12-3-2007

Amplicon length heterogeneity (ALH)-PCR generated bacterial community profiling : a novel application for the forensic examination of soil

Todd Martin Crandall
Florida International University

DOI: 10.25148/etd.FI14061539

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

AMPLICON LENGTH HETEROGENEITY (ALH)-PCR GENERATED BACTERIAL COMMUNITY PROFILING; A NOVEL APPLICATION FOR THE FORENSIC EXAMINATION OF SOIL

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

FORENSIC SCIENCE

by

Todd Martin Crandall

2009

To: Dean Kenneth Furton
College of Arts and Sciences

This thesis, written by Todd Martin Crandall, and entitled Amplicon Length Heterogeneity (ALH)-PCR Generated Bacterial Community Profiling; A Novel Application for the Forensic Examination of Soil, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Jose Almirall

DeEtta Mills

Giri Narasimhan

Kalai Mathee, Major Professor

Date of Defense: December 3, 2007

The thesis of Todd Martin Crandall is approved.

Dean Kenneth Furton
College of Arts and Sciences

Dean George Walker
University Graduate School

Florida International University, 2009

ACKNOWLEDGMENTS

I would like to thank my wife Brooke first for standing by me with encouragement, support and patience throughout this challenging time. I also appreciate her sacrificing her time and personal pursuits to help me pursue higher education. I am thankful to my major advisor, Dr. Mathee for pushing me to complete this thesis and not give up as well as for expanding my cultural horizons. I appreciate the assistance and advice of my committee and their lab personnel. Specifically, I thank Dr. Mills, Dr. Almirall, Dr. Narasimhan, Dr. Entry, Liliana Moreno, Robert Sautter, Melissa Doud, Dr. Jeannette Perr and Dr. Chengyong Yang.

I am extremely thankful to my sons, Jonah, Ezra, Reuben and Jude, for their silent motivation, endless unconditional love and constant cheering me onward. Finally, I thank my Heavenly Father for His careful watch, getting me through all the challenges of recent years, both personal and professional.

ABSTRACT OF THE THESIS

AMPLICON LENGTH HETEROGENEITY (ALH)-PCR GENERATED BACTERIAL COMMUNITY PROFILING; A NOVEL APPLICATION FOR THE FORENSIC EXAMINATION OF SOIL

by

Todd Martin Crandall

Florida International University, 2009

Miami, Florida

Professor Kalai Mathee, Major Professor

Current forensic comparisons of soil most often rely upon physical characterizations. We hypothesized that bacterial community profiles obtained by Amplicon Length Heterogeneity-Polymerase Chain Reaction (ALH-PCR) of the 16S *rRNA* genes would provide discriminating data for soil comparisons. Dual extractions and replicate amplifications were performed on each soil. Chemical characterization by elemental analysis, pH, moisture content, percent Carbon and percent Nitrogen were performed. Supervised classification of the microbial community profiles using a Support Vector Machine (SVM) learning tool was over 95 % accurate labeling a microbial community profile to its originating soil type. By comparison, the chemical analysis data yielded accuracies between 40 and 77 %. The results of this study support the application of this method in the comparison of casework size soil samples. Results of this study may also justify the future development of a database of microbial community profiles for inferring the possible origin of unknown soil samples.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

INTRODUCTION

The goal of all forensic examinations is to be able to compare samples of evidence in such a way that one can determine with a great degree of certainty whether or not the samples came from the same source. The principle behind forensic sample comparison can be found in a statement by French doctor Edmund Locard. The Locard Exchange Principle states that when two objects come into contact, there is a division and transfer of material (104). Forensic science is concerned with finding aspects of the transferred materials which are unique or rare, making any comparison matches significant.

Forensic geologists applies his or her knowledge to compare earth materials from known and questioned samples in order to provide resolution in a public forum. A forensic geologist is an expert in geology trained in a multitude of disciplines. They apply their knowledge and abilities to assist criminal investigations in one of two ways. The first is a comparative analysis where a soil sample from evidence is compared to soil samples related to a suspect or their alibi. The second is to determine the origin or source of a soil sample. This second scenario requires extensive experience and knowledge as the examiner must identify the rare minerals in the sample and then find areas where those minerals are known to exist (79). To date, forensic geologists have provided valuable insights into many criminal investigations and have at times produced dramatic results. Although exact numbers are not given, it is safe to say that forensic geologists examine thousands of samples annually in North America alone (79). The presence of unnatural objects in soil such as glass, paint, asphalt and brick fragments can be very valuable as they can render higher levels of discrimination.

Soils are naturally complex systems. The addition of unnatural objects which must also be identified for comparison studies increases the complexity of any soil analysis. For these reasons, forensic soil examinations are approached as each individual case dictates. Knowledge of how to approach these cases can only come from years of experience and training.

Pioneers of Forensic Geology

The idea of using soil to elucidate the whereabouts of an individual was introduced about 1890 when Sir Arthur Conan Doyle wrote about how Sherlock Holmes could tell by the color and consistency of splashes on his trousers, in what part of London he had received them (29). The first book addressing forensic soil examination, *Handbuch für Untersuchungsrichter* (1893), was written by Hans Gross who is currently acknowledged as the founder of criminal investigation (9, 48). Gross wisely supported employing the microscopist and the mineralogist in the study of "dust, dirt on shoes and spots on cloth" saying that "dirt on shoes can often tell us more about where the wearer of the shoes had last been than toilsome inquiries" (50). Soil was first used as evidence in a criminal case in 1904 when Georg Popp examined soil deposits on murder suspect Karl Laubach's trousers (79). Popp was trained in chemistry but his laboratory also performed microscopic evaluations of food, mineral water and bacteria. Popp determined Laubach had been at the place where the victim's body was found and the route he had taken on the way home after discovering that two layers of soil on the suspect's pants compared with the sites. When police confronted Laubach with the evidence, he confessed to the crime. Since 1904, many criminal investigations have been aided by the examination of soils by forensic geologists.

Forensic Geological Methodology

Current techniques in forensic soil analysis focus on some of the physical properties of soil such as color hue, particle size and density distributions, mineral content, the presence of any foreign or unnatural objects as well as chemical properties of some minerals and clays (91). Various highly skilled microscopic techniques are used to identify naturally occurring rocks and minerals (79, 80). In order for a significant link to be established between two soil samples, the uniqueness of a particular sample compared to the surrounding area must also be examined (91). This makes sample collection an important consideration.

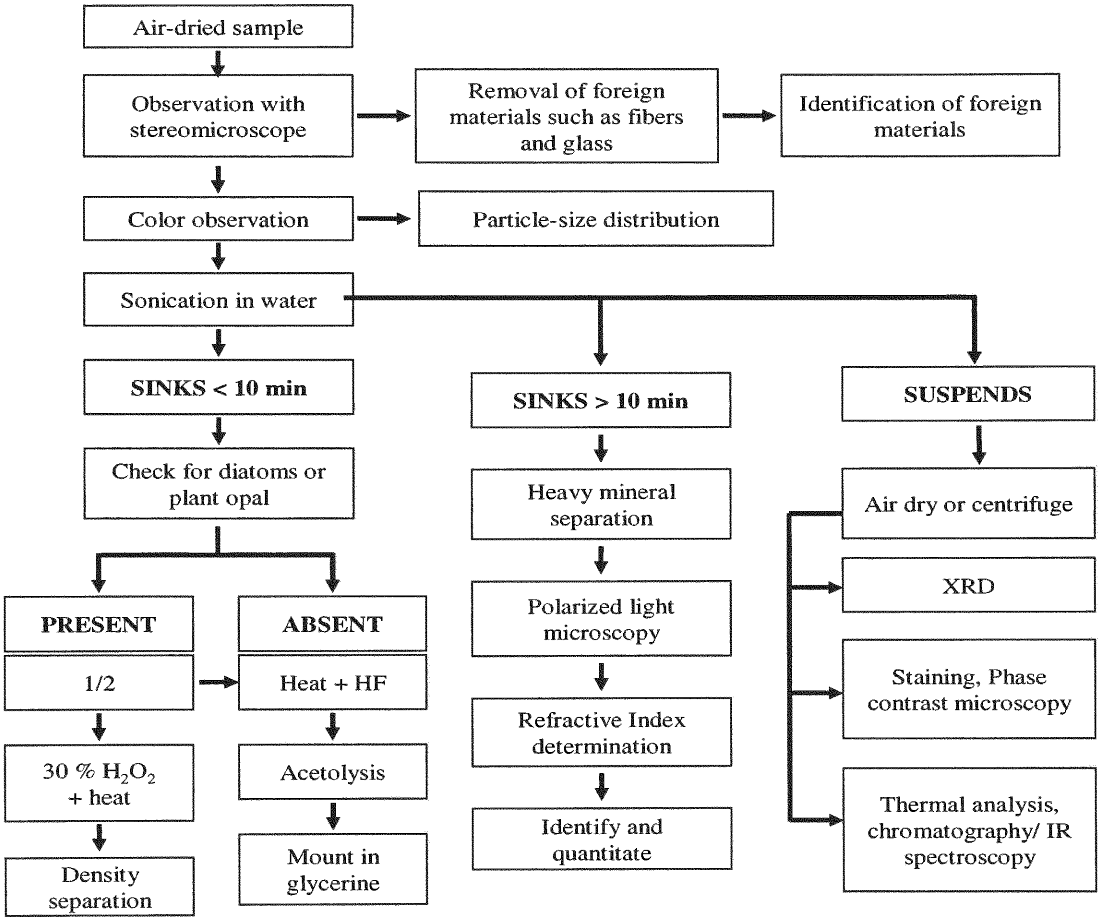
Sample Collection. The method of collecting reference samples for comparison to the evidence sample is determined by the surface on which the evidence sample was deposited (91). If the evidence sample was deposited on a shoe or a vehicle tire, it most likely came from a soils surface layer and reference samples should be collected accordingly. Soil evidence deposited on a shovel requires more thorough consideration as the soil may show high degrees of spatial variation both vertically and horizontally. Opposed to random sampling, examiners are trained to look for samples consistent in color and texture with the evidence sample so as not to miss a potential match (79). Perhaps the most famous case where proper sample collection played a crucial role occurred in the kidnapping and murder of the Coors® beer heir, Adolph Coors III (8). On February 9, 1960 Joseph Corbett Jr. kidnapped Coors. A resident of Jefferson County Colorado discovered Coors' truck on a bridge still running. That same day a note arrived addressed to Mrs. Adolph Coors demanding a $500,000 ransom. Eight days later, in Atlantic City New Jersey, Corbett burned his car which he had falsely registered under the name of Walter Osborne (8). Over six months later, hunters in Douglas County Colorado discovered Coors' key chain below a landfill which led to the discovery of most of Coors' body just a few days later. In the six or seven months before Coors' body was found however, investigators carefully collected samples from Corbett's burned vehicle revealing four layers of soil deposited under the fenders. In their effort to discover the location of the body, over 360 soil samples were taken along the western front of the Rocky Mountains near Denver. Although the body was discovered while the study was still in progress, analysis of the soil layers from the fender well and the reference samples revealed that the most recently deposited soil had come from the dump where the car was burned in New Jersey. The second most recent layer of soil had come from the site where Coors' remains were discovered. Corbett Jr. had driven from Colorado to New Jersey without picking up new soil or losing any soil which was already on the vehicle. The third most recent layer came from the area near Coors' ranch and the innermost and oldest layer was not identified to a particular location but was assumed to have come from an area near Denver. Joseph Corbett Jr. was linked to the vehicle and the vehicle was linked to both the kidnapping/murder site and the site where Coors' body was dumped, all the result of careful collection and analysis of the soil on the vehicles fenders. This evidence aided in the conviction of Corbett Jr. of murder. He was given a life sentence but was paroled in 1978 after having served eighteen years (69).

Proper sample collecting has some universal rules. Dry soil samples are placed in clean, leak-proof plastic containers. Wet soils must be dried before being stored for analysis as moisture can cause changes in minerals and changes in color, therefore, they are collected in paper or cloth containers (79). Samples must always be collected with a permit when necessary if they are to be admitted as evidence in a court of law. Samples must also be properly documented to maintain the chain of custody, which is a written record attached to the sample describing the people responsible for the sample's care as it is transferred. Following proper collection, documentation and storage of a soil sample, analysis begins as outlined in Figure 1.

Physical Characterization of Forensic Soils

Color. Color determination is one of the descriptive measures for identifying minerals. The color of a soil is determined by the primary constituents as well as cementing agents and particle size. Generally, the larger the particle size, the lighter the color of the soil. Color can vary largely depending on the cementing agent or coating in the soil, for example, iron oxide will produce a dark red-black color while carbonates and salts will whiten a soil (91). In 1996, a study of 300 soil samples taken in close proximity to each other showed that over one-half of the samples could be distinguished from the others based on color alone (61). Geologists and engineers determine soil and mineral

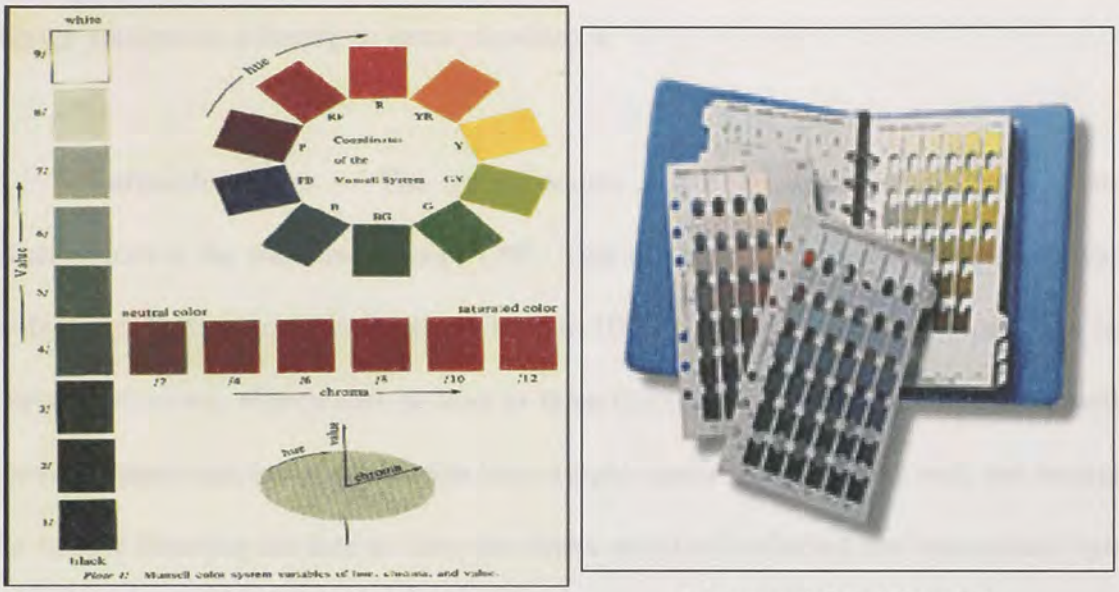
Figure 1. General Schematic of Soil Examination Sequence. The starting point for any soil examination may vary depending on the fraction of the soil thought to be most informative. XRD = X-ray Diffraction, IR = Infrared. Modified from Skip Palenik, Microtrace. Inc.



Air-dried sample
Observation with stereomicroscope
Removal of foreign materials such as fibers and glass
Identification of foreign materials
Color observation
Particle-size distribution
Sonication in water
SINKS < 10 min
SINKS > 10 min
SUSPENDS
Check for diatoms or plant opal
PRESENT
ABSENT
1/2
Heat + HF
30 % H_2O_2 + heat
Acetolysis
Density separation
Mount in glycerine
Heavy mineral separation
Polarized light microscopy
Refractive Index determination
Identify and quantitate
Air dry or centrifuge
XRD
Staining, Phase contrast microscopy
Thermal analysis, chromatography/ IR spectroscopy

color by comparison to a color standard called the Munsell Color System. The system characterizes a color on three scales, chroma, value and hue. These three values constitute a color's Munsell notation (Figure 2). The system is recognized as the standard method of color specification in multiple fields of color technology and science.

Figure 2. The Munsell Color System and Munsell Washable Soil Color Charts. The diagram on the left show the three axis for color determination; hue, value and chroma.



www.gretagmacbeth.com/index/products

Sugita and Marumo performed a validation study of soil discrimination using the Munsell color scale. Of 73 samples taken from three different classes of soil, 70 % of the soils were distinguishable by color after drying. Additional preparations of the lighter, clay fraction of the soil like decomposition of organic matter, moistening and removal of iron oxide, followed by color characterization allowed them to differentiate 97 % of the samples (100). They also found that color determination after ashing, the point where a substance loses its capacity to ignite, did not enhance discrimination power which contradicted the results published by Dudley (31).

Light Microscopy. The microscope has been the most effective instrument in forensic soil examination (80). It is the instrument for finding unusual minerals and foreign substances adhering to items of evidence.

Stereomicroscope. The most widely used microscope in forensic soil examinations is the stereomicroscope (79). This is a binocular microscope that can vary in its magnification power from about 10 X to 100 X. Because the stereomicroscope has two sets of lenses, objects can be seen in three dimensions. The light source is usually above the specimen but some models have a light source in the base as well just beneath the sample allowing the user to view the object with both reflected and transmitted light. Stereomicroscopes are convenient because the distance from the objective to the sample is enough to allow manipulation of the sample during observation. Special grids are used to assist in the counting and measuring of particles. The number or percentage of each type of grain is recorded and used as a significant measure of comparison. Any foreign objects such as fibers, paint chips, hair or plastic can be removed and examined further by specialists and provide great value to the analysis (79).

One case which benefited greatly from stereomicroscopic analysis occurred in Sydney, Australia in 1960 (93). Eight-year-old Graeme Thorne's family had just won the Sydney Opera House lottery. They were photographed and featured in the news. On July 7, Stephen Bradley kidnapped Graeme on his way to school, placed him in his trunk and called Graeme's family to demand a ransom. Five weeks later, Graeme's body was discovered on a vacant lot covered by some overgrowth. The body was wrapped in a rug that provided many clues. Microscopic examination showed that the rug had soil with trace amounts of pink, lime stock mortar, dog hair from a Pekinese and foliage from two trees, smooth cypress which was common ornamental and Squarrossa false cypress which was quite rare. Police began to scan the surrounding neighborhoods looking for houses with pink mortar while carrying branches from the two cypress trees. After weeks of searching, they found a house which had all three characteristics. The previous occupant had moved his family out of the country on the day of the murder. That person was Stephen Bradley. Police found Bradley and his family and brought him back to Australia where he was tried in March of 1961. Bradley was convicted of the murder and sentenced to life in prison (42).

Under the stereomicroscope, many characteristics of the grains in a sample can be observed including polish, texture, weathering, color, rounding and surface coating. Identification of the minerals in a soil sample is an extremely significant factor in determining the source of an unknown sample and in comparative analysis. Minerals can also be identified by their optical properties.

Polarized Light Microscopy. Optical properties of minerals and glass are best measured using a polarized light microscope (PLM). One measure that can be determined using a PLM is refractive index or RI value (91). RI is a measure of the ratio of the velocity of light as it through a vacuum compared to its velocity through another medium. A compound microscope is used with a rotating stage which houses a plane polarizing filter. Light which passes the filter will only be vibrating in one orientation. Under these conditions, a mineral or rock which transmits light can be cut, polished and mounted onto the stage. The object is then immersed in an oil of known RI. If the oil and the object have different refractive properties, the observer will see the light bend by observing what is called the Becke line as the distance is increased between the objective and the stage. If the object has a higher RI than the medium, the Becke line will bend toward the object and vice versa. The procedure is repeated using oils of different RI value until the Becke line disappears at which point the object and the oil have the same RI (26).

RI values can also be obtained using one oil and a Mettler hot stage in the single variation method. The oil must have known RI values at different temperatures. The sample is mounted, and the temperature is increased until the Becke line disappears. The temperature is recorded and the RI value is determined (92). RI values of liquids change depending on the color of light used in the microscope source. This phenomenon is called dispersion. Measures of dispersion, the difference in RI values at various wavelengths of light, and RI are the most-used properties in the identification of glass (79). Varying the wavelength of the light source and the temperature of the oil medium is called the double variation method and it produces RI accuracy levels of ± 0.001 (36).

Plane-polarized light can also be used to examine the crystal structure of individual rocks and minerals. An additional plane-polarizing filter is inserted above the objective lens within the microscope tube. This polarizer can be rotated to 90° from the orientation of the other. At this point the filters are in a North-South and East-West orientation and the object is being viewed under crossed-polars. An isotropic crystal, which has a uniform crystal lattice structure and thus only one RI value, will refract light in the same direction regardless of the orientation of the crystal. This type of crystal will always appear dark or "extinct" under crossed-polars. Any mineral that does not go extinct in some orientation under crossed-polars is anisotropic. Knowing whether the crystal is isotropic and then determining its RI value are strong indicators of the mineral's identity (79).

Electron Microscopy. Minerals, rocks and fossils can be observed in much greater detail using an electron microscope. An electron microscope uses a high energy beam of electrons as its source and a series of magnetic lenses to focus that beam into extremely small areas. The object must be covered in an electrically conductive powder such as carbon or gold. The small wavelength of the electrons allows for much higher resolution than traditional light microscopy and clear magnifications from 25 X to 650,000 X.

Scanning electron microscope (SEM). One of the strongest advantages of the scanning electron microscope is its automation capabilities. One can select for a specific classification of particle and allow the SEM to find the minerals of interest (70). The technique allows an observer to search characterize individual particles based on their size, morphology and surface characteristics such as scratches, pits and mineral growth. The coupling of an Energy Dispersive X-ray Spectrometer (EDS) to the SEM allows the operator to single out individual particles and then analyze the elemental composition of the surface of the particle using electron bombardment (discussed in chemical analysis section). The physical characteristics of the particle in addition to the elemental composition strengthen its classification and the ability of the examiner to discriminate soil samples. SEM-EDS for the analysis of forensic soils was evaluated by McVicar and

Graves (70). They found that the technique was reliable, fast, and accurate in discriminating soils from different sources and in recognizing replicates of the same soil.

A civil suit in southern Alabama in 1980 had a motorcycle rider who lost his leg in an accident pitted against a local dealer and a national motorcycle company (55). He claimed that he had purchased a helmet visor the night before the accident and that as he came down into a low lying, foggy area, the visor clouded and he was unable to move it out of his line of vision. SEM analysis of the inside of the visor revealed scratches containing tiny grains of feldspar, a mineral not found in southern Alabama. The closest place where the mineral was shown to exist was 150 miles away. Other examinations demonstrated that the visor was not new as he claimed and that it did not have the optical characteristics of the visors sold at the local dealer. After disclosing the evidence to the rider and his lawyer, they dropped their suit (55).

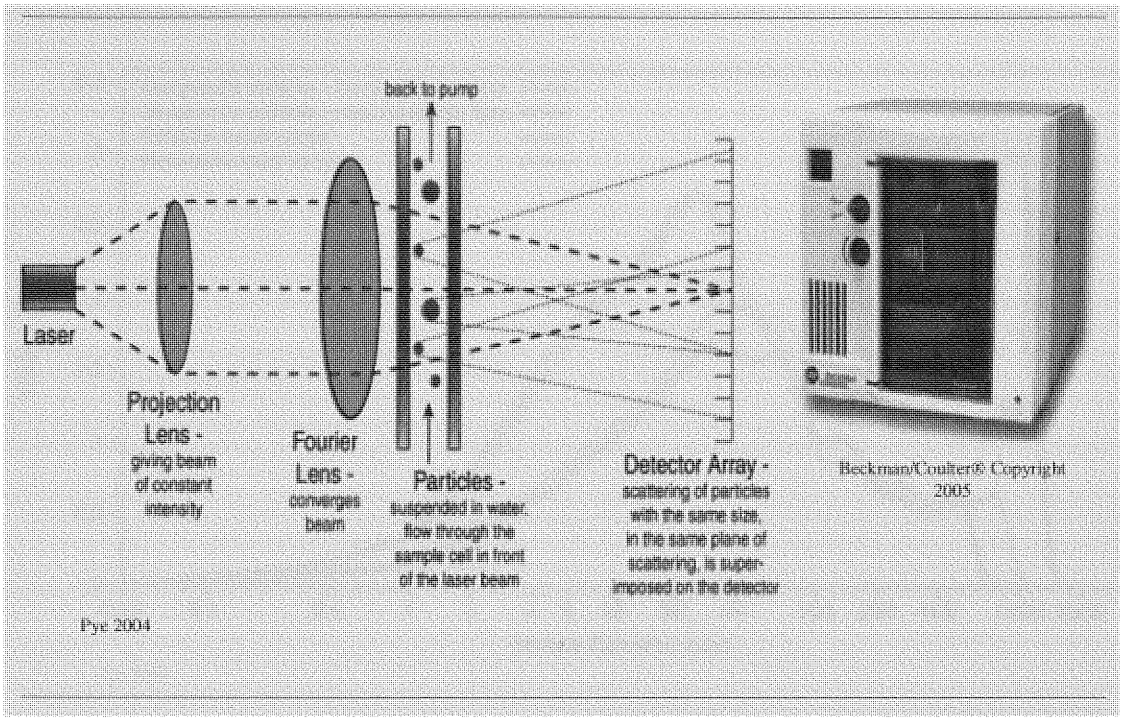
Particle-size Distribution. One unique aspect of soil is the frequency of individual ranges of particle size resulting from various erosion forces. There are multiple methods of determining the particle-size distribution of a soil including wet and dry sieving using a series of nested mesh sieves and a shaker, laser diffraction, Coulter counting and microscopic image analysis. Samples are first either sonicated in water to break up aggregates or they are treated with dilute hydrochloric acid, followed by hydrogen peroxide to remove carbonates and organic cementing agents (79). Nested sieves decrease in the size of their openings from top to bottom. The soil is shaken through and the mass of soil falling within each size range is recorded as a percent of the total sample.

Recently, Chazottes *et al* described the most variable particle size range as the 1-0.063 mm range in random sampling of two different soils (19). Sugita and Marumo analyzed 73 soil samples from a 300 square kilometer area by wet and dry sieving into three fractions ranging from 0-2 mm (99). Using the frequency of particles in each fraction they were able to distinguish 87.9 % of the samples. Using a particle size analyzer, which determines the amount of each particle size by measuring transmittance of light while centrifuging the suspended soil sample, they analyzed the fine particle fraction. Addition of this data increased their discrimination to 95.9 % (99).

Dudley was the first to use a Coulter counter in particle size distribution analysis of sand and silt fractions (99). A Coulter counter uses soil suspended in a weak electrolyte, as voltage is applied across a small sensing zone, resistance is detected proportional to the volume of each non-conductive particle. An output is generated in a short amount of time detailing particle sizes and counts (30).

Laser diffraction in combination with wet sieving was used by Wanogho to analyze the fine particle fraction < 0.063 mm and was able to distinguish their samples (110). This technique relies on the fact that particles scatter light with different intensities depending on their size (Figure 3). In a validation study by Pye *et al*, glass-bead control standards on this instrument yielded accuracy levels of 0.03 % comparing the mean particle size to the true value (85). The authors also showed that the technique was only reproducible if the original soil sample was homogeneous. A representative sample could be obtained from a single sub-sample only when analyzing soils that were better sorted due to the constant effects of environment like wind and rain. For the many soils that are not well sorted, multiple sub-samples were necessary and an average and

Figure 3. Coulter™ LS230 Laser Granulometer and the Multisizer™3 Coulter Counter®. As particles flow through the sample cell they scatter laser light in a manner characteristic of their size.


back to pump
Laser
Projection Lens - giving beam of constant intensity
Fourier Lens - converges beam
Particles - suspended in water, flow through the sample cell in front of the laser beam
Detector Array - scattering of particles with the same size, in the same plane of scattering, is super-imposed on the detector

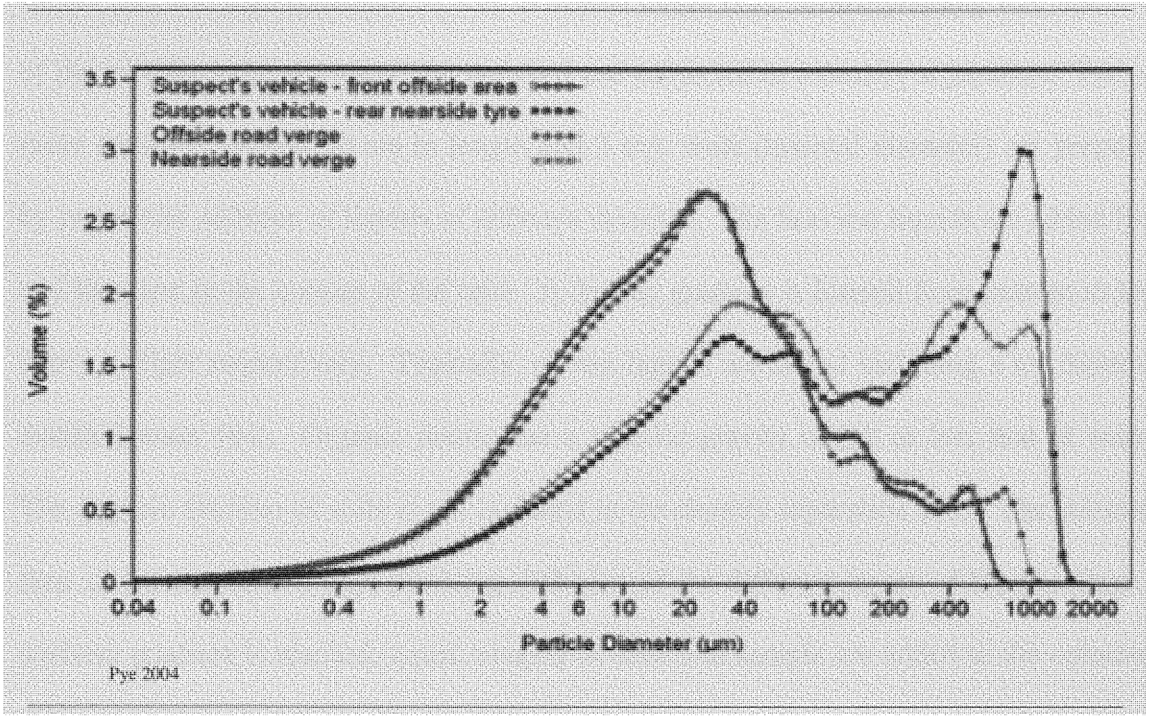
Pye 2004

Beckman/Coulter® Copyright 2005

standard deviation were recorded. Laser diffraction analysis of particle size distributions in random soil samples was shown to be a rapid and accurate method of characterizing soil.

In one described case, Pye *et al* applied laser diffraction to a hit and run case where the suspect's car had veered off the road onto the shoulder and the median before killing one pedestrian and injuring another (85). Reference samples from the side of the road and samples taken from the mud deposited on the suspect's car were analyzed by laser diffraction and determined to be a close match (Figure 4). results in separate analyses even if the method was standardized (79).

Figure 4. Laser Granulometer Output from Hit and Run Case. Data output generated from two evidence and two reference soils demonstrating a likelihood that the suspect's vehicle veered off the road at the location of sampling.



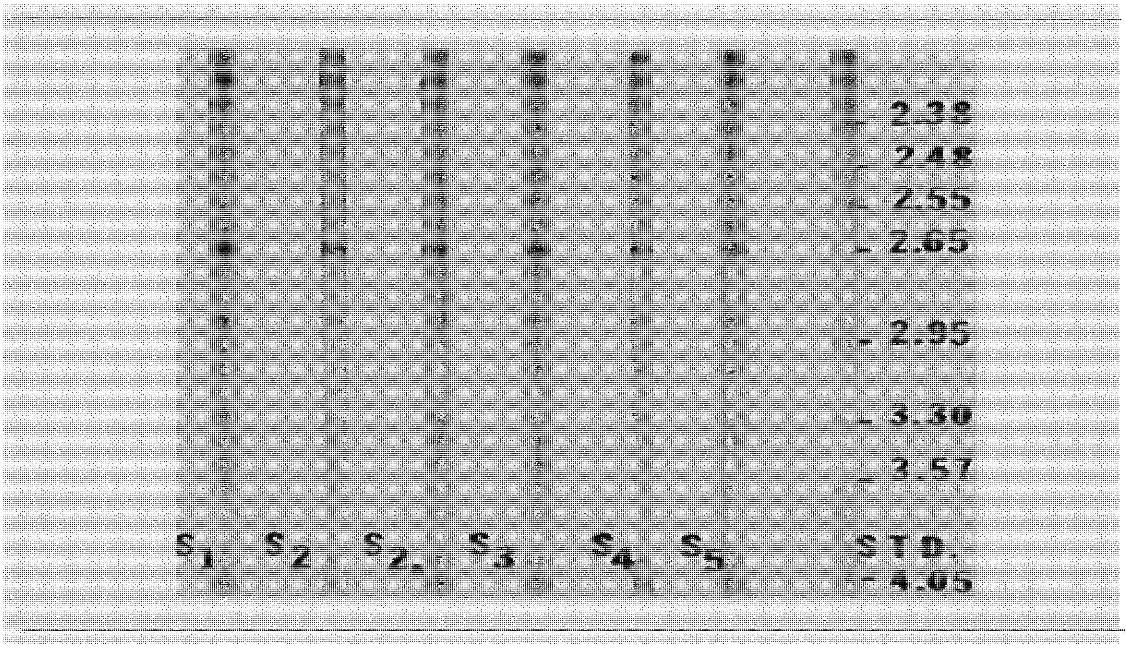
Pye 2004

Particle size distribution is a discriminatory characteristic in the analysis of forensic soil samples. When it is used in conjunction with microscopic determinations of comparison, it can strengthen the interpretation of the evidence.

Density Distribution. In many cases, the most discriminating materials in soil samples are those minerals of high density. Typically, they are separated out using a heavy liquid like bromoform, which has a specific gravity of between 2.88 and 2.90, meaning it is 2.88 times denser than water at 4° C (91). Soils contain particles of different size and density. A technique which can perform a separation based on size and subsequently on density will reveal whether two soils are similar. In performing a forensic examination of soil by density distribution, soils are first pulverized with a rubber tool and then separated by size in nested sieves (79). Corresponding size fractions of the samples to be compared are weighed and subjected to a density separation. Two or more columns, around 30 cm in length and 5 mm wide (internal diameter) are prepared by sequential addition of liquids of different densities. The heaviest liquids are added first and they rest in the bottom of the tube. A density gradient is formed from bottom to top, heavy to light respectively. Typically, eleven liquids are added including a top layer of distilled water (84). Organic matter has a specific gravity of 0.9 or less that of water. The organic fraction of soils is not included in the density separation (80). The columns then sit for one to two days allowing diffusion of the different layers of liquid. Once a uniform gradient is reached, equal weights of each soil sample are added to the columns. Because the columns are being used for a comparative analysis, they must be exactly the same temperature and they must be prepared exactly the same way. In a few hours, each particle will have settled to a depth in the column where its density equals that of the liquid. Comparisons are based on the concentration of particles at the different densities or levels in the column. The columns can easily detect differences in density of 0.01 g/cm^3 (79). The density separations are backlit with white light and photographed side-by-side which can be an effective visual for courtroom presentation (Figure 5).

The forensic value of density distributions is widely debated. According to Murray and Tedrow, 80 % of soils are composed of quartz which has a specific gravity of 2.65 (80). The traditional organic density gradient columns have a density range from 2.89 to 1.5. This is too low to effectively separate rarer, denser minerals of higher

Figure 5. Photographic Representation of a Density Distribution. One reference and five questioned soil samples following a density-gradient separation (Petraco, 2000). Sample **S1** came from the crime scene. Samples **S2** and **S2A** were from soil stains on the victims clothing. Samples **S3-5** are soils from the suspects clothing. **STD** is the standard containing soil particles of pre-determined density or specific gravity for comparison purposes.



forensic interest. A highly dense aqueous salt preparation was proposed by Petraco and Kubic in 2000 which has a density range up to s.g. 4.05. However, a case was made by Murray that the technique itself has inherent problems which may produce different results in separate analyses even if the method was standardized (79). These problems include, but are not limited to; (*i*) particles of different densities adhering to each other settle at a level between their respective densities, (*ii*) the current density range being too low for effective separation of particles of forensic interest, (*iii*) small variations in coating and/or fluid or solid inclusions possibly causing the same minerals to exhibit different densities, (*iv*) porous particles trapping air causing the particle to be more buoyant and (*v*) the way a sample is collected and prepared leading to observable differences in the column. Murray states that it is questionable whether or not the technique is even worth doing because it does not contribute to the identifying of the diverse rock, fossil and mineral combinations found in soils which are the parameters of most value to a forensic soil examination.

Chemical Characterization of Forensic Soils

Minerals, rocks, fossils, sand, silt and clay can also be characterized based on their chemical composition. Several techniques are currently being used for forensic soil examinations including Scanning Electron Microscope-Energy Dispersive X-Ray Spectroscopy (SEM-EDS), Fourier Transform Infrared spectroscopy (FTIR), Gas Chromatography, High Performance Liquid Chromatography (HPLC), Cathodoluminescence, X-Ray Diffraction (XRD) laser Raman spectroscopy and Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES). Others are still being validated such as capillary electrophoresis (CE) which has been used as a separation technique for anions chloride, nitrate, sulfate and phosphate which are commonly found in soil (18).

An **Energy Dispersive X-ray Spectrometer (EDS)** coupled to a scanning electron microscope will analyze x-rays produced by electron bombardment of a particles surface (70). The x-rays produced by these secondary electrons will have wavelengths characteristic of the elements from which they came. The amount of x-ray radiation at a particular wavelength is indicative of the relative amount of the element present in the surface of the object. A preliminary study of automated SEM-EDS analysis of forensic soils indicated that equal or better discrimination could be obtained faster, and less expensive than traditional optical microscopic characterizations (70). However, sample preparation is not yet standardized and can play a significant role in the repeatability of the analysis (17).

Fourier Transform Infrared (FTIR) spectroscopy is a technique used to analyze a soil's organic components as well as pesticides, polymers and a limited range of inorganics (91). A soil sample is mounted onto a KBr pellet. Light is passed through the sample at a range of wavelengths. A detector measures absorbance at each wavelength and a spectrum is generated. Absorbance at the different wavelengths will characterize the compound. A study by Cox *et al* showed that when Munsell color notation failed to discriminate between four reference soils of different origins, their percent organic fractions determined by FTIR analysis was able to distinguish the soil (23).

Raman Spectroscopy is another method of characterizing a soils organic fraction (39). This technique however, unlike FTIR, does not require a thin section of the sample. The spectra obtainable by Raman spectroscopy represent molecular and vibrational information complementary to IR spectroscopy. Raman has better resolution than traditional IR but does not yet have the spectral reference libraries (91). Bestwick and Espinoza demonstrated that Raman spectroscopy could identify all symmetrical inorganic and organic compounds in aqueous extractions of the soil (5). They also demonstrated that the combination of FTIR and Raman spectroscopy could differentiate samples which could not be discriminated using color and density gradient characterizations (5).

Gas Chromatography (GC) is a well established method of forensic examination of volatile organic residues from scenes of arson and suspected environmental crime scenes (2). GC is a separation technique based on migration of compounds at different rates through a stationary phase (49). This stationary phase coats the column through which the compounds travel. An inert gas at pressure is applied to a column and acts as the mobile phase pushing the compounds through the column to the detector. The amount of time a compound stays in the column (retention time) is dependent upon its boiling point and therefore the temperature of the column as well as the chemistry of coating in the column itself. Current gas chromatography allows for the analyst to select a temperature program and column suited to the specific mixture he/she is trying to separate. Temperature is the most determining factor in a gas chromatographic method. Coupled to a mass spectrometer, gas chromatography (GC-MS) is the gold standard for identifying organic compounds from complex mixtures like those found in soils.

High Performance Liquid Chromatography (HPLC) coupled to a mass analyzer is an effective method of separating and identifying non-volatile organic residues found in natural environments (4). A study by Siegel and Precord showed that coupling a UV detector with two different wavelengths to a reverse phase-HPLC column produced quantitative data which could discriminate their soil samples (96). Both GC and HPLC have well established standards for the analysis of environmental contamination and arson suspected samples, however, as is the case with many of the newer, more sophisticated soil analysis techniques, no studies have conclusively demonstrated their value as a single discriminatory technique for the analysis of soil samples.

Cathodoluminescence (CL) is a technique which uses a defocused electron beam to cause a mineral to illuminate visible light. A thin section of a sample is prepared by mounting it with epoxy. The sample is bombarded with electrons until it glows. The wavelengths emitted by the mineral characterize its trace elemental composition (79). In 1995, a Smithsonian Institution research report detailed a study where the FBI gave ten pairs of soil samples to researchers at the National Museum of Natural History Mineral Sciences department (112). The goal was a blind determination of which samples were replicates. Scientists in the study used cathodoluminescence to effectively match the soil samples stating that "the advantage of CL is that a given mineral typically emits only certain colors. If you have some knowledge of which minerals produce certain colors, finding a match is not difficult (112)." In the Smithsonian study described above, X-ray Diffraction (XRD) was also used to characterize the same soil samples (112).

X-ray diffraction is performed on the tiny particles in the clay fraction of soil (90). This is advantageous to crime scene scenarios as this size fraction is often the most recovered in transfers (19). After sieve separation, the clay is dried and the powder is smeared in acetone on a quartz plate. An electron beam is directed at the sample and the arrangement of the atoms in the crystalline structure determines how the x-ray will be diffracted. Every crystalline material has a distinctive diffraction pattern. The pattern can be shown on film or using electron detectors to produce an image. The Smithsonian Institute's research report stated that because of the predominance of quartz and feldspar in the clay, examination by XRD gave almost identical results for all twenty soil samples (112). Discriminating data could only be obtained by comparing amounts of the rarer minerals in the samples. Murray states that XRD of the clay fraction of soils can render hard to interpret results when the sample is composed of more than one mineral. However, XRD is currently the principal tool in the modern identification of clay minerals (79).

XRD aided forensic scientists after two women disappeared in Adelaide Hills, Australia (38). A day after they disappeared, the suspect was arrested when police found a bloody shovel with soil-like material on the blade. XRD patterns of the soil on the shovel and a waterlogged soil-regolith found in samples from a local quarry were compared and determined to be identical. Three weeks after the disappearance of the two women, foxes found the bodies in the same quarry.

Inductively Coupled Plasma-Optical Emission Spectrophotometry (ICP-OES) is a powerful technique for determining the elemental composition of soils or glass of forensic interest (86, 103). It is a method capable of analyzing a sample for multiple elements in a short amount of time and it has a high dynamic range measuring element concentrations from 1 to 1,000 parts per billion . Soils are digested in acid and introduced as a slurry into an extremely hot argon plasma where the compounds are atomized and ionized (103). Radiation is emitted, separated into a range of wavelengths characteristic of the elements being analyzed and converted into an electrical signal by a photomultiplier tube. The intensity of emitted light is directly related to the quantity of that element in the digest.

A 2005 study by Pye *et al* on the precision of ICP-OES, specifically how reproducible the method was in replicate analyses of the same soil sample (86) determined that second digestions of the same soil produced a relatively precise output (CV ~ 5 %) but that when standard reference materials were analyzed over long periods of time and on different instruments, precision dropped (CV ~ 11 %).

Obviously there are multiple physical and chemical methods of soil characterization currently accepted for forensic examinations. According to the Forensic Science Handbook, "(the) discriminating power in soil examination lies in the number and kinds of minerals available" (91). To date, most if not all techniques employed in forensic laboratories for soil characterization have been directed toward this goal of counting and identifying minerals. However, most of these analyses had their origins in other scientific disciplines and in the highly scrutinized arena of forensic science, they have shown important limitations. For these reasons, although the complexity of soil provides many opportunities to obtain useful forensic evidence, there are few simple standard procedures which can be applied to all cases (91).

Limitations of Current Forensic Soil Examination

In order for any forensic examination to have value, it must satisfy certain conditions. The first and foremost are the **rules of evidence**. After all, why perform an analysis at all if it will not be admitted into court? The rules of evidence depend on whether the case falls under federal or state jurisdiction and on the state in which the crime occurred. The main purpose is to protect the jury from being misled (91). Two landmark cases have set the precedent in state courts concerning the admissibility of scientific evidence or testimony. The first case ***Frye v. the United States,*** occurred in 1923 when James Frye, who was tried and convicted for second degree murder, appealed in hopes that the results of a crude precursor to the modern polygraph would be admitted as evidence of his innocence (*Frye vs. United States*, 293 F. 1013, D.C. Cir. 1923). The results of the polygraph were not admitted and the courts explanation as to why became the Frye rule which states that "while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." The second case, ***Daubert v. Merrell Dow Pharmaceuticals,*** went all the way to the United States Supreme Court in 1993 (*Daubert v. Merrell Dow Pharmaceuticals*, Inc., 509 US 579, 113 S. Ct. 2786, 1993). In this case, two individuals were suing the pharmaceutical company on the premise that their drug, Benedictin, an anti-nausea drug, caused birth defects when taken by pregnant women. The court decided that the federal rules of evidence superceded the Frye rule. ***Federal Rule 702***, adopted in 1975, refers to the admissibility of expert testimony in federal court (43). After being amended in 2000 it states, "(i)f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training or education, may testify thereto in the form of an opinion or otherwise, if (*i*) the testimony is based upon sufficient facts or data, (*ii*) the testimony is the product of reliable principles and methods, and (*iii*) the witness has applied the principles and methods reliably to the facts of the case" (Federal Rules of Evidence, Article VII, Opinions and Expert Testimony, Rule 702). The Daubert decision also declared the judge as the "gatekeeper", as he/she is ultimately responsible for deciding whether evidence satisfies all the criteria (97). Daubert also added the provision that the technique or method used to obtain the expert opinion must have been (*i*) tested, (*ii*) peer-reviewed and published, (*iii*) generally accepted in the scientific community, (*iv*) deemed to have an acceptable rate of error and (*v*) subjected to standards controlling technique operation (97).

The word "reliably" in the third provision of Federal Rule 702 means that not only must the examiner be proficient in his/her technique, but also that the interpretations and conclusions they make, they must also prove them to be accurate (Federal Rules of Evidence, Article VII, Opinions and Expert Testimony, Rule 702). Most if not all current techniques for the forensic examination of soil cannot satisfy this reliability requirement in and of themselves. This is not only due to the fact that soil characterizations are based on class characteristics but it is also due to the inability of the techniques to demonstrate a rate of error when conclusions are drawn from their results. Forensic geologists have attempted to overcome this just by adding more methods of analysis to a forensic examination until there are enough results in total to carry some significance when conclusions are drawn. Obviously, in the forensic setting, this is not the ideal as it makes for long analysis times and expensive costs of employing the only people qualified to make significant conclusions using these methods. There are only two ways to overcome this obstacle. One is to find a way to assign individualizing characteristics to soil, which would probably require more analyses than are available even outside the area accepted in the forensic community. Currently the only fields of forensic science that offer individualized or inclusive evidence are human DNA, toolmarks, latent prints and ballistics. The other is to extend current databases to a point where analysts could be tested using blind samples and their conclusions could be used to determine the accuracy of the technique. The power of databases in forensic science is evidenced by the human identification through DNA phenomenon.

The Human "DNA Fingerprint"

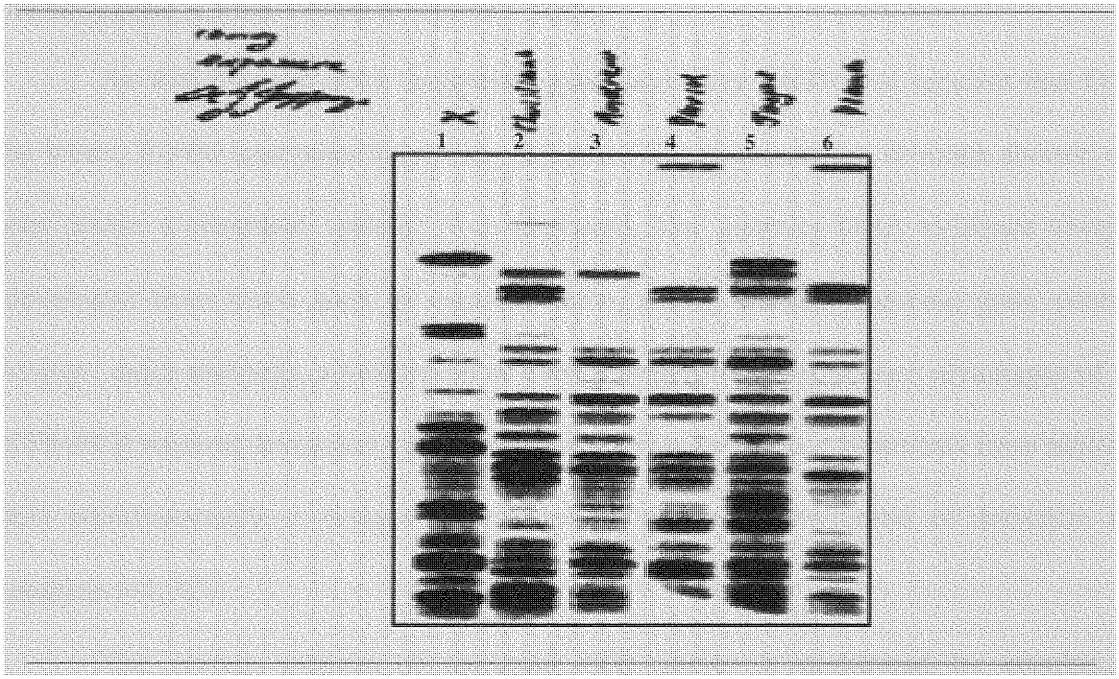
Before DNA analysis became prevalent in suspect and criminal identification, blood group markers were employed. ABO blood grouping, which assesses the antigens on red blood cells was developed by Karl Landsteiner in 1900 (62), and Human Leukocyte Associated Antigen (HLA) typing, pioneered by George Snell, Jean Dausset and Baruj Benacerraf, which identifies the surface proteins on white blood cells, could be combined to exclude suspects from about 97% of cases (98). Still, as was the case with almost all forms of forensic evidence, this was only supportive and could not be used as a direct link from suspect to crime scene.

Minisatellite DNA. In the mid nineteen-eighties, two major breakthrough discoveries lead to the current deluge of technology and hype surrounding DNA analysis. In 1985, Sir Alec Jeffreys discovered regions in the human genome which varied significantly in length between individuals (57). These areas, which he called minisatellites, contained tandemly repeated sequences of DNA or "stutter DNA". The core, repeating units were anywhere from 16 to 64 bases in length for the myoglobin gene he was studying (57). The variation between two individuals was due to one person often having different numbers of these repeated elements than another individual, anywhere from three to twenty-nine, thus the minisatellite was also dubbed a Variable Number of Tandem Repeats (VNTR) locus. In order to detect the genotype of an individual at a particular VNTR locus, Jeffreys used a technique called restriction fragment length polymorphism (RFLP). The technique requires the use of restriction enzymes to cut the regions of DNA which flanked the VNTR (57). Fragments are then separated based on their length in an agarose gel and transferred to a nylon membrane. The membrane was subsequently washed with a radioactive probe, a labeled segment of DNA which would hybridize only to the repeated sequence of the VNTR. Exposure on film revealed the location of the VNTR in the gel and the size was determined (Figure 6). Sizes of the restriction fragments differed between individuals and thus Jeffreys coined the phrase "DNA fingerprint" (58). The first paternity case in which DNA fingerprinting was admitted into evidence occurred in England (Sarbah v. The Home Office, 1985).

Christiana Sarbah and her son Andrew were reuniting after he had been in Ghana visiting his father. Immigration officials at Heathrow Airport suspected Andrew's passport was forged. It wasn't until an intervention by a Member of Parliament that Andrew was permitted to return to his home. Although lawyers amassed tons of evidence from photographs, statements and results of other genetic tests, the case remained open. Finally, the lawyers contacted Jeffreys after being informed that his test could determine maternity. The comparison of DNA fingerprints of the mother Christiana, Andrew, and three of his siblings demonstrated that the alleged mother and child in question share many DNA fragments of comparable length (Figure 6). Children also shared fragments with the boy in question which were not common to the mother, an event highly unlikely among unrelated individuals. The results proved maternity so conclusively that the immigration office announced it would accept this as definitive evidence for all future cases and the police announced their hope in the technique for identifying criminals (71).

Polymerase Chain Reaction (PCR). The second breakthrough was the discovery of a way to synthesize large quantities of DNA *in vitro*. In 1986, Kary Mullis described the Polymerase Chain Reaction (77). The technique is capable of taking DNA from just a few cells and producing millions of copies of a specific sequence. The technique was further enhanced by the discovery and commercialization of a thermo-stable DNA polymerase from *Thermophilus aquaticus* which allowed the copying process to be automated and to take place in a closed system safe from contamination (78). This has enhanced the utility of crime scene samples to a great degree allowing for

Figure 6. Autoradiograph of the First Casework Human DNA Fingerprint. DNA fingerprints of (Lanes 2 to 6) Christiana, Andrew, David, Joyce, Diana and an unrelated individual "x" (Lane 1) produced by RFLP and Southern Blot hybridization (Jeffreys *et al* 1985). The likelihood of son Andrew Lane 3) having so many DNA fragments of comparable length to the alleged mother, Christiana (Lane 2) and siblings (Lanes 4-6) and yet being unrelated is extremely low.



investigators to produce an individual's DNA profile from tiny deposits of biological material such as a blood stain the size of a pin head, whereas before they were trained to collect blood from stains the size of a quarter (60). In the years following Jeffreys discovery of the VNTR loci and Mullis' PCR, forensic DNA profiling moved to a PCR-based system.

Microsatellites. Peter Gill of Britain's Forensic Science Service devised a DNA profiling method using new loci called microsatellite DNA (45) which were determined to be a valuable source for DNA typing in 1991 by Edwards *et al* (35). The microsatellites were found to contain shorter repeat DNA segments but were also composed of highly polymorphic DNA. Microsatellite regions contain short tandem repeats (STR's) of DNA sequences with a core repeated sequence that range anywhere from two to seven base pairs in length (12). The total length of STR loci used in human identification is from 100 up to 450 base pairs.

An RFLP-based DNA fingerprint generated from VNTR loci requires that DNA be intact, 20,000 base pairs long (12). Because STR fragments are so much shorter, one can obtain a complete profile from degraded, low quantity DNA evidence samples developed from items such as cigarette butts, eating utensils, chewing gum, postage stamps, razor shavings, a toothbrush or even a fingerprint (60). Other advantages of the STR-PCR method developed by Gill were the ability to produce fragments from multiple loci simultaneously, a process called multiplexing and automation of detection and analysis (44, 45).

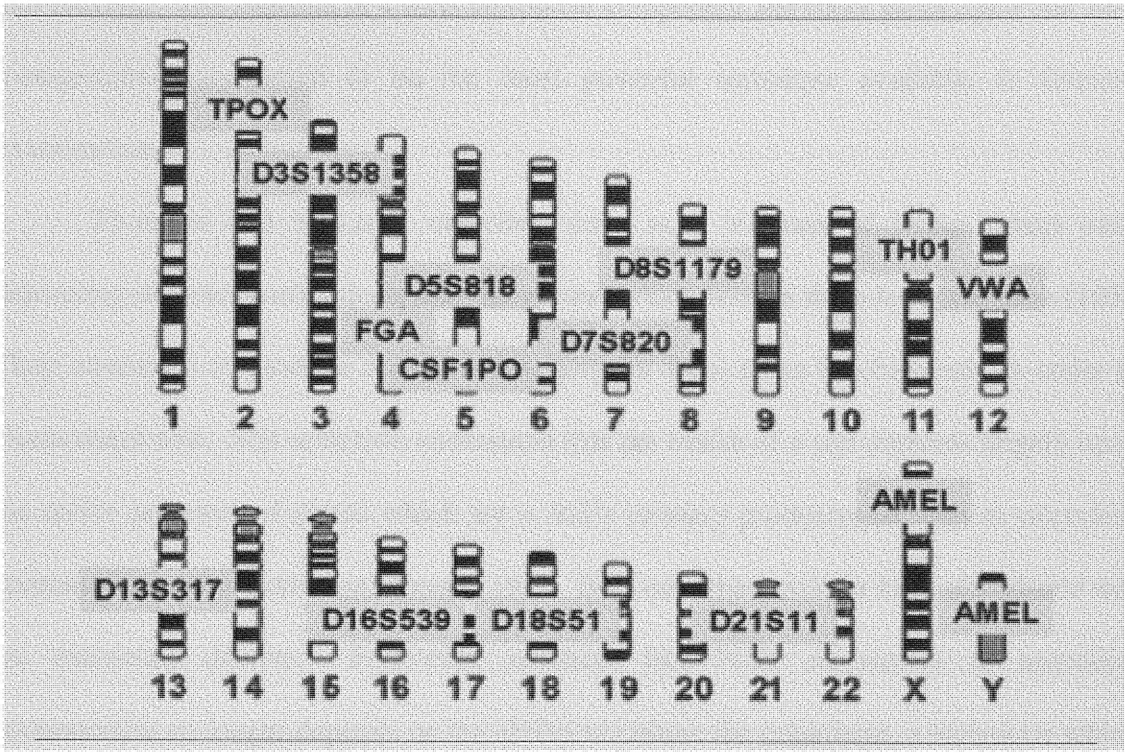
The current method of detecting STR alleles is performed by attaching a fluorescent dye or fluorophore to the 5' end of one of each PCR primer pair (82). Once PCR has been run, the fragments, each with the fluorophore attached, are separated based on their length in a capillary electrophoresis (CE) apparatus and detected by a UV laser which excited the fluorophore and causes fluorescence. This is a largely automated system of detection and is much faster than the earlier slab gel method. CE is able to achieve one base pair resolution with standard deviations less than 0.117 base pairs (14). STR fragments can be effectively separated in about 30 minutes and the DNA profile, which is comprised of numbers of tandem repeats at each locus, can be exported in a tabular format. These tables are compiled to create a database of profiles.

DNA Databases and Forensics

In 1994 the Federal Bureau of Investigation was given funds by U.S. Congress to create the Combined DNA Index System (CODIS) (37). The FBI adopted thirteen core STR loci for each individual profile (13). The thirteen loci and their location in the human genome are shown in Figure 7. The combination of alleles in an individual's DNA profile produced from thirteen STR loci results in random match probabilities of one in a trillion (11, 51). When numbers like this are pronounced in a courtroom, there is no longer room for reasonable doubt. This is what makes human DNA identification individualized, inclusive evidence. The combination of a well established, reliable technique like STR analysis and the CODIS database are what gives the forensic investigator the ability to declare a random match probability, likelihood ratio or probability of inclusion. A significant portion of recent forensic scientific research has been to establish allele frequencies for the 13 core STR loci from various populations. This is a massive effort to further validate the statistical methods a forensic DNA examiner employs to provide significance to a conclusion when they find samples that compare. Currently there are commercially available kits for typing the core 13 STR loci, the sex marker Amelogenin as well as additional STR loci in a single multiplex PCR reaction (Penta D and Penta E in Promega's PowerPlex® 16 kit, www.promega.com; and D2 and D19 in Applied Biosystem's Identifiler®, www.appliedbiosystems.com).

In a forensic context, the DNA database only exists in human identification applications. There are no databases of this type currently available to forensic soil examination. Most of the techniques were adopted from other scientific disciplines with

Figure 7. The 13 Core STR Loci Used in the CODIS Database. Names and chromosomal locations of the core STR loci are shown as well as the sex marker Amelogenin (http://www.cstl.nist.gov/biotech/strbase/).



different purposes for evaluating soil characteristics. The only database available to forensic geologists is their own knowledge and experience which may be localized to a particular region and relies upon a range of skills that require years to develop. Some new techniques in soil science, especially those that are DNA based, may lend themselves to the creation of powerful databases of soil profiles.

Biological Characterization of Soil

Variations in soil are not limited to inert physical characteristics like particle size distribution, color, mineral content, and density distribution. Soil is a much more complex system composed also of living and decaying invertebrates, plants, fungi, algae, and bacteria. The analysis of a soils biological component and more particularly the soils metagenome may prove to be the most powerful technique available to forensic examination.

Thornton and McLaren described soils as a biomass of living things each depositing diverse biochemicals into the ground which remain there for some time (105). They attempted to produce a chemical fingerprint from these biochemicals in order to characterize soil. Soil scientists understand that the health of a soil depends largely upon the active presence of plants which are a soil's primary producers as well as microbial communities, which are decomposers (95). Soil microbes exist in high abundance and are extremely diverse. Following advances like PCR, assessment of microbial communities no longer depends on the ability to culture the organisms, a technique which has been estimated to characterize less than one percent of the microbes in soil (89). Molecular based methods of characterizing microbial communities in soil have revealed that the communities are dominated by organisms previously unknown (46, 64, 101, 111). The microbial community profiles produced by these methods can also be used for discriminating between soils of interest.

Some of the molecular, DNA-based methods of assessing bacterial communities are denaturing gradient gel electrophoresis (DGGE), terminal-restriction fragment length polymorphism (T-RFLP) and length heterogeneity PCR (LH-PCR) and they will be discussed in the following pages. Most of these techniques target the 16S rRNA genes, which transcribe a ribosomal RNA component critical to prokaryotic protein translation (21, 22, 68, 74, 83, 88).

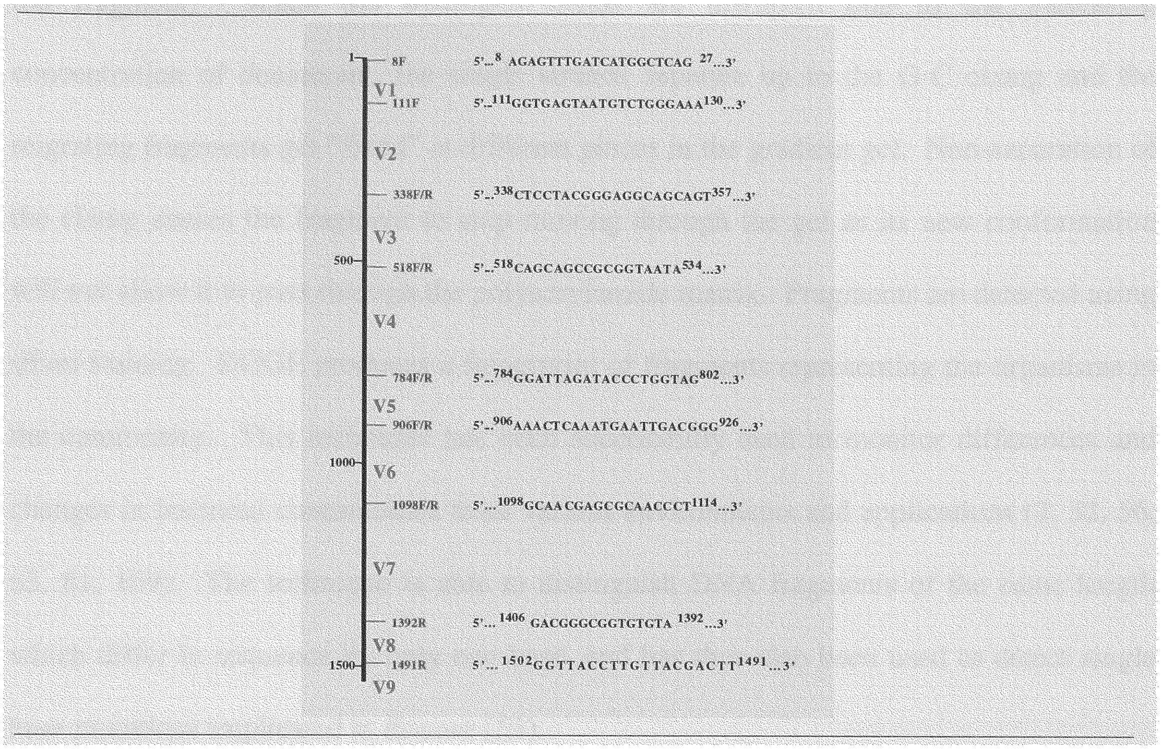
16S rRNA

Ribosomes are critical for protein production in all living organisms. Ribosomal RNA complexes with various proteins to make a functioning ribosome. The 16S rRNA component is an optimal genetic marker for characterizing bacteria for many reasons. The gene has conserved regions which are functional domains where selective pressure will not allow the bacterial DNA to mutate (108). Otherwise the translation process would be interrupted and the bacteria would die. These functional centers have a conserved sequence common to all bacteria. The gene also has variable regions interspersed between the conserved regions (Figure 8). This design lends itself to PCR analysis as primers can be derived from flanking conserved regions to amplify fragments of DNA in the variable regions, allowing molecular detection techniques to exploit length and sequence variations for the characterization of individual bacteria and bacterial communities. The 16S rRNA genes are considered to be one of the best targets for identifying bacteria, bacterial communities and for establishing evolutionary phylogeny (3, 34, 102). Once a target gene has been selected and primers identified, an appropriate method of detection must be determined.

DNA-based Assessment of Soil Microbial Community Structure

The most widely used DNA techniques for evaluating bacterial community structure involve amplification of a target gene and subsequent separation of DNA fragments using one of the techniques described below.

Figure 8. A Linear Schematic of the 16S rRNA Gene from *E. coli.* Alternating conserved (blue) and variable (red) regions as well as location and sequence of primers are shown. Natural variation in length and sequence can be targeted to putatively identify bacterial species.


1
500
1000
1500
V1
V2
V3
V4
V5
V6
V7
V8
V9
8F
111F
338F/R
518F/R
784F/R
906F/R
1098F/R
1392R
1491R
5'...8 AGAGTTTGATCATGGCTCAG 27...3'
5'..111GGTGAGTAATGTCTGGGAAA130...3'
5'..338CTCCTACGGGAGGCAGCAGT357...3'
5'..518CAGCAGCCGCGGTAATA534...3'
5'..784GGATTAGATACCCTGGTAG802...3'
5'..906AAACTCAAATGAATTGACGGG926...3'
5'...1098GCAACGAGCGCAACCCT1114...3'
5'... 1406 GACGGGCGGTGTGTA 1392...3'
5'...1502GGTTACCTTGTTACGACTT1491...3'

Denaturing Gradient Gel Electrophoresis (DGGE) is a technique that separates DNA fragments based on their sequence and moblity through a polyacrylamide gel with a gradient of denaturant. The gradient is usually composed of varying concentrations urea and formamide that denatures DNA. A tube containing 100 % formamide, a tube containing 100 % urea and a proportioning valve between them allow the person preparing the gel to control the concentration of formamide in the gel from bottom to top. One PCR primer is generated to have a GC-clamp composed of about thirty bases of GC-rich nucleotides on its 5' end (81). As the PCR-amplified fragment moves through the increasing concentrations of denaturant, its hydrogen bonding is disrupted. Hydrogen bonding between double-stranded DNA is characteristic of the sequence of base pairs in the fragment. When the hydrogen bonds are disrupted due to the increasing concentration of denaturant, the single strands separate up to the G-C clamp and the migrating fragments get "fixed" at different places in the gradient gel. Non-separation of the clamp causes the fragment to stop moving through the gel as its new conformation will not allow it to pass through the polyacrylamide matrix. Fragments are detected using silver staining. DGGE produces a fingerprint of fragments representing the organisms in the community. This technique has been successfully used to monitor differences and changes in bacterial communities from various environments and applications (7, 32, 56, 63, 81, 109). The technique is able to distinguish DNA fragments of the same length which differ in sequence by only one base, and has thus also been used to detect single base mutations implicated in disease (53).

In the forensic examination context, DGGE has some disadvantages including high cost of the longer primers, tedious gel preparation and separation times, lower reproducibility across laboratories and lack of automation.

Terminal Restriction Fragment Length Polymorphism (T-RFLP), a variation of the method used by Jeffreys, is a commonly used PCR-based method for profiling a microbial community. The method not only allows for direct comparisons of diversity between communities but also gives a semi-quantitative assessment of each fragment in the profile. A T-RFLP profile is generated by fluorescently tagging the 5' of a primer used in PCR, a method advantageous to RFLP because only targeted fragments are detected reducing the data to a manageable volume. The PCR-amplified fragments are then digested with restriction enzymes. Variations in fragment length depends upon the location of the restriction site relative to the 5' end of the PCR primer for each particular organism in the community. The fragments are then separated on high resolution sequencing gels or genetic analyzers. The technique uses a fluorescent scanner or laser to illuminate the DNA fragments in the gel and an internal standard of known size is run in the background to facilitate sizing. Fragment sizes obtained by T-RFLP can be used in some measure to infer the contributors to a community profile based on the large databases of ribosomal DNA sequences (67), however they do not confirm their identity. Effective evaluations of bacterial community diversity and structure from ocean sediments, remediation soils and other environments have been made using T-RFLP (6, 66, 67, 74, 94, 101).

From the forensic standpoint, T-RFLP is advantageous because analysis is much faster than DGGE allowing higher throughput and more reproducible profiles. In some cases T-RFLP produces too many fragments which can complicate the analysis (74). Other disadvantages are the possibility of incomplete restriction enzyme digestions and the fact that additional purification steps are necessary during the preparation of the samples.

Length Heterogeneity (LH-PCR) is becoming an increasingly popular method of bacterial community analysis (74, 88, 101, 106). This technique exploits natural variations in length between the conserved regions of the 16S rRNA genes. A 5' fluorescent label is attached to one of the primers in the PCR. Amplicons or PCR products of different length are separated on sequencing gels or in a capillary electrophoresis system. LH-PCR is a powerful technique for community profiling because like T-RFLP, it yields quantitative data based on the abundance of each amplicon in the profile. Advantages of LH-PCR over T-RFLP include high throughput as many samples can be analyzed in a shorter amount of time, with a higher level of reproducibility and it employs a technically less complex process (74). Application of LH-PCR to environmental samples found the technique to be highly efficient and reliable (74, 88, 101, 106). The amplicon lengths can also be used to make general inferences about the members associated with a community when they are compared to ribosomal RNA sequence and amplicon length databases (74). They do not however, specifically identify the microorganism. In the forensic arena, LH-PCR is a promising technique for microbial community profiling as it can provide fast turnaround times, lower costs of analysis, highly reproducible profiles and a pattern useful in quick screening of samples for comparison purposes.

DNA Profiling of Soil for Forensic Examinations

DNA-based profiling of soil has potential use in forensic comparisons for the identification and discrimination of suspect soil samples. Current methods of molecular-based DNA profiling of soils have based much of the conclusions they make upon ecological measures of soil health like diversity indices. However, DNA profiling of soils based on community structure has some inherent advantages over tradition forensic soil examination. Murray states that the number one reason for inconclusive results from a forensic soil examination is sample size and that most soil comparisons require at least a cup of soil (79). A recent study by Horswell *et al* evaluated the use of T-RFLP generated bacterial community profiles in a simulated forensic examination of soil (54). They found that reproducible profiles were obtainable for samples taken from a shoe sole or soiled clothes. They also found that the profile was representative of the site from which it was collected. Mills, King, Miller and Mathee have demonstrated similar results using the LH-PCR method (75).

Bioinformatics and Soil Comparison. The data produced by LH-PCR is a highly complex matrix detailing relative abundances of each fragment length for multiple fragments. Past researchers using this method have chosen to reduce the complexity of their data set to binary matrices, ecological indices, amplicon length data, or to select only a few of the data points for comparison (25, 74, 88, 101, 106).

Support Vector Machine (SVM) Learning Tools. There are computational tools available which can handle highly dimensional data such as that generated from LH-PCR profiling. Support Vector Machines are machine learning based computational tools used for supervised classification (24) that learn how to classify samples after being given a labeled training set of data. A SVM takes each profile and treats it as a feature vector in Euclidean space. It uses a kernel function to mimic the mapping of each vector in dimensional space and then produces a separator, or support vector, for each pair of classifiers in the data matrix. Once the support vectors have been generated, test samples are classified using a simple discriminant function based upon which side of the support vectors they fall. SVM's are sophisticated classification tools and have demonstrated their effectiveness in dealing with highly dimensional data such as that produced in gene expression microarrays (10, 41), LH-PCR data (113), face recognition (52) and text characterization (59).

Hypothesis and Objectives

The usefulness of microbial community profiles in forensic comparisons is determined by the geographic distribution of the organisms within a particular ecological niche. Current understanding is that soil type is the primary determinant of bacterial community structure and therefore will provide a start point for classifying our samples (47, 65). There are six different soil types characterized by the United States Department of Agriculture in Miami-Dade County Florida (Figure 9). We hypothesize that core microbial communities from each of the soil types will be significantly distinguishable by LH-PCR analysis and subsequent data interpretation. Specifically, this project will:

i. Evaluate the bacterial community structure of three Miami-Dade county soil types by LH-PCR.

ii. Test the robustness of the technique by sampling areas of known environmental insult.

iii. Evaluate other physical and chemical properties of soil samples to determine their utility as markers for forensic comparison.

iv. Create a database of microbial community profiles and chemical data from soil and analyze the accuracy of an SVM based supervised classification using various combinations of the biological and chemical data.

This project is a practical validation of the use of bacterial community profiles produced by LH-PCR as a rapid comparison method in forensic soil examination. Use of this technique may prove to have many advantages over traditional forensic geology including time of analysis, reproducibility of results, reliability and cost. If significant evidence is produced from this study, a database of soil community profiles may be justified which would give a forensic soil examiner the ability to use a profile from a suspect sample to infer the origin of the soil with some degree of accuracy. This would greatly enhance the utility of common soil evidence in a forensic context.

MATERIALS AND METHODS

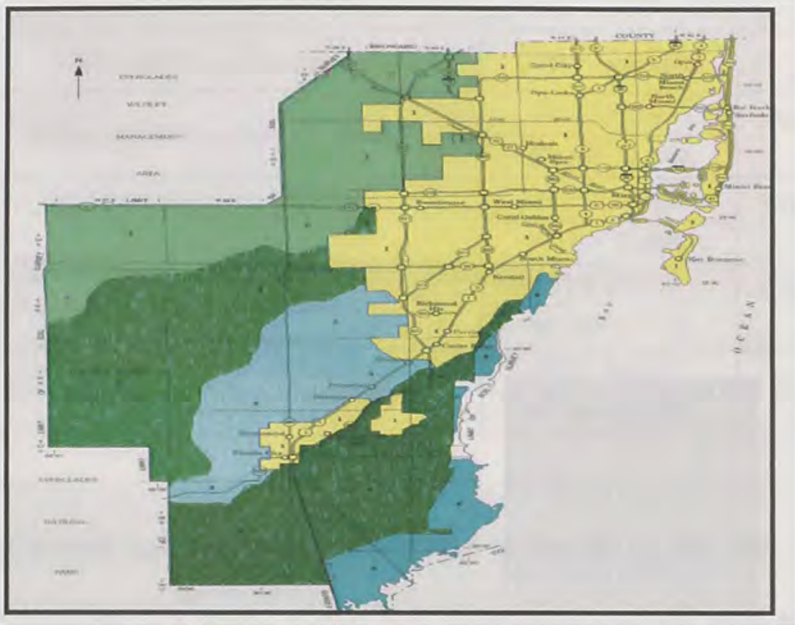
In order to evaluate the hypotheses set forth in this paper, and determine the robustness, reliability and reproducibility of the selected techniques for their routine application in the forensic arena, the following procedures were applied to all soils sampled in this study.

Bacterial Community DNA Profiling of Three Miami-Dade County Soil Types

Soils were sampled from three soil types in Miami-Dade County as described by the USDA (Figure 9). DNA was extracted and quantified. PCR and high resolution capillary electrophoresis (CE) on an ABI Prism 310 (Applied Biosystems; Foster City, CA) generated community profiles from the first three variable regions of the bacterial 16S rRNA gene. Gene products were analyzed for length (bp) and abundance (peak height) using GeneScan®2.1 and Genotyper® software (Applied Biosystems; Foster City, CA).

Sample collection. Soil samples were collected from three of the six different soil types that have been previously characterized by the USDA, Krome association (light blue), Rock outcrop-Biscayne-Chekika association (lime green) and Lauderhill-Dania-Pahokee association (mint green) respectively (Figure 9). Within the Krome association soil, an area considered to be pristine soil and an area of natural remediation from petroleum contamination was identified.

Figure 9. A General Soil Map of Miami-Dade County describing six different soil types each represented by a different color. Tan – Urban Land Udorthents assoc., mint green – Lauderhill-Dania-Pahokee assoc., lime green – Rock outcrop-Biscayne-Chekika assoc., forest green – Perrine-Biscayne-Pennsuco assoc., light blue – Krome assoc., aqua blue – Perrine-Terra Ceia-Pennsuco assoc. (USDA, Natural Resource Conservation Service, University of Florida-Soil Science Department)



In order to account for any seasonal changes, samples were collected in February, 2004 and in August, 2004, South Florida's dry and wet season. Soils were sampled on a completely random block pattern design. Blocks 100 m x 100 m were chosen in grassy areas well within the soil type boundaries. Within each block, three circular plots 2 m in diameter were measured and then three cores at random locations within the plot were taken. GPS coordinates were measured and recorded for each circle plot (table 1). Cores were taken using 50 ml conical tubes. These cores are 2.5 cm in diameter and vary in depth according to the thickness of the topsoil (average: 3-7 cm). Soil samples were

Table 1. Global Positioning System (GPS) and Miami City Street Coordinates for Sites. GPS coordinates for all sites sampled in this study as well as Miami street addresses are shown below.

Site	USDA Soil type description	GPS coordinates	Street coordinates
1	Krome association (contaminated)	N 25° 31.082'	SW 217 Ave 270 St
		W 080° 32.966'	
2	Krome association (pristine)	N 25° 31.031'	SW 217 Ave 270 St
		W 080° 32.968'	
3	Rock outcrop-Biscayne-Chekika association	N 25° 36.568'	SW 231 Ave 168 St
		W 080° 32.943'	
4	Lauderhill-Dania-Pahokee association	N 25° 45.203'	SW 177 Ave 12 St
		W 080° 28.907'	

collected using aseptic techniques and transported to the lab on wet ice. Soil samples were then homogenized in a sealed plastic bag, large chunks of soil were pulverized and small stones were removed. Aliquots of ~1 g were transferred into 1.5 ml microcentrifuge tubes. The spatula used for transfer was cleaned and sterilized with ethanol between each sample. Aliquots were labeled and stored in the –80 °C freezer. To account for any seasonal differences in community structure, sampling was performed in February, 2004 and August, 2004, Florida's dry and wet seasons respectively.

Extraction of the Soil Metagenome. Total DNA from each of the soil samples were extracted using the FastDNA™ Spin Sample Kit for Soil (Cat # 6560-000, QBiogene, Vista, CA). Duplicate extractions of each soil sample were performed. The manufacturer's protocol was employed with a slight modification for soils suspected to be rich in organic content or contaminated (72). Briefly, 500 mg of the homogenized soil was transferred to a multimix 2 matrix tube containing ceramic beads which aid in mechanical lysis. Sodium Phosphate buffer (978 µl) and MT buffer (122 µl) were added to each tube and the tubes were capped and vortexed. Tubes were then placed in the Fast Prep instrument (Qbiogene, Vista, CA) and run at intensity level 5.5 for 20 seconds. Samples were immediately put on ice for 2 minutes. At this point, extraction tubes from soils suspected to be high in organic content or contamination were placed in a 70 °C water bath for 30 minutes. Samples were then placed back in the Fast Prep instrument for 10 seconds at intensity 5.5, and then were placed on ice again for two minutes. Samples were centrifuged at 16,300 g for 5 minutes and the supernatant transferred to a clean 2 ml centrifuge tube using a micropipette. Protein Precipitation Solution (250 µl) was added to each tube and tubes were inverted 10 times. Samples were centrifuged again for 5 minutes at 16,300 g to pellet the precipitate. The supernatant containing the DNA was transferred to a clean 1.5 ml centrifuge tube. Binding matrix was vortexed until it was resuspended and then added (1 ml) to each sample. Vortexing was frequently repeated to ensure suspension of the Binding Matrix. Tubes were then placed on a rotator at medium speed for 3 minutes in order for the DNA to bind to the silica. The samples were then placed on counter for 10 minutes in order for the Binding Matrix to settle. The supernatant above the Binding Matrix was then carefully removed using a micropipette. Binding Matrix was then gently resuspended by tapping the tube with one finger and then transferred to a SPIN^{TM} Filter and catch tube. These were centrifuged at 16,300 g for 1 minute. In this step, salt bridges allowed the extracted DNA to remain bound to the silica while other residual proteins and solutions wash through. The flow-through was discarded and the column was placed back in the catch tube. SEWS-M (Salt/Ethanol

Wash Solution) was then added (500 µl) to each filter and the tubes were centrifuged at 16,300 g for 1 minute. Flow-through was discarded and the filter was placed back into the emptied catch tube. This wash step was repeated once. Tubes containing filters were centrifuged at 16,300 g for 2 minutes to dry the filter of any residual SEWS-M. The filter was removed and placed in a new catch tube and allowed to air dry for 5 minutes. DNA was eluted by adding 100 µl of 65 °C diethylpyrocarbonate (DEPC) water to the filter. Gentle tapping with the finger was used to re-suspend the Binding Matrix above the filter. Tubes were centrifuged at 16,300 g for 1 minute. DNA was eluted once more by adding 50 µl of 65 °C DEPC water, suspending the Binding Matrix again and centrifuging for 1 minute at 16,300 g. The catch tubes containing the DNA were labeled and stored at -20 °C.

DNA Quantification. DNA was quantified using a DyNAQuant 2000 fluorometer (Hoefer Scientific Instruments, San Francisco, CA). The instrument was turned on 30 minutes prior to use. In order to prepare the proper amount of working solution, the number of readings being taken was estimated and one calibration standard reading was added for every five unknown sample readings (ie. 25 samples + 5 standards = 30 readings). The number of total readings was multiplied by two to get the final volume in milliliters of working solution (60 ml). The working solution was prepared by adding one tenth volume 10 X Tris NaCl EDTA (TNE) buffer (6 ml), 1 µl Hoescht dye for every ml 10 X TNE (6 µl), and then adding filtered reverse osmosis (RO) water until the final volume is reached. Readings were obtained by first transferring the working solution to a calibrated dispenser and then setting the dispenser volume at 2 ml. The solution was given a final swirl. Working solution was dispensed (2 ml) into a clean glass cuvette making sure no bubbles are present as they affect volume and concentration. The cuvette containing 2 ml of solution was placed into the reading well and the fluorometer was zeroed. At this point, 2 μl 100 ng/μl Calf-thymus DNA standard was added to the cuvette, the cap was placed on the cuvette and it was inverted five times to mix. After 10 seconds, the fluorometer was calibrated by entering the concentration of the standard at 120 ng/μl. This was done to compensate for DNA which was not of bacterial origin which would otherwise overestimate the concentration of template DNA going into PCR. The cuvette was washed using the vacuum funnel with the water jet adapter by pouring filtered RO water into the funnel. The cuvette was allowed to dry for ~5 seconds atop the apparatus and then 2 ml of working solution was again dispensed into the cuvette. The cuvette was placed in the reading well and the instrument was zeroed. Samples of unknown concentration were then added to the cuvette (2 μl), and the cap placed. The capped cuvette was inverted five times and replaced in the reading well. After ~10 seconds, the read button was pushed and a reading was recorded. The calibration step was repeated every five samples.

The PCR protocol used in this study calls for 1 μl of template DNA at a concentration of 10 ng/μl. DNA extracts were diluted to this concentration using the following formula: $V_1C_1 = V_2C_2$ where C_1 is the concentration of the DNA coming directly from the quantification procedure, V_2 is 100 μl, C_2 is 10 ng/μl, and solving for V_1 gives the amount of concentrated DNA to deliver in a total volume of 100 μl. After delivering the DNA to a new tube, the volume was brought up to 100 μl with DEPC H_2O

and the tubes were labeled for the soil extraction they came from and their concentration, 10 ng/µl. Tubes were stored at -20 °C.

PCR Amplification of 16S rRNA Gene. Three variable regions (V1, V1+V2, and V3) of the 16S rRNA gene were PCR amplified for each soil DNA extraction. All PCR reactions were performed in duplicate for each soil extraction and for each variable region. PCR amplification was performed on a DNA engine Opticon 2® (MJ Research, Waltham, MA) using the following volumes and concentrations; 9.9 µl DEPC H_2O, 2.0 µl 10X PCR Buffer, 2.0 µl 25 mM $MgCl_2$, 2.0 µl 2.5 mM dNTPs, 0.5 µl 20 uM forward and reverse primers, 0.1 µl Amplitaq Gold LD DNA Polymerase (Applied Biosystems, Foster City, CA) 2.0 µl 1% bovine serum albumin (BSA) and 1.0 µl of 10 ng/ml DNA. PCR was performed with the following program; 95 °C for 11 minutes (initial denaturation of DNA and activation of polymerase), 95 °C for 1 minute (denaturation of dsDNA), 55 °C for 1 minute (primers anneal), 72 °C for 1 minute (extension), 72 °C for 10 minutes (final extension), and a 15 °C hold indefinitely. The program cycles 25 times through the 1 minute denaturation, annealing and extension stages.

DNA Analysis by Agarose Gel Electrophoresis. PCR products were screened using agarose gel electrophoresis. Depending on the variable region being targeted, PCR products were expected anywhere from about 80 base pairs to about 330 base pairs in length. Following amplification, 5 µl of each PCR product were loaded in a 1.5 % agarose gel in 1X Tris-Borate EDTA (TBE) buffer next to a 100 bp molecular ladder

(New England Biolabs, Beverly, MA) and separated by electrophoresis at 5 V per cm. This concentration was effective in resolving bands in the low molecular weight range (50-1000 base pairs). The agarose gel was photographed after staining with ethidium bromide and illumination in a UV light box.

ALH analysis of PCR products. PCR products were analyzed for Amplicon Length Heterogeneity (ALH) using an ABI® Prism 310 (Applied Biosystems, Foster City, CA) high-resolution genetic analyzer. Samples were prepared by adding 0.5 μl of the PCR product to 9.5 μl of a 24:0.25 ratio mixture of highly deionized formamide (a denaturant used to maintain DNA in a single strand conformation) and GeneScan^{TM} 500 ROX size standard (PE Biosystems, Foster City, CA), heating for 2 minutes at 95 °C and cooling on ice for 5 minutes. Samples were electrophoresed for 28 minutes when separating fragments from the V1 + V2 region or for 24 minutes for V1 and V3 fragments. Fragments were separated by capillary electrophoresis in a matrix of POP-4 polymer (Applied Biosystems, Foster City, CA). Matrix DS-30_6FAM_HEX_NED_ROX with its respective filter, D, were set before beginning the run.

GeneScan® Analysis of DNA Fragments. Using GeneScan® 2.1 (Applied Biosystems, Foster City, CA), DNA fragments were analyzed for size (bp) and for abundance (peak height). Analysis range was set at 2,800 to 10,000 and the peak threshold was set at 50 Relative Fluorescent Units (RFU) for each dye. The local southern sizing method was used. An internal sizing standard, GeneScan^{TM} 500 ROX was added to each sample in order to correctly size amplicons. Size calling was checked for linearity based on the migration and size calling of DNA fragments of known length in the internal size standard. The internal size standard consisted of fluorescent labeled DNA fragments of lengths between 75 and 500 bp.

Exporting Fragment Data Using Genotyper®. After sizing of amplicons was completed using Genescan® 2.1, peaks were filtered using Genotyper® (Applied Biosystems, Foster City, CA) software. Filtering was performed based upon expected sizes (bp) of DNA fragments generated by the PCR and upon the fluorophore label attached to the DNA. A table of called peaks including their size (bp) and their peak height was created and exported into a Microsoft® Excel spreadsheet (Figure 14). There were four total replicates per region per soil sample. Any peak not present in three out of four replicates was considered an artifact and not included as true data. All peaks were confirmed by checking the raw data again in GeneScan®. At this point, decisions were made and applied to all runs concerning any ambiguous peak calling due to rounding fragment sizes to the nearest integer. These decisions resulted in a profile analysis sizing key. The ABI® 310 is able to separate fragments with ± 0.01 bp pair resolution. Synonymous peaks in replicates usually called within a range of about 0.3 bp. When the range spanned the rounding area of bp calling, sizing became difficult. For example, peaks representing an amplicon approximately 78 bp long in four replicates of one sample sized at 78.41, 78.50, 78.56 and 78.65. The size calling system in Genotyper® has to round to the nearest integer in order for the data to be manageable. Based upon the average sizes of the peaks representing the same amplicon, a decision was made to call

all peaks which fell in this size range 79 bp long even though Genotyper® would export some of them as a 78 bp fragment. These decisions had to be consistently applied to all profiles to ensure accurate interpretation.

Combining, Normalizing and Pruning Microbial Community Profile Data. The resulting data was pruned by eliminating any amplicons not present in at least three of the four replicate profiles. The "pruned" data was then pasted into rows combining the data from V1, V3 and V1 + V2 regions (113). All data in each row represented PCR's from a single soil extraction. There were a total of four rows (replicates) for each soil sample. All data in each row was normalized (individual peak fluorescence divided by total fluorescence of all peaks in row) so that each fragment length now had a relative abundance unit attached. At this point another "pruning" was performed removing any peak representing less than 1 % (0.01) of the total profile (113). The resulting microbial community profile is our final data output including fragment length in bp and relative abundance of each peak in the profile from each soil replicate. A combined total of 864 profiles derived from nine soil samples taken from four sites for each of two seasons, two extractions per sample, two amplifications per extraction for three separate variable regions.

Assessing the Soils Physical and Chemical Properties

Physical and chemical properties including moisture content, pH, percent carbon and percent nitrogen, elemental composition and contamination by semi and non-volatile organic compounds were examined for each soil sample. These factors are known influences on bacterial community structure and were monitored to determine whether they could be used in discriminating samples from different soil types.

Total % Carbon and Total % Nitrogen. Each sample of soil was digested and analyzed for total carbon (C) and total nitrogen (N) using a ThermoFinnigan EA 1112 Flash NC analyzer courtesy of Dr. James Entry, USDA Kimberly, Idaho. Briefly, an approximate 10 g sub-sample of each soil was dried at 70 °C for 72 hours and subsequently sieved through a 1,000 micron steel mesh. A 0.2 g sub-sample was placed in a tin foil cup, sealed and placed in a ThermoFinnegan EA 1112 Flash NC Soil Analyzer and assessed for total % C and total % N.

Soil Moisture Content and pH Determination. Soil samples were weighed before and after drying overnight at 80 °C. Moisture content for each sample was recorded according to the EPA's moisture content (θ) calculator:

$\theta = (W_w - W_d / P_w) / [(W_w - W_d / P_w) + (W_d / P_s)$, where θ equals soil porosity assumed equivalent to the moisture content, W_w and W_d are the wet and dry weight of the soil and P_w and P_s are the densities of water and solids. Bulk density reported in g/ml was obtained from an EPA soil survey of Dade County (http://soils.usda.gov). The pH was determined by adding an equal volume of de-ionized water in ml per gram dry soil. After calibration the pH meter was set in the slurry and a reading was taken.

Elemental Composition of Soils by Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES). Each soil sample was analyzed for total Phosphorous (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Manganese (Mn), Iron (Fe), Copper (Cu), Boron (B) and Zinc (Zn) according to USEPA method 3051 XPHCL (USEPA, 1986). Briefly a 0.25 g sub-sample of previously sieved soil was placed in a 250 ml XP-1500 CEM MARS # 61535 digestion vessel. After the addition of nine ml of concentrated $HNO_{3,}$ the mixture was incubated for 8 hours. Three ml of 12 M HCl was then added and the solution was swirled for 1 minute. The vessel was sealed to 20 p.s.i and placed in a CEM MARS 5 61535 microwave oven. The vessel was then pressurized to 600 p.s.i. and temperature was increased to 165 °C for 3 minutes. At this point pressure and temperature were increased to 750 p.s.i. and 175 °C for 5 minutes. The mixture was then allowed to cool and transferred to a 50 ml volumetric flask. The volumetric flask was washed five times with 3 ml of micropure water in order to ensure that all nutrients were transferred to the flask. The volume in the flask was brought up to 50 ml with micropure water. A 15 ml sub-sample was filtered through 0.45 micron filter and 1 ml of filtrate was analyzed for total concentrations of each aforementioned element on a Perkin Elmer Optical Spectrometer Optima 4300 DV ICP (PE Biosystems, Foster City, CA).

Detection of Diesel Range Organics (DRO). An assessment of the volatile and semi-volatile organic compounds present in the soil samples tested the robustness of our microbial community profiling technique. The presence or absence of these compounds in the soil samples was determined according to the standard set by the American Society for Testing and Materials (ASTM, West Conshohocken, PA) E 1618-01 for testing ignitable liquid residues by Gas Chromatography-Mass Spectrometry (GC-MS). Briefly, aliquots of each soil sample were placed in a container and heated in a furnace at 80 °C for 14 hours. All volatile hydrocarbons were adsorped to an 8 mm X 8 mm strip of activated charcoal suspended in the headspace of the container. Pure carbon disulfide (CS_2) was added (200 μl) to extract the compounds from the activated charcoal strip in a 2 ml glass vial. After vortexing for one minute, vials were placed in a rack with the charcoal strip submersed in the solvent for 30 minutes. The vials were then centrifuged at 8,000 g for 5 minutes. The extraction was transferred to a clean 2 ml glass vial with inserts for the Varian 8200 (Supelco; Bellefonte, PA) auto-sampler. The extractions were injected splitless with an injection volume of 1 μl at 280 °C on a 1079 temperature programmable injector. The gas chromatographer was a Varian Star 3400 CX (Varian Chromatography Systems Walnut Creek, CA) with a DB-5 GC (Supelco Bellefonte, PA) fused silica capillary column of length 30 m, diameter 0.25 mm and film thickness 0.25 μm. The carrier gas was He at a flow rate of 1 ml/min. The column temperature program was; initial 45 °C, ramp 5 °C/min to 75 °C, hold for 2.5 min, ramp 5 °C/min to 115 °C, ramp 11 °C/min to 250 °C, and then ramp 50 °C/min until a final temperature of 300 °C is reached and held for 2 min. The total run time is 28.77 min. The gas chromatographer was interfaced to a Varian Saturn 2000 Ion Trap Mass Spectrometer

(Varian Chromatography Systems Walnut Creek, CA). An ion scan range of 40-450 m/z was used. Ions were analyzed using extracted ion profiling after the full range of ions were collected by the mass spectrometer. The presence or absence of diesel fuel was determined based on pattern recognition, following the ASTM guidelines. Diesel fuel has a distinct pattern of alkanes with regard to their retention times and relative abundance. Recognition of this pattern along with two key markers necessary to identify diesel, namely pristane and phytane, were the two factors used in determining the presence or absence of diesel.

Traditional Ecological Indices. Measures commonly taken to assess the health of an ecosystem include richness, evenness, diversity and Hmax (16, 22, 33, 88). Richness is a measure of the minimum number of species represented in each sample. In the case of microbial community profiles, richness is represented by the total number of distinct amplicons or peaks. Evenness is a measure of the distribution of abundance of each amplicon in a sample. Shannon-Weaver diversity index (H') is a measure of biodiversity which takes into account species richness and species evenness (114). Hmax is the maximum theoretical diversity of species for all samples. Each sample's richness, evenness, diversity and Hmax was determined from the microbial community profile data using a Microsoft Excel macro.

Statistical Analysis of Physical Properties, Chemical Data and Traditional Ecological Indices. One-way analysis of variance (ANOVA) was used to determine whether there were significant differences ($p < 0.05$) between samples from different soil types as well as between subplots within each soil type for each individual parameter. Initial analysis was performed by categorizing samples only according to soil type and subsequently by both soil type and season. One-way ANOVA was done using the statistical software package SPSS (Version 10.0 for Windows SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor Chicago, Illinois 60606) website: http://www.spss.com. Samples from different soil types and samples from subplots within each site were also analyzed for significant differences in traditional ecological indices.

Supervised Classification Using Support Vector Machine Learning Tools. ALH-PCR generated profiles were used to create a database used for supervised classification of test samples. Training sets consisting of n-1 samples were generated by merging all amplicon data for the variable region(s) of interest into one merged file. The single spreadsheet creates a column for every possible fragment length present in the original spreadsheets. Test sets were one complete set of replicates left out of the training set. This is a more robust method of creating training and test sets for supervised classification. The SVM classifier was implemented using the LibSVM package. This package is available for download at http://www.csie.ntu.edu.tw/~cjlin/libsvm for academic use. The program was run using three kernel functions: *linear* characterized by $\mathbf{K(X,Y) = (X \bullet Y + 1)^d}$, with d = 1 for a polynomial in the first degree, *radial basis function* characterized by $\mathbf{K(X,Y) = exp(-\gamma||X\text{-}Y||^2)}$, and *sigmoid* characterized by $\mathbf{K(X,Y) = tanh(\gamma\ (X \bullet Y) + \theta)}$. Default parameters were used in each case. Preparing data for the SVM classifier required merging of data from multiple files as well as proper labeling of training and test sets. Programs to perform these tasks were written in Java™

(Sun Microsystems; Santa Ana, CA) by Chengyong Yang PhD, School of Computing and Information Science, Florida International University. Once classification had been done on test samples, the outputs were imported into a spreadsheet and sample results were compared to their correct label (soil type and season) to determine accuracy of the classifier.

RESULTS

This study was based on the hypothesis that microbial community profiles from a sample of soil in addition to data obtained from physical and chemical analyses would yield a unique statistical unit distinguishable from samples originating from another soil type. Specifically, this study generated bacterial community DNA profiles from three Miami-Dade County soil types using ALH-PCR to target natural length variation in the 16S rRNA genes and assessed physical and chemical properties of each soil sample. The sampling of one soil type was replicated in an area of known contamination by diesel fuel. This was done to test the robustness of the techniques applied in this study, namely DNA extraction, ALH-PCR and microbial community profile classification by presenting samples challenged by environmental insults. This was followed by statistical analysis of individual parameters to detect any significant difference between soil types. Supervised classification of soil profiles using SVM computational tools was used to determine the overall accuracy of databases of soil microbial community profiles compared to data from physical and chemical analyses in correctly classifying unknown test samples.

Total Percent Carbon (C) and Percent Nitrogen (N) Between Sites. Each soils total percent Carbon and Nitrogen by mass were examined to determine their potential value as a forensic marker (Table 2). Mean percent C was significantly different ($p < 0.05$) in Bonferroni post-hoc comparisons for all but one comparison between soil types, Krome association compared to Lauderhill-Dania-Pahokee association. Lauderhill-Dania-Pahokee soils had the highest percent C followed by Krome and Rock outcrop-

Table 2. Between Site Comparisons % Carbon, % Nitrogen, Soil Moisture and pH. Mean values for total % C and % N, soil moisture, and pH values as well as results from One-way ANOVA to determine whether differences between soil types were significant ($p < 0.05$) are shown. Mean values for sites which have the same letter in superscript were not significantly different using Bonferroni post-hoc comparisons.

Soil type	pH	Moisture	% carbon	% nitrogen
Krome Association n = 36	6.44 ± (0.76)^{a}	0.43 ± (0.23)^{c}	16.83 ± (4.70)^{f,h}	0.76 ± (0.30)^{i}
Rock outcrop Biscayne Chekika n = 18	6.99 ± (1.19)^{a,b}	0.52 ± (0.19)^{c,d}	11.60 ± (2.77)^{g}	0.54 ± (0.17)^{j}
Lauderhill-Dania-Pahokee n = 18	7.19 ± (1.12)^{b}	0.26 ± (0.06)^{e}	18.43 ± (1.84)^{h}	0.67 ± (0.18)^{i,j}
ANOVA	$F = 4.29$ $P < 0.13$ $R^2 = 0.06$	$F = 8.47$ $P < 0.01$ $R^2 = 0.20$	$F = 17.21$ $P < 0.01$ $R^2 = 0.33$	$F = 5.30$ $P < 0.01$ $R^2 = 0.13$

Biscayne-Chekika.

Mean percent N was significantly different only for the comparison between Krome and Rock outcrop-Biscayne-Chekika soils. Krome soils had the highest percent N followed by Lauderhill-Dania-Pahokee and Rock outcrop-Biscayne-Chekika soils had the lowest percent N.

Soil Moisture and pH Determination Between Sites. Soil moisture as well as pH were measured to determine whether differences existed between sites. Soil moisture and pH values were significantly different ($p < 0.05$) for comparisons between Krome and Lauderhill-Dania-Pahokee soils. Significant differences in moisture were also seen between Rock outcrop-Biscayne-Chekika and Lauderhill-Dania-Pahokee soils.

Seasonal Comparisons Between Sites for % Carbon, % Nitrogen, Soil Moisture and pH. When data was broken down by season, no significant differences were seen in comparisons between sites for Nitrogen content (Table 3). Significant differences were seen in percent C for the wet season when comparing Krome and Rock outcrop-Biscayne-Chekika soils as well as between Rock outcrop-Biscayne-Chekika and Lauderhill-Dania-Pahokee soils. No significant differences were seen in Carbon content during the dry season. The only significant difference in soil moisture was in dry season comparisons of Rock-outcrop-Biscayne-Chekika and Lauderhill-Dania-Pahokee soils. During the dry season, no significant differences were seen in pH between any of the sites. Wet season pH data was significantly different for comparisons between Krome and Rock outcrop-Biscayne-Chekika soils in addition to comparisons between Krome and Lauderhill-Dania-Pahokee soils.

Seasonal Comparisons Within Sites for % Carbon, % Nitrogen, Soil Moisture and pH. Measures for pH increased significantly ($p < 0.05$) for all sites from dry season to wet season (Table 3). No significant differences were seen between seasons for measures of % C, % N, or for soil moisture. Breaking data down by season resulted in lower significance levels (increased P) and a lower F statistic in ANOVA for all comparisons except for pH. Within group variance in mean pH values decreased

Table 3. Mean Values for Soil Type and Season and One-way ANOVA Results for % Carbon, % Nitrogen, Soil Moisture and pH. Mean values for total % C and % N, soil moisture, and pH values as well as results from One-way ANOVA to determine whether differences between soil types were significant ($p < 0.05$) are shown. Mean values for samples which have the same letter in superscript were not significantly different using Bonferroni post-hoc comparisons.

Soil type and season	Carbon	Nitrogen	Moisture	pH
K dry n = 18	$16.46 \pm (5.85)^{a,c,d}$	$0.73 \pm (0.29)^{e}$	$0.54 \pm (0.29)^{f,i,k}$	$5.73 \pm (0.21)^{l,q}$
K wet n = 18	$17.20 \pm (3.32)^{a,c}$	$0.80 \pm (0.31)^{e}$	$0.32 \pm (0.09)^{g,k,h}$	$7.14 \pm (0.28)^{m}$
RBC dry n = 9	$12.32 \pm (2.56)^{a,b,d}$	$0.57 \pm (0.18)^{e}$	$0.69 \pm (0.09)^{f,i}$	$5.85 \pm (0.18)^{l,q,o}$
RBC wet n = 9	$10.88 \pm (2.93)^{b,d}$	$0.50 \pm (0.17)^{e}$	$0.34 \pm (0.03)^{f,g,k,h}$	$8.14 \pm (0.15)^{n}$
LDP dry n = 9	$18.75 \pm (2.21)^{a,c}$	$0.73 \pm (0.19)^{e}$	$0.26 \pm (0.05)^{g,h,k}$	$6.10 \pm (0.03)^{o,q}$
LDP wet N = 9	$18.11 \pm (1.43)^{a,c}$	$0.62 \pm (0.15)^{e}$	$0.25 \pm (0.05)^{h,g,k}$	$8.28 \pm (0.08)^{p,n}$
ANOVA	$F = 6.92$ $P < 0.01$ $R^2 = 0.34$	$F = 2.42$ $P < 0.05$ $R^2 = 0.16$	$F = 9.35$ $P < 0.01$ $R^2 = 0.47$	$F = 360.49$ $P < 0.01$ $R^2 = 0.97$

K- Krome association soil, R B C- Rock-outcrop-Biscayne-Chekika association soil, L D P- Lauderhill-Dania-Pahokee association soil.

dramatically when seasons were accounted for (samples labeled for soil type only, F = 4.29; samples labeled for soil type and season, F = 360.49). Interestingly, mean values for soil moisture decreased, although not significantly, from dry season to wet season for all three soil types. Mean pH values ranged between 5.73 and 8.28 when seasons were accounted for.

Subplots Comparisons Within Sites for % Carbon, % Nitrogen, Soil Moisture and pH. Soil samples were also compared within soil types and within the same season to determine whether significant differences existed without crossing soil type boundaries (Table 4). No significant differences were seen between subplots of the same soil type for percent carbon for soils sampled during either season. During the dry season, the first subplot from the Krome diesel site was significantly different in nitrogen content when compared to the third subplot from the Krome pristine site. For soil moisture measures taken during the dry season, all subplots from the Krome diesel site were significantly lower ($p < 0.05$) than all subplots of the Krome pristine site. During the wet season, pH comparisons between the first subplot of the Krome diesel site and the first subplot of the the Krome pristine site showed a significant difference. All other comparisons between subplots of the same soil type showed no significant difference for any of these measures.

Elemental Composition of Soils by ICP-OES. Elemental analysis of each soil sample was performed using ICP-OES to determine whether samples from different soil types could be distinguished based on their elemental composition. Accuracy levels for this method, calculated from the mean of five replicate analyses of a quality control standard for each element, were generally between 89 and 98 %, which is within the

Table 4. Within Site, Between Subplots Comparisons of % C, % N, Moisture and pH. Means and standard deviations are shown for each subplot within the soil types.

Dry season						
Site[a]	**% C**			**% N**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	11.57 (±4.38)	21.78 (±13.00)	13.84 (±1.87)	.29 (±0.18) [a]	.76 (±0.20)	.66 (±0.28)
K_p	15.85 (±0.58)	17.14 (±1.35)	18.61 (±0.68)	.71 (±0.11)	.88 (±0.13)	1.09 (±.09) [b]
RBC	10.54 (±0.96)	11.91 (±1.97)	14.50 (±3.05)	.52 (±0.24)	.54 (±0.21)	.63 (±0.11)
LDP	18.39 (±2.61)	19.58 (±2.94)	18.29 (±1.57)	.74 (±0.28)	.76 (±0.25)	.68 (±0.08)
Site[a]	**Moisture**			**pH**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	.18 (±0.04) [c]	.33 (±0.06) [c]	.22 (±0.06) [c]	5.85 (±0.19)	5.91 (±0.10)	5.80 (±0.03)
K_p	.80 (±0.05) [d]	.80 (±0.02) [d]	.79 (±0.02) [d]	5.51 (±0.18)	5.65 (±0.26)	5.67 (±0.22)
RBC	.72 (±0.12)	.71 (±0.08)	.65 (±0.07)	5.85 (±0.17)	5.92 (±0.21)	5.79 (±0.23)
LDP	.25 (±0.06)	.27 (±0.06)	.26 (±0.04)	6.12 (±0.04)	6.10 (±0.04)	6.09 (±0.01)

Wet season						
Site[a]	**% C**			**% N**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	12.68 (±3.10)	18.53 (±2.74)	19.50 (±5.13)	.44 (±0.25)	.95 (±0.50)	.85 (±0.26)
K_p	18.08 (±0.53)	16.25 (±1.16)	18.16 (±1.86)	.92 (±0.15)	.70 (±0.23)	.93 (±0.24)
RBC	8.34 (±1.74)	10.26 (±1.64)	14.04 (±1.83)	.36 (±0.05)	.46 (±0.03)	.69 (±0.18)
LDP	18.53 (±0.70)	18.00 (±2.54)	17.80 (±0.92)	.58 (±0.19)	.60 (±0.21)	.68 (±0.07)
Site[a]	**Moisture**			**pH**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	.26 (±0.12)	.33 (±0.15)	.38 (±0.07)	7.50 (±0.12)^{e}	7.14 (±0.21)	7.19 (±0.22)
K_p	.37 (±0.04)	.34 (±0.07)	.38 (±0.04)	6.78 (±0.09)^{f}	7.24 (±0.26)	7.00 (±0.25)
RBC	.34 (±0.02)	.32 (±0.04)	.36 (±0.01)	8.22 (±0.11)	8.20 (±0.03)	8.02 (±0.19)
LDP	.21 (±0.02)	.29 (±0.06)	.24 (±0.02)	8.26 (±0.07)	8.27 (±0.12)	8.30 (±0.08)

[a]Sites are represented by Krome-diesel (K_d), Krome-pristine (K_p), Rock-outcrop Biscayne Chekika (RBC) and Lauderhill Dania Pahokee (LDP)

*** Values followed by a different letter in superscript were significantly different ($p < 0.05$).**

expected performance capability of ICP-OES. **Table 5** shows high accuracy levels for the elements compared to a Certified Reference Material (CRM). Only Fe and Mn

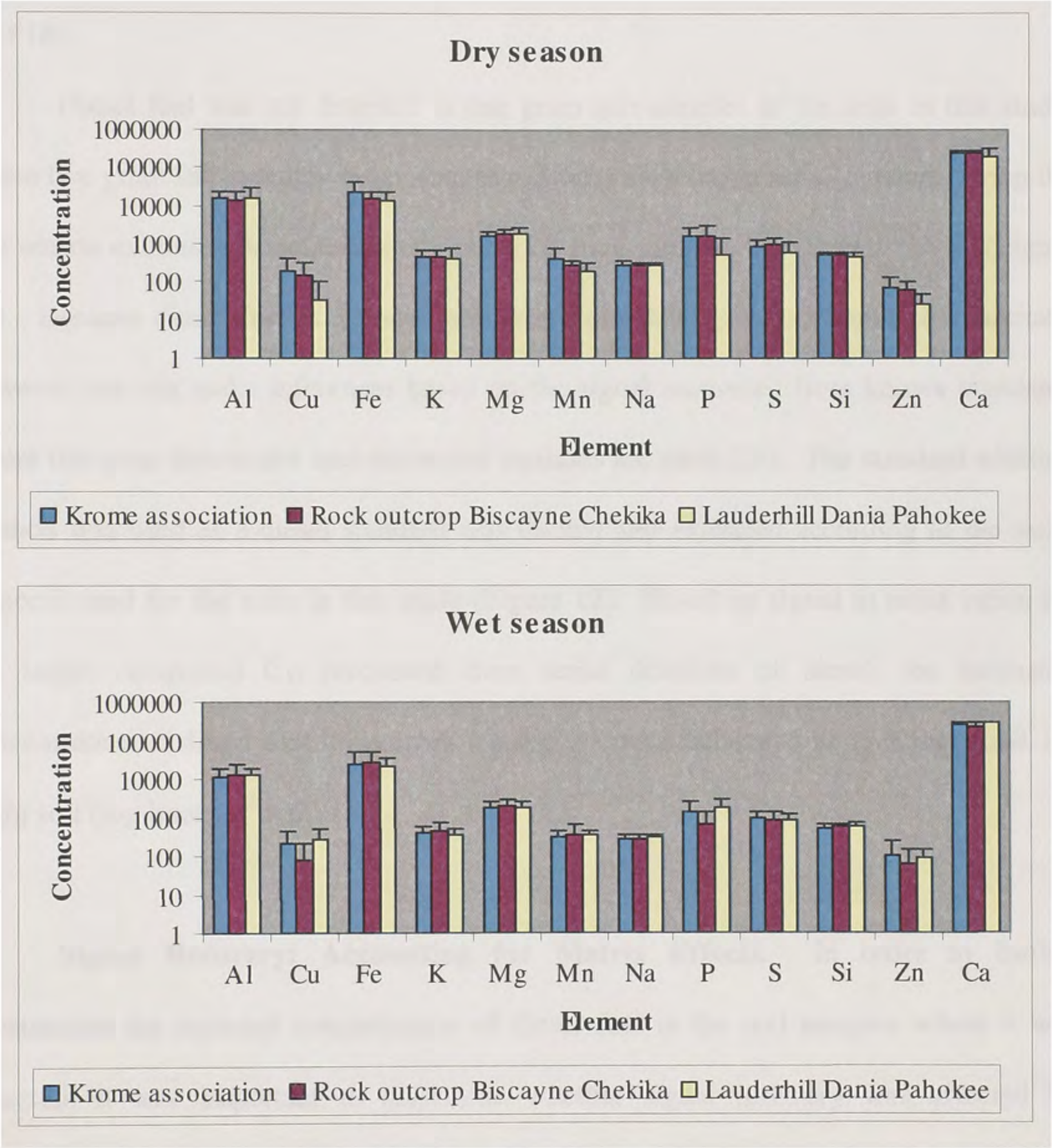
Table 5. Accuracy Levels Obtained by ICP-OES of a Certified Reference Material Standard for Elements Shown. Mean % accuracy and standard deviation for the elements Al, B, Ca, Fe, K, Mg, Mn, Na, P, S, Si and Zn based on five replicate analysis of quality control standards.

Element	**Al**	**B**	**Ca**	**Fe**	**K**	**Mg**
% Accuracy Mean ± S. D.	98.21 ± 1.67	93.12 ± 4.46	96.62 ± 2.29	65.19 ± 2.66	96.91 ± 1.81	97.91 ± 1.92
Element	**Mn**	**Na**	**P**	**S**	**Si**	**Zn**
% Accuracy Mean ± S. D.	77.25 ± 2.79	97.92 ± 1.42	92.91 ± 6.53	90.93 ± 8.89	97.32 ± 2.22	89.77 ± 10.25

showed significant deviation from the certified values. Precision measures were not able to be determined because although multiple soil samples were taken from each site, no replicate analyses of the individual soil samples were conducted. After trying non-treatment of the data, removing Ca data due to its abundance, normalizing to the Al concentration, normalizing to the Fe concentration and log normalization to Al or Fe, , One-Way ANOVA showed no significant differences ($p < 0.05$) in elemental composition for any of the sites compared in this study. A graph of the concentration in parts per million of each element per gram soil is shown in Figure 10. Calcium had the highest mean concentration followed by Iron, Aluminum and Magnesium.

Detection of Diesel Range Organics (DRO). In order to determine the effects of an environmental insult on the ability to differentiate microbial community profiles from different soil types, soils were sampled from an area of prior contamination by

Figure 10. Elemental Composition Per Gram Soil. Mean concentration and standard deviation on a logarithmic scale for Aluminum, Calcium, Copper, Potassium, Magnesium, Manganese, Sodium, Phosphorus, Sulfur, Silicon and Zinc obtained by **ICP-OES**. There were no significant differences ($p < 0.05$) between the soil types for any of the elements profiled (ANOVA).

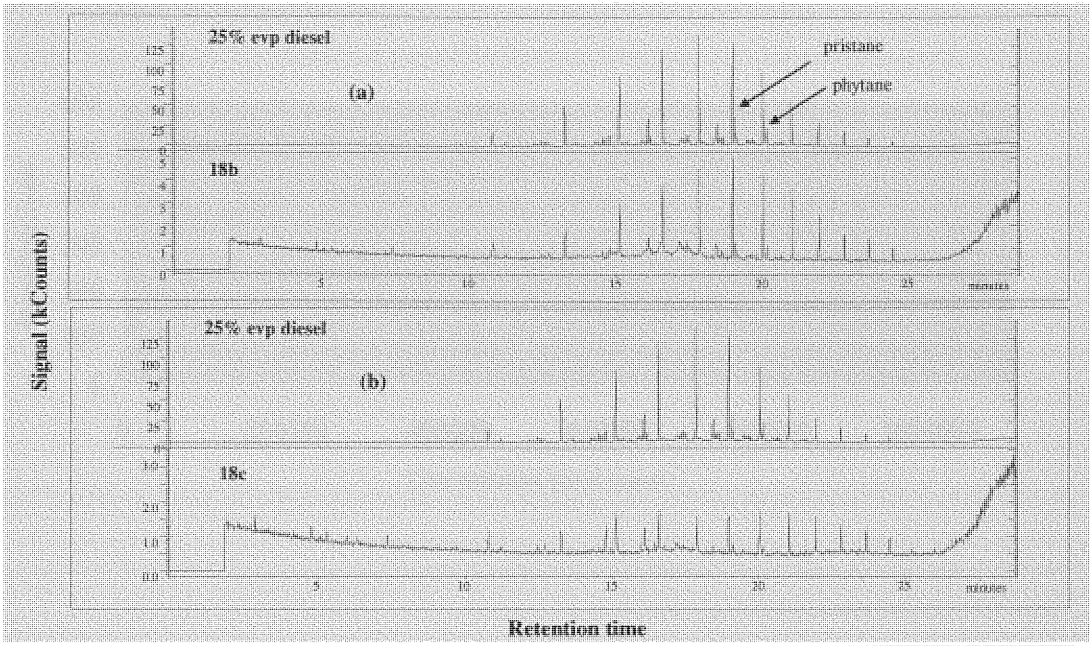


diesel fuel. All soils sampled in this study were tested for diesel fuel contamination. Detecting diesel fuel is a method of pattern recognition when comparing questioned samples to known standards, identifying target compounds by their retention times and mass spectra and recognizing the specific markers pristane and phytane (ASTM standard E-1618).

Diesel fuel was not detected in one gram sub-samples of the soils in this study. When five gram sub-samples were used, two Krome association samples taken during the wet season exhibited detectable levels of diesel fuel, samples "18 b" and "18 c" (Figure 11). Because diesel fuel is a complex mixture of alkanes, it is difficult to quantitate, however one can make inferences based on the signal recovered from known standards where the same instrument and extraction methods are used (28). The standard addition method was used as a diesel standard was diluted and extracted according to the same protocol used for the soils in this study (Figure 12). Based on signal to noise ratios for the target compound C_{17} recovered from serial dilutions of diesel, the estimated concentration of diesel fuel in samples 1 b and 1 c was calculated at 19.3 mg/g and 1.2 mg/g soil (log linear $R^2 = 0.915$).

Signal Recovery: Accounting for Matrix Effects. In order to further substantiate the reported concentration of diesel fuel in the soil samples where it was detected, it was important to determine whether signal recovery was affected by adsorption of diesel to the soil matrix from which samples "18 b" and "18 c" were collected. Parallel extractions were conducted after spiking various volumes of diesel fuel onto a Kimwipe™ as well as onto autoclaved soils from the same site where the

Figure 11. Diesel Detected in Two Krome Association Soils. Gas chromatographs of diesel range organic compounds extracted from two soil samples (5 g) of the Krome association as well as that of a 25 % evaporated diesel standard run on the same day, sample names are bold. Although the ratios of the alkanes have changed in the sample in panel (b), the pattern is still recognizable and the markers pristane and phytane are seen.

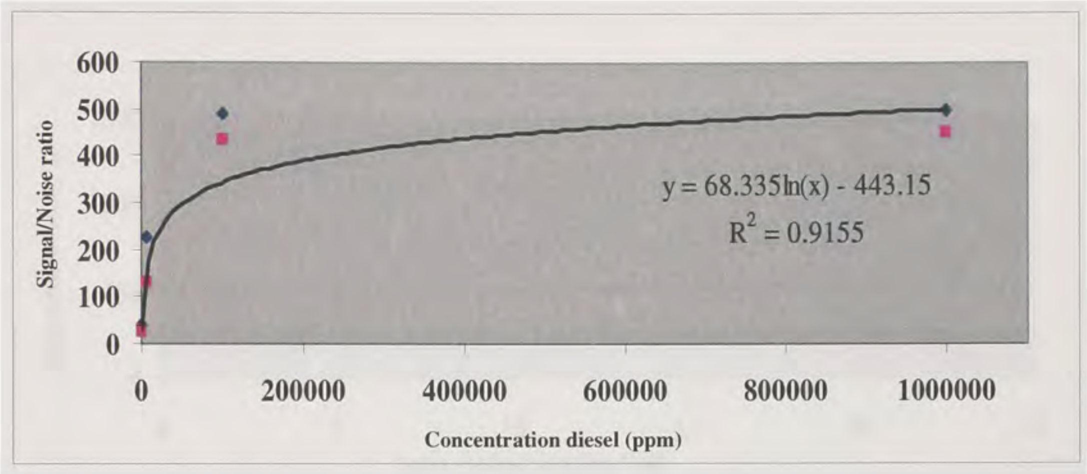


diesel fuel was detected. Signal recovery based on signal to noise ratios for C_{17} are graphically displayed in Figure 13. The calibration curves do not show a decrease in diesel recovery from the autoclaved soil compared to the Kimwipe™, especially in the portion of the curve where the concentration range falls for the two samples discussed previously.

Extraction of the Soil Metagenome. In order to develop a microbial community profile of the soils sampled, total soil DNA had to first be extracted. Duplicate extractions of total DNA were performed on each soil sample to account for any possible extraction bias and to determine the reproducibility of the methods chosen. Amplifyable

Figure 12. Graph of Signal Recovery from Diesel Spiked on Autoclaved Soil. Signal: noise ratios for target compounds C_{17} and C_{18} were graphed to determine the amount of diesel in the two soil samples from the Krome Association. The derived equation was based on signal recovery for C_{17}.

**□ - signal to noise ratio for C_{17}, ◇- signal to noise ratio for C_{18}



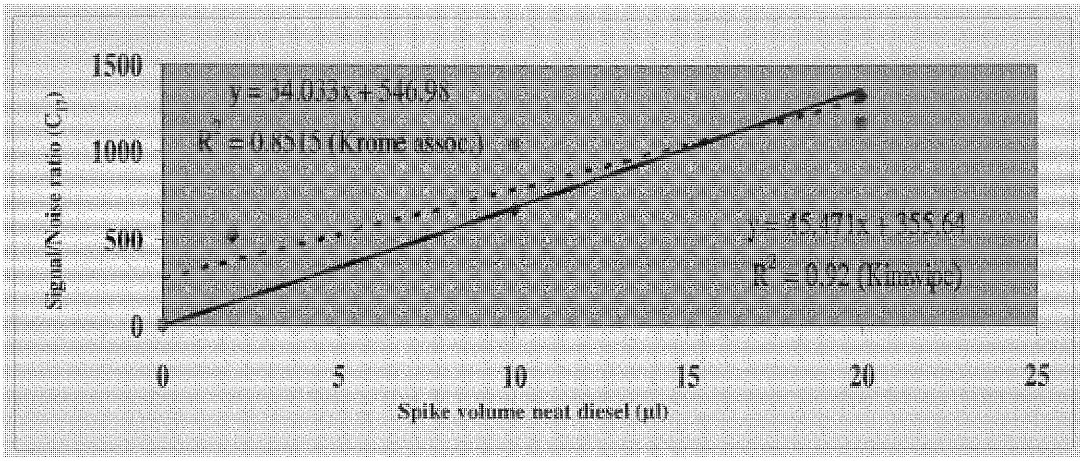
DNA was isolated from each soil sample taken demonstrating the robustness of the FastDNA™ Spin Sample Kit for Soil in removing inhibitors common to the soil matrix. When extractions were suspected to harbor PCR inhibitors as seen by low amplification products, a fresh aliquot of BSA in subsequent amplifications yielded PCR products consistently. During the extraction, addition of a heating step after initial lysis consistently resulted in a darker lysate and subsequently a darker final elution.

DNA Quantification. Total DNA was extracted from each soil sample. Hoescht dye was effective in estimating total DNA isolated from each soil. DNA concentrations

Figure 13. Soil Matrix Effects on Recovery of Diesel Standards. Graphical representation of effects of matrix adsorption when extracting diesel fuel standards spiked on a Kimwipe™ vs. diesel spiked on autoclaved soil from Krome association. Recovery of diesel was not lessened by adsorption of diesel to the soil matrix.

□ - signal to noise ratio (C_{17}) for Krome Association soil (autoclaved), ◇- signal to noise ratio (C_{17}) for Kimwipe

___ - Kimwipe, --- Krome Association soil (autoclaved)



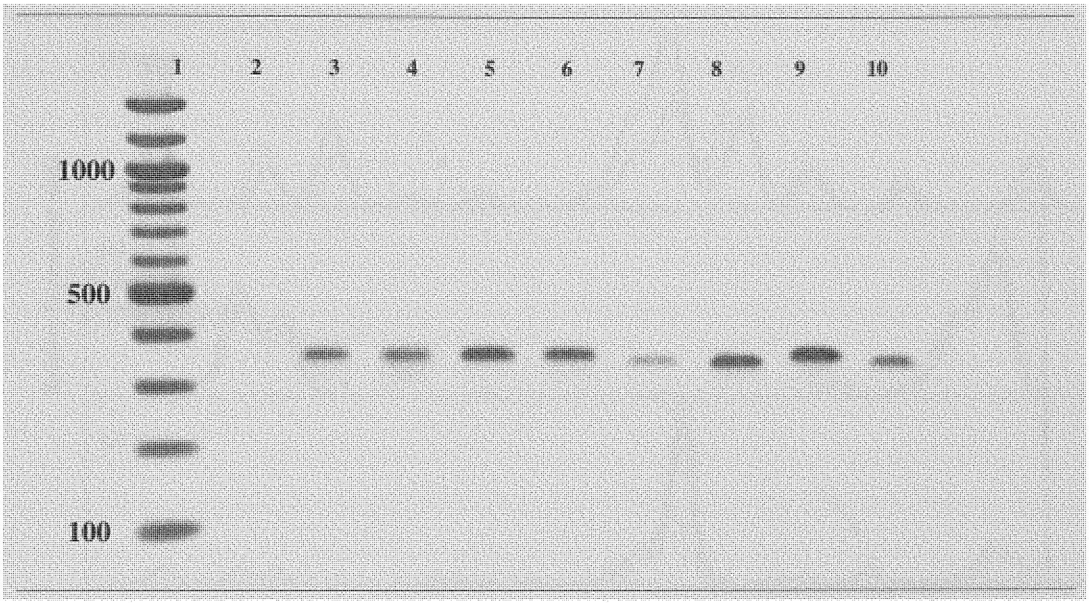
for the soil extractions generated in this study ranged from 2 to 353 ng/µl (final volume = 100 µl).

PCR Amplification of 16s rRNA Genes. Microbial community profiles were successfully generated from all soil DNA extracts generated in this study using PCR techniques. A common phenomenon encountered in DNA profiling using PCR-based techniques is the addition of a final, non-template Adenine to the 3' end of extension products. The temperature program for PCR was designed to encourage complete non-template addition by adenine. The 10 minute final extension step in the PCR thermocycling program was effective at eliminating –A products as evidenced by the lack of a consistent, less pronounced peak, 1 bp shorter than main amplicons. This was an important check in this study in order to not overestimate the diversity of the microbial communities profiled by producing and using both –A and +A amplicons which can occur especially in samples where an overabundance of template DNA is added (20). Any anomalous peaks generated by the PCR process were eliminated upon data review in downstream applications.

DNA Analysis by Agarose Gel Electrophoresis. Amplification products were effectively confirmed by electrophoresis in a 1 % agarose gel with a 100 bp DNA reference ladder. Amplifying the V1 range yielded amplicons in the 55 to 100 bp range, V3 amplicons ranged from ~ 169 to 200 bp and the combined range V1 + V2 yielded amplicons in the range of ~ 310 to 365 bp (Figure 14).

ALH Analysis of PCR Products. Effective length-based separation of PCR products was achieved using the ABI® Prism 310 high-resolution genetic analyzer. Amplicons differing in length by as little as one base-pair were discretely resolved as long as their intensity ranged from ~ 50 to 6,000 relative fluorescent units. Amplicon length heterogeneity-PCR was demonstrated to be a reproducible technique for all soils queried. Replicate amplifications of the two extractions from each soil sample produced highly similar profiles for each variable region (V1, V3 and V1 + V2) amplified based on visual comparison alone (Figure 15).

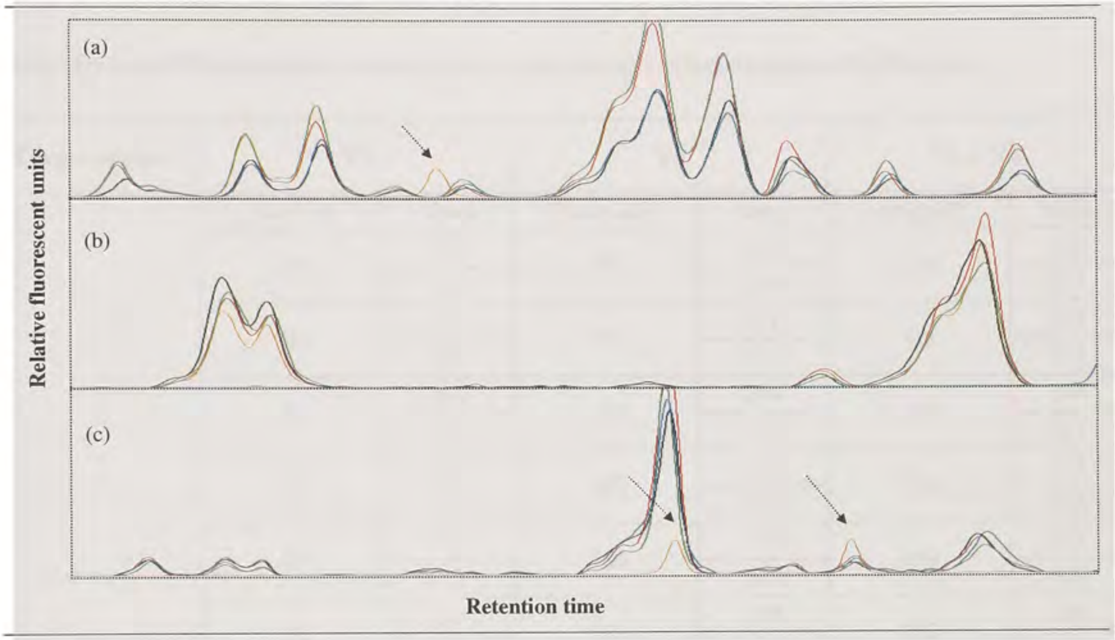
Figure 14. Amplification Products of the V1 + V2 Domain Confirmed by Agarose Gel Electrophoresis. Agarose (1.5 %) gel electrophoresis of bacterial isolates of Krome Association soils. Lane 1; 100 base pair ladder (Promega; Madison, WI), lane 2; sterile water, lanes 3 to 10; V1 + V2 domain of isolates from eight separate samples of Krome Association soils. Differences in length for some isolates are apparent.



Gene Scan Analysis of DNA Fragments. In all separations, migration through the capillary was checked for linearity by plotting peaks of known size against time of their detection, an automatic function of GeneScan®. All data used in subsequent analysis was confirmed to have a fit to curve R^2 value of at least 0.99 for the internal sizing standard. Fragment length and peak heights were imported into Genotyper®.

Exporting Fragment Data Using Genotyper®. Synonymous peaks in replicates called within a range of about 0.3 base pairs. The rounding function of Genotyper® caused some synonymous peaks to be sized one bp different, a function which must be used to reduce to volume of the exported table to manageable size. This function made it

Figure 15. Reproducibility of ALH Profiles of the Three Variable Domains. Each panel shows replicate amplifications from two extractions of the same soil sample separated on an ABI® Prism 310. The single orange peak in panel (a) and the two orange peaks in panel (c) belong to the internal size standard GeneScan™ 500 ROX (indicated by arrow). Panel shows the (a) V1, (b) V3, and (c) V1 + V2 domains, respectively. Each represent the total microbial community from this Rock outcrop-Biscayne-Chekika soil.



difficult to determine the true peak size for amplicons which presented this challenge. For this reason, a sizing key was used to group synonymous peaks of different called size into one amplicon size (Table 6). The sizing key was applied to all samples in this study.

Filtering of peaks to make the exported table more manageable was executed in Genotyper. For the V1 region, all amplicons less than 55 bp in length were eliminated prior to exportation from Genotyper. Peaks up to 155 bp in length were filtered out of the export table for amplifications of the V3 region and peaks up to 255 bp in length were filtered before table export in amplifications of V1 + V2.

Table 6. Sizing Key Used to Make Accurate and Global Amplicon Size Calls. The sizing key developed in this study was used to compensate for the rounding function in sizing DNA fragments using Genotyper®, the key retains the reproducibility of peaks within an amplification and ensures the most conservative estimate of diversity.

Target region	**V1**		**V3**		**V1 + V2**	
	final call	sized	final call	sized	final call	sized
	57	56	170	169	310	310
		57		170		311
	60	60	172	171	315	314
		61		172		315
	64	63	175	174	317	316
		64		175		317
	67	66	185	184	326	325
		67		185		326
	69	69	186	186	328	327
		70		187		328
	70	70	187	187	330	330
		71		188		331
	72	71	194	193	337	336
		72		194		337
	78	78	195	195	340	339
		79		196		340
	81	80			341	340
		81				341
	83	82			342	341
		83				342
	85	84			343	343
		85				344
	85	85			350	349
		86				350
	87	87			353	353
		88				354
	91	90			357	356
		91				358

Combining, Normalizing and Pruning Microbial Community Profile Data. Careful consideration was given to the method by which data could be combined from individual amplifications of variable regions without unduly biasing the contribution of each individual amplicon. Allowing each variable region an equal contribution to the overall profile overestimates the abundance of some low intensity amplicons in profiles with a low total RFU yield when the regions are combined and normalized.

After considering multiple options, the decision was made to combine variable region data sets by pasting the amplicon length and peak height data into a single row representing the amplicons derived from the same extraction. This is the truest depiction as it retains the empirical peak heights. Tables from each variable region containing pruned amplicon length and peak height were pasted together in this fashion and then normalized for the entire row. The final product was converted into a text file which was used in the SVM classifier.

Traditional Ecological Indices Comparisons Between Sites. Values for Richness, Evenness, Diversity and Hmax were generated from the microbial community profiles in order to determine whether one could differentiate between samples from different soil types using these traditional indices. Mean values obtained for each measure as well as the results of One-way ANOVA are displayed in Table 7. No significant differences ($p < 0.05$) were seen in Bonferroni post-hoc comparisons for Richness or for Hmax between soil types. Comparisons for diversity showed significant differences only for the comparison between Krome association and Lauderhill-Dania-Pahokee association soils. For evenness, significant differences were seen between Krome association and Rock outcrop-Biscayne-Chekika association soils as well as between Krome and Lauderhill-Dania-Pahokee soils.

Table 7. Ecological Indices for Richness, Evenness, Diversity and Hmax. Mean values and One-way ANOVA results for the ecological indices were derived from the combined microbial community profile data for each soil type. Means and standard deviations are shown. Values for sites followed by the same letter are not significantly different in Bonferroni post-hoc comparisons.

Soil type	Richness (S)	Diversity (H)	Evenness (E)	H max
Krome Association **n = 36**	22.19 ± (3.07)[a]	2.51 ± (0.15)[b]	0.81 ± (0.04)[d]	3.09 ± (0.14)[g]
Rock outcrop-Biscayne-Chekika **n = 18**	20.67 ± (2.17)[a]	2.53 ± (0.11)[b,c]	0.84 ± (0.03)[e,f]	3.02 ± (0.11)[g]
Lauderhill-Dania-Pahokee **n = 18**	21.89 ± (1.84)[a]	2.64 ± (0.13)[c]	0.86 ± (0.03)[f]	3.08 ± (0.09)[g]
ANOVA	$F = 2.11$ $P < 0.05$ $R^2 = 0.11$	$F = 5.40$ $P < 0.01$ $R^2 = 0.14$	$F = 10.81$ $P < 0.01$ $R^2 = 0.24$	$F = 1.96$ $P < 0.15$ $R^2 = 0.05$

Seasonal Comparisons of Traditional Ecological Indices Between Sites. When traditional ecological indices data from the soils were broken down by season, no significant differences were seen ($p < 0.05$) between the sites for measure of richness in either the wet or dry season (Table 8). Seasonal mean values for richness ranged between ~ 20 and 23 for all sites. Significant differences were seen in diversity for dry season comparisons between Krome and Rock outcrop-Biscayne-Chekika soils and between Krome and Lauderhill-Dania-Pahokee soils. Seasonal mean diversity measures ranged between 2.40 and 2.70 for all sites. Significant differences were seen in dry season comparisons for evenness between Krome and Rock outcrop-Biscayne-Chekika

Table 8. Seasonal Mean Values and One-way ANOVA Results for the Ecological Indices Richness, Evenness, Diversity and H_{max}. Values for ecological indices were derived from the combined microbial community profile data for each soil type and season. Means and standard deviations are shown. Values for sites followed by the same letter are not significantly different in Bonferroni post-hoc comparisons.

Soil type	Richness (*S*)	Diversity (*H*)	Evenness (*E*)	H_{max}
K dry **n = 18**	$20.72 \pm (2.35)^{a}$	$2.40 \pm (0.13)^{c,g}$	$0.81 \pm (0.03)^{h,k,j,l}$	$3.02 \pm (0.11)^{m}$
K wet **n = 18**	$23.67 \pm (3.05)^{b}$	$2.61 \pm (0.09)^{d,g}$	$0.82 \pm (0.04)^{h,i,j,l}$	$3.16 \pm (0.13)^{n}$
RBC dry **n = 9**	$20.44 \pm (2.30)^{a}$	$2.56 \pm (0.13)^{e,d,g}$	$0.85 \pm (0.03)^{i,k,j,l}$	$3.01 \pm (0.11)^{m}$
RBC wet **n = 9**	$20.89 \pm (2.15)^{a,b}$	$2.50 \pm (0.09)^{c,d}$	$0.82 \pm (0.03)^{h,k,i,j,l}$	$3.03 \pm (0.11)^{m,n}$
LDP dry **n = 9**	$20.78 \pm (1.30)^{a,b}$	$2.58 \pm (0.11)^{f,d,g}$	$0.85 \pm (0.03)^{j,k,i,l}$	$3.03 \pm (0.06)^{m,n}$
LDP wet **n = 9**	$23.00 \pm (1.66)^{a,b}$	$2.70 \pm (0.11)^{g,d}$	$0.86 \pm (0.03)^{k,i}$	$3.13 \pm (0.07)^{m,n}$
ANOVA	$F = 4.67$ $P < 0.05$ $R^2 = 0.26$	$F = 10.89$ $P < 0.01$ $R^2 = 0.45$	$F = 8.56$ $P < 0.01$ $R^2 = 0.39$	$F = 4.42$ $P < 0.01$ $R^2 = 0.25$

K- Krome association soil, R B C- Rock-outcrop-Biscayne-Chekika association soil, L D P- Lauderhill-Dania-Pahokee association soil.

soils and between Krome and Lauderhill-Dania-Pahokee soils. Significant differences were also seen in wet season comparisons between Krome and Lauderhill-Dania-Pahokee soils. The seasonal mean values for evenness ranged between 0.81 and 0.86 for all sites. No significant differences were seen in Hmax during the dry or wet season between any of the site comparisons. Seasonal Hmax values ranged between 3.02 and 3.16 for all sites sampled.

Seasonal Comparisons of Traditional Ecological Indices Within Sites. Values for richness increased for all three sites during the transition from the dry season to the wet season, however this change was only significant ($p < 0.05$) in Krome association soils. Krome and Lauderhill-Dania-Pahokee soils had a significant increase in their mean diversity values from dry season to wet season. No significant differences were seen in evenness for any site during the transition in seasons. Even though all sites demonstrated an increase in Hmax from the dry season to the wet season, the only significant change in mean Hmax values between seasons was observed in the Krome soils.

Based on the results of ANOVA, the most distinguishing index for pulling apart the soils based on the soil types in this study was evenness, which had the highest *F* statistic, the lowest *P* value and the highest R-squared value using soil type only labels. Using soil type and season labels, the diversity index was the most apt at distinguishing between the sites and seasons.

Within Soil type Differences in Ecological Indices. One-way ANOVA was used to determine whether significant differences existed for within site comparisons between the subplots (Table 9). No significant differences ($p < 0.05$) were seen between subplots of any site sampled during either season for any of the ecological indices data produced in this study.

Table 9. Subplots Comparisons Based on Ecological Indices for the Dry and Wet Seasons. Means and standard deviations are shown for each subplot within the soil types.

Dry season						
Site[a]	**Diversity**			**Evenness**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	2.27 (±0.14)	2.45 (±0.10)	2.32 (±0.95)	.79 (±0.03)	.83 (±0.04)	.78 (±0.03)
K_p	2.49 (±0.09)	2.39 (±0.12)	2.50 (±0.10)	.79 (±0.01)	.78 (±0.03)	.81 (±.05)
RBC	2.64 (±0.05)	2.51 (±0.13)	2.54 (±0.17)	.88 (±0.02)	.84 (±0.02)	.84 (±0.04)
LDP	2.45 (±0.09)	2.66 (±0.03)	2.61 (±0.08)	.82 (±0.04)	.86 (±0.01)	.86 (±0.01)
Site[a]	**Richness**			**Hmax**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	18.0 (±2.0)	19.0 (±0.0)	20.0 (±1.0)	2.89 (±0.12)	2.94 (±0.00)	2.99 (±0.05)
K_p	23.7 (±2.1)	21.3 (±1.2)	22.3 (±1.5)	3.16 (±0.09)	3.06 (±0.05)	3.10 (±.07)
RBC	20.3 (±0.6)	20.3 (±3.6)	20.7 (±2.9)	3.01 (±0.02)	3.00 (±0.18)	3.02 (±0.14)
LDP	20.0 (±1.0)	21.7 (±1.2)	20.7 (±1.5)	2.99 (±0.05)	3.07 (±0.06)	3.02 (±0.08)

Wet season						
Site[a]	**Diversity**			**Evenness**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	2.62 (±0.09)	2.71 (±0.03)	2.47 (±0.09)	.81 (±0.02)	.82 (±0.01)	.80 (±0.03)
K_p	2.58 (±0.04)	2.63 (±0.04)	2.67 (±0.04)	.84 (±0.04)	.86 (±0.01)	.85 (±0.01)
RBC	2.44 (±0.10)	2.57 (±0.03)	2.49 (±0.11)	.83 (±0.01)	.83 (±0.03)	.81 (±0.04)
LDP	2.74 (±0.10)	2.74 (±0.14)	2.60 (±0.02)	.88 (±0.04)	.87 (±0.03)	.84 (±0.02)
Site[a]	**Richness**			**Hmax**		
	Plot I	**Plot II**	**Plot III**	**Plot I**	**Plot II**	**Plot III**
K_d	26.0 (±1.7)	27.3 (±0.6)	22.0 (±2.6)	3.26 (±0.07)	3.31 (±0.02)	3.09 (±0.12)
K_p	22.0 (±4.4)	21.3 (±0.6)	23.3 (±2.1)	3.08 (±0.19)	3.06 (±0.03)	3.15 (±.09)
RBC	19.3 (±2.1)	22.0 (±2.0)	21.3 (±2.1)	2.96 (±0.11)	3.09 (±0.09)	3.06 (±0.10)
LDP	23.0 (±1.7)	24.0 (±2.0)	22.0 (±1.0)	3.13 (±0.08)	3.18 (±0.09)	3.09 (±0.05)

[a]Sites are represented by Krome-diesel (K_d), Krome-pristine (K_p), Rock-outcrop Biscayne Chekika (RBC) and Lauderhill Dania Pahokee (LDP).

***No significant differences between subplots based on $p < 0.05$ level.**

Supervised Classification Using Support Vector Machine Learning Tools. Microbial community profiles generated in this study were classified using a Support Vector Machine-based learning tool. Databases were created using a separate classifier for the Krome diesel site and the Krome pristine site. However, accuracy of the SVM classifier increased when these two sites were kept under the same label despite the unbalanced data set. For this reason, all subsequent data treatment was done with the Krome diesel and Krome pristine sites coupled together. All possible combinations of the variable regions were used in order to determine the most discriminating (accurate) data combinations and labeling schemes. Supervised classification was performed on all four replicates for each of the 72 soils sampled in this study for a total of 288 profiles. Classifications were executed again using labels for both soil type and season to determine whether adding seasonal labels would improve the accuracy of the classifier (Table 10). When samples were labeled by soil type only, V1 provided the most accurate classification compared to V3 and V1 + V2 regions. When samples were labeled for both soil type and season, classification accuracies to the correct soil type increased for all data combinations. Under these labeling conditions, the V1 + V2 region provided the most discriminating data for the SVM classifier with an accuracy of 91.7 %. Correctly classifying samples to both their soil type and season proved challenging for the SVM (data in parentheses) and resulted in much lower accuracies (lowest = 68.4 % for V1 region). The most accurate combination of data for the SVM classifier was the combined profile for all three regions profiled when samples were labeled with both their soil type and season yielding an accuracy of 96.9 %.

Table 10. Determining the Optimal Labeling Scheme and Data Combination for the SVM Classifier. Classification accuracies derived from SVM outputs using all possible combinations of data pertaining to variable region, chemical data and season (n=72). Prediction accuracies in parentheses indicate accuracy of the classification to both soil type and season.

Soil type only labels				Soil type and season labels			
Variable region(s)	**Linear**	**RBF**	**Sigmoid**	**Variable region(s)**	**Linear**	**RBF**	**Sigmoid**
V1	83.0	82.3	83.0	**V1**	87.9 (68.4)	90.6 (71.5)	87.9 (68.4)
V3	77.8	79.2	77.8	**V3**	90.6 (77.1)	91.7 (78.5)	90.6 (77.1)
V1 + V2	79.5	79.5	79.5	**V1 + V2**	90.6 (77.1)	91.7 (79.2)	90.6 (77.1)
V1 & V3	83.0	82.3	83.0	**V1 & V3**	87.9 (68.4)	90.6 (71.5)	87.9 (68.4)
V1 & V1 + V2	83.0	82.3	83.0	**V1 & V1 + V2**	87.9 (68.4)	90.6 (71.5)	87.9 (68.4)
V3 & V1 + V2	77.8	79.2	77.8	**V3 & V1 + V2**	90.6 (77.1)	91.7 (78.5)	90.6 (77.1)
V1, V3 & V1 + V2	92.4	92.0	92.4	**V1, V3 & V1 + V2**	96.9 (68.4)	96.5 (71.5)	96.9 (68.4)
				Elemental profile	41.7 (25.0)	45.8 (19.4)	41.7 (25.0)
				Elemental Profile + pH, H2O, % C, % N	73.6 (73.6)	76.4 (76.4)	73.6 (73.6)

SVM Classifications Using Physical and Chemical Data. Classifications based on the elemental profiles obtained by ICP-OES had much lower prediction accuracy (between 41.7 % and 45.8 %). Addition of the other chemical data (pH, moisture, % C and % N) drastically improved the accuracy of the classifier (45.8 % to 76.4 %). This data set was just as accurate at predicting both the correct site and season. However the accuracy was still approximately 20 % lower that the most accurate combination of microbial community profile data.

Distinguishing Between Subplots Using the SVM Classifier. Microbial community profiles were relabeled to determine whether the SVM classifier could accurately predict which subplot within a site a soil belonged to. Prediction accuracies were between 31.6 % and 40.0 %, considerably less than the SVM classifications to distinguish between soils of differing soil type (Table 11). The highest accuracy was for all three combined variable region data, consistent with the between site classifications. The large drop in accuracy indicates that the most discriminating labeling scheme for the database SVM classifier is the soil type and season labels.

Table 11. Supervised Classification of Soil Subplots Using SVM Learning Tool. Prediction accuracies for SVM classifiers based on soil subplots are listed for each variable region and for each possible combination of variable regions. Subplots were anywhere from 2 to 100 meters apart.

SVM Prediction accuracy (%) to Subplots			
Variable region(s)	**Linear**	**RBF**	**Sigmoid**
V1	38.5	38.9	38.5
V3	34.4	36.5	34.4
V1 + V2	34.4	34.0	34.4
V1 & V3	35.1	34.4	35.1
V1 & V1 + V2	31.6	34.0	31.6
V3 & V1 + V2	34.0	34.7	34.0
V1, V3 & V1 + V2	38.9	40.0	39.6

DISCUSSION

Soil examination is a challenging discipline of forensic science because there are numerous valid approaches one can undertake in order to develop a conclusion. Specific techniques are applied only when the nature of the soil evidence dictates their use. No single technique, however, is currently standard practice applied to all soil evidence (79). This poses a major challenge to forensic soil comparison because of the range of expertise examiners must possess in order to approach soil evidence in the laboratory (54).

Soil is ubiquitous in nature, a common form of evidence, and due to its complexity, can provide valuable clues to a case (91). Bacteria are ubiquitous in soil. Recent technological advances including PCR have allowed scientists to profile natural bacterial communities inhabiting soil based upon their DNA (64, 111). In recent years, DNA-based profiling of microbial communities has become a rapid and inexpensive process (107), requiring only two or three days to process multiple samples (~ 30), and has become quite routine, requiring relatively little training. The profiles derived from these methods can be used to distinguish between soils of interest (40, 76, 88, 113).

We set out to validate the use of ALH-PCR bacterial community profiling strictly in the forensic context and simultaneously compare it to some of the techniques which have been in use for years in soil science and forensic soil examination. We produced four replicate bacterial community profiles for each of three variable domains from 72 soil samples. We then evaluated them for their physical and chemical properties using validated techniques and performed statistical analyses to determine the levels of discrimination which can be achieved using these techniques. We also performed a supervised classification of our microbial community profiles and our soil physical/chemical profiles using highly sophisticated SVM-based learning tools capable of delineating highly dimensional data matrices like those obtained by ALH-PCR.

Soil Comparisons for Total Percent Carbon (C) and Percent Nitrogen (N)

Although significant differences were seen for some comparisons between soil type and between subplots within a site when data was labeled by season, no breakdown of the data resulted in significant differences being observed for all possible comparisons for soil percent C or for soil percent N. The C:N ratios generated in this study were ~ 20 for Krome and Rock-outcrop Biscayne Chekika soils and ~ 30 for Lauderhill Dania Pahokee soils. Reddy *et al* reported C:N ratios in wetland soils between 15 and 25 (87), Aitkenhead and McDowell compared soil C:N ratios and dissolved organic carbon (DOC) flux and observed C:N ratios of 20.97 for warm conifer forests and 32.40 for swamp forests, both of which were represented by Florida biomes(1). The C:N ratios produced in this study correspond well with their findings.

Soil Moisture and pH Comparisons

Moreno *et al* found no significant differences in moisture between soil types and therefore concluded that moisture content was not affecting changes in the microbial community (76). In this study, significant differences in soil moisture were seen for two of the three possible soil type comparisons as well as for wet/dry season comparisons within two of three soil types. All three soil types showed an increase in soil moisture in the dry season compared to the wet season. This seems counterintuitive, however, the soil moisture calculation takes into account a soils moist bulk density, which estimates the pore space available in a soil for water and for roots, and is not based on how much water has fallen during recent rains (http://soils.usda.gov). While significant differences in soil moisture did not demonstrate forensic value for this study, it leaves open the possibility that soil moisture may account for some differences in the microbial communities profiled.

No breakdown of soil pH data produced significant differences for all comparisons between or within soil types. Soil pH values in this study ranged between ~ 6.4 and 7.2, and when pH values were broken down by season they ranged from ~ 5.7 to 8.3. Moreno et al reported pH values between 6.4 and 7.6 for the soil types they surveyed (76). The USDA soil survey of Dade County reported pH ranges of 7.4 to 8.4 for Krome and Rock outcrop Biscayne Chekika soils and 5.6 to 7.8 for Lauderhill Dania Pahokee soils (http://soils.usda.gov). The pH values generated in this study for Krome and Rock outcrop Biscayne Chekika soils are generally about one pH unit lower than those reported by the USDA soil survey. The pH values for each season were all generated in one day. The only plausible explanation that can account for this discrepancy is experimental error.

Elemental Analysis of Soil

Of all the elements profiled, calcium had the highest concentration in the analyzed soils, consistent with Moreno *et al.* The calcium concentration accounted for about 25 % of the soils total composition, which was about ten times the concentration of iron, the next most abundant element. This was not unexpected because of the way this region of the Everglades Trough was formed when the underlying limestone dissolved, lowering the land below the water table. Limestone is composed mainly of calcium carbonate.

This study used 0.25 g samples of soil to analyze for elemental composition. The results of ANOVA showed no significant differences between or within soil types based on any of the elements queried. Relative standard deviations for the elements ranged between 46 and 84 %, reflecting the high degree of heterogeneity within the soil types sampled. This is difficult to see in the graphic because, in order to show the concentrations obtained for Ca, the elemental profiles had to be put on a logarithmic scale, making the standard error bars appear smaller compared to the means (Figure 10). Despite the high relative standard deviations, mean concentrations of each element were quite similar when comparing sites.

When elemental profiles were used as feature vectors in the SVM classifier, prediction accuracies ranged from 41.7 % to 45.8 %, which is roughly 10 % higher than one would expect by chance alone. It isn't possible to make any supported conclusion regarding this finding due to the high relative standard deviations observed for each element within a soil type. It is fair to say that the combination of the elements into feature vectors revealed enough of a pattern between the soil types to slightly increase the accuracy of the classifier. However, in a forensic context, neither ANOVA or SVM classification of the elemental profiles obtained from these soils proved useful in their distinction.

ALH-PCR is Robust, Reproducible and Reliable

In his book, *Forensic DNA Typing*, Butler uses the term robust when describing a method in which successful results are obtained a high percentage of the time and few samples, if any, need to be repeated (15). In this context, ALH-PCR certainly qualifies. Few samples from this study needed to be rerun and of those that needed re-analysis, most required an increased input of DNA or an increased addition of BSA.

According to Butler, "a reproducible method means that the same or very similar results are obtained each time a sample is tested" (15). Microbial community profiles were very reproducible using this method (Figures 15) both by visual comparison and by subsequent data analysis. This corroborates with findings from prior studies by Mills, Suzuki, and Ritchie (74, 88, 101).

Although some of the resolution of the ALH-PCR technique was lost due to sizing challenges (Table 3), once the sizing key was applied to the profiles, reproducibility of the amplicon length data was restored. Sizing may be improved by using an internal sizing standard with more size fragments or perhaps by using a more stringent polymer like POP-6 to obtain better resolution of peaks. Another idea to improve sizing might be accomplished by comparing amplicons to an molecularly dense human STR ladder with a different fluorophore and assigning each amplicon an allele/microvariant designation. If amplicons are designated this way, they would be more consistent in replicate analyses between laboratories. The allele designations could then be converted back into base-pair values for database purposes. An amplicon ladder could also be developed based on some of the more commonly seen microbes from soil ALH profiles and areas of the ALH profile which were the most difficult to resolve. This amplicon ladder could be labeled with a spectrally resolved fluorophore so as not to interfere with amplicons from the soil sample, thus giving the examiner an additional reference to make appropriate sizing calls. These options should be considered for future studies using this technique.

Butler describes a reliable method as "one in which the obtained results are accurate and correctly reflect the sample being tested" (15). Based on the resolution of the ABI® Prism 310 genetic analyzer (± 0.1 bp), we can say with some confidence that our data sets are accurate. There are ALH databases available for referencing amplicon sizes to all the bacteria capable of producing a fragment of the same length. However, the only way to truly test the reliability of data generated by this technique is through inter-laboratory validations using replicates of the same soils. This would be an important step toward validating this technique for forensic application.

A Potential Pitfall Associated with ALH-PCR at V1 + V2

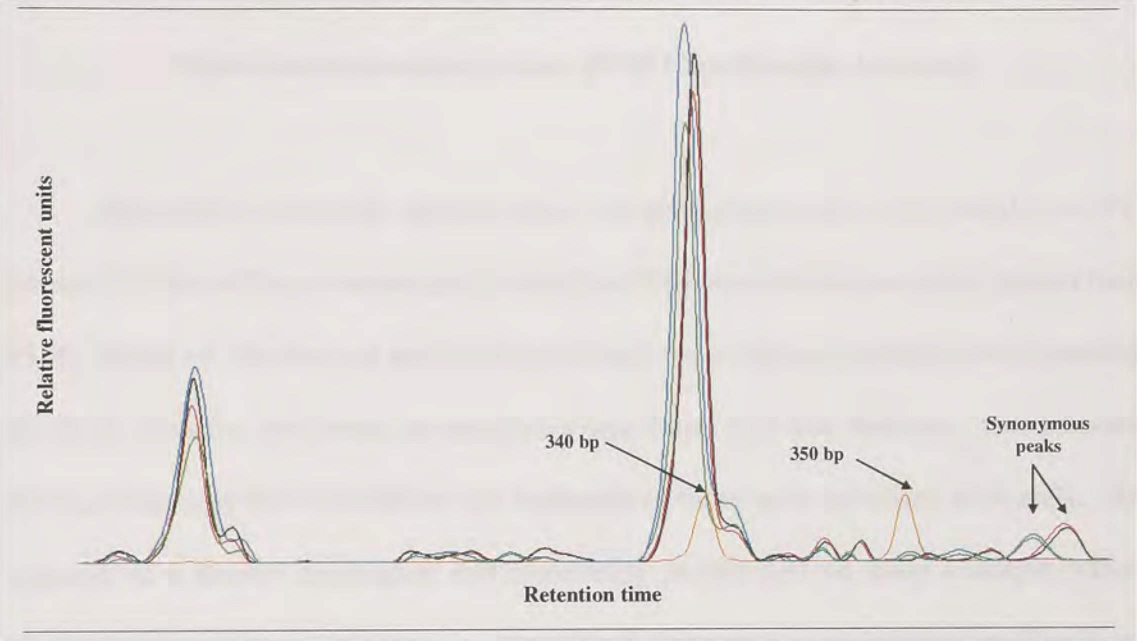
Although the profiles generated in this study were highly reproducible, a sizing issue did occasionally arise, primarily when analyzing the V1 + V2 region. Amplifying this region of the 16S rRNA gene produces DNA fragments between ~ 300 and 360 bp.

In some samples, some of the larger size fragments from analyses run on separate days did not co-migrate and fell out of the ± 0.5 bp window for calling synonymous peaks (Figure 16). When this occurred, samples were reanalyzed and found to be reproducible to a portion of the prior runs. This finding corroborates with the statement by Butler; "(i)n general, the greater the molecular weight of the PCR products, the larger the measurement error" (15). Based on this finding, this author suggests reference and questioned soil samples be analyzed on the same instrument and on the same day, in addition to duplicate analysis, in order to ensure that sizing is reproducible.

Traditional Ecological Indices Comparisons

We determined that significant individual differences ($p < 0.05$) did exist in limited comparisons between soil types for richness, Hmax, diversity and evenness after One-way ANOVA tests. Mills et al found that post-hoc comparisons of ecological indices indicated that while some comparisons of soil treatments were significantly different, no single index could distinguish all treatments (73). Results of this study corroborate their findings. ANOVA models explained only ~ 25 % of within group variance for both richness and Hmax. These values were higher for diversity (45 %) and evenness (39 %). We also observed that R-squared values increased in all cases when samples were labeled for soil type and season, suggesting that changes in the bacterial community profile correlated to seasonal variation as opposed to random variations. This finding is important because it supports findings in the biological characterizations which

Figure 16. Electropherogram of V1 + V2 Region where Migration of Larger DNA Fragments is Not Reproducible. Four replicate amplifications of the same soil sample separated on an ABI® Prism 310. The two orange peaks belong to the internal size standard GeneScanTM 500 ROX and their sizes are 340 and 350 bp respectively. Synonymous peaks co-migrated only up until ~ 340 bp for some samples analyzed on different days. Amplicons of larger size demonstrate higher measurement error as stated by Butler (14).



demonstrated that the microbial community profiles from each soil type were best distinguished when they were labeled separately for each season. Although these ecological indices showed significant differences in some comparisons between soil types, this finding would most likely not carry weight in a courtroom. This is mostly due to the fact that these are single point characteristics. Although these characteristics summarize a great deal of data from the microbial community profile, the fact that the values for each ecological index could easily be seen in a very different soil type doesn't make it a very powerful test in the discrimination of soils. By contrast, an ALH-PCR generated microbial community profile analyzed using an SVM-based machine learning tool retains multiple points of discrimination and each point has a semi-quantitative value. A comparison of this type in a forensic context is much more powerful and conclusive results would carry substantially more weight in a courtroom.

Diesel Contamination Lessens SVM Classification Accuracy

Microbial community profiles from variable regions V1, V3, and V1 + V2, labeled for both soil type and season, yielded an SVM classification accuracy greater than 95 %. Based on this data set and labeling scheme, we examined in detail how accurately the SVM classifier performed on samples where diesel fuel was detected. Our research revealed that only 66 % (37/56) of the replicates of those soils classified accurately. An example of a Krome association soil community profile derived from a sample where diesel fuel was detected is shown in Appendix I along with representative profiles from all three soil types surveyed. The lower accuracy suggests that matches from soils suspected to be contaminated should be interpreted with caution.

SVM Demonstrates Classification Based on Soil Type is Highly Accurate

Initially, microbial community profile data from "Krome pristine" and "Krome diesel" soils were labeled separately for the SVM classifier and the highest accuracy they generated was 94.4 %, using a combined profile derived from variable regions V1, V3, V1 + V2, and separate season labels. Under this scenario, nine of the sixteen misclassifications involved the SVM predicting a "Krome pristine" soil came from a "Krome diesel" soil and vice versa. When the Krome sites were combined, microbial community profiles from the combined variable regions labeled for both soil type and season, generated a classification accuracy up to 96.9 %, despite the unbalanced data set. This finding corroborates with the work done by Girvan *et al* which found that geographically distinct farms within the same soil type had almost identical community profiles despite the distance separating them (> 65 km) and different land use practices (47). It also shows that even within the same soil type, the SVM was able to find patterns that could distinguish two soils at a relative high sensitivity.

The high accuracy of the SVM classifier using the microbial community profiles is significant because these classifications were made using a DNA database. The accuracy of the SVM classifier is even more significant when you ponder the fact that this tool analyzes abundance data for approximately 30 data points simultaneously. The SVM algorithms have a C parameter which enable them to effectively ignore small numbers of extreme outliers which might otherwise confound the classifier. In addition, a potential forensic application of the SVM has recently been demonstrated by its accurate discrimination of closely related strains of *Bacillus anthracis* (27).

In conclusion, this study further validates the application of ALH-PCR in the forensic examination of soil. The technique provides a rapid, robust, reliable and reproducible avenue for producing conclusive soil comparisons. This technique can be performed on laboratory equipment already existing in any forensic biology lab and requires little training to perform. Analysis and interpretation of these profiles can easily become automated because the output is numerical (fragment length, abundance), just like that of human STR profiles. The most powerful possibility based on this study is the development of an ALH-PCR soil microbial community profile database from soils across the United States. A properly controlled database of this sort could provide accurate localizing of unknown soil samples from criminal investigations to small areas and aid in the cause of justice.

LITERATURE CITED

1. **Aitkenhead, J. A., and W. H. McDowell.** 2000. Soil C:N Ratio as a Predictor of Annual Riverine DOC Flux at Local and Global Scales. Global Biogeochemical Cycles **14:**127-138.

2. **Almirall, J., and K. Furton.** 2004. Analysis and Interpretation of Fire Scene Evidence. CRC Press, Boca Raton, FL.

3. **Arahal, D. R., F. E. Dewhirst, B. J. Paster, B. E. Volcani, and A. Ventosa.** 1996. Phylogenetic Analyses of Some Extremely Halophilic Archaea Isolated from Dead Sea Water, Determined on the Basis of Their 16S rRNA Sequences. Applied and Environmental Microbiology **62:**3779-3786.

4. **Bellar, T. A., and W. Buddle.** 1988. Determination of Nonvolatile Organic Compounds in Aqueous Environmental Samples Using Liquid Chromatography/Mass Spectrometry. Analytical Chemistry **60:**2076-2083.

5. **Bestwick, M. L., and E. O. Espinoza.** 2001. The Forensic Analysis of Soil by Raman Spectroscopy, p. 37-39, Northwest Association of Forensic Scientists, Bend, Oregon.

6. **Blackwood, C. B., T. Marsh, S. H. Kim, and E. A. Paul.** 2003. Terminal Restriction Fragment Length Polymorphism Data Analysis for Quantitative Comparison of Microbial Communities. Applied and Environmental Microbiology **69:**926-932.

7. **Boon, N., W. De Windt, W. Verstraete, and E. M. Top.** 2002. Evaluation of Nested PCR-DGGE (Denaturing Gradient Gel Electrophoresis) with Group-specific 16S rRNA Primers for the Analysis of Bacterial Communities from Different Wastewater Treatment Plants. FEMS Microbiology Ecology **39:**101-112.

8. **Boyd, S.** April 18, 2000, posting date. Adolph Coors, III Murder Investigations Records 1960. Douglas County Sheriff's Office. [Online.]

9. **Brian, L.** 1992. The Encyclopedia of Forensic Science. Headline Book Publishing, London, U. K.

10. **Brown, M. P. S., W. N. Grundy, D. Lin, N. Christianini, C. Sugnet, T. S. Furey, M. Ares, and D. Haussler.** 2000. Knowledge-based Analysis of Microarray Gene Expression Data Using Support Vector Machines. Proceedings of the National Academy of Sciences of the United States of America, **97:**262-267.

11. **Budowle, B., B. Shea, S. Niezgoda, and R. Chakraborty.** 2001. CODIS STR Loci from 41 Sample Populations. Journal of Forensic Sciences **46:**453-489.

12. **Budowle, B., J. Smith, T. Moretti, and J. DiZinno.** 2000. DNA Typing Protocols: Molecular Biology and Forensic Analysis. Eaton Publishing, Natick, MA.

13. **Budowle, B., T. R. Moretti, S. J. Niezgoda, and B. L. Brown.** 1998. CODIS and PCR-based Short Tandem Repeat loci: Law Enforcement Tools. Presented at the Second European Symposium on Human Identification:73-88.

14. **Buel, E., M. B. Schwartz, and M. J. LaFountain.** 1998. Capillary Electrophoresis STR Analysis: Comparison to Gel-based Systems. Journal of Forensic Sciences **46:**164-170.

15. **Butler, J. M.** 2005. Forensic DNA Typing; Biology, Technology and Genetics of STR Markers, 2nd ed. Elsevier, Burlington, MA.

16. **Casamayor, E. O., C. Pedros-Alio, G. Muyzer, and R. Amann.** 2002. Microheterogeneity in 16S Ribosomal DNA-Defined Bacterial Populations from a Stratified Planktonic Environment Is Related to Temporal Changes and to Ecological Adaptations. Applied and Environmental Microbiology **68:**1706-1714.

17. **Cengiz, S., A. C. Karaca, I. Cakir, H. B. Uner, and A. Sevindik.** 2004. SEM-EDS Analysis and Discrimination of Forensic Soil. Forensic Science International **141:**33-37.

18. **Cengiz, S., and O. Sakul.** 2001. Capillary Electrophoresis in Forensic Soil Analysis. Trace Elements and Electrocytes **18:**87-91.

19. **Chazottes, V., C. Brocard, and B. Peyrot.** 2004. Particle Size Analysis of Soils Under Simulated Scene of Crime Conditions: the Interest of Multivariate Analyses. Forensic Science International **140:**159-166.

20. **Clark, J. M.** 1988. Novel Non-templated Nucleotide Addition Reactions Catalyzed by Procaryotic and Eucaryotic DNA Polymerases. Nucleic Acids Research **16:**9677-9686.

21. **Clement, B. G., L. M. Nicholas, and C. L. Kitts.** 1998. Presented at the The International Symposium on Microbial Ecology, Halifax, Nova Scotia, Canada.

22. **Cocolin, L., M. Manzano, C. Cantoni, and G. Comi.** 2001. Denaturing Gradient Gel Electrophoresis Analysis of the 16S rRNA Gene V1 Region to Monitor Dynamic Changes in the Bacterial Population Durng Fermentation of Italian Sausages. Applied and Environmental Microbiology **67:**5113-5121.

23. **Cox, R. J., H. L. Peterson, J. Young, C. Cusik, and E. O. Espinoza.** 2000. The Forensic Analysis of Soil Organic by FTIR. Forensic Science International **108:**107-116.

24. **Cristianini, N., and J. Shawe-Taylor.** 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, New York, N. Y.

25. **Crosby, L. D., and C. S. Criddle.** 2003. Understanding Bias in Microbial Community Analysis Techniques due to *rrn* Operon Copy Number Heterogeneity. BioTechniques **34:**1-9.

26. **Dixon, W. C.** 1979. Applications of Optical Microscopy in the Analysis of Asbestos and Quartz, ACS Symposium Series, No. 120, Analytical Techniques in Occupational Health Chemistry.

27. **Doran, M., D. S. Raicu, J. D. Furst, R. Settimi, R. Schipma, and D. P. Chandler.** 2007. Oligonucleotide Microarray Identification of *Bacillus anthracis* Strains Using Support Vector Machines. Bioinformatics **23:**487-492.

28. **Douglas, G. S., W. A. Burns, A. E. Bence, D. S. Page, and P. Boehm.** 2004. Optimizing Detection Limits for the Analysis of Petroleum Hydrocarbons in Complex Environmental Samples. Environmental Science and Technology **38:**3958-3964.

29. **Doyle, A. C.** 1956. The Complete Sherlock Holmes, vol. 1. Doubleday, New York, N. Y.

30. **Dudley, R. J.** 1977. The Particle Size Analysis of Soils and Its Use in Forensic Science. The Determination of Particle Size Distributions Within the Silt and Sand fractions. Journal of Forensic Science Society:219-229.

31. **Dudley, R. J.** 1975. The Use of Colour in the Discrimination Between Soils. Journal of Forensic Science Society **15:**209-218.

32. **Duineveld, B. M., A. S. Rosado, J. D. van Elsas, and J. A. van Veen.** 1998. Analysis of the Dynamics of Bacterial Communities in the Rhizosphere of the Chrysanthemum via Denaturing Gradient Gel Electrophoresis and Substrate Utilization Patterns. Applied and Environmental Microbiology **64:**4950-4957.

33. **Dunbar, J., S. Takala, S. M. Barns, J. A. Davis, and C. R. Kuske.** 1999. Levels of Bacterial Community Diversity in Four Arid Soils Compared by Cultivation and 16S rRNA Gene Cloning. Applied and Environmental Microbiology **65:**1662-1669.

34. **Dunbar, J., L. O. Ticknor, and C. R. Kuske.** 2001. Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction

Fragment Profiles of 16S rRNA Genes from Bacterial Communities. Applied and Environmental Microbiology **67:**190-197.

35. **Edwards, A., A. Civitello, H. A. Hammond, and C. T. Caskey.** 1991. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. American Journal of Human Genetics **49:**746-756.

36. **Emmons, R. C.** 1929. The Double Variation Method of Refractive Index Determination. American Mineralogist **14:**414-426.

37. **F.B.I.** 2004. CODIS: The Combined DNA Indexing System, Quantico, VA.

38. **Fitzpatrick, R. W.** 2004. Presented at the SuperSoil 2004, Sydney, Australia.

39. **Franciosco, O., C. Ciavatta, S. Sanchez-Cortes, V. Tugnoli, L. Sitti, and C. Gessa.** 2000. Spectroscopic Characterization of Soil Organic Matter in Long-Term Amendment Trials. Soil Science **165:**495-504.

40. **Franklin, R. B., and A. L. Mills.** 2003. Multi-scale Variation in Spatial Heterogeneity for Microbial Community Structure in an Eastern Virginia Agricultural Field. FEMS Microbiology Ecology **44:**335-346.

41. **Furey, T., N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, and D. Haussler.** 2000. Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data. Bioinformatics **16:**906-914.

42. **Garton, S.** 1993. "Feed Him to the Sharks", vol. 12. Geelong, Victoria, British Colombia, Canada.

43. **Giannelli, P.** 1993. "Junk Science: The Criminal Cases". Journal of Criminal Law and Criminology **84:**105-128.

44. **Gill, P.** 2002. Role of Short Tandem Repeat DNA in Forensic Casework in the UK, Past, Present and Future Perspectives. BioTechniques **32:**366-385.

45. **Gill, P., C. P. Kimpton, A. Urquhart, N. Oldroyd, E. S. Millican, S. K. Watson, and T. J. Downes.** 1995. Automated Short Tandem Repeat (STR) Analysis in Forensic Casework--a Strategy for the Future. Electrophoresis **16:**1543-1552.

46. **Giovannoni, S. J., T.B. Britschgi, C.L. Moyer, and K.G. Field.** 1990. Genetic Diversity in Sargasso Sea Bacterioplankton. Nature **345:**60-63.

47. **Girvan, M., J. Bullimore, J. Pretty, A. M. Osborn and A. S. Ball.** 2003. Soil Type is the Primary Determinant in the Composition of the Total and Active Bacterial Communities in Arable Soils. Applied and Environmental Microbiology **69:**1800-1809.

48. **Grassberger, R.** 1956. Pioneers in Criminology. XIII. Hans Gross (1847-1915). Journal of Criminal Law, Criminology and Police Science **47:**397-405.

49. **Grob, R. L., and E. F. Barry.** 2004. Modern Practice of Gas Chromatography. John Wiley and Sons Inc., Hoboken, N. J.

50. **Gross, H.** 1907. Handbook for Examining Magistrates as a System of Criminology. Calcutta, New York, N. Y.

51. **Hammond, H. A., L. Jin, Y. Zhong, C. T. Caskey and R. Chakraborty.** 1994. Evaluation of 13 Short Tandem Repeat Loci for Use in Personal Identification Applications. American Journal of Human Genetics **55:**175-189.

52. **Heiseley, B., T. Serre, M. Pontil, and T. Poggio.** 2001. Component-based Face Detection. Presented at the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition:657-662.

53. **Hope, M.** June 2004. Density Gradient Gel Electrophoresis, p. 38-39, Microbiologist.

54. **Horswell, J., S. J. Cordiner, E. W. Maas, T. M. Martin, K. B. W. Sutherland, T. Speir, B. Nogales, and A. M. Osborn.** 2002. Forensic Comparison of Soils by Bacterial Community DNA Profiling. Journal of Forensic Sciences **47:**350-353.

55. **Isphording, W.** 2005. Trial by Science or Trial by Jury? Justice or Mercy? Presented at The Geological Society of America-Southeastern Section, 54th Annual Meeting **37:**47.

56. **Iwanmoto, T., K. Tani, K. Nakamura, Y. Suzuki, M. Kitagawa, M. Eguchi, and M. Nasu.** 2000. Monitoring Impact of *in situ* Biostimulation Treatment on Groundwater Bacterial Community by DGGE. FEMS Microbiology Ecology:129-141.

57. **Jeffreys, A. J., V. Wilson, and T. S. L.** 1985. Hypervariable 'Minisatellite' Regions in Human DNA. Nature **314:**67-73.

58. **Jeffreys, A. J., V. Wilson, and S. L. Their.** 1985. Individual-specific 'Fingerprints' of Human DNA. Nature **316:**76-79.

59. **Joachims, T.** 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Presented at The 10th European Conference on Machine Learning:137-142.

60. **Jones, P.** 2004. DNA Forensics: from RFLP to PCR-STR and Beyond. Forensic Magazine **1:**14-17.

61. **Junger, E. P.** 1996. Assessing the Unique Characteristics of Close-proximity Soil Samples: Just How Useful is Soil Evidence? Journal of Forensic Sciences **41:**27-34.

62. **Kantha, S. S.** 1995. Is Karl Landsteiner the Einstein of the Biomedical Sciences? Medical Hypotheses **44:**254-256.

63. **Leung, K., and E. Topp.** 2001. Bacterial Community Dynamics in Liquid Swine Manure During Storage: Molecular Analysis Using DGGE/PCR of 16S rDNA. FEMS Microbiology Ecology **38:**169-177.

64. **Liesack, W., and E. Stackebrandt.** 1992. Occurence of Novel Groups of the Domain Bacteria as Revealed by Analysis of Genetic Material Isolated from an Australian Terrestrial Environment. Journal of Bacteriology **174:**5072-5078.

65. **Litchfield, C. D., G. Gromicko, L. E. Dansey, and J. E. G. Minkley.** 1992. Influence of Soil Type on the Microbial Degradation of PCP and PAHs., p. 275-292. *In* J. S. Srivastava, and D. Hayes (ed.), Gas, Oil, and Environmental Biotechnology VI. Institute of Gas Technology, Des Plaines, IL.

66. **Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1997. Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. Applied and Environmental Microbiology **63:**4516-4522.

67. **Marsh, T. L., P. Saxman, J. Cole, and J. Tiedje.** 2000. Terminal Restriction Fragment Length Polymorphism Analysis Program, a Web-Based Research Tool for Microbial Community Analysis. Applied and Environmental Microbiology **66:**3616-3620.

68. **Massol-Deya, A. A., D. A. Odelson, R. F. Hickey, and J. M. Tiedje.** 1995. Bacterial Community Fingerprinting of Amplified 16S and 16-23S Ribosomal DNA Gene Sequences and Restriction Endonuclease Analysis (ARDRA), p. 3.3.1:1-8. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), Molecular FEMS Microbial Ecology Manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.

69. **McPhee, J.** The Gravel Page, p. 44-52, The New Yorker, vol. 71.

70. **McVicar, M. J., and W. J. Graves.** 1997. The Forensic Comparison of Soils by Automated Scanning Electron Microscopy. Journal of the Canadian Society of Forensic Sciences **30:**241-261.

71. **Micklos, D. A., G. A. Freyer.** 2003. DNA Science: A First Course, 2nd ed. CSHL Press, Woodbury, NY.

72. **Mills, D. K.** 2000. Molecular Monitoring of Microbial Populations during Bioremediation of Contaminated Soils. Ph.D. dissertation. George Mason University, Fairfax, VA.

73. **Mills, D. K., J. Entry, J. Voss, P. Gillevet, and K. Mathee.** 2006. An Assessment of the Hypervariable Domains of the 16S rRNA Genes for their Value in Determining Microbial Community Diversity: the Paradox of Traditional Ecological Indices. FEMS Microbial Ecology **57:**496-503.

74. **Mills, D. K., K. Fitzgerald, C. D. Litchfield, and P. M. Gillevet.** 2003. A Comparison of DNA Profiling Techniques for Monitoring Nutrient Impact on Microbial Community Composition During Bioremediation of Petroleum Contaminated Soils. Journal of Micrbiological Methods **54:**57-74.

75. **Mills, D. K., S. King, S. Miller, and K. Mathee.** 2003. Personal communication.

76. **Moreno, L., D. K. Mills, J. Entry, R. Sautter, and K. Mathee.** 2006. Microbial Metagenome Profiling Using Amplicon Length Heterogeneity-Polymerase Chain Reaction Proves More Effective Than Elemental Analysis in Discriminating Soil Specimens. Journal of Forensic Science **51:**1315-1322.

77. **Mullis, K. B.** 1986. Specific Enzymatic Amplification of DNA *in vitro*: the Polymerase Chain Reaction. Quantitative Biology **51:**263-273.

78. **Mullis, K. B.** 1990. The Unusual Origin of the Polymerase Chain Reaction. Scientific American **262:**56-65.

79. **Murray, R. C.** 2004. Evidence from the Earth: Forensic Geology and Criminal Investigation. Mountain Press, Missoula, MT.

80. **Murray, R. C., and J. C. F. Tedrow.** 1975. Forensic Geology. Rutgers University Press, Rutgers, N.J.

81. **Muyzer, G., E. C. DeWaal, and A. G. Uitterlinden.** 1993. Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S rRNA. Applied and Environmental Microbiology **59:**695-700.

82. **Nazarenko, I. A., S. K. Bhatnagar, and R. J. Hohman.** 1997. A Closed Tube Format for Amplification and Detection of DNA Based on Energy Transfer. Nucleic Acids Research **25:**2516-2521.

83. **Olive, D. M.** 1999. Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. Journal of Clinical Microbiology **37:**1661-1669.

84. **Petraco, N., and T. Kubic.** 2000. A Density Gradient Technique for Use in Forensic Soil Analysis. Journal of Forensic Sciences **45:**872-873.

85. **Pye, K., and S. Blott.** 2004. Particle Size Analysis of Sediments, Soils and Related Particulate Materials for Forensic Purposes Using Laser Granulometry. Forensic Science International **144:**19-27.

86. **Pye, K., S. J. Blott, and D. S. Wray.** 2005. Elemental Analysis of Soil Samples for Forensic Purposes by Inductively Coupled Plasma Spectrometry-Precision Considerations. Forensic Science International **160:**178-192.

87. **Reddy, K. R., J. R. White, A. L. Wright, and T. Chua.** 1999. Influence of Phosphorus Loading on Microbial Processes in the Soil and Water Column of Wetlands, p. 256, Cell Growth: Control of Cell Size. Cold Spring Harbor Laboratory Press, New York, N.Y.

88. **Ritchie, N. J., M. E. Schutter, R. P. Dick, and D. D. Myrold.** 2000. Use of Length Heterogeneity PCR and Fatty Acid Methyl Ester Profiles to Characterize Microbial Communities in Soil. Applied and Environmental Microbiology **66:**1668-1675.

89. **Rozsak, D. B., and R. R. Colwell.** 1987. Survival Strategies of Bacteria in the Natural Environment. Microbiological Reviews **51:**365-379.

90. **Ruffel, A., and P. Wiltshire.** 2004. Conjunctive Use of Quantitative and Qualitative X-ray Diffraction Analysis of Soils and Rocks for Forensic Analysis. Forensic Science International **145:**13-23.

91. **Saferstein, R.** 2002. Forensic Science Handbook, 2nd ed. Prentice Hall, Upper Saddle River, N.J.

92. **Saferstein, R.** 1995. Physical Properties: Glass and Soil, p. 101-103, Criminalistics: An Introduction to Forensic Science. Prentice Hall, Upper Saddle River, N. J.

93. **Sanders, N.** 1983. Crimes of Passion: TV, Popular Literature and the Graeme Thorne Kidnapping, 1960. Australian Journal of Cultural Studies **1:**56-59.

94. **Scala, D. J., and L. J. Kerkhof.** 2000. Horizontal Heterogeneity of Denitrifying Bacterial Communities in Marine Sediments by Terminal Restriction Fragment Length Polymorphism Analysis. Applied and Environmental Microbiology **66:**1980-1986.

95. **Schloter, M., O. Dilly, and J. C. Munch.** 2003. Indicators for Evaluating Soil Quality. Agriculture, Ecosystems and Environment **98:**255-262.

96. **Siegel, J. A., and C. Precord.** 1985. The Analysis of Soil Samples by Reverse Phase High-performance Liquid Chromatography Using Wavelength Ratioing. Journal of Forensic Sciences **30:**511.

97. **Simard, L. S., and W. G. Young.** 1994. Dauberts Gatekeeper: The Role of the District Judge in Admitting Expert Testimony, Tulane Law Review.

98. **Staats, J.** 1970. Bibliography of George D. Snell. Transplantation Proceedings **2:**174-185.

99. **Sugita, R., and Y. Marumo.** 2001. Screening of Soil Evidence by a Combination of Simple Techniques: Validity of Particle Size Distribution. Forensic Science International **122:**155-158.

100. **Sugita, R., and Y. Marumo.** 1996. Validity of Color Examination for Forensic Soil Identification. Forensic Science International **83:**201-210.

101. **Suzuki, M., M. S. Rappe, and S. J. Giovannoni.** 1998. Kinetic Bias in Estimates of Coastal Picoplankton Community Structure Obtained by Measurements of Small-Subunit rRNA Gene PCR Amplicon Length Heterogeneity. Applied and Environmental Microbiology **64:**4522-4529.

102. **Suzuki, M. T., M. S. Rappé, Z. W. Haimberger, H. Winfield, N. Adair, J. Stöbel, and J. G. Stephen.** 1997. Bacterial Diversity among Small-Subunit rRNA Gene Clones and Cellular Isolates from the Same Seawater Sample. Applied and Environmental Microbiology **63:**983-989.

103. **SWGMAT.** 2005. Elemental Analysis of Glass, p. 84, Forensic Science Communications, vol. 7.

104. **Thornton, J. I.** 1997. The General Assumptions and Rationale of Forensic Identification. *In* D. K. D. Faigman, M. Saks and Joseph Sanders (ed.), Modern Scientific Evidence: The Law and Science of Expert Testimony, vol. 2. West Publishing Co., St. Paul, MN.

105. **Thornton, J. I., and A. D. McLaren.** 1975. Enzymatic Characterization of Soil Evidence. Journal of Forensic Sciences **20:**693-700.

106. **Tiirola, M. A., J. E. Suvilampi, M. S. Kulomaa, and J. A. Rintala.** 2003. Microbial Diversity in a Thermophilic Aerobic Biofilm Process: Analysis by Length Heterogeneity PCR (LH-PCR). Water Research **37:**2259-2269.

107. **Torsvik, V., R. Sorheim, and J. Goksoyr.** 1996. Total Bacterial Diversity in Soil and Sediment Communities-a Review. Journal of Industrial Microbiology **17:**170-178.

108. **Van de Peer, Y., S. Chapelle, and R. De Wachter.** 1996. A Quantitative Map of Nucleotide Substitution Rates in Bacterial rRNA. Nucleic Acids Research **24:**3381-3391.

109. **van Elsas, J. D., G. F. Duarte, A. Keijzer-Wolters, and E. Smit.** 2000. Analysis of the Dynamics of Fungal Communities in Soil via Fungal-specific PCR of Soil DNA Followed by Denaturing Gradient Gel Electrophoresis. Journal of Micrbiological Methods **43:**133-151.

110. **Wanogho, S., G. Gettinby, and B. Caddy.** 1987. Particle Size Distribution Analysis of Soils Using Laser Diffraction. Forensic Science International **33:**117-128.

111. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA Sequences Reveal Numerous Uncultured Microorganisms in a Natural Community. Nature **345:**63-65.

112. **Webb, J. A.** 1995. Crime Scene Soil Samples Get Close Scrutiny from Museum Scientists. Smithsonian Institute Research Reports **82:**51-53.

113. **Yang, C., D. K. Mills, K. Mathee, Y. Wang, K. Jayachandran, M. Sikaroodi, P. Gillevet, J. A. Entry, and G. Narasimhan.** 2005. An Ecoinformatics Tool for Microbial Community Studies: Supervised Classification of Amplicon Length Heterogeneity (ALH) Profiles of 16S rRNA. Journal of Microbiological Methods **65:**49-62.

114. **Yang, Y.-H., J. Yao, S. Hu, and Y. Qi.** 2000. Effects of Agricultural Chemicals on DNA Sequence Diversity of Soil Microbial Community: A Study with RAPD Marker. Microbial Ecology **39:**72-79.

APPENDIX I

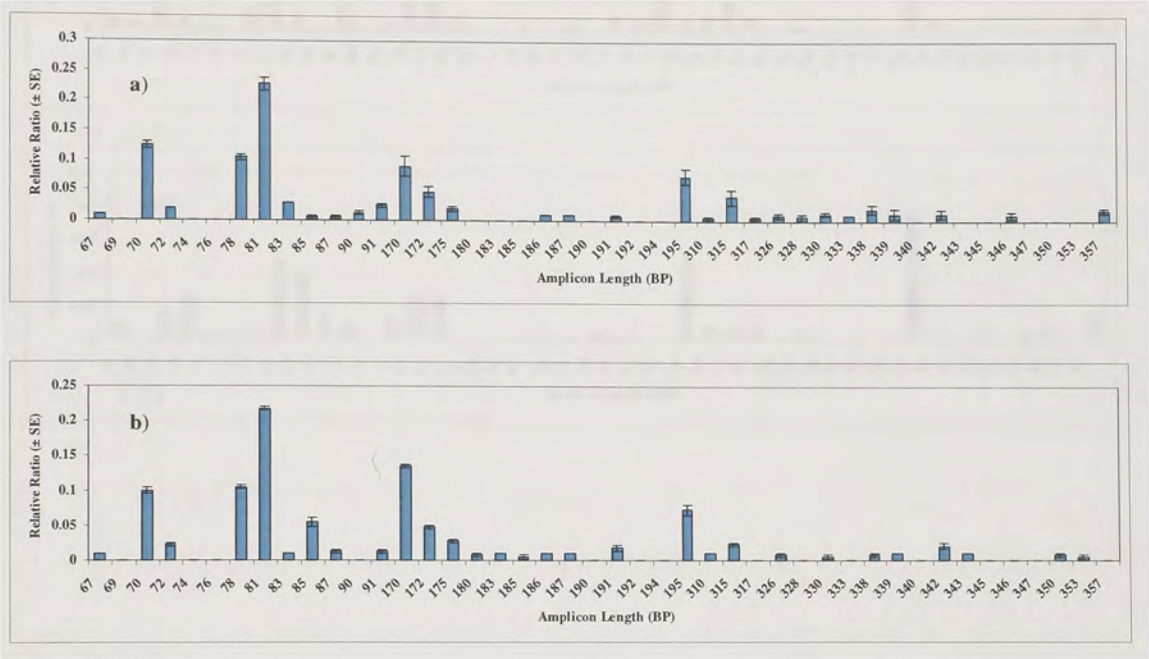
Not unlike many of the disciplines in forensic science, forensic soil analysis is at its foundation, a comparative science. Typically, a questioned soil sample such as one collected from a suspect's clothing or vehicle is compared to soils collected from areas associated with the crime scene in addition perhaps to an area linked to the suspect's alibi. The goal is to be able to determine with some degree of certainty whether the soils under comparison could have a common origin.

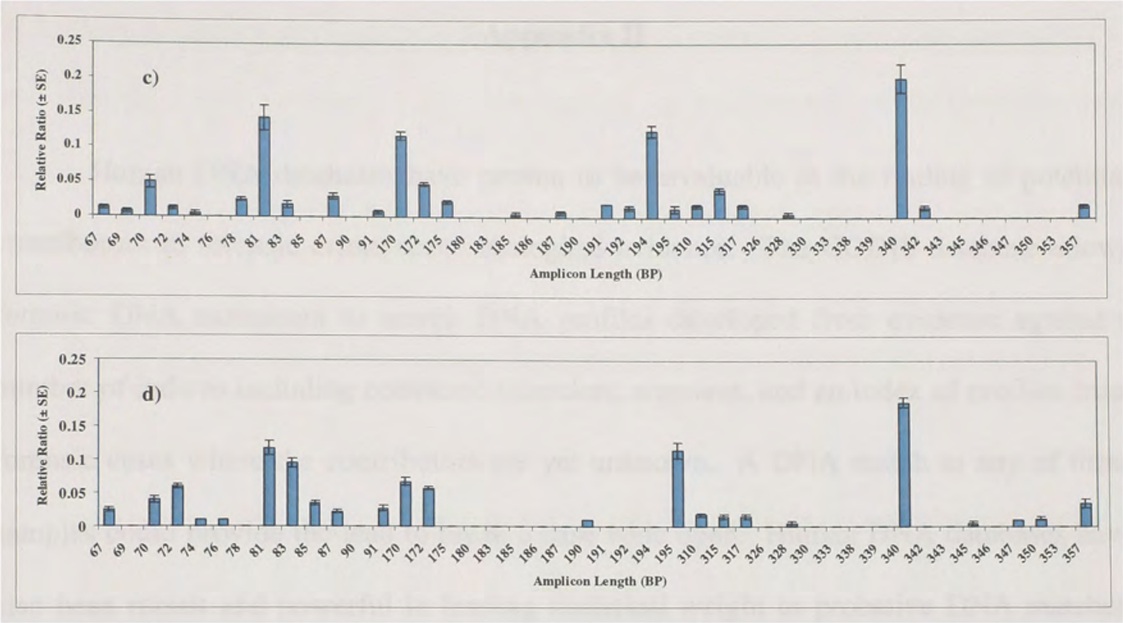
Traditionally, the determination that two soils compare is based on the isolation and identification of rare minerals within each soil sample. Finding and identifying these minerals in a sample of soil is usually performed by an expert geologist with many years of experience. Often, their expertise are localized to a confined geographical area. With the recent advent of DNA profiling and more specifically, DNA profiling of the soil metagenome, a more universal approach to forensic soil comparisons may be on the horizon.

Graphical Representation of Microbial Community Profiles

With little additional training, a few reagents and instrumentation already in place in most forensic laboratories, a forensic DNA scientist can easily extract and profile the DNA metagenome from soil. This procedure will produce multiple quantitative data points which can be used for the discrimination of soils from different soil types. **Figure 17** shows a graphical representation of four soil samples, one from each of the three soil types queried in this study as well as one additional sample from one soil type confirmed to contain diesel fuel. The x-axis shows approximate base pair values for each amplicon generated by PCR of the 16S rRNA genes the four representative soils. For visual comparison purposes, a place on the x-axis is held for each possible amplicon regardless of whether the sample yielded that particular product. Amplicons less than 1 % of the overall signal produced from the entire microbial community profile were filtered, even in samples where they were reproducible. The y-axis shows the abundance for each amplicon relative to the overall bacterial community profile. Standard error bars represent fluctuations in abundance from the duplicate extractions and replicate amplifications of each soil sample. Visual comparison demonstrates that the two soil samples from the same soil type (one containing diesel) in panels "a" and "b" are more similar to each other than to those from different soil types, shown in panels "c" and "d". Although limited, the comparisons of the graphs from these four representative samples support the results obtained by supervised classification of the entire set of microbial community profiles using the support vector machine-learning tool. Specifically, soils from the same soil type were consistently more similar than soils of different soil type and that this applied despite environmental insult. In addition, soils of different soil type origin were sufficiently different in their microbial community structure to predict their soil type of origin with a high degree of accuracy.

Figure 17. Concatenated Microbial Community Profiles. Each panel represents the average ratios for each amplicon derived from replicate amplifications of two extractions of the same soil sample. The y-axis shows the average contribution of each amplicon relative to the overall bacterial community profile. Amplicon length is shown on the x-axis. All three variable domains queried are shown. Panels "a" and "b" represent Krome association soils, diesel was detected in the soil sample represented in panel "b". Panel "c" represents a Rock outcrop-Biscayne-Chekika association soil, "d" represents a Lauderhill-Dania-Pahokee association soil.




c)
Relative Ratio (± SE)
0.25
0.2
0.15
0.1
0.05
0
67 69 70 72 74 76 78 81 83 85 87 90 91 170 172 175 180 183 185 186 187 190 191 192 194 195 310 315 317 326 328 330 333 338 339 340 342 343 345 346 347 350 353 357
Amplicon Length (BP)
d)
Relative Ratio (± SE)
0.25
0.2
0.15
0.1
0.05
0
67 69 70 72 74 76 78 81 83 85 87 90 91 170 172 175 180 183 185 186 187 190 191 192 194 195 310 315 317 326 328 330 333 338 339 340 342 343 345 346 347 350 353 357
Amplicon Length (BP)

Appendix II

Human DNA databases have proven to be invaluable in the finding of potential contributors to forensic crime scene biological evidence. The CODIS database allows forensic DNA examiners to search DNA profiles developed from evidence against a number of indices including convicted offenders, arrestees, and an index of profiles from forensic cases where the contributors are yet unknown. A DNA match to any of these samples could provide the lead to break a case wide open. Human DNA databases have also been robust and powerful in lending statistical weight to probative DNA matches, aiding a jury in their interpretation of a case.

Soil DNA Databasing for Determination of Provenance

When a soil sample from evidence has a questioned origin, there is no current DNA database available as a tool to a forensic soil examiner. This study has demonstrated that using a limited database of soil microbial community DNA profiles, one can determine from what geographical area (delineated by soil type) a soil originated with a high degree of accuracy. Expanding a database like this could prove to be a useful tool available to a forensic soil examiner. The current consensus is that soil type is the most influential factor determining the microbial community structure. Whether all representative soil types in a state or across the nation can be distinguished in a DNA database containing their microbial community profiles remains to be determined. Much of the answer will depend on the proper labeling of data (possibly separating seasons) and the quality of data and interpretation. Recent developments in molecular biology have revealed that soils have a vast diversity of microbes which could provide the distinguishing power to differentiate soils of uncommon soil type. The instructions below detail how to use a Java tool for merging spreadsheets of microbial community profiles into one large data matrix used to train the SVM classifier. The procedure for testing the prediction accuracy of the classifier and retrieving test results are also described.

Java and SVM SOP

Save the Java and lib-SVM2.8 folders in a directory you can access using the "run" and "cmd" functions from the start menu (for me they are saved in a folder with my name "Todd" in the C drive, this info will help you alter the commands you see below for your computer).

Merging data files using Java:

- Select run from the Start menu
- Type "cmd" and hit enter
- You'll see C:\Documents and Settings\Admin> or something similar
- Type "cd.." and hit enter to move one directory up (don't forget the two periods), repeat. You should see C:\>
- Type "cd todd" and hit enter (don't forget the space in between cd and the name of your folder)

- Type “cd java” and hit enter, you should now see C:\todd\java>
- Type “java FileDataMerge”, entire line should read “C:\todd\java>java FileDataMerge”, hit enter.
- At this point a window will ask you how many files you want to merge. Select the number of excel files to merge and hit enter (the excel files must be saved as tab Text (Tab delimited) (*.txt) files prior to merging. Files can be selected by double clicking on the file name. Select each file.
- At this point a window will appear asking you to name the output files. Type in an output name like “test” and hit enter.

At this point, you will create your SVM training sets.

- You should see “C:\todd\java>”, type

  “java RepCrossValidation test” and hit enter (don’t forget the spaces).
- You will see the total number of your samples, in this case 288.
- You will see “C:\todd\java>” again, type “test 1 71”, this will run the SVM classifier for the number of samples you have not including replicates. In my case there were four replicates for each sample, so for a total of 288 samples where there are four replicates of each, you have 72 samples. Although you typed 1 and 71 (one less than your number of samples), the SVM will test all your samples and output the predictions into the java folder beginning with “0” and going to “71”.

- For each sample, there will be three separated outputs files, one for each kernel function, ie “testlin0.out”, “testrbf0.out” and “testsig0.out”, you have to open each output individually and transcribe the predictions into a spreadsheet to calculate the overall accuracy. NOTE: Each time the classifier is run the previous folders are replaced with the results of the last classification.