, producing a proportion of a single bacterial group in relation to the total bacteria assayed. Proportions were square-root transformed to help balance the dataset (Oksanen, 2011) and these values were then used

to compare: 1) soil among habitats (inter-habitat variability), 2) soils from the same habitat at different times of the year (temporal variability), 3) soil 10 feet from the main collection site collected on the same day (surface spatial heterogeneity), and 4) different depths in soil from the same habitat (depth heterogeneity).

Reproducibility of Real-time PCR Profiles

Reproducibility of DNA extractions was assessed based on the normalized bacterial proportions of replicate extractions from the first collection period. Technical reproducibility was assessed based on C_T values from replicate PCRs of a single extraction from each habitat. Values were averaged, and standard deviation and coefficient of variation (CV) calculated. CV values from the individual bacterial groups in each habitat were also averaged to assess which exhibited the lowest and highest reproducibility.

Statistical Analyses

Data were analyzed using the Vegan package (Oksanen, 2011) for R statistical software v. 2.12.1 (R Foundation for Statistical Computing). Soil samples that were collected in replicates (allowing for variance to be examined) were first analyzed using ADONIS, which partitioned variation based on dissimilarity. Permutation tests were performed to inspect the significance of these partitions, producing a p-value that indicated if there were any statistical differences in bacterial abundance either among habitats or within a single habitat temporally and spatially. Inter-habitat variability was assessed by examining data from replicate extractions of soil from the main site collected

during the first collection period. Intra-habitat temporal variability was assessed by examining data from replicate extractions of the main collection site of each habitat at different times of the year. Depth heterogeneity was assessed by examining data from replicate extractions at different depths below the main collection site of the Agricultural Field and Woodlot.

NMDS was used to separate bacterial proportions in multidimensional space and visualize multivariate patterns among observations. Data were first ordered in a dissimilarity matrix based on Bray-Curtis dissimilarity (Bray and Curtis, 1957), which was calculated by the following equation:

$$(\sum \left| X_{ij} - X_{ik} \right|) / (\sum (X_{ij}, + X_{ik}))$$

where j and k represent the two samples being compared, based upon variables, $i = 1$ to N (Faith *et al.* 1987). Dissimilarities were then plotted in two dimensions in such a way that the ordination distance between samples in the final configuration correlated as close as possible to the rank-order of their dissimilarities. Habitats were also compared in a pairwise manner using this same method.

RESULTS

Amplification of the recA Gene

The first bacteria assayed were within the order *Rhizobiales*. Soil DNAs, and control DNA from *R. leguminosarum bv. trifolii* strain USDA 2370, *R. etli* strain USDA 9032, *R. tropici* strains USDA 9030 and 9039, *S. meliloti* strain USDA 1002, and *B. japonicum* strain USDA 6 amplified using universal primers specific for the *recA* gene (Appendix A). However, when amplification of *S. meliloti* strain USDA 1002 and *R. leguminosarum bv. trifolii* strain USDA 2370 was attempted in real-time PCR with the strain specific probes, only the control DNA amplified. Further, the control DNAs for these strains did not amplify with the same efficiency when multiplexed as did their singleplex reactions. When control DNAs were multiplexed at different concentrations, the strain in lower abundance either crossed the cycle threshold much later than expected or not at all. Based on these results, specific primer sets were developed for each group of bacteria.

Primer Screens for Bacterial Groups of Interest

Amplification results for the group specific 16S primer screens are shown in Table 2. Soil DNA from the Agricultural Field amplified with all primer sets. *R. leguminosarum*, *R. etli*, and *R. tropici* amplified in soil DNA from all habitats, but the primers cross-reacted with control DNA from other bacterial groups and were therefore not used in the final assay. Only *Bradyrhizobium japonicum*, *Acidobacteria* group 1, and *Burkholderia* amplified in all four habitats with no cross-reactivity (Appendix B). *Agrobacterium* amplified in all habitats except the Woodlot.

Table 2. *16S rRNA* amplification of soil DNA
Results of group specific 16S primer screens for the presence of different bacteria in all habitats. Four groups were chosen for the assay.

	Habitats			
Bacteria	**Agricultural Field**	**Marsh**	**Woodlot**	**Yard**
R. leguminosarum[*]	+	+	+	+
R. etli[*]	+	+	+	+
R. tropici [*]	+	+	+	+
S. meliloti	+	-	-	-
B. japonicum[†]	+	+	+	+
B. elkanii	+	-	-	-
Acidobacteria group 1[†]	+	+	+	+
Burkholderia [†]	+	+	+	+
Agrobacterium[†]	+	+	-	+

– indicates no amplification
+ indicates amplification
* Amplification in all soil types, but cross-reactivity with other species
† Selected for assay

Optimization of Species Specific Primers/Probes

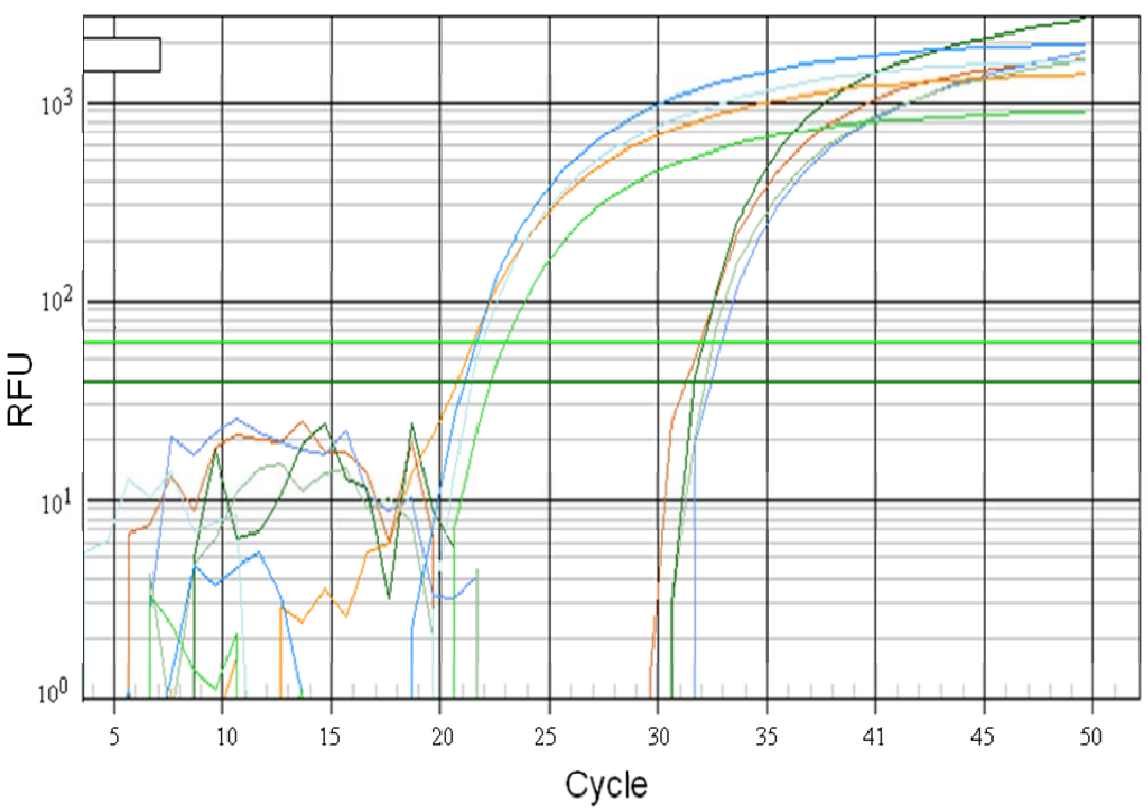
All primer and probe sets selected for the assay amplified and were specific for their intended bacteria. Optimal reaction conditions for amplification are listed in Table 1. Efficiencies, slopes, and R^2 values are shown in Table 3. Efficiencies for primers and probes ranged from 90.0% (*Agrobacterium*) to 101.1% (*Burkholderia*), and R^2 values ranged from 0.985 (*B. japonicum)* to 0.996 (*Burkholderia*). *B. japonicum* was detectable down to a 1 in 1000 dilution (~6,000 genome copies) and *Burkholderia* was detected down to a 1 in 100,000 dilution (~3 genome copies), as were *Agrobacterium* and *Acidobacteria* group 1 (genome copies unknown). DNAs from all 156 soil extracts either amplified initially in real-time PCR or amplified after re-extraction (exemplified in Figure 4).

Table 3. Standard curve data for bacterial groups

Serial dilutions of control DNA were used to create standard curves and calculate efficiencies, slopes, and R^2 values. Efficiencies should fall between 90–105% with R^2 values > 0.980 (Bio-Rad Laboratories).

Bacteria	Efficiency (%)	Slope	R^2 Value
B. japonicum	100.9	-3.299	0.985
Acidobacterium group 1	95.4	-3.437	0.994
Burkholderia	101.1	-3.295	0.996
Agrobacterium	90.0	-3.587	0.986

Figure 4. Typical real-time PCR profile



Cycle number is along the x-axis and relative fluorescence units (RFU) are on the y-axis. The threshold values (horizontal green lines) are set to exclude background fluorescence and cross the amplification curves during the exponential phase of the reaction. Shown are the amplification curves for *B. japonicum* (right) versus *Acidobacteria* group 1 (left) in the Marsh. Replicates (which represent separate extractions) group together, while abundance of the two bacterial groups varies by ~500 fold.

Reproducibility of Real-time PCR Profiles

CV values for normalized bacterial proportions are listed in Table 4. Values from multiple extractions of a single habitat ranged from 9.2% for *Burkholderia* in the Agricultural Field to as high as 100.1% for *B. japonicum* in the Woodlot. Overall, *B. japonicum* varied the most across all habitats, with an average CV of 62.5%, while *Acidobacteria* group 1 varied the least, with an average CV of 13.3%. No statistics were generated for *Agrobacterium* in the Woodlot, as there was no amplification.

CV values of technical replicates ranged from 18.8% for *Agrobacterium* in the Marsh to 101.9% for *B. japonicum* in the Agricultural Field. Averaged CV values varied the most in replicate reactions for *B. japonicum* (89.8%), while *Agrobacterium* (47.3%) varied the least.

Table 4. Reproducibility of real-time PCR profiles
Normalized bacterial proportions were used to assess variation of soils extracted multiple times. Linear C_T values were used to assess variation of technical replicates.

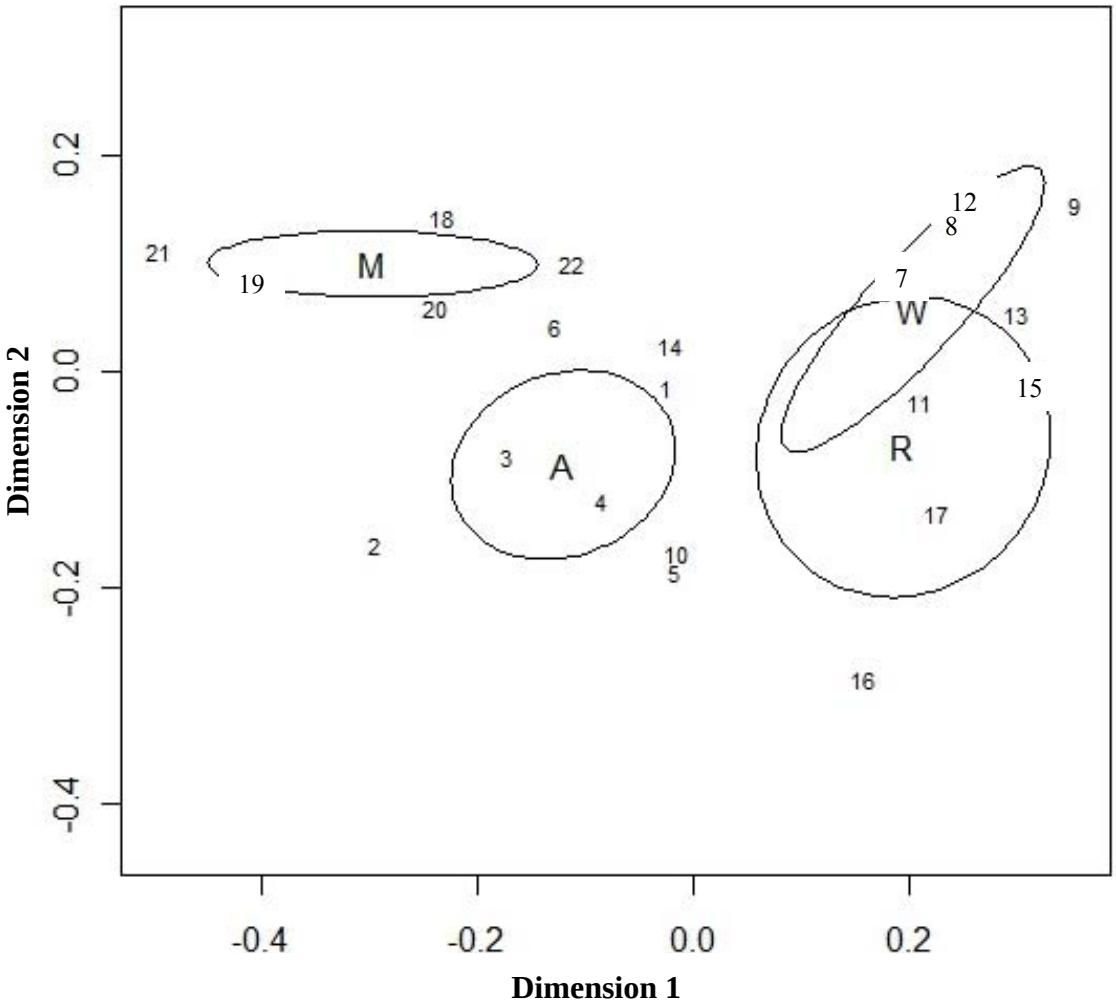
Bacteria	CV of Multiple Extractions				Average
	Agricultural Field	Woodlot	Yard	Marsh	
B. japonicum	43.5	100.1	52.4	54.0	62.5
Acidobacterium group 1	12.0	9.3	9.5	22.2	13.3
Burkholderia	9.2	17.4	25.0	9.5	15.3
Agrobacterium	58.0	NA*	74.3	34.9	55.8
Bacteria	**CV of Technical Replicates**				**Average**
	Agricultural Field	**Woodlot**	**Yard**	**Marsh**	
B. japonicum	101.9	67.9	99.6	NA*	89.8
Acidobacterium group 1	32.6	96.7	48.6	71.8	62.4
Burkholderia	68.6	51.4	37.0	99.8	64.2
Agrobacterium	90.9	NA*	32.2	18.8	47.3

CV = coefficient of variation
*Did not amplify

Habitat to Habitat Variability

At least one habitat differed significantly from the others ($p = 0.005$) when examining replicate extractions from the main collection site of all habitats. The NMDS plot in Figure 5 shows that the Woodlot and Yard overlapped, while samples from the Marsh and Agricultural Field were isolated to their respective habitats.

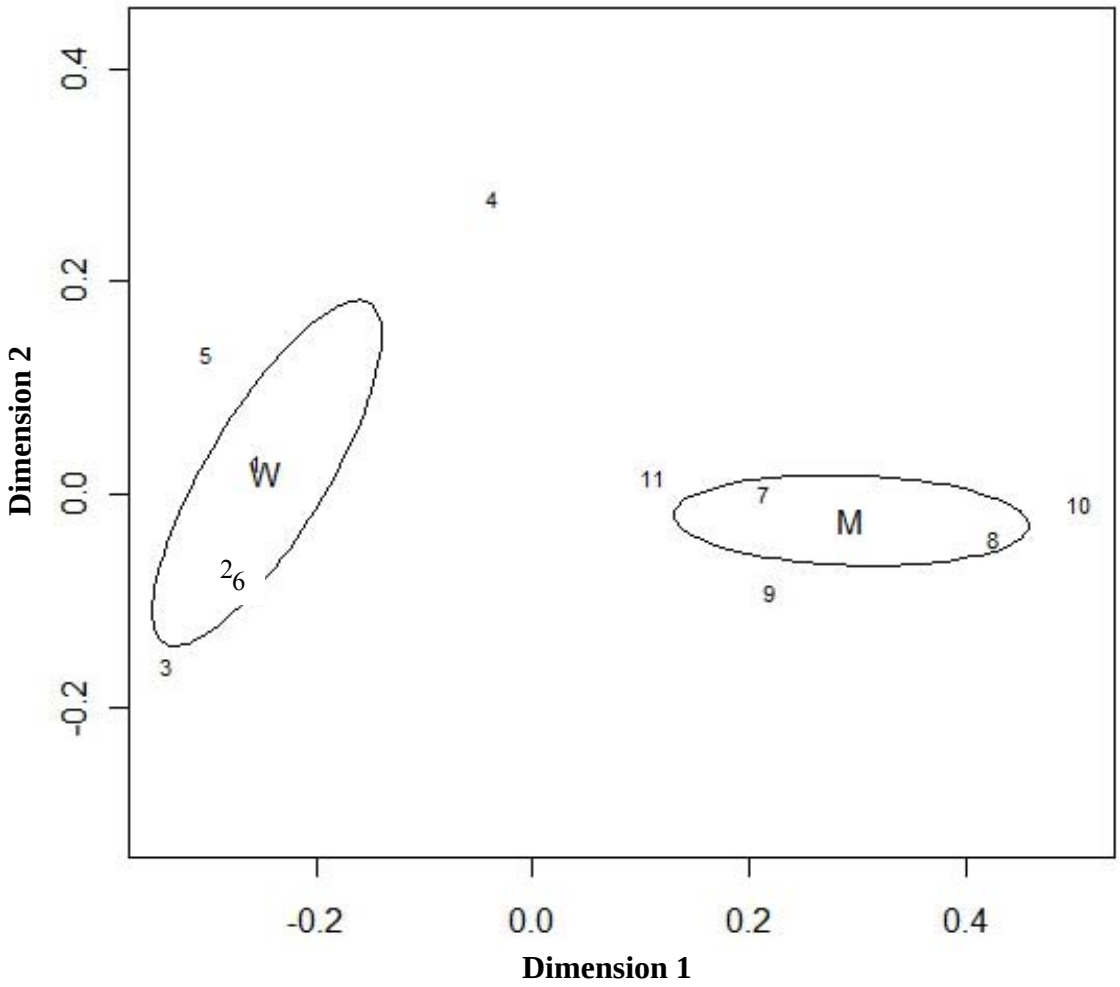
Figure 5. Replicate extractions from main collection sites



NMDS plot showing the 95% confidence ellipses around the samples for each habitat. The Marsh (M) (18–22) and the Agricultural Field (A) (1–6) formed distinct clusters while the Woodlot (W) (7–12) and the Yard (R) (13–17) overlapped. Units on each axis are arbitrary and represent distances between pairs of communities that maintain rank-order of the dissimilarities.

Five of six (~83%) pairwise comparisons between replicate extractions had complete separation of habitats. As an example, Figure 6 shows the Woodlot and the Marsh. Separation was also seen between the Yard and Marsh, and the Agricultural Field and the other habitats (Appendix C). The Yard and the Woodlot could not be separated.

Figure 6. Pairwise comparison of Woodlot and Marsh

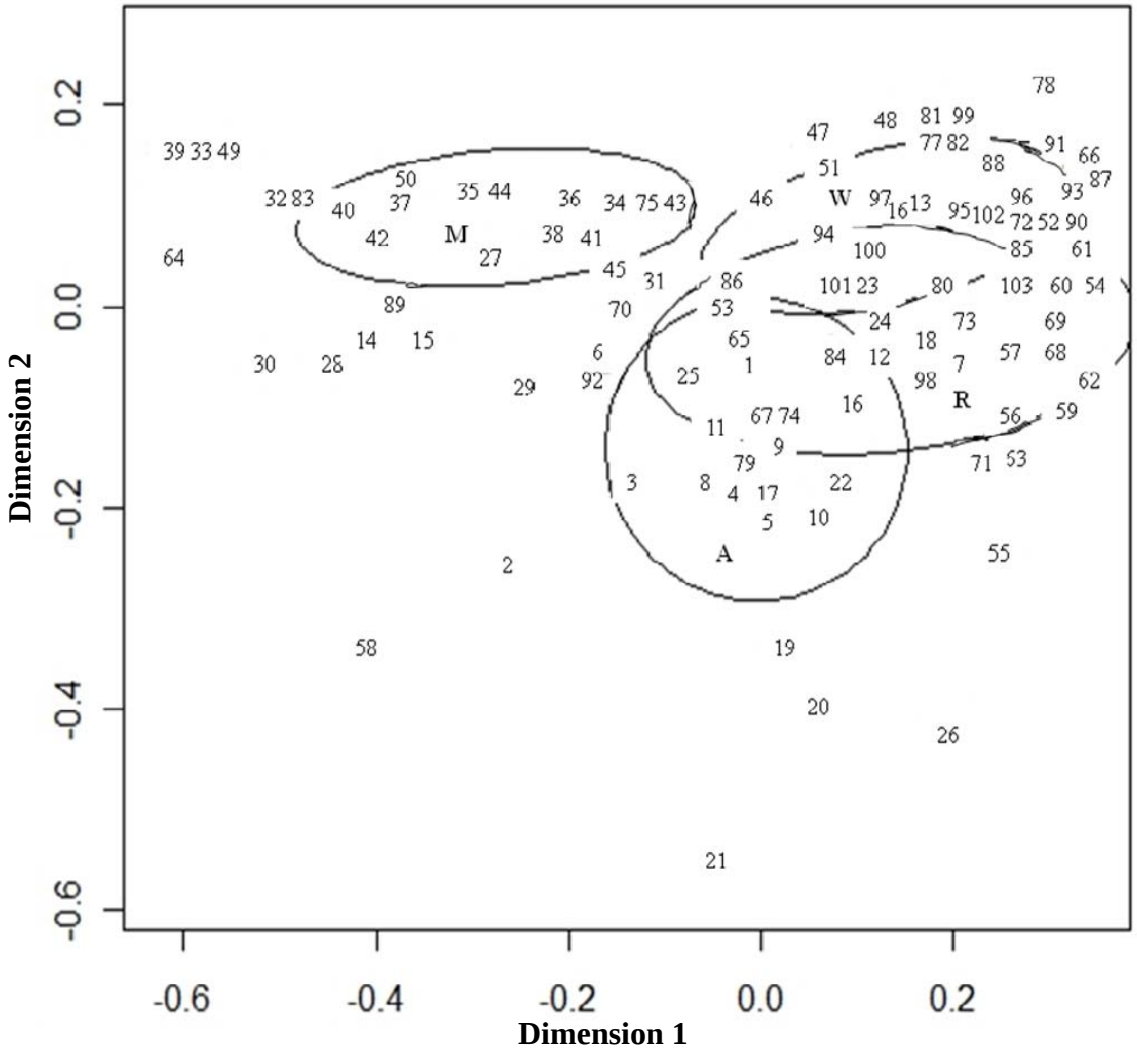


NMDS plot showing the 95% confidence ellipses around replicate extractions of the Woodlot (1–6) and Marsh (7–11). Habitats were separated when compared in a pairwise manner.

Intra-habitat Temporal Variability

At least one habitat was significantly different from the others ($p = 0.005$) when analyzing data that incorporated samples from all habitats at different times of the year. The NMDS plot in Figure 7 incorporates the temporal data for all habitats. Data from the Marsh are located in a distinct region of the plot, while most of the data for the Agricultural Field, Woodlot, and Yard overlap.

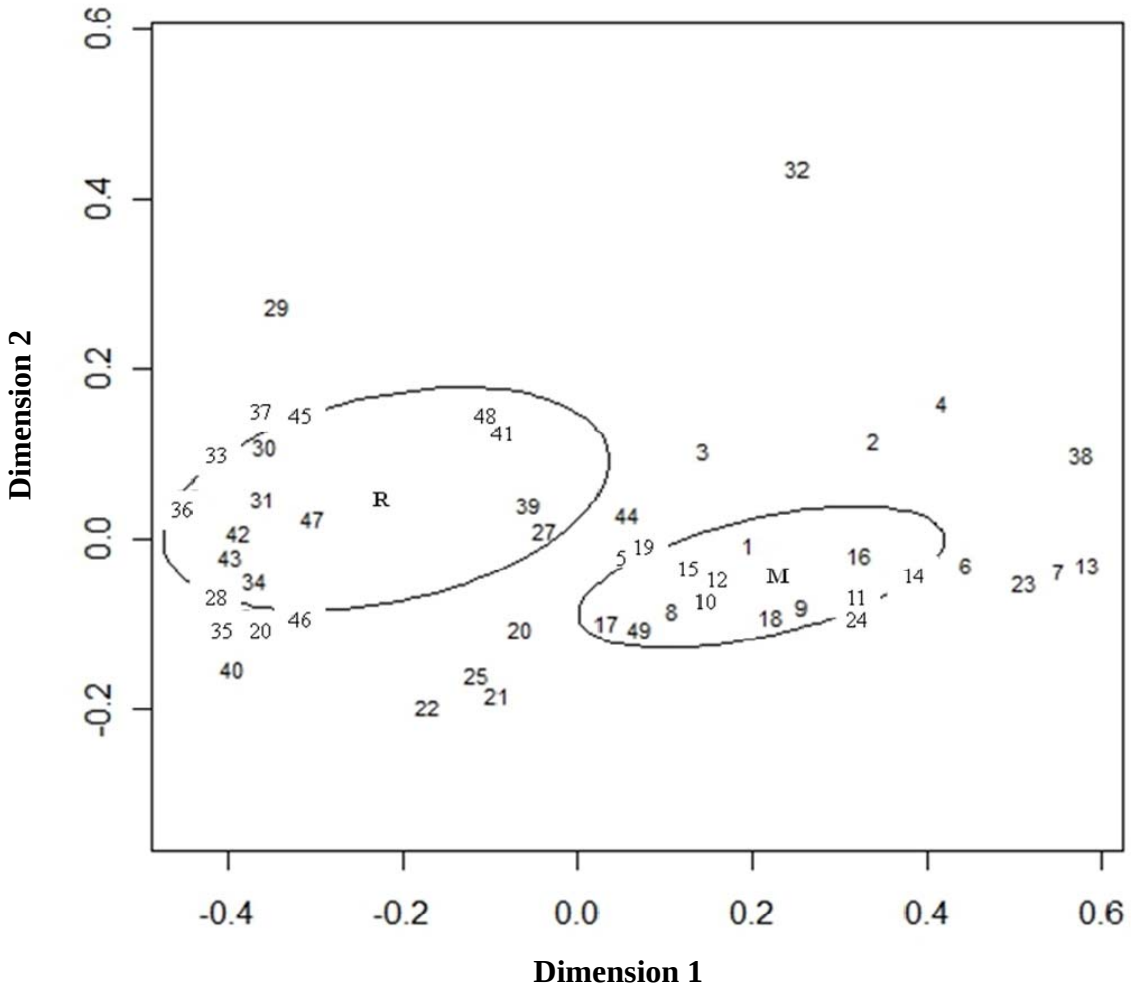
Figure 7. Multidimensional scaling results for temporal data



NMDS plot showing the 95% confidence ellipses around the samples for the different habitats. Data from the Marsh (27–51) tended to separate more than data from the Agricultural Field (1–26), Woodlot (76–103), and Yard (52–75).

Four of six (66%) pairwise comparisons of temporal data for each habitat had complete separation, as exemplified in Figure 8 for the Yard and Marsh. Separation was also achieved between the Marsh and the rest of the habitats, as well as between the Agricultural Field and the Woodlot (Appendix C). The Yard could not be separated from the Woodlot or the Agricultural Field.

Figure 8. Pairwise comparison of temporal data for the Yard and Marsh



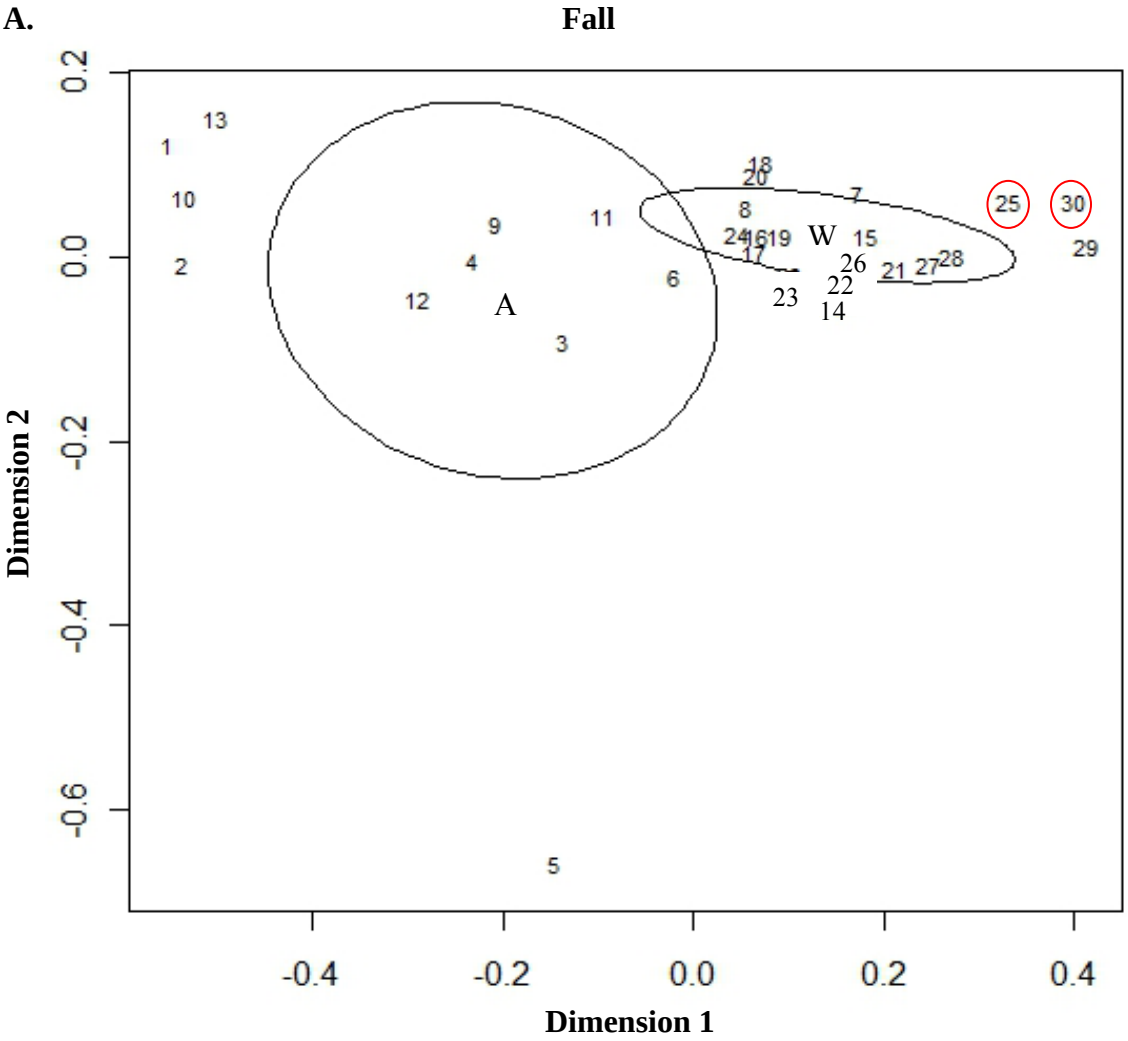
NMDS plot showing the 95% confidence ellipses around the samples for the Yard (26–49) and the Marsh (1–25). Habitats were separated when compared in a pairwise manner.

Replicate samples from the Woodlot ($p = 0.83$) and the Yard ($p = 0.48$) showed no significant difference in the microbial community between the fall and spring. However, significant seasonal differences were detected in the Marsh ($p = 0.02$) and the Agricultural Field ($p = 0.02$).

Intra-habitat Spatial Variability

There was no significant difference in relative abundance among samples taken at different depths from the Agricultural Field (Fall: $p = 0.26$; Spring: $p = 0.12$) or the Woodlot (Fall: $p = 0.06$; Spring: $p = 0.06$). These two habitats partially separated in both the fall and the spring (Figure 9). Besides one Woodlot sample collected from the surface, samples that were collected 8–10 inches in the ground from this habitat in the fall were among the samples with the greatest distance from the confidence ellipse. Samples from this depth were also more clay-like upon visual inspection.

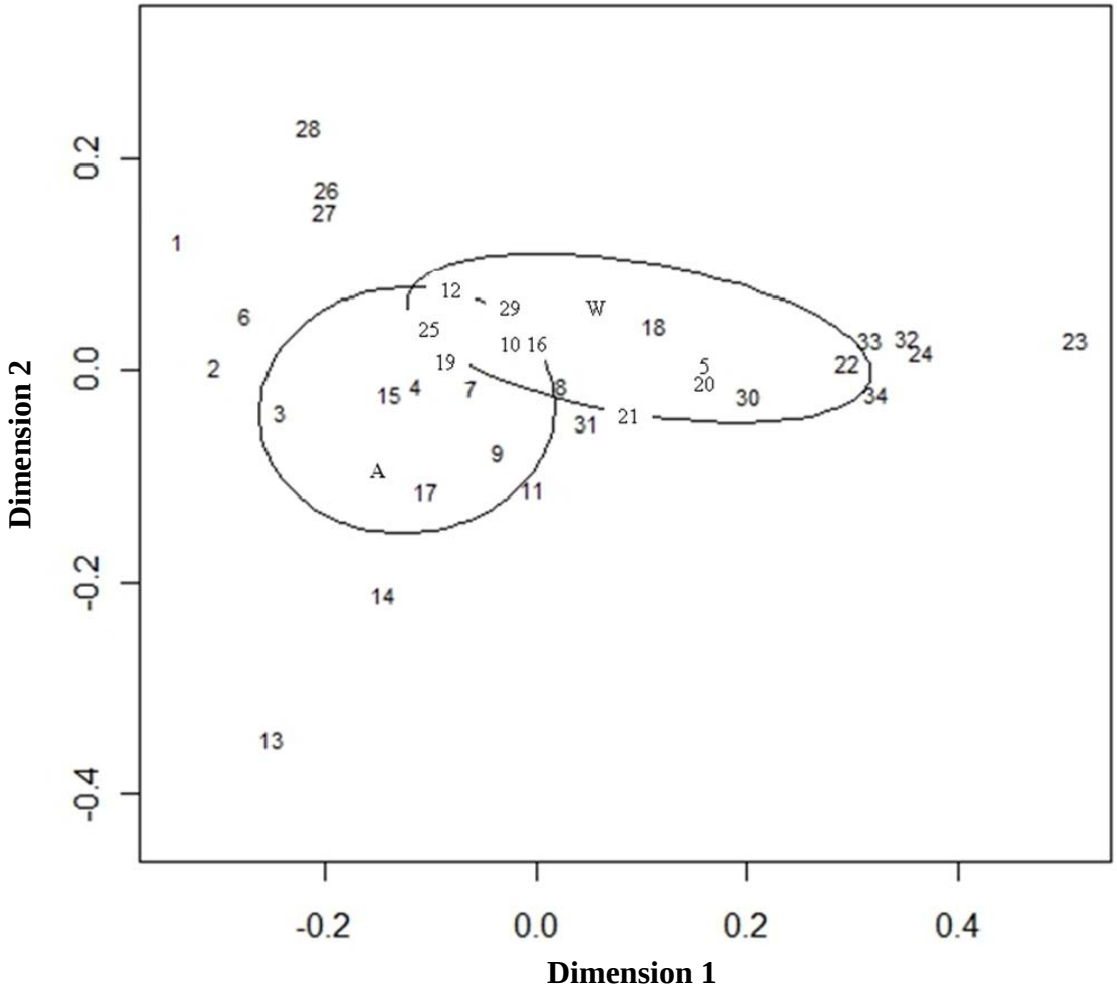
Figure 9. Pairwise comparison of depth data for the Agricultural Field and Woodlot



NMDS plot showing the 95% confidence ellipses around bacterial proportions in the Agricultural Field (1–12) and the Woodlot (13–30) in the fall. Samples 25 and 30 in the fall (circled) represent samples from 8–10 inches in depth. These were the only samples from this depth that were the farthest from the confidence ellipses.

Figure 9 (cont'd)

B. **Spring**

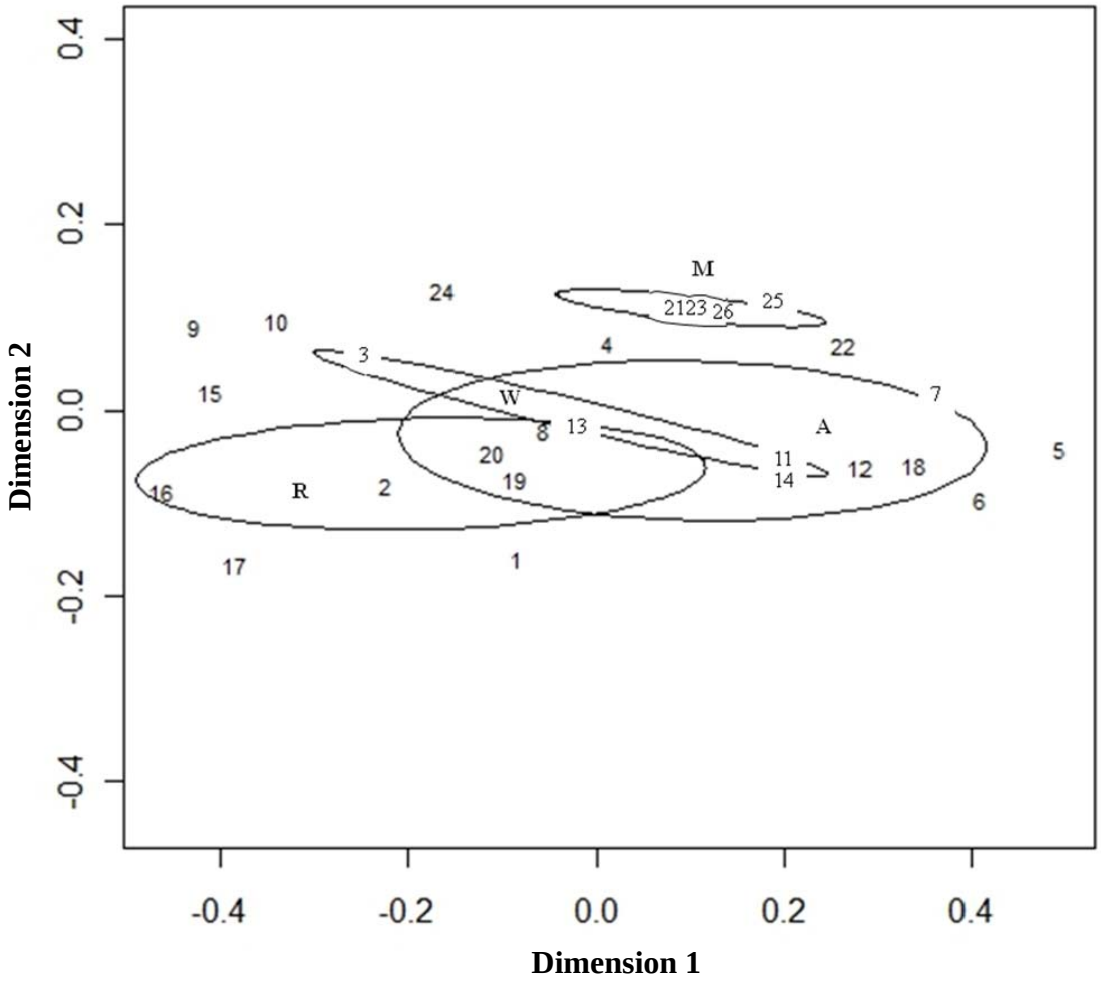


NMDS plot showing the 95% confidence ellipses around bacterial proportions in the Agricultural Field (1–15) and the Woodlot (16–34) in the spring.

Figure 10 shows that the Marsh tended to separate from the other habitats in the fall when incorporating soils collected different directions around the main collection site. Habitats were not separated in the spring. This dataset could not be analyzed with ADONIS as replicate samples were not collected.

Figure 10. Multidimensional scaling results for distance data in fall and spring

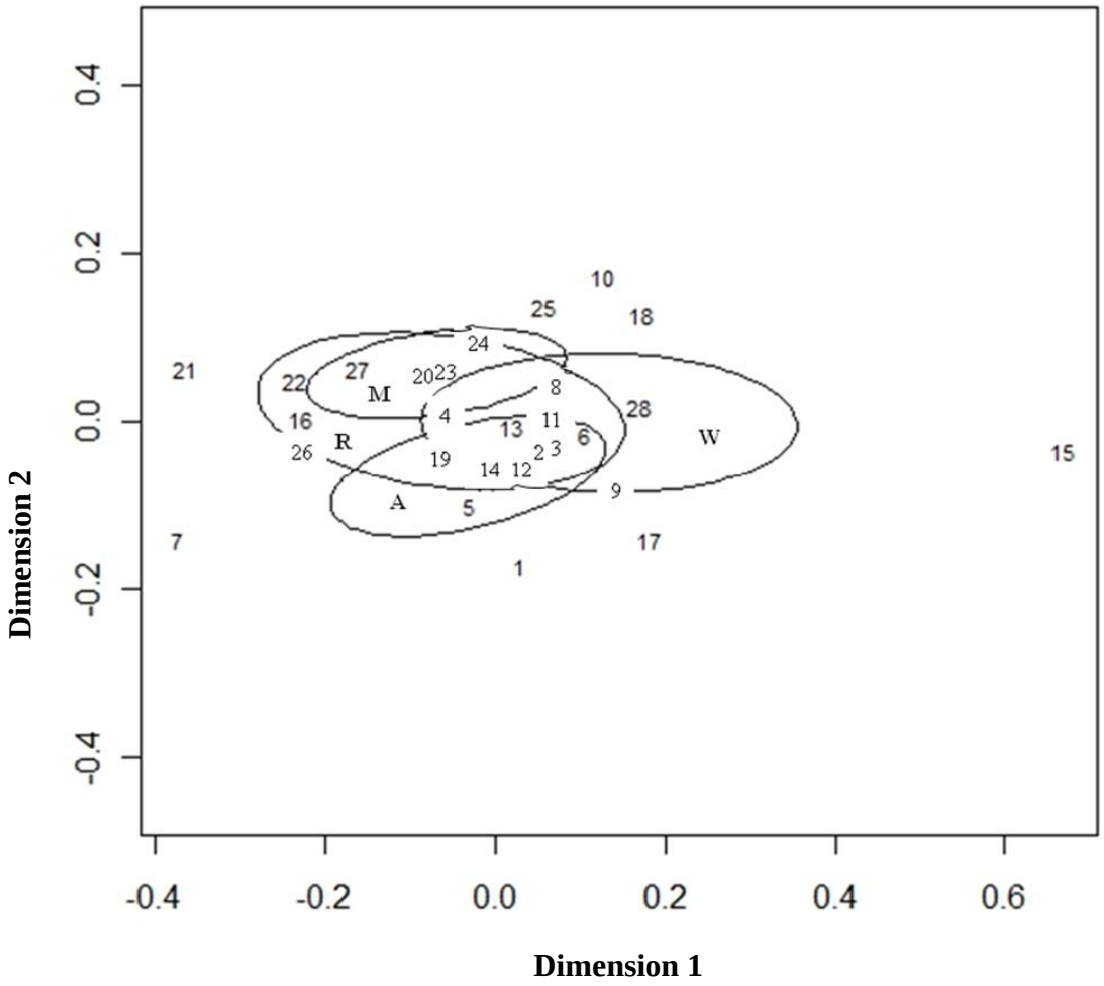
A. **Fall**



NMDS plot showing the 95% confidence ellipses around the bacterial proportions for the different habitats. Only the Marsh separated from the other habitats. Agricultural Field = (1–7); Woodlot = (8–14); Yard = (15–20); Marsh = (21–26).

Figure 10 (cont'd)

B. **Spring**

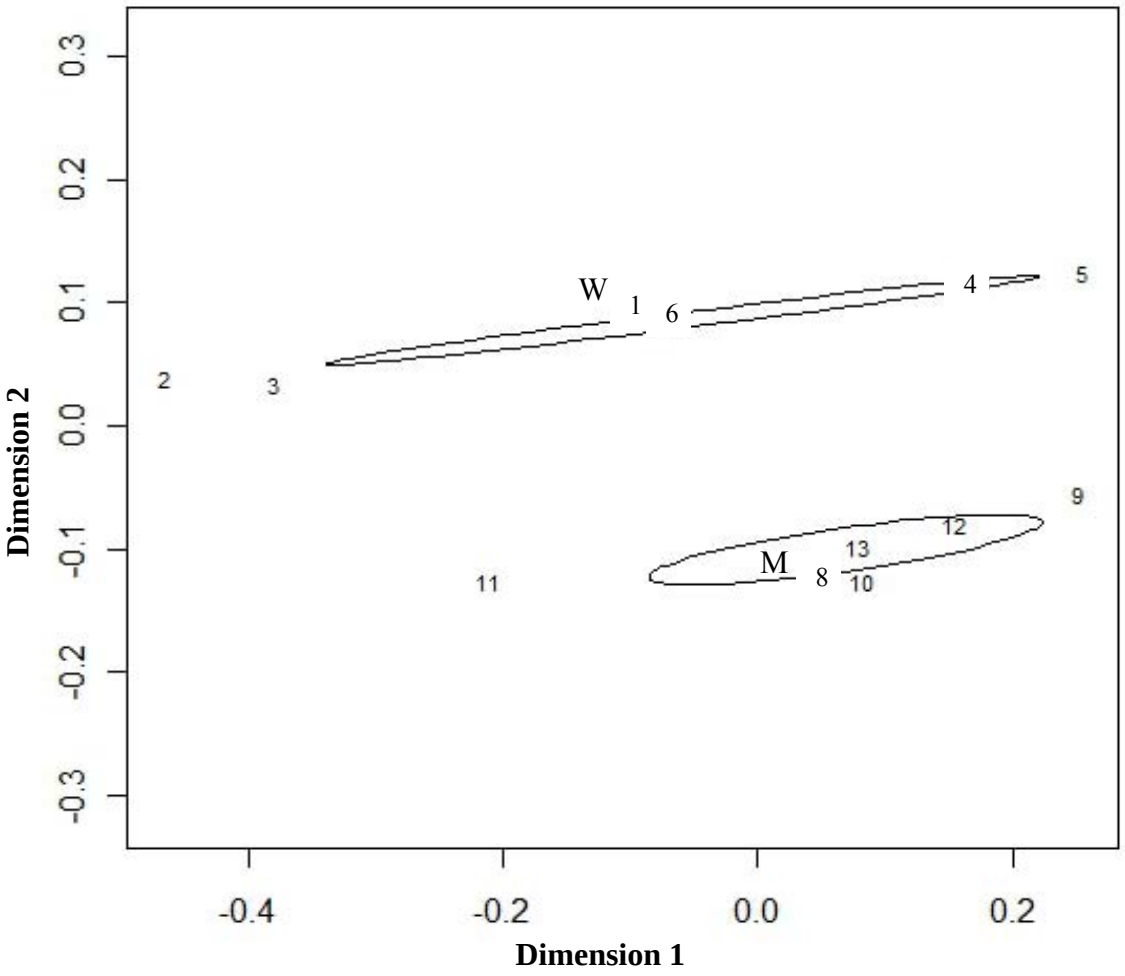


NMDS plot showing the 95% confidence ellipses around the bacterial proportions for the different habitats. Habitats could not be separated. Agricultural Field = (1–7); Woodlot = (8–15); Yard = (16–21); Marsh = (22–28).

Six of twelve (50%) pairwise comparisons between samples collected different directions around the main collection site had complete separation of habitats. Figure 11 shows the Woodlot and the Marsh in the fall. Other pairwise comparisons that had complete separation in the fall included the Marsh and all other habitats, and the Woodlot and the Yard. The Agricultural Field did not separate from the Woodlot or the Yard in

the fall. The Agricultural Field and Marsh, and the Yard and Marsh formed distinct clusters in the spring (Appendix C).

Figure 11. Pairwise comparison of distance data for the Woodlot and Marsh in the Fall



NMDS plot showing the 95% confidence ellipses around the samples for the Woodlot (1–7) and the Marsh (8–13) in the fall. Habitats were completely separated when compared in a pairwise manner.

DISCUSSION

Analysis of soil for forensic purposes has existed for over a century. While different characteristics of soil may be examined, the main goal of any soil identification technique is to decide if questioned and known soils could have the same origin. Bacterial analysis has shown great promise as one method for doing so. Earlier research at Michigan State University focused on the effectiveness of T-RFLP analysis for the typing of forensic soil samples. Meyers and Foran (2008) studied all bacteria and concluded that time had a greater effect on heterogeneity within a habitat than spatial distance. Lenz and Foran (2010) tested the genus *Rhizobium* and again showed that habitats varied temporally. Habitat heterogeneity was also more defined, with several habitats consistently forming distinct clusters. While informative, T-RFLP is not quantitative, preventing the assessment of relative abundance values that could potentially help further differentiate habitats. The goal of the current research was to assess the differences in the abundance of bacterial groups commonly found in soil using real-time PCR. Since this abundance varies among soils, every habitat should have a unique real-time profile if enough bacterial groups are assayed. Furthermore, while levels of bacteria may fluctuate spatially and temporally within a habitat, their abundance in relation to other bacteria in the same sample may remain similar.

Real-time PCR has commonly been used to analyze bacteria in soil; however, many studies (Skovhus *et al.*, 2004; Lee *et al.*, 2009) have utilized absolute quantitation methods, based on standard curves, to determine the amount of one particular bacterial group or strain. Furthermore, multivariate statistics are not generally used to examine real-time PCR data. Lee *et al.* (2009) employed NMDS as a multivariate technique to

assess methanogenic community dynamics in three anaerobic batch digesters treating different wastewaters, but used absolute quantitation. Occasionally, real-time PCR is used to assess relative abundance in relation to the total bacteria assayed. Skovhus *et al.* (2004) determined the relative abundance of *Pseudoalteromonas* species in marine environments in relation to the total Eubacterial rDNA detected, but again used absolute quantitation methods to make those determinations. The current research is the first instance where multivariate techniques were used to examine relative abundance values from real-time PCR data.

Initial research into the *recA* gene aided in several decisions regarding the final assay, including the use of a singleplex rather than multiplex design, targeting groups of bacteria instead of a specific strain, and assaying the *16S rRNA* gene for better wide scale sequence differentiation. In regards to the multiplex design and the level of specificity of the assay, several problems were noted with the strain specific probes in real-time PCR. First, attempting to amplify equal concentrations of several control DNAs in a single reaction led to reduced efficiency compared to their singleplex counterparts. This likely resulted from the PCR reagents needing to amplify multiple gene targets instead of just one. Similarly, attempting to amplify different concentrations of multiple control DNAs in a single reaction likely resulted in preferential amplification of the DNA in higher abundance, leading to the DNA in lower abundance either amplifying much later than expected, or not all. This indicated that prevalent soil bacteria might also be preferentially amplified, which would not allow for a true assessment of their abundance. Multiplex assays are commonly employed in forensics; however, the ability to amplify several targets in a multiplex design depends on the interactions of the primers and

probes, and overall amplification efficiency. Further optimization may eventually lead to multiplexing of the current assay. Second, lack of amplification in real-time PCR for the soil DNA probably meant the specific strains were not present, since amplification in regular PCR indicated no inhibition. This suggested that primer sets targeting groups of bacteria that are present in nearly all habitats and vary in abundance would be more appropriate, as a single strain might not exist/be found in multiple habitats. To accomplish this, the *16S rRNA* gene was thought to be the most appropriate target for the assay, due to the abundance/availability of sequence data.

Designing primers for the different bacteria required alignment of gene sequences to determine both conserved regions within a group of interest and variable regions among groups. Only a few could be compared at a time when manually aligning sequences from GenBank®, limiting the ability to design species/group specific primers. For the most part, this was resolved using the SILVA database within ARB software. While other genes are sometimes better for discriminating closely related species (Daffonchio *et al.*, 2003; Clarridge, 2004), ARB software was used to compare 16S sequences from *Bacteria*, *Archaea*, and *Eukarya* all at once, increasing primer specificity.

Problems with cross-reactivity and lack of amplification in multiple habitats prevented the use of some of the primer sets in the final assay. Since the abundance of *S. meliloti* and *B. elkanii* was zero for all habitats other than the Agricultural Field, these bacteria were not targeted. Primers specific for *R. leguminosarum*, *R. etli*, and *R. tropici* cross-reacted with control DNA from other bacteria, meaning non-target species could also be amplified. Although the 16S primer sets for these three bacteria were designed using *ARB* software, each primer set contained a universal primer that may have limited

their specificity. The need to use a universal primer when targeting these bacteria stems from the nature of the real-time assay. In order for the assay to work properly, the region of the gene being targeted could not be much larger than ~300 base pairs. ARB software usually found variable regions suitable for primer design, but if they were a large distance apart, they were not useful. While small amplicon size was initially considered advantageous, it actually somewhat hindered the ability to design specific primers.

B. japonicum, Burkholderia, Acidobacteria group 1, and *Agrobacterium* were finally selected for the assay due to their presence in all habitats and the specificity of the primers/probes. Bacterial phylogenetics is very complicated, and many genera/species are superficially defined. Therefore, targeting a single species and preventing non-specific amplification can be problematic when developing an assay to differentiate bacterial groups. *Burkholderia* have a complex taxonomy with many closely related species, so primers were designed that assayed the entire genus. In fact, testing a larger group of bacteria most likely allowed for more frequent detection during real-time PCR analysis for all soils. This was probably true for *Acidobacteria* group 1 and *Agrobacterium* as well.

Optimization of 16S amplification was essential to prevent possible bias for any particular bacterial group, and to ensure accurate and reproducible species/group quantification. Primer and probe combinations showed some variation in regards to efficiency, R^2 value of the standard curve, and consistency across replicate reactions between bacterial groups. Efficiency was assessed based on the slope of the standard curve, which can be affected by differences in reaction conditions (e.g. primer/probe concentration) or pipetting error. Suboptimal reaction conditions typically make

efficiencies decrease, while pipetting error normally leads to an apparent increase in efficiency (>100%) (Bio-Rad Laboratories). All primer and probe combinations were between 90 – 105% efficiency; however, the values obtained for *Agrobacterium* (90.0%) and *Acidobacteria* group 1 (95.4%) suggest that perhaps further optimization was needed for these groups.

The lower R^2 value and higher detection limit with *B. japonicum* was presumably due to the strain used as the control DNA (USDA 6). This particular strain had a few mismatched base pairs in the reverse primer near the 3' end, shifting the C_T values such that low concentration standards could not attain fluorescence above background, and therefore did not cross the cycle threshold. This decreased the number of points on the standard curve, and although only two points are needed to generate a line with an R^2 value of 1, incorporation of additional standards likely would have improved the overall linearity of the data. Cross-referencing with other *B. japonicum* strains showed that most did not have this mismatch. Amplification of this bacterium in the actual soil samples did not appear to be problematic. The lower R^2 value for *Agrobacterium* could again be from suboptimal reaction conditions that led to decreased amplification efficiency. Also, when reaction conditions were optimized, annealing temperature and primer/probe concentration were tested, but amplification efficiency at different starting concentrations of DNA was not. It is possible that amplification efficiency was not the same for the different standards used to generate the curve (Bio-Rad Laboratories), which could have lowered the R^2 value for both *Agrobacterium* and *B. japonicum*.

The reproducibility of real-time PCR profiles generated from multiple extractions was important to assess once reaction conditions were optimized. This was calculated using relative abundance of bacteria within each sample rather than the raw C_T values because total DNA concentrations were not determined before real-time PCR was performed; therefore differences in raw C_T values could simply be a result of varying amounts of starting DNA. The high CV values could stem from differences in extraction efficiency. If soil type affected the extraction efficiency more in one habitat than another, differences in the real-time PCR profiles among habitats would not be an accurate representation of the true abundance in the soil. Furthermore, multiple extractions from separate sub-samples of soil may have caused varying amounts of inhibitors to be present in the extract, possibly affecting amplification efficiency. However, since inhibitors of PCR typically target *Taq* polymerase, the amplification of each group should be affected the same, and since relative abundance values were used to assess reproducibility, this concern was most likely not an issue. In the current research, if the bacterial group amplified, inhibition was assumed to be low; however, spiking the soil extracts with a known amount of control DNA could have helped determine the effects of inhibition.

The high variability in technical replicates was not expected as it was thought multiple extractions would be much more variable than sampling from the same DNA extract several times. However, the highest CV value (62.5%) from multiple extractions was close to the lowest CV value (62.4%) for technical replicates. This variation can most likely be attributed to the way in which the real-time PCR components were dispensed in the reaction plate, and the number of replicates. Although real-time PCR has been shown to be highly reproducible, its sensitive nature means that even the

smallest differences in reaction component concentration (template DNA, primers, probe, etc.) can translate into large differences in C_T value. This is especially true when one considers the logarithmic nature of PCR and the fact that the C_T values need to be converted to their linear form to more accurately represent the variation in the dataset. Heid *et al.* (1996) examined the raw C_T values of 10 replicate reactions of DNA from bacterial isolates and found real-time PCR produced CV values of less than 1%. Likewise, Livak and Schmittgen (2001) ran 96 replicate reactions on a single plate that produced CV's of less than 1% when using raw C_T values. However, when C_T values were converted to their linear form using the 2^{-Ct} transformation, CV's went up to 13.5% percent. In the current research, only three replicates were used, which may not have been a good representation of the true variation. While 96 replicates is not necessary, increasing to five or six, as was done when assessing variability in multiple extractions, may have been more informative.

Choosing the most appropriate statistical analysis for this study posed quite a challenge. NMDS was eventually selected to analyze the normalized data because it does not seek to find any particular relationship between variables, which can negatively impact the robustness of other multivariate statistical techniques. The construction of a dissimilarity matrix helped assess differences in bacterial communities. In order to provide a visual representation of these differences in multidimensional space, NMDS generated an ordination in which the distances between all pairs of samples are in rank-order agreement with their dissimilarities. In other words, if any given pair of samples has a dissimilarity less than some other pair, then the first pair should be no further apart

in the ordination plot than the second pair. The degree to which distances agreed in rank-order with the dissimilarities was assessed through the generation of a monotone regression line that represented hypothetical distances that were in perfect rank order. The difference between these hypothetical distances and the actual distances is referred to as the "stress" of the dataset. The aim was to find coordinates in multidimensional space (ordination plot) that would minimize this value because stress decreases as the rank-order agreement between distances and dissimilarities improves.

Increasing the dimensionality used in NMDS (2-D vs. 3-D) can sometimes help lower the stress value; however, two dimensions were chosen to analyze the data in this study because only four variables were being examined, and increasing the number of dimensions did not lead to better separation. The number of variables and small sample size of the datasets also likely explain why the 95% confidence ellipses often did not include many data points. Typically, multivariate analysis is performed on datasets that include hundreds or even thousands of different variables. Each peak in the T-RFLP profiles analyzed by Lenz and Foran (2010) represented a different variable, possibly explaining better separation of habitats compared to plots generated from the real-time PCR data. In the current study, confidence ellipses simply helped provide a visual representation of the trends associated with the dataset. Finally, ADONIS was used to attribute statistical confidence to the results obtained, by partitioning the variation in the dissimilarity values. Permutation tests inspected the significance of these partitions, producing a p-value that indicated if there were significant differences within the dataset; a feature that is important for Daubert considerations.

Assessment of the normalized data for inter-habitat variability indicated that at least one habitat was significantly different from the rest. This is exemplified by the separation of the Agricultural Field and the Marsh in NMDS, and most likely stems from low levels of *B. japonicum* in the Marsh and low levels of *Agrobacterium* in the Agricultural Field. As expected, this created very distinct relative abundance values that contributed to the distance calculated using NMDS. In contrast, the Woodlot and the Yard did not have large differences in bacterial abundance, making the proportions all very similar, and causing samples from these habitats to group close together in the NMDS plot. The inability to differentiate the Woodlot and Yard may also stem from difficulties in using NMDS to analyze a large number of dissimilar samples all at once. This was also seen by Lenz and Foran (2010). Differences in the distances between highly dissimilar habitats (e.g., the Agricultural Field and Marsh) may have caused them to occupy a distinct region of the plot, while similar habitats (although still having some differences) were indistinguishable.

Separation of habitats was more regularly achieved when done in a pairwise manner, which more closely reflects a true forensic situation. Lenz and Foran (2010) also found that pairwise comparisons helped separate soil samples into their respective habitats. This simplification of the data set may have allowed habitats that could not be separated when all habitats were compared (likely resulting from large distances between highly dissimilar habitats), to separate when only two habitats were examined. In the current research, habitat replicates separated 5 out of 6 times (~83%) when examined in a pairwise manner. Lack of separation between the Woodlot and Yard may be attributed to several factors including similarity in bacterial abundances discussed earlier, or the fact

that *Agrobacterium* did not amplify in the Woodlot. An abundance of zero for this genus could have affected distance determination in NMDS.

The normalized data for intra-habitat temporal variability indicated that relative bacterial abundance in the Marsh and Agricultural Field varied significantly at different times of the year. Rotation crops, such as those seen in the Agricultural Field, are common in agriculture, and this change in plant life likely resulted in significant differences ($p = 0.02$) in bacterial abundance. Sensitivity to changes in weather was most likely the biggest contributor to the variation observed in the Marsh. During several collection periods in the winter months, the soil was either completely encased in or below a layer of ice. More importantly, the surrounding area consisted of very moist soil with varying water levels throughout the year. Differing levels of moisture in the soil at different times of the year could certainly have caused variation in the bacterial population. Castro *et al.* (2010) found that bacterial abundance changed significantly with varying amounts of precipitation, wherein *Acidobacteria* increased in dry environments while *Proteobacteria* increased in wet environments. Spring run-off could have also changed the bacterial abundance as the top layer of soil was washed away from the collection site.

Significant differences in bacterial abundance over time for both the Agricultural Field and the Marsh, and the complexity of the dataset, could explain the lack of separation in the NMDS plots incorporating temporal data for all the habitats. However, when examined on a monthly basis, the Marsh was the only habitat that consistently occupied its own region in the NMDS plot. Although this habitat was highly variable in its bacterial abundance at different times of the year, it consistently had low levels of *B.*

japonicum, creating distinct abundance values. Again, pairwise comparisons aided in the ability to separate habitats 4 out of 6 (66%) times, even when temporal data were included.

The normalized data for *distance* heterogeneity could not be assessed by ADONIS because replicate samples were not collected from the four cardinal locations. Instead, assessment of heterogeneity could only be inferred from NMDS. Plots incorporating the distance data for each habitat in both the fall and spring did not show the same separation as the habitat replicates used to assess inter-habitat variability, indicating possible intra-habitat spatial variability. Variation at different distances from the main collection site is also very important to a forensic investigation since the exact location from which the question soil originated is usually only known if soil collected from a shoe or tire can be traced back to an impression that was left at the scene. Interestingly, while soil samples taken from around the Marsh were the most diverse upon visual inspection, they still tended to separate from the rest of the habitats. The soil collected north of the main site was ten feet away from the shore, and was much drier. There were also differences in the surrounding plant life. The west soil was under a dock and did not receive sunlight. In contrast, within habitat samples from the Agricultural Field, Yard, and Woodlot were very similar visually and had the same surrounding plant life, but did not separate from the rest of the habitats. If replicate samples had been collected, ADONIS could have been used to help elucidate distance heterogeneity. Pairwise comparisons did not show as much separation compared to those examining inter-habitat and temporal variability, with complete separation occurring only 6 out of 12 (50%) times.

Assessment of the normalized data for *depth* heterogeneity indicated that bacterial abundance did not vary significantly at different depths in the Agricultural Field or the Woodlot. Variation at different depths can be important to a forensic investigation, if it is suspected the soil may have come from beneath the surface. NMDS plots incorporating depth data for both the Agricultural Field and Woodlot showed that the habitats could still be partially separated. This again supports what was seen with earlier research (Lenz and Foran, 2010) where separation was more regularly achieved when done in a pairwise manner, thus decreasing the complexity of the data set. Interestingly, the nearly significant p-value (0.06) for the Woodlot may be attributed to the section of soil from eight to ten inches in the ground. This section was more clay-like in appearance, while samples closer to the surface were darker in color. In fact, these samples were among the farthest away from the confidence ellipse in Figure 9, possibly correlating differences in bacterial abundance with differences in soil type.

There are several findings from this study that warrant further investigation. Future research would need to include the incorporation of additional bacteria into a multiplex design, to not only increase resolution of the different habitats, but to increase throughput. Amplification efficiencies should be investigated using soil DNA, to account for the effects of any inhibitors that might be present. If efficiencies are still between 90–105% after optimization, inhibitors would not be a concern when determining relative abundance. Finally, the use of a liquid handling robot would help ensure consistent and precise results, limiting any variability that might stem from pipetting technique.

Recently, some areas of forensic science have been scrutinized for methods that lack repeatability or any attribution of confidence to the results obtained. Therefore, the

legal implications of this assay must be considered. The use of ADONIS helped apply a more rigorous examination to the dataset, providing a p-value that would allow a forensic scientist to determine how significant the results may be, and is important for Daubert considerations. The 95% confidence ellipses used in NMDS also helped in this regard. However, not enough research has been conducted to say that a particular soil sample came from a certain area, to the exclusion of all other soil in the world. In addition, since larger groups of bacteria (phyla, genera, etc.) were targeted, it is feasible that different species within these groups could produce the same real-time PCR profile. This brings into question whether samples with the same profile are truly the same, and has obvious implications in regards to the argument from the defense. Additional research would certainly need to be conducted before this assay could be incorporated into a forensic setting.

Although this assay shows promise, sequencing technologies may be considered the next step in the future of forensic soil analysis. Sequencing techniques provide very robust analyses of the microbial community by generating thousands of short overlapping sequences that can be used to not only determine which bacteria are present in a sample, but also the relative abundance of the bacteria detected. Furthermore, recent advances in sequencing technologies have allowed the cost per sample to decrease drastically. In fact, many studies have already been conducted that have used this technique to examine microbial communities in several different types of samples, including those found in waste water treatment facilities, human tissue samples, and soil. For example, Will *et al.* (2010) were interested in how bacterial abundance changed at different depths in the soil. In this study, the microbial composition at different depths was examined for three

habitats using pyrosequencing. The most abundant bacterial groups from over 650,000 generated sequences were examined. This analysis allowed for very small changes in bacterial composition to be detected, and revealed that bacterial species varied significantly ($p<.00001$) at different depths. Fierer *et al.* (2010) examined bacteria transferred from a person's fingers to a computer keyboard using pyrosequencing. This resulted in roughly 800–1,500 sequences per sample and revealed that bacteria differed not only from person to person, but also from finger to finger on the same person. The level of sensitivity and the amount of data generated in both these studies indicate that sequencing could be very useful for wide scale soil comparisons, which would be required if microbial profiling of soil is to be considered a viable tool in the forensic community.

Conclusions

The results of the current study suggest that real-time PCR could be a useful tool for analyzing forensic soil samples. Enough inter-habitat variability was detected to allow several habitats to be separated while only examining *B. japonicum, Burkholderia, Acidobacteria* group 1, and *Agrobacterium*. Assessment of intra-habitat variability indicated that the Agricultural Field and the Marsh exhibited temporal variability, but it appears this mostly depends on the amount of perturbation the soil is subjected to. Depth in the soil did not seem to affect microbial abundance, which is beneficial if questioned soil may have come from beneath the surface. Additional samples would have to be collected around the main site of each habitat before any conclusions could be made on distance heterogeneity. Generally, the Marsh tended to be isolated from the other

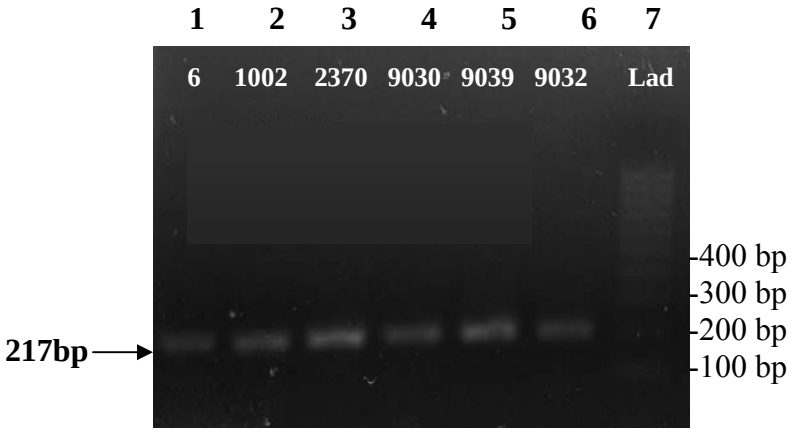
habitats even when examining soils collected at different times of the year, which shows promise. Overall, it appears the largest contributor in the ability to separate habitats stems from, 1) the number of bacteria in the assay, since this greatly effects multivariate statistical techniques, and 2) large differences in the relative abundance of each bacterial group, as exemplified in the Marsh which had very low levels of *B. japonicum*. Finally, the use of pairwise comparisons allowed for greater separation of soils, showing increased relevance and practicality for incorporation of this assay in a forensic setting.

Still, there are several factors that need to be considered before the assay could be implemented. The ability to determine relative amounts of bacteria, while very good at producing unique profiles for different habitats, may be more sensitive to temporal and spatial variability than other microbial techniques, making it difficult to say with confidence where/when a soil may have originated. Traditional methods are still the gold standard for forensic soil analysis, but only through further research testing different soil types and habitats will microbial profiling be considered an additional tool for the forensic soil analyst.

APPENDICES

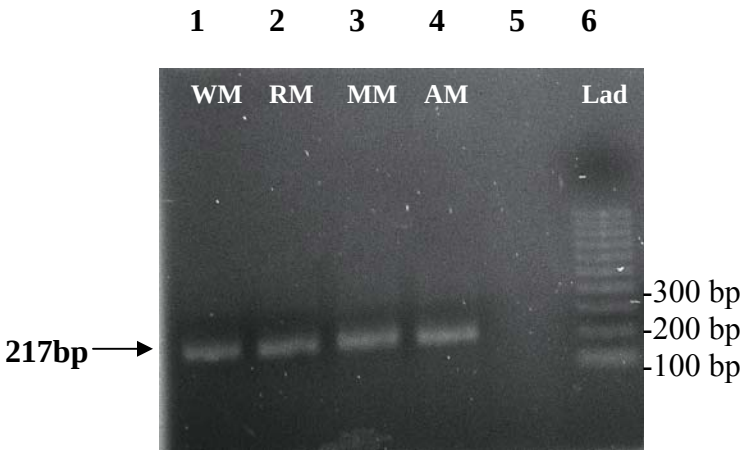
APPENDIX A:

Figure 12. Universal rhizobial *recA* amplification results with control DNA



Lanes 1–6 are the different control DNAs that amplified. Lane 7 is a 100bp DNA ladder.

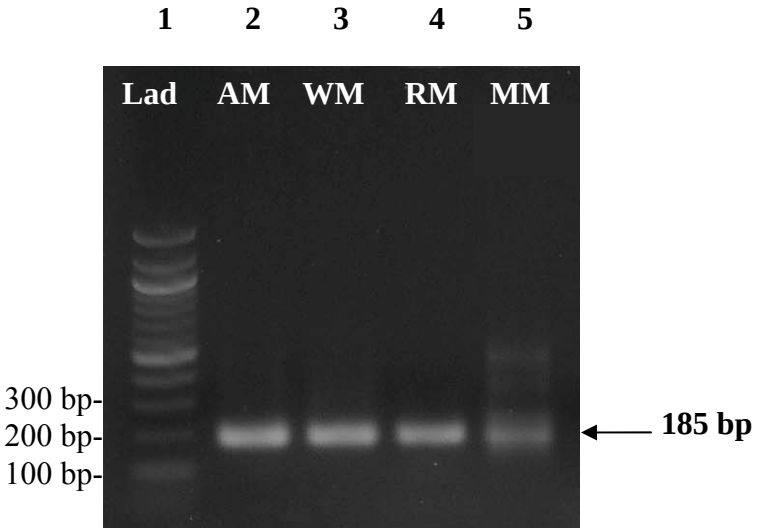
Figure 13. Universal rhizobial *recA* amplification results with soil DNA



Lanes 1–4 are the amplicons from the different habitats that amplified. Lane 6 is a 100bp DNA ladder. Lane 7 is a positive control with control DNA and Lane 8 is a negative control.

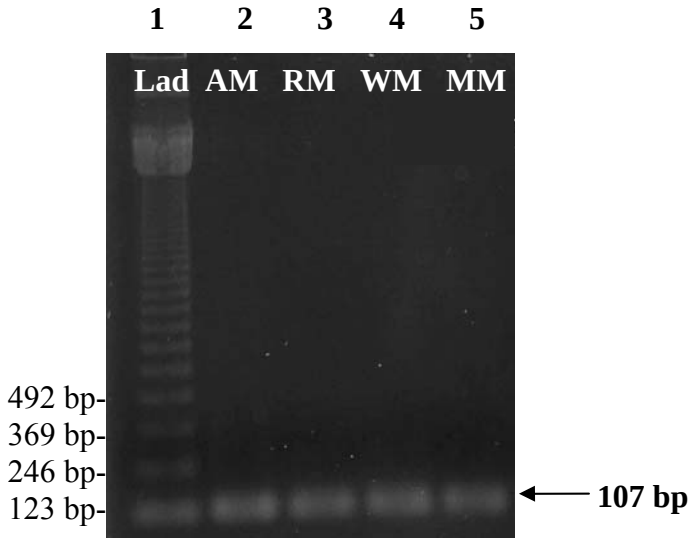
APPENDIX B:

Figure 14. *B. japonicum* amplification results with soil DNA



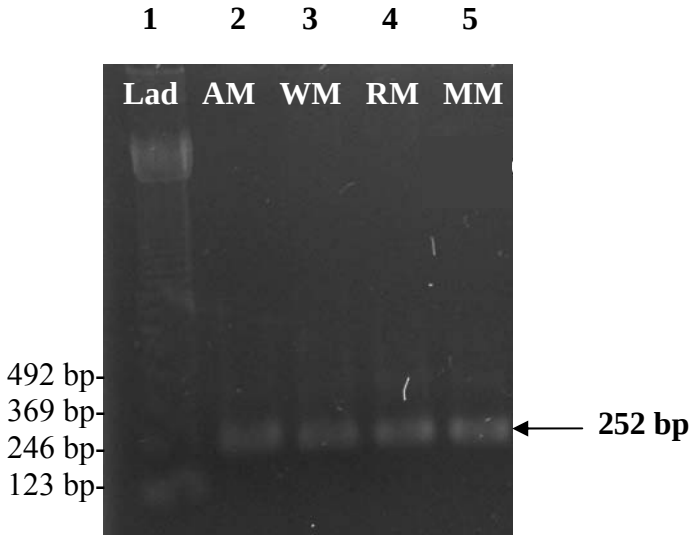
Lanes 2–5 are the amplicons from the different habitats that amplified using the *B. japonicum* primer set. Lane 1 is a 100bp DNA ladder. Non-specific binding seen in lane 5 (Marsh) was not observed with the incorporation of the probe in real-time.

Figure 15. *Acidobacteria* group 1 amplification results with soil DNA



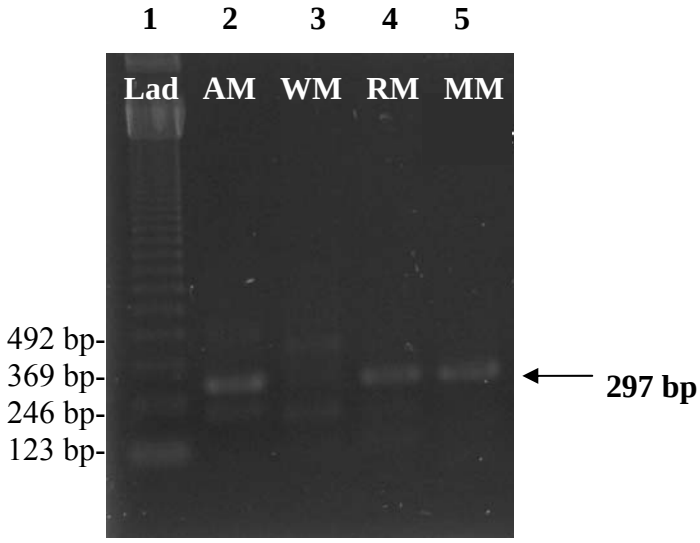
Lanes 2–5 are the amplicons from the different habitats that amplified using the *Acidobacteria* group 1 primer set. Lane 1 is a 123bp DNA ladder.

Figure 16. Genus *Burkholderia* amplification results with soil DNA



Lanes 2–5 are the amplicons from the different habitats that amplified using the *Burkholderia* primer set. Lane 1 is a 123bp DNA ladder.

Figure 17. Genus *Agrobacterium* amplification results with soil DNA

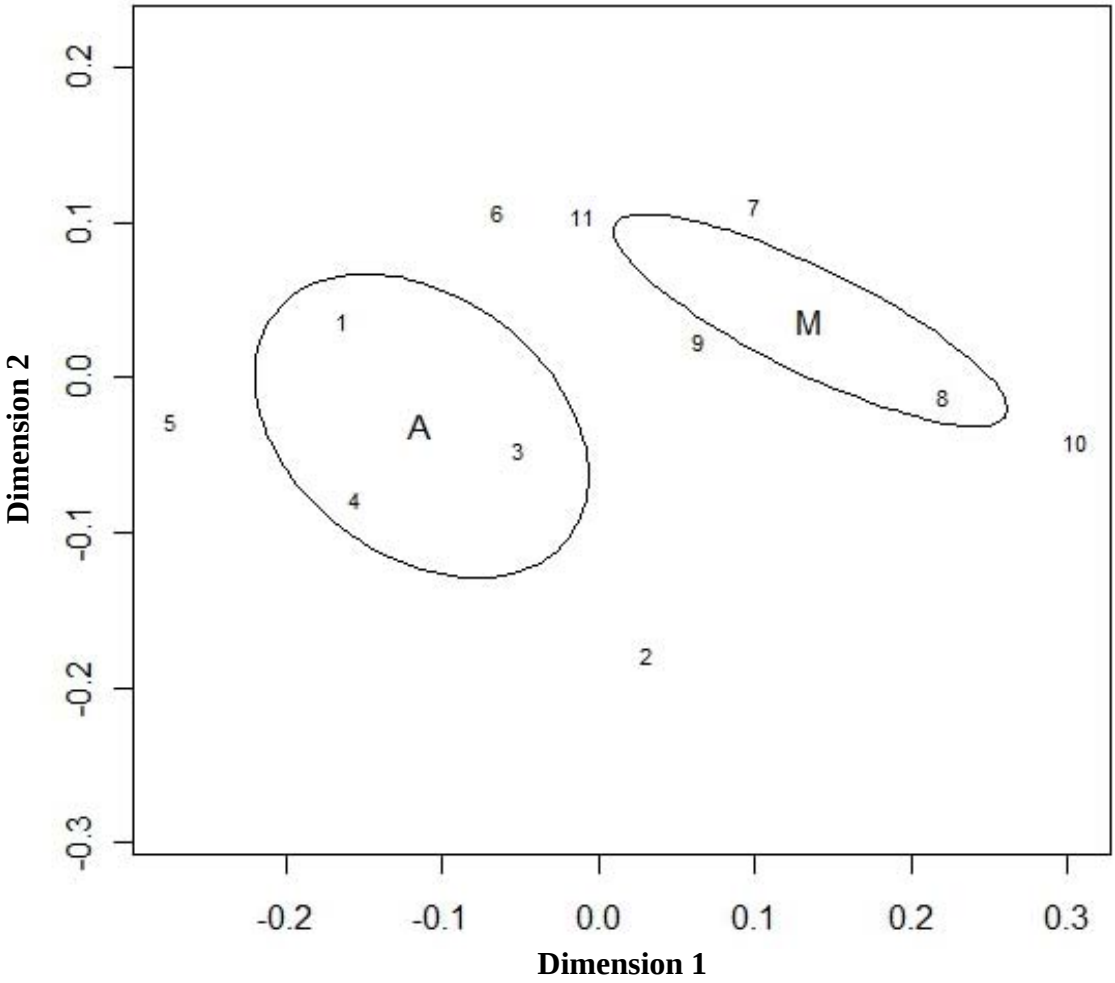


Lanes 2–5 are the amplicons from the different habitats that amplified using the *Agrobacterium* primer set. No amplification was observed in Lane 3 (Woodlot). Lane 1 is a 123bp DNA ladder.

APPENDIX C:

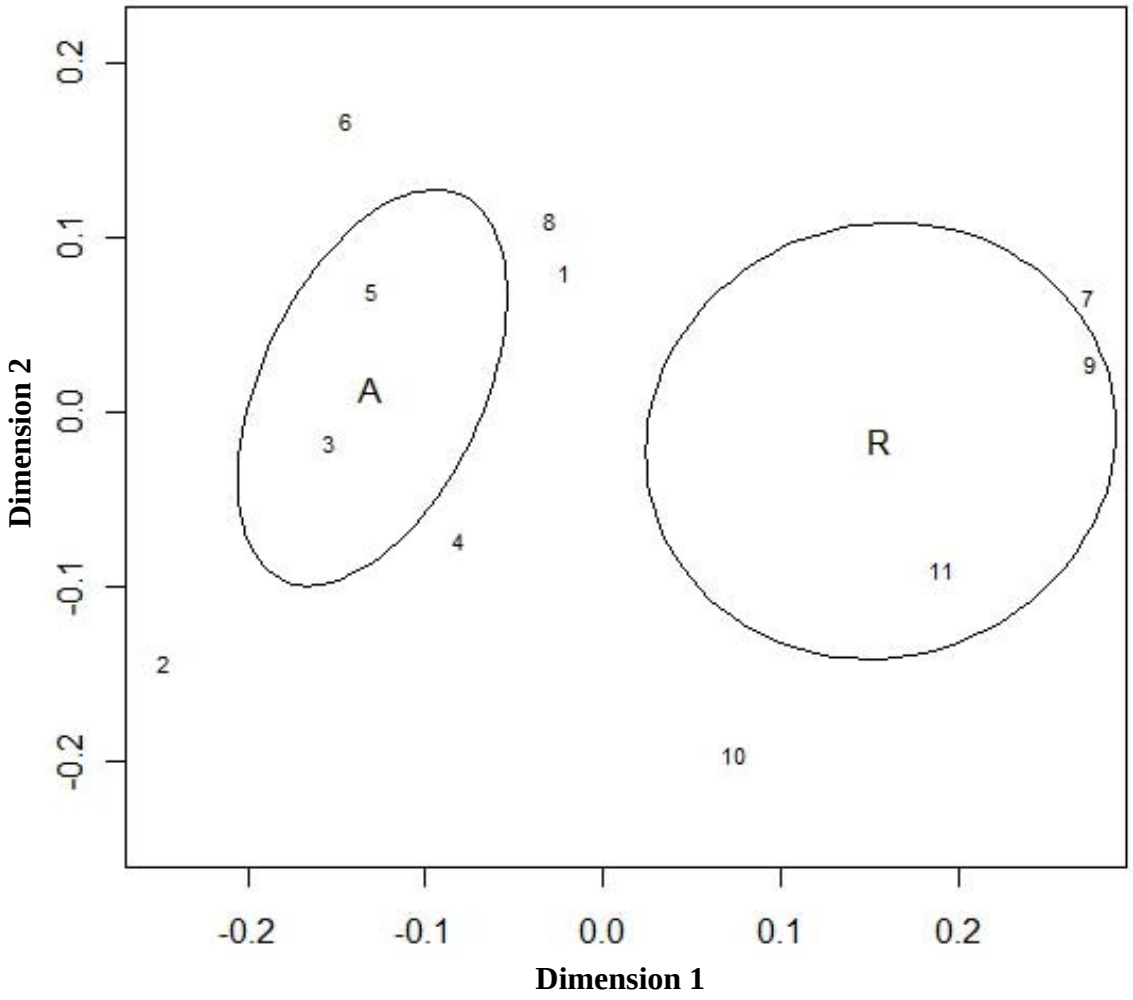
Habitat to Habitat Variability

Figure 18. Pairwise comparison of the Agricultural Field and Marsh



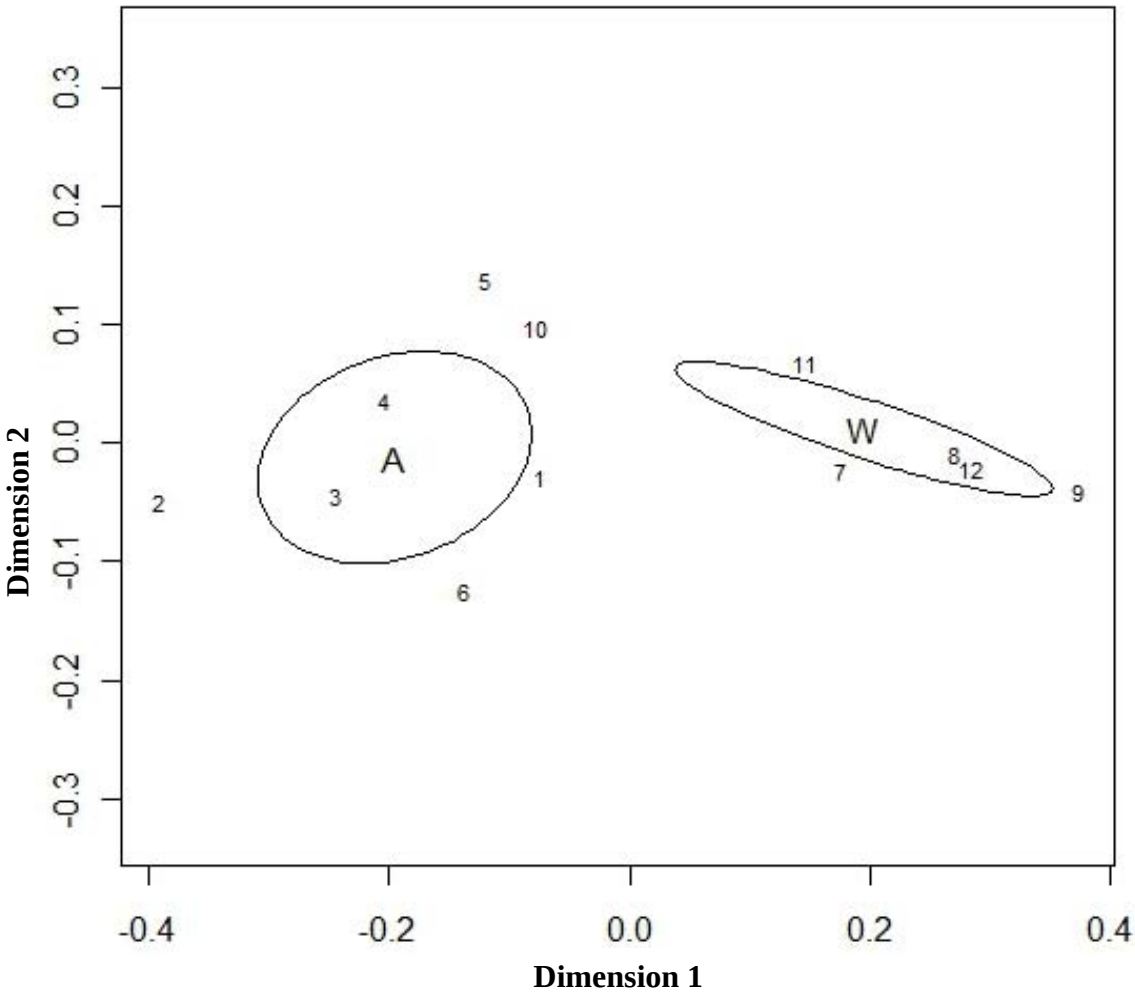
NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–6) and the Marsh (7–11). Habitats were separated when compared in a pairwise manner.

Figure 19. Pairwise comparison of the Agricultural Field and Yard



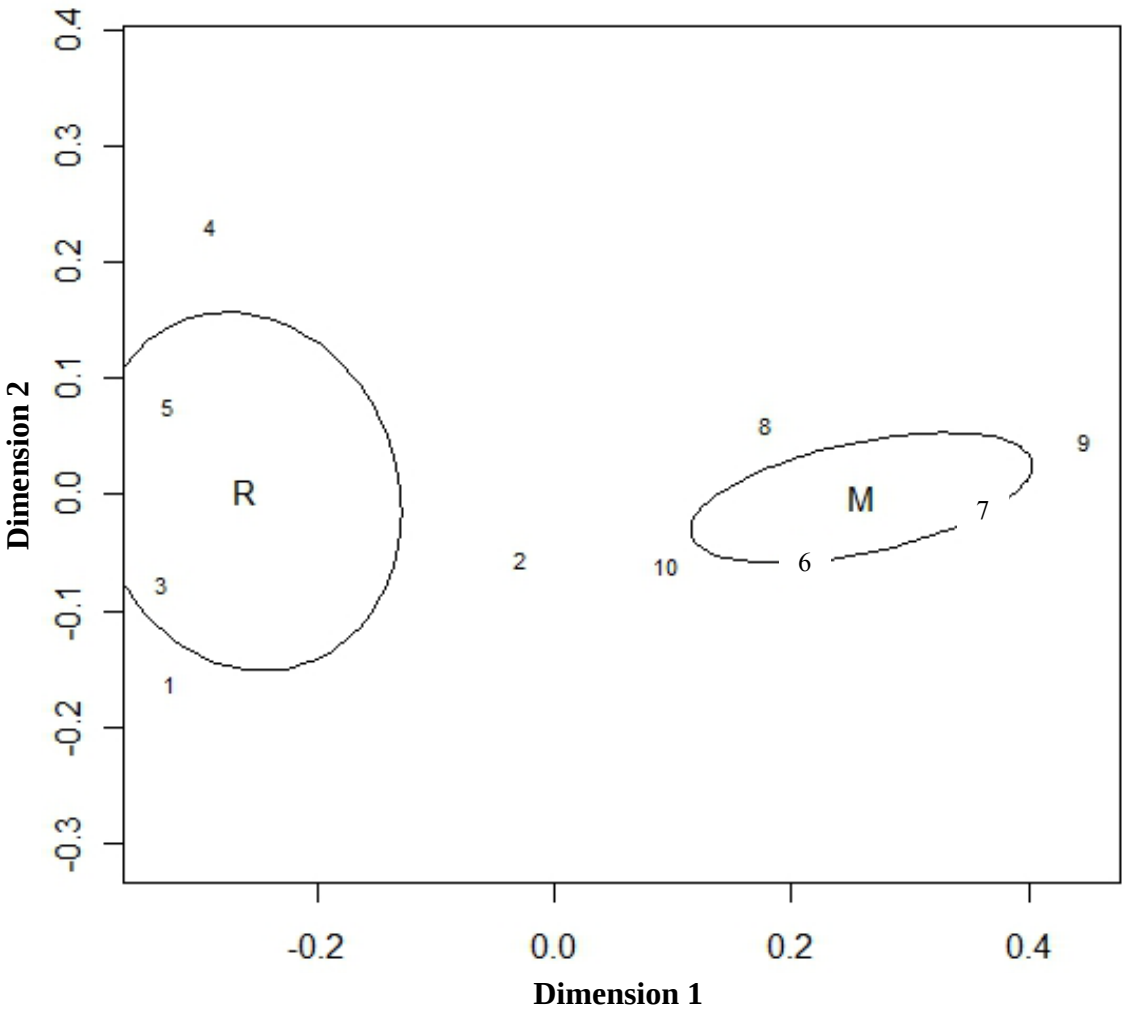
NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–6) and the Yard (7–11). Habitats were separated when compared in a pairwise manner.

Figure 20. Pairwise comparison of the Agricultural Field and Woodlot



NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–6) and the Woodlot (7–12). Habitats were separated when compared in a pairwise manner.

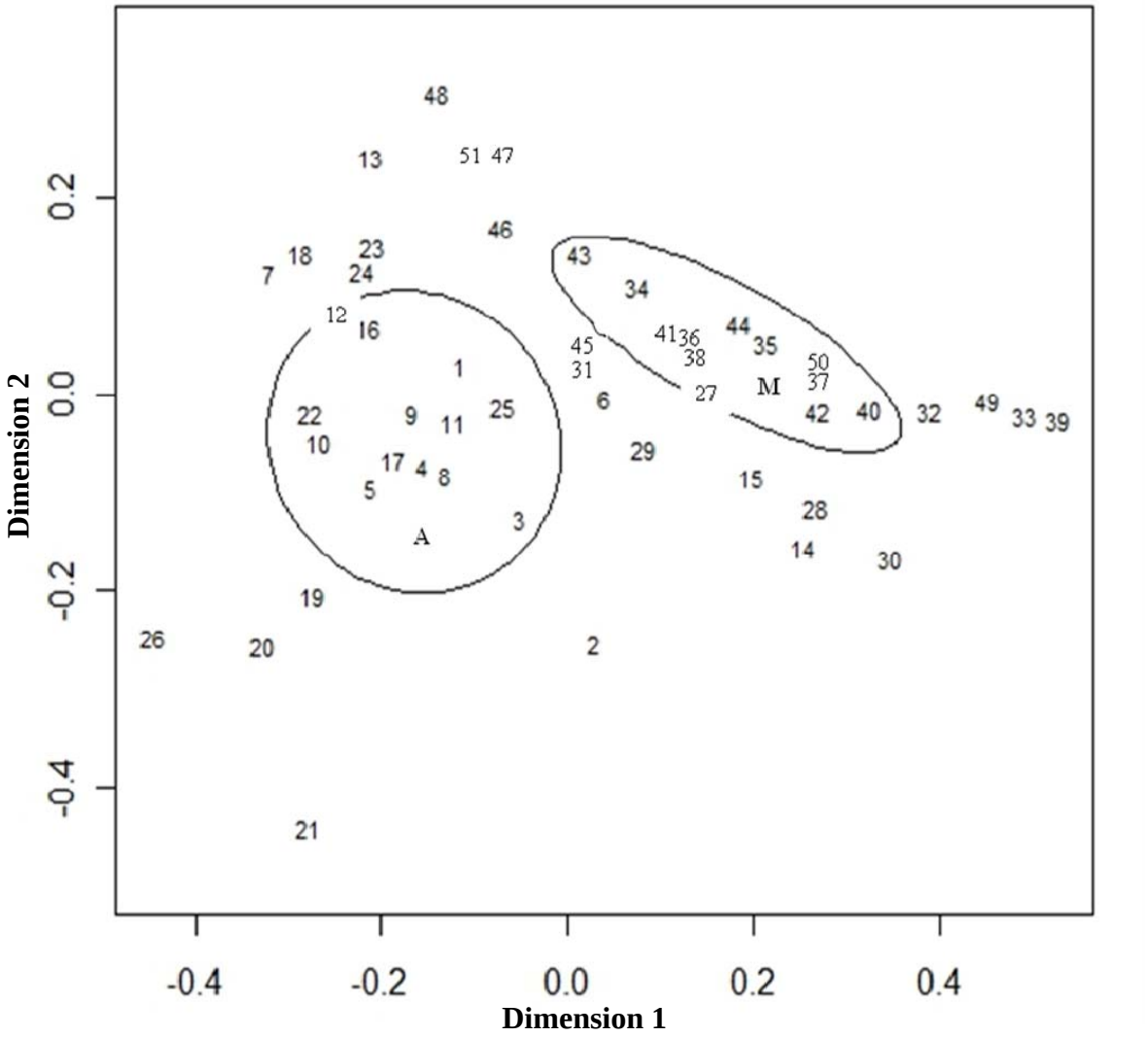
Figure 21. Pairwise comparison of the Yard and Marsh



NMDS plot showing the 95% confidence ellipses around the samples for the Yard (1–5) and the Marsh (6–10). Habitats were separated when compared in a pairwise manner.

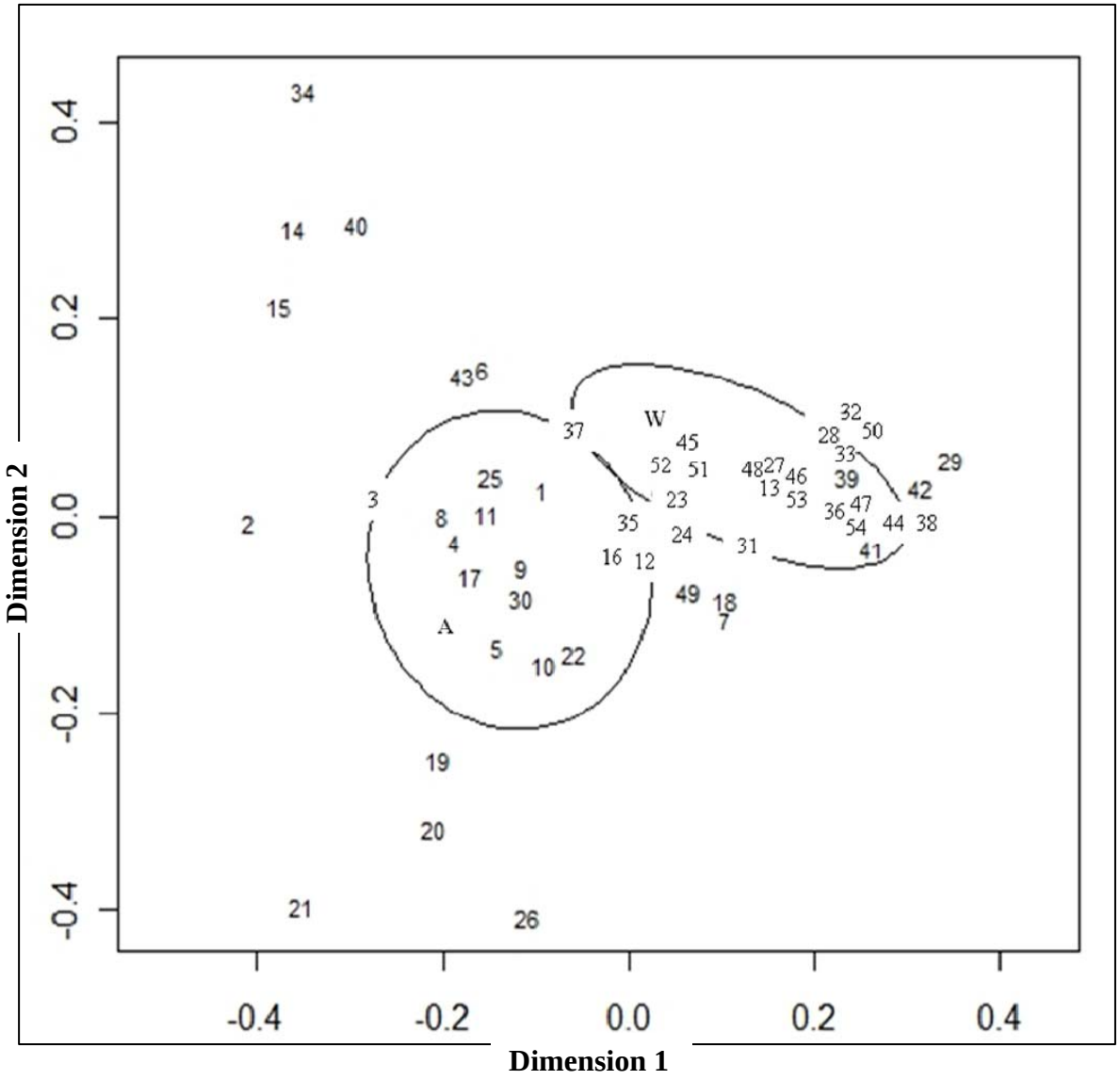
Intra-habitat Temporal Variability

Figure 22. Pairwise comparison of temporal data for the Agricultural Field and Marsh



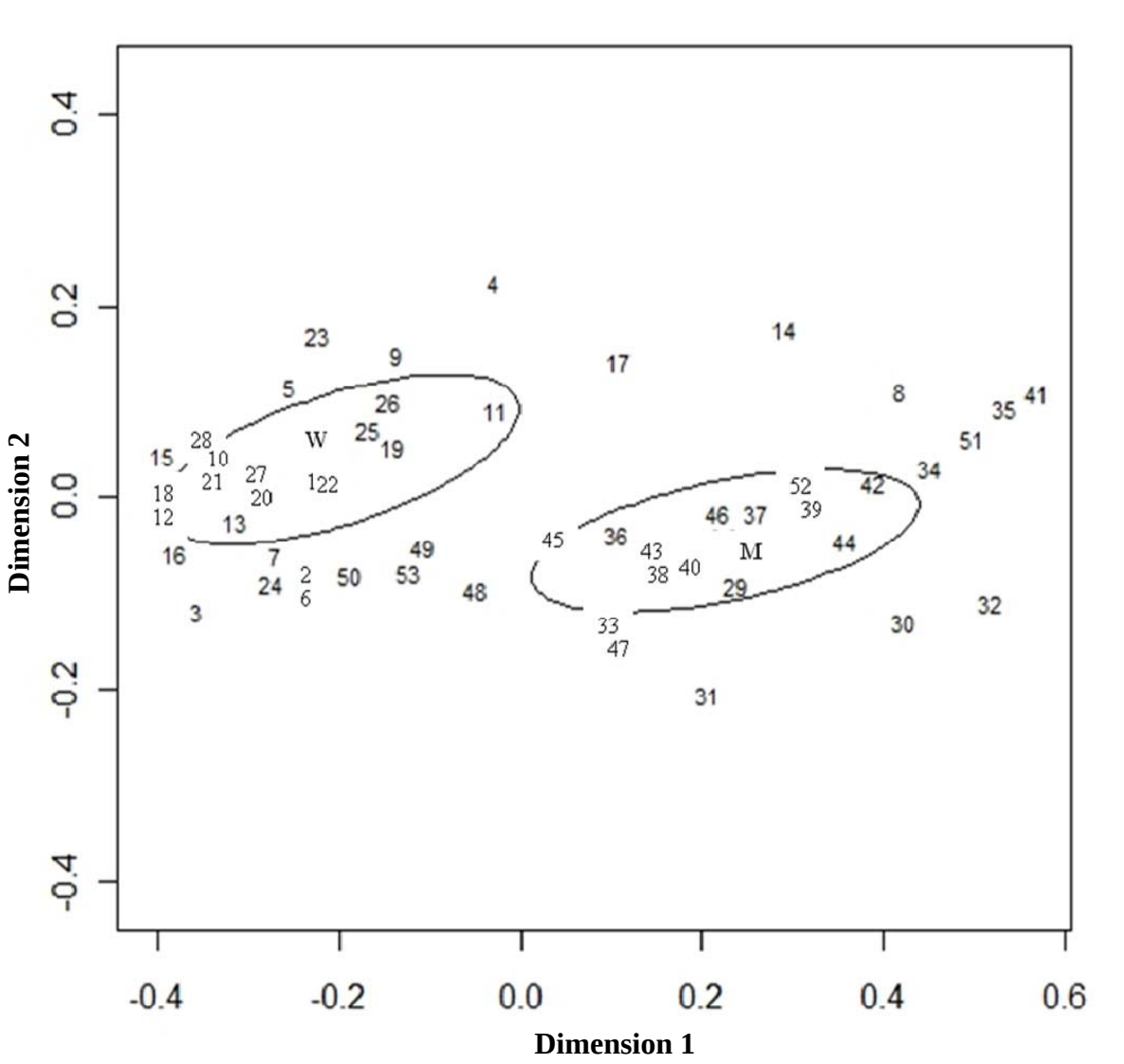
NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–26) and the Marsh (27–51). Habitats were separated when compared in a pairwise manner.

Figure 23. Pairwise comparison of temporal data for the Agricultural Field and Woodlot



NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–26) and the Woodlot (27–54). Habitats were separated when compared in a pairwise manner.

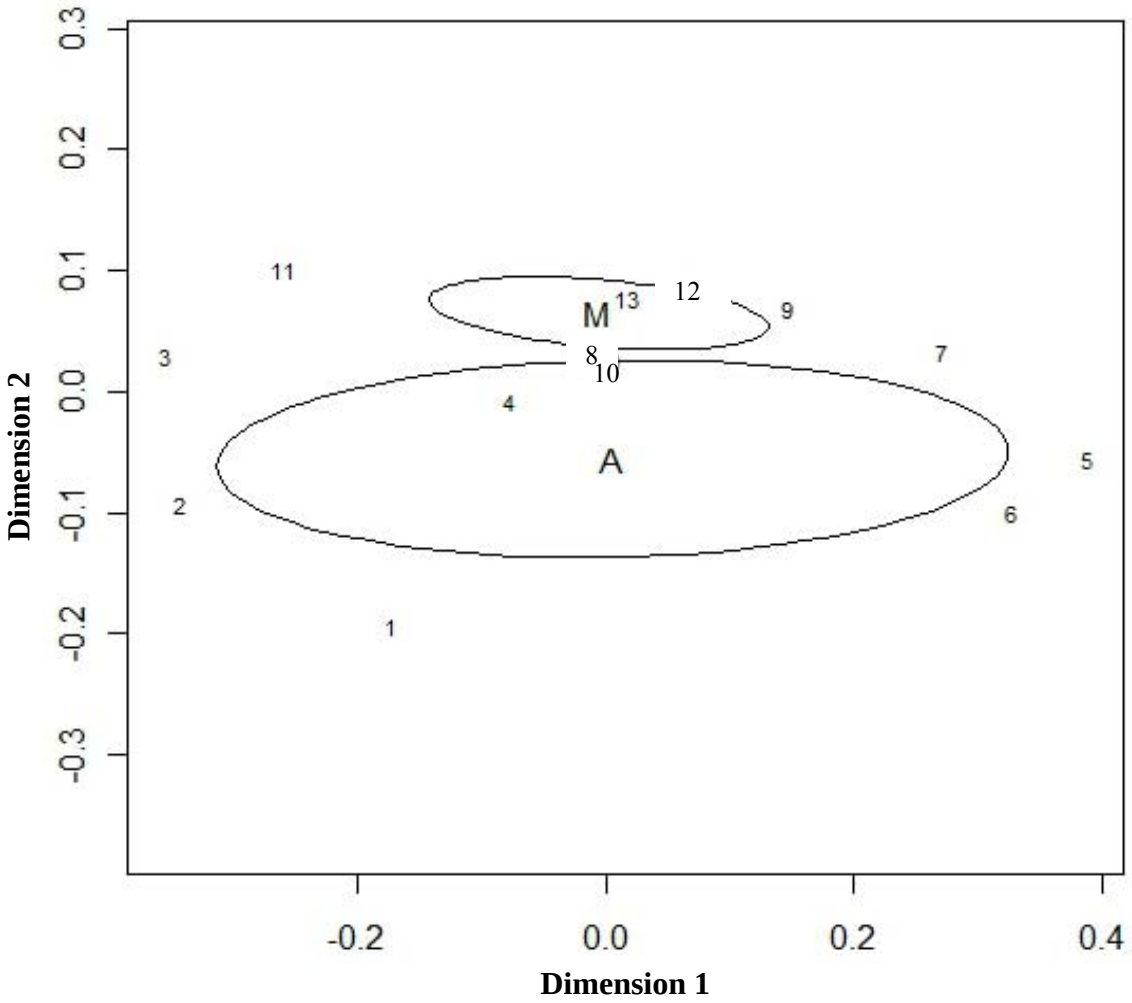
Figure 24. Pairwise comparison of temporal data for the Woodlot and Marsh



NMDS plot showing the 95% confidence ellipses around the samples for the Woodlot (1–28) and the Marsh (29–53). Habitats were separated when compared in a pairwise manner.

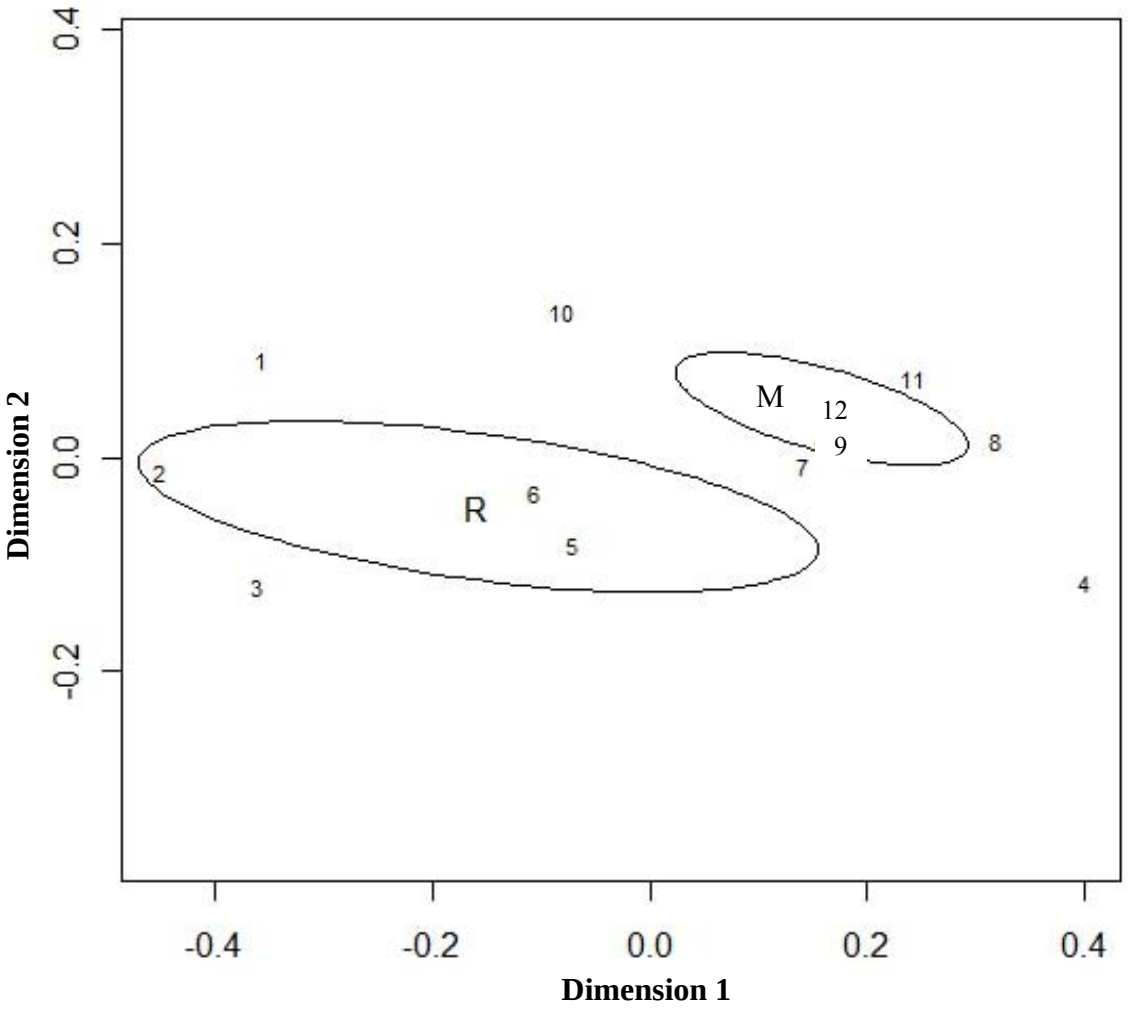
Intra-habitat Spatial Variability-Distance

Figure 25. Pairwise comparison of distance data for the Agricultural Field and Marsh in the Fall



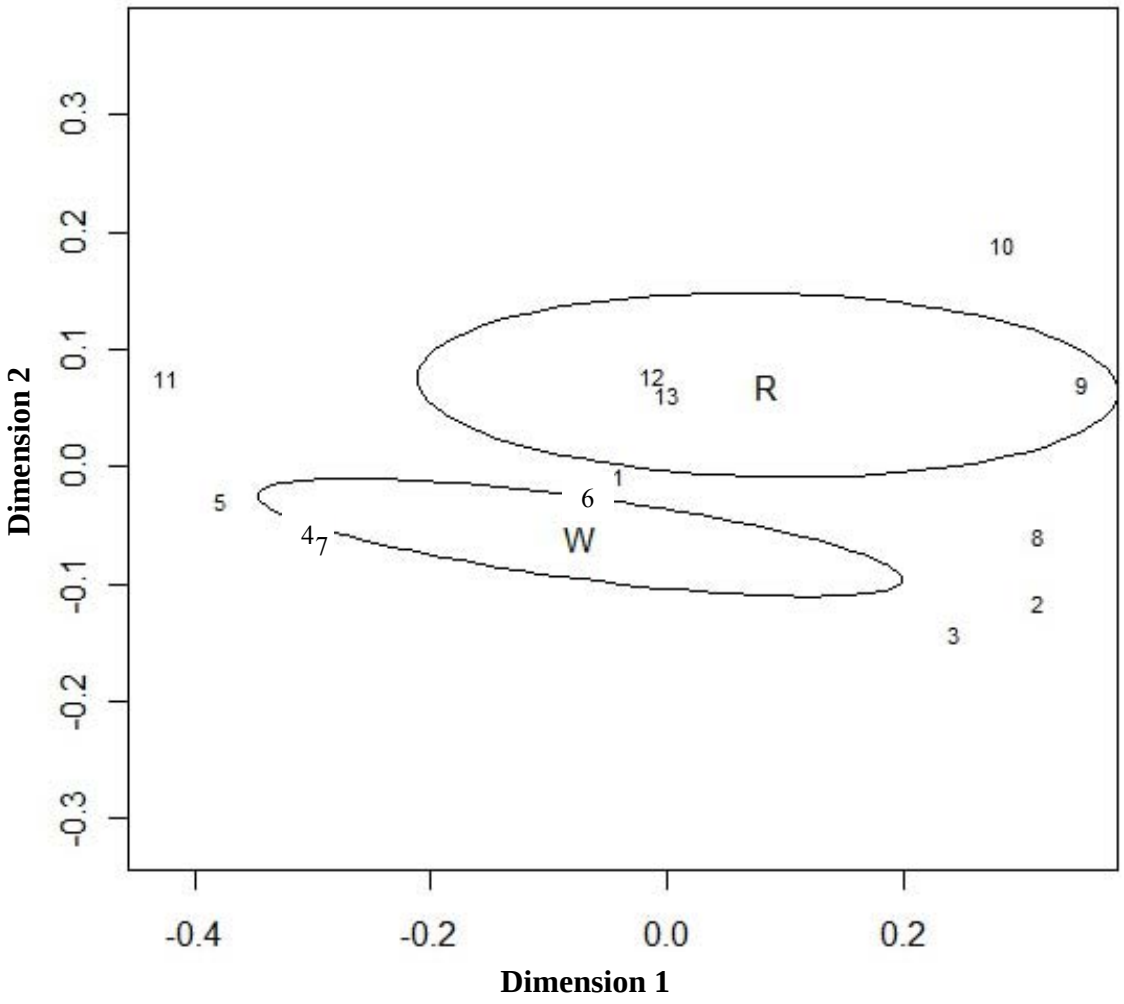
NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–7) and the Marsh (8–13). Habitats were separated when compared in a pairwise manner.

Figure 26. Pairwise comparison of distance data for the Yard and Marsh in the Fall



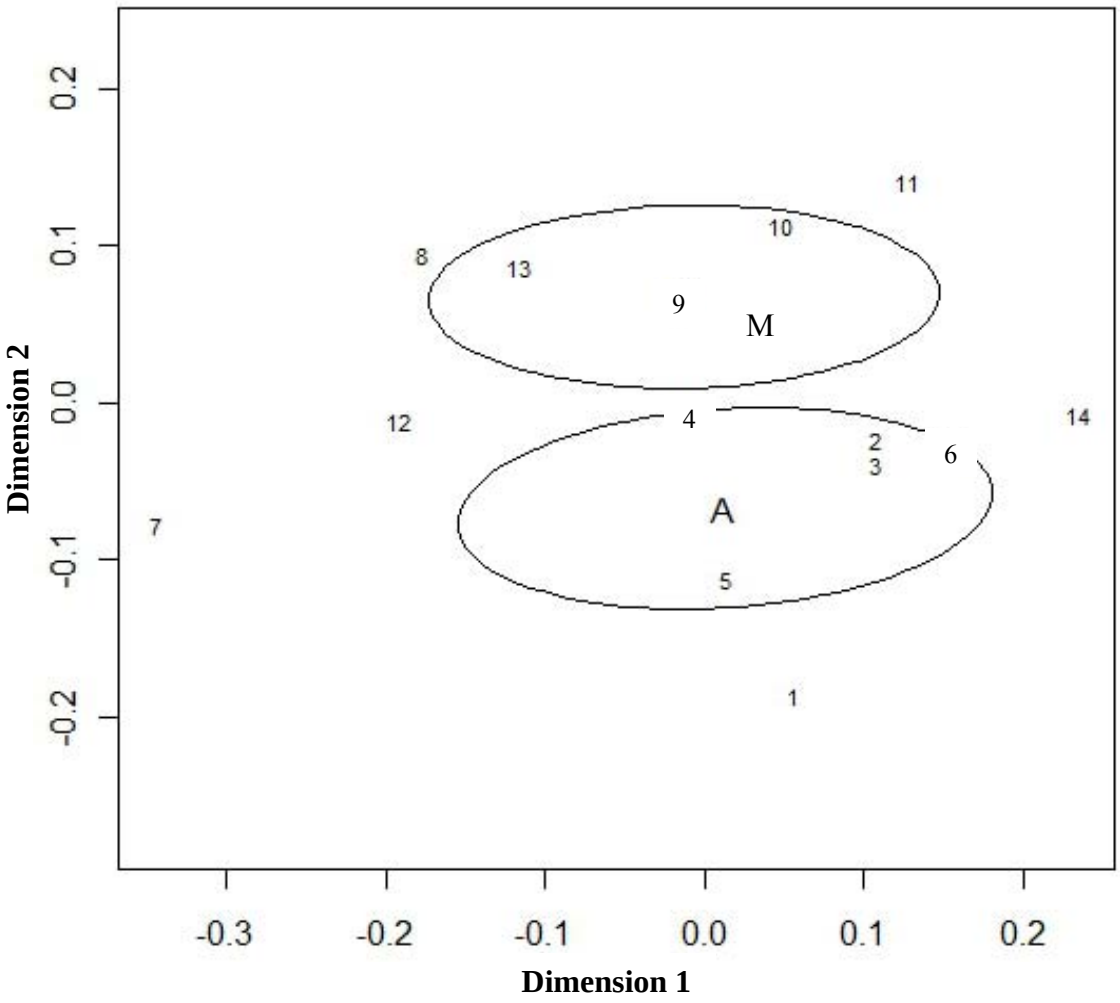
NMDS plot showing the 95% confidence ellipses around the samples for the Yard (1–6) and the Marsh (7–12). Habitats were separated when compared in a pairwise manner.

Figure 27. Pairwise comparison of distance data for the Woodlot and Yard in the Fall



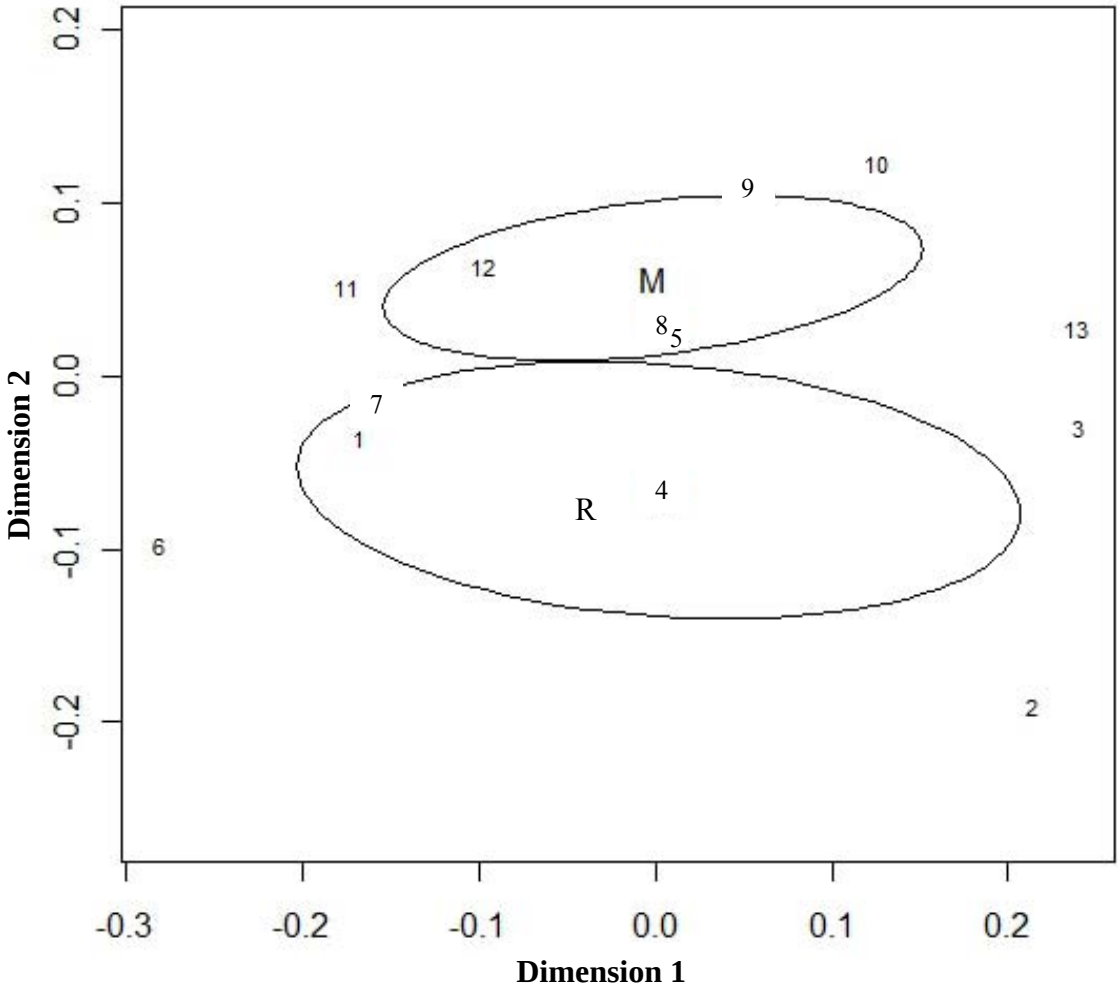
NMDS plot showing the 95% confidence ellipses around the samples for the Woodlot (1–7) and the Yard (8–13). Habitats were separated when compared in a pairwise manner.

Figure 28. Pairwise comparison of distance data for the Agricultural Field and Marsh in the Spring



NMDS plot showing the 95% confidence ellipses around the samples for the Agricultural Field (1–7) and the Marsh (8–14). Habitats were separated when compared in a pairwise manner.

Figure 29. Pairwise comparison of distance data for the Yard and Marsh in the Spring



NMDS plot showing the 95% confidence ellipses around the samples for the Yard (1–6) and the Marsh (7–13). Habitats were separated when compared in a pairwise manner.

REFERENCES

REFERENCES

Anderson M. A new method for non-parametric multivariate analysis of variance. Austral Ecology 2001;26(1):32–46.

Applied Biosystems. Real-Time PCR Vs. Traditional PCR. <http://www6.applied biosystems.com/support/tutorials/pdf/rtpcr_vs_tradpcr.pdf>. Accessed 2011 June 10.

Barns SM, Takala SL, Kuske CR. Wide Distribution and Diversity of Members of the Bacterial Kingdom *Acidobacterium* in the Environment. Applied and Environmental Microbiology 1999;65(4):1731–1737.

Bell C, McIntyre N, Cox S, Tissue D, Zak J. Soil Microbial Responses to Temporal Variations of Moisture and Temperature in a Chihuahuan Desert Grassland. Microbial Ecology 2008;56(1):153–167.

Bio-Rad Laboratories. Real-Time PCR Fundamentals. < http://www3.bio-rad.com/B2B/ vanity/gexp/content.do?language=English?language=English&ccatoid=-36529& pcatoid=-35468&com.broadvision.session.new=Yes&root=%2fProduct+Family %2fGX%2fHome&country=US&BV_SessionID=@@@@0016780371.1313463 920@@@@&BV_EngineID=ccccadfeggmkdkhcfngcfkmdhkkdflm.0> Accessed 2011 June 10.

Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs 1957;27(4):326–349.

Bull PA, Parker A, Morgan RM. The forensic analysis of soils and sediment taken from the cast of a footprint. Forensic Science International 2006;162:6–12.

Marumo Y. Forensic Examination of Soil Evidence. Japanese Journal of Forensic Science and Technology 2002;7(2):95–111.

Carter DO, Yellowlees D, Tibbett M. Cadaver decomposition in terrestrial ecosystems. Naturwissenschaften 2007;94:12–24.

Castro HF, Classen AT, Austin EE, Norby RJ, Schadt CW. Soil Microbial Community Responses to Multiple Experimental Climate Change Drivers. Applied and Environmental Microbiology 2010;76(4):999–1007.

Cengage G. 2006. Popp, Georg. World of Forensic Science. Ed. Lerner KL and Lerner BW. <http://www.enotes.com/forensic-science/popp-georg>. Accessed 2011 June 10.

Cernohlavkova J, Jarkovsky J, Nesporova M , Hofman J. Variability of soil microbial properties: Effects of sampling, handling and storage. Ecotoxicology and Environmental Safety 2008;72(8):2102–2108.

Clarridge JE. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clinical Microbiology Reviews 2004;17(4):840–862.

Croft DJ, Pye K. Multi-technique comparison of source and primary transfer soil samples: an experimental investigation. Science and Justice 2004;44(1):21–28.

Coenye T, Vandamme P. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. Environmental Microbiology 2003;5(9):719–729.

Daffonchio D, Cherif A, Brusetti L, Rizzi A, Mora D, Boudabous A et al. Nature of Polymorphisms in 16S-23S rRNA Gene Intergenic Transcribed Spacer Fingerprinting of *Bacillus* and Related Genera. Applied and Environmental Microbiology 2003;69(9):5128–5137.

Dudley RJ. The use of density gradient columns in the forensic comparison of soils. Medicine and Science Law 1979;19(1):39–48.

Duodu S, Bhuvaneswari TV, Gudmundsson J, Svenning MM. Symbiotic and saprophytic survival of three unmarked *Rhizobium leguminosarum* biovar *trifolii* strains introduced into the field. Environmental Microbiology 2005;7(7):1049–1058.

Eichorst SA, Breznak JA, Schmidt TM. Isolation and Characterization of Soil Bacteria That Define *Terriglobus* gen. nov., in the Phylum *Acidobacteria*. Applied and Environmental Microbiology 2007;73(8):2708–2717.

Escobar MA, Dandekar AM. Agrobacterium tumefaciens as an agent of disease. Trends in Plant Science 2003;8(8):380–386.

Faith DP, Minchin PR, Belbin L. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 1987;69(1/3):57–68.

Felske A, Rheims H, Wolterink A, Stackebrandt E, Akkermans ADL. Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. Microbiology 1997;143:2983–2989.

Franklin RB, Mills AL. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiology Ecology 2003;44(3):335–346.

Fuka MM, Engel M, Hagn A, Munch JC, Sommer M, Schloter M. Changes of Diversity Pattern of Proteolytic Bacteria over Time and Space in an Agricultural Soil. Microbial Ecology 2009;57(3):391–401.

Gremion F, Chatzinotas A, Harms H. Comparative 16S rDNA and 16S rRNAsequence analysis indicates that *Actinobacteria* might be a dominant part of the metabolically active bacteria in heavy metal contaminated bulk and rhizosphere soil. Environmental Microbiology 2003;5(10):896–907.

Gruntzig V, Nold SC, Zhou J, Tiedje JM. *Pseudomonas stutzeri* nitrite reductase gene abundance in environmental samples measured by real-time PCR. Applied Environmental Microbiology 2001;67:760–768.

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium 1999;41:95–98.

Heid CA, Stevnes J, Livak KJ, Williams PM. Real Time Quantitative PCR. Genome Research 1996;6:986–994.

Heuer H, Krsek M, Baker P, Smalla K, Wellington EMH. Analysis of Actinomycete communities by specific genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Applied and Environmental Microbiology 1997;63(8):3233–3241.

Higuchi R, Dollinger G, Walsh PS, Griffith R. Simultaneous amplification and detection of specific DNA sequences. Biotechnology 1992;10(4):413–417.

Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. Proceedings of the National Academy of Science 1991;88:7276–7280.

Horrocks M, Coulson SA, Walsh KA. Forensic palynology: Variation in the pollen content of soil on shoes and in shoeprints in soil. Journal of Forensic Sciences 1999;44:119–122.

Horrocks M. Sub-sampling and preparing forensic samples for pollen analysis. Journal of Forensic Sciences 2004;49(5):1024–1027.

Horrocks M, Walsh KA. Pollen on grass clippings: putting the suspect at the scene of the crime. Journal of Forensic Sciences 2001;46(4):947–949.

Hristova RK, Lutenegger MC, Scow MK. Detection and quantification of methyl *tert*-butyl ether degrading strain PMI by real-time TaqMan PCR. Applied Environmental Microbiology 2001;67:5154–5160.

Janssen PH. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Applied and Environmental Microbiology 2006;72(3):1719–1728.

Jordan DC. Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a Genus of Slow-Growing, Root Nodule Bacteria from Leguminous Plants. International Journal of Systematic Bacteriology 1982;32(1):136–139.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences 1985;82(20):6955–6959.

Lee C, Kim J, Hwang K, O'Flaherty V, Hwang S. Quantitative analysis of methanogenic community dynamics in three anaerobic batch digesters treating different wastewaters. Water Research 2009;43(1):157–165.

Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Research 1993;21(16):3761–3766.

Lee SH, Cho JC. Distribution Patterns of the Members of Phylum *Acidobacteria* in Global Soil Samples. Journal of Microbiology and Biotechnology 2009;19(11):1281–1287.

Lenz EJ, Foran DR. Bacterial profiling of soil using genus-specific markers and multidimensional scaling. Journal of Forensic Sciences 2010;55(6):1437–1442.

Livak KJ, Flood SJA, Marmaro J, Giusti W, Deetz K. Oligonucleotides with Fluorescent Dyes at Opposite Ends Provides a Quenched Probe System Useful for Detecting PCR Product. PCR Methods and Applications 1995;4:357–362.

Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $^{-\Delta\Delta C}{}_{T}$ Method. Methods 2001;25:402–408.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar et al. ARB: a software environment for sequence data. Nucleic Acids Research 2004;32(4):1363–1371.

Marumo Y. Determination of free oxides in soils by x-ray fluorescence spectrometry. Japanese Journal of Soil Science and Plant Nutrition 1989;60:99–105.

Marumo Y. Morphological examination of grass fragments in forensic soil comparison. Proceedings of the International Symposium on the Forensic Aspects of Trace Evidence 1991;22.

Meyers MS, Foran DR. Spatial and temporal influences on bacterial profiling of forensic soil samples. Journal of Forensic Sciences 2008;53(3):652–660.

Muyzer G, Smalla K. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 1998;73:127–141.

National Research Council. Strengthening Forensic Science in the United States: A Path Forward. Washington DC: The National Academies Press, 2009.

Nute HD. An improved density gradient system for forensic science soil studies. Journal of Forensic Sciences 1975;20(4):668–73.

Oksanen J. Multivariate Analysis of Ecological Communities in R: vegan tutorial. Oulu, Finland: Univ. of Oulu, 2011.

Parke JL, Gurian-Sherman D. Diversity of the *Burkholderia cepacia* Complex and Implications for Risk Assessment of Biological Control Strains. Annual Reviews of Phytopathology 2001;39:225–258.

Petraco N, Kubic T. A density gradient technique for use in forensic soil analysis. Journal of Forensic Sciences 2000;45(4):872–873.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 2007;35(21):7188–7196.

R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Development Core Team, 2011.

Rawlins BG, Kemp SJ, Hodgkinson EH, Riding JB, Vane CH, Poulton C et al. Potential and Pitfalls in Establishing the Provenance of Earth-Related Samples in Forensic Investigation. Journal of Forensic Sciences 2006;51(4):832–845.

Ruffell A, Wiltshire P. Conjunctive use of quantitative and qualitative X-ray diffraction analysis of soils and rocks for forensic analysis. Forensic Science International 2004;145:13–23.

Sait M, Davis KER, Janssen PH. Effect of pH on Isolation and Distribution of Members of Subdivision 1 of the Phylum *Acidobacteria* Occurring in Soil. Applied and Environmental Microbiology 2006;72(3):1852–1857.

Spain AM, Krumholz LR, Elshahed MS. Abundance, composition, diversity and novelty of soil Proteobacteria. The ISME Journal 2009;3(8):992–1000.

Stokes KL, Forbes SL, Tibbett M. Freezing skeletal muscle tissue does not affect its decomposition in soil: evidence from temporal changes in tissue mass, microbial activity and soil chemistry based on excised samples. Forensic Science International 2009;183(1-3):6–13.

Suzuki MT, Taylor LT, DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Applied Environmental Microbiology 2000;66:4605–4614.Wanogho S, Gettiney G, Caddy B, Robertson J. Some factors affecting soil sieve analysis in forensic science.1.Dry sieving. Forensic Science International 1987;33:129–137.

Skovhus TL, Ramsing NB, Holmstrom C, Kjelleberg S, Dahllof I. Real-Time Quantitative PCR for Assessment of Abundance of *Pseudoalteromonas* Species in Marine Samples. Applied and Environmental Microbiology 2004;70(4):2373–2382.

Wang Y, Qian PY. Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. PLoS One 2009;4(10):e7401.

Weir BS. 2010. The current taxonomy of rhizobia. <http://www.rhizobia.co.nz/taxonomy/rhizobia.html>. Accessed 2011 June 10.

Widmer F, Fliessbach A, Laczko E, Schulze-Aurich J, Zeyer J. Assessing soil biological characteristics: a comparison of bulk soil community DNA-, PLFA-, and Biolog™-analyses. Soil Biology & Biochemistry 2001;33:1029–1036.

Will C, Thurmer A, Wollherr A, Nacke H, Herold N, Schrumpf M et al. Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. Applied and Environmental Microbiology 2010;76(20):6751–6759.

Woese CR, Stackebrandt E, Macke TJ, Fox GE. A phylogenetic definition of the major eubacterial taxa. Systematic Applications of Microbiology 1985;6:143–151.

Zadora G, Brozek-Mucha Z. SEM-EDX—a useful tool for forensic examinations. Materials Chemistry and Physics 2003;81:345–348.