INVESTIGATION OF DNA PROFILING METHODS FOR FORENSIC EXAMINATION OF SOIL EVIDENCE.

By

James M. Waters, B.Btec. (Hons.)

Being a thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

The Flinders University of South Australia

School of Biological Sciences

Bedford Park, South Australia

September 2004

Table of Contents

Summary

In this thesis, an investigation of the potential of two DNA based profiling techniques for the analysis of forensic soil evidence is presented. These profiling techniques were:

- The previously described technique, Terminal Restriction Fragment Length Polymorphism Analysis (TRFLP) of 16S DNA (Chpt 3.1) and
- A semi-novel profiling technique, Arbitrarily Amplified DNA (AAD) profiling (Chpts 3.2 and 3.3), analysed by both:
    - *conventional* length polymorphism of DNA fingerprints (AADLP) (Chpt 3.2) and
    - *a completely novel method*, DNA sequence similarity (AADSS) (Chpts 3.2 and 3.3), which was investigated using
        - Southern Hybridisation (Chpt 3.2) and
        - Microarray Technology (Chpt 3.3).

These methods were successful at distinguishing samples of soil to varying degrees. TRFLP analysis was capable of generating low but significant differences in similarity statistics between replicate and distinct soil profiles, while both AAD analyses (AADLP and AADSS) generated large and significant differences in similarity statistics between replicate and distinct soil profiles. The potential for significant differences between similarity statistics to be generated enables classification of soil samples as either having common origins (matching, if there is no significant difference) or not (excluded from matching, if there is a significant difference).

The affect of several technical, environmental and practical variables on the biological profiles generated using these techniques was investigated further. These variables included sampling and processing of soils (assessed with TRFLP, AADLP and AADSS), time *in situ* (TRFLP and AADSS), time *ex situ* under a number of storage conditions, as well as spatial variations of microbial communities over small distances (AADSS only).

The most capable method for distinguishing soils of different origin was AADSS, closely followed by AADLP, with TRFLP obtaining only marginal success relative to the other two methods. The preferred format of the AADSS technique is the microarray technology as it is capable of generating a good deal more data from soil DNA profiles than Southern hybridisation. However, as both are capable of distinguishing soils, the low cost Southern hybridisation technique may provide a suitable entry point technology for many forensic laboratories.

The molecular mechanism of the arbitrary amplification profiling system (AAD) was investigated (Chpt 3.4), allowing insight into the way these profiles are generated and potential ways to control the process in order to optimise profiles for various purposes. Many potential improvements and developments are suggested which may further enhance the utility of the techniques presented in this thesis.

The findings presented in this thesis demonstrate the potential for biological profiling of soil communities as a relatively simple, high resolution, objective tool that permits stringent statistical analysis, is not reliant on expert interpretation and is complementary to existing strategies for the forensic examination of soil evidence.

Declaration

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text.

James M Waters
September 2004

I believe that this thesis is properly presented, conforms to the specifications for a thesis and is sufficient to be, *prima facie*, worthy of examination.

Leigh A Burgoyne
September 2004

Acknowledgements

This thesis would not have been completed without the help and support of many people, too numerous to mention individually. There are however, a few people who should be thanked and acknowledged:

Professor Leigh A. Burgoyne, my supervisor, for his passionate support and involvement throughout the entire duration of this research project. I have been considerably expanded by our many discussions, scientific and otherwise;

Professor David E. A. Catcheside, my co-supervisor, for his help editing manuscripts and this thesis, for supplying a laboratory bench and a great working environment;

Dr. Paul Kirkbride, Assistant Director of Forensic Science Service, South Australia and Manager of operational programs and research of National Institure of Forensic Science (NIFS), my co-supervisor, for his help and enthusiasm with the project.

members of the Catcheside laboratory; Fred Bowring, P. Jane Yeadon, Lin Koh, J. Paul Rassmussen, Steve Henderson, Graham Eariss and Russell Stainer for their seemingly endless source of wisdom, support, entertainment and friendship;

my family, for their unwavering support, love and faith

and my partner Megan, who knows why.

Abbreviations

A_{600}	absorbancy at 600nm
AFLP	alternate fragment length polymorphism
AP-PCR	arbitrarily primed polymerase chain reaction
bp	base-pair(s)
BSA	bovine serum albumin
cDNA	copied *or* cloned deoxyribonucleic acid
CTAB	hexadecyltrimethylammonium bromide
CV	coefficient of variance
DGGE	denaturation gradient gel electrophoresis
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
A	adenine
C	cytosine
G	guanine
T	thymine
dNTP	deoxynucleoside triphosphate
dsDNA	double stranded deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
FAM	6-carboxyfluorescein
Kbp	kilobase pairs
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PLFA	phospholipid fatty acid
RAPD	random amplified polymorphic DNA
rDNA	ribosomal deoxyribonucleic acid
RFLP	restriction fragment length polymorphism
RFU	relative fluorescence unit
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SD	standard deviation
SDS	sodium dodecyl sulphate *or* sodium lauryl sulphate
SMIPS	structurally mediated inter-primer selectivity
SI	similarity index
TAE	Tris acetate and EDTA
Taq	*Thermus aquaticus*
TGGE	temperature gradient gel electrophoresis
TRF	terminal restriction fragment
TRFLP	terminal restriction fragment length polymorphism
UV	ultraviolet

1 INTRODUCTION AND BACKGROUND

Soil analysis is an important part of many fields of academia and industry. The body of research encompassing soil science is far too extensive to summarise in a single book, let alone one short chapter. Thus, the information contained within this chapter is a small sample, focused on soil analysis for forensic purposes using biological techniques.

This thesis aims to develop and validate a suitable procedure with the potential for analysing and comparing minute quantities of soil, for direct application in the forensic sciences. Biological information contained within soils, specifically DNA, will be utilised to provide a means of distinguishing between soils and achieving this aim. Organisms inhabiting soil are tremendously diverse with the bacterial fraction alone containing up to 4000 completely different genomes of standard soil bacteria (Torsvik, Goksoyr, *et al.* 1990). If an *Escherichia coli* sized genome (4.3 x 10^6 base pairs) was assumed for each of these bacteria, the soil bacterial "metagenome" may contain up to 1.7 x 10^{10} base pairs, five-fold the size of the Human genome (~3 x 10^9 base pairs). This measure of diversity does not take into account all other classes of organisms including, fungi, plant debris, insects, nematodes, etc. which contribute extensively to the diversity of DNA contained in soil. It is this biological diversity that is to be exploited by creating DNA based "fingerprints" of soils that may be used for comparison and classification of soils. Comparing soils will be of use for evidential purposes, for example matching soil at a crime scene to soil taken from a suspect's property or person. Classification of soils will be useful in preliminary investigations where locating a potential origin of a soil sample is required.

Ideally, methods aiming to compare or classify soils for any purpose should be informative, reproducible, sensitive, cost effective, rapid and easy to perform. Analytical techniques performed for a specific purpose must also lend themselves to practical applications. In forensic comparisons for example, sample collection and processing must be convenient, swift and adaptable to the small size of the samples often collected, which may be taken from tyres, clothing, footwear or beneath a fingernail.

1.1 Current Forensic Techniques for Soil Analysis are Mostly Physical.

Comparing soil samples in forensic science is of recognised importance but does not attract as much attention from forensic scientists as other fields of investigation. It has been noted that this is due to beliefs that homogeneous soil is distributed widely over the earth and analytical techniques are complicated and difficult to perform, requiring high-level training. Forensic procedures that have been used for soil comparison have focused almost entirely on the physical and chemical characteristics of soils and soil particles. Comparing soil colour in a variety of conditions is perhaps the simplest of analytical methods but it has been capable of distinguishing 97% of soils from others tested (Sugita and Marumo 1996). Other methods used for forensic comparison of soils include microscopic examination of mineral deposits (Petraco 1994b; Petraco 1994a) and density gradient techniques able to separate minerals and other soil components, based on densities, into a distinctive pattern of bands suspended in a column of solution (Nute 1975). More recently, advances in technology has seen the development of isotopic ratio testing (Croft and Pye 2003) and X-ray fluorescence spectroscopy (Hiraoka 1994) to profile soils.

These current forensic analyses are more directed towards classifying a soil by identifying the soil type and inferring a general area where it may have originated, which in some instances may be many kilometres square, or several separate areas. There is currently no high-resolution method for matching soils as humans are matched by the individualisation techniques currently available. Generally forensic soil samples are only useful for matching purposes in rare cases when anthropomorphic substances are found in the sample, such as flecks of paint, glass, oil or blood. An opportunity exists for a simple, sensitive, rapid and completely objective method for analysis of trace soil samples for forensic investigations that is independent of a requirement for extensive specialist knowledge. Comparisons of the microbial content of soils have the potential to satisfy these needs.

1.2 A Biological Approach to Forensic Soil Analysis

This thesis investigates an alternative approach to forensic soil comparison based on the biological content of soils. Such an approach could provide an

alternative or complementary tool for forensic scientists when comparing soil samples. Biological techniques have previously been used for the analysis of soil. However, this is still very much a developing field of soil science and there has been limited application of such techniques.

Soil microbiology can be analysed by phenotypic or metabolic profiles using commercially available media (Garland and Mills 1994; Winding 1994; Haack*, et al.* 1995), phospholipid fatty acid (PLFA) profiles (Song*, et al.* 1999; Calderon*, et al.* 2000; Green and Scow 2000; Calderon*, et al.* 2001) or DNA profiles. Phenotypic profiling of soil microbial communities relies on culturing organisms and therefore is only representative of a very small proportion of the total community as it is now widely accepted that only 0.1-0.01% of soil bacteria grow in culture (Faegri*, et al.* 1977; Hopkins*, et al.* 1993). Superficially this by itself is an advantage as it simplifies the soil micro-organism population to useful levels for comparison. However, without DNA sequence based comparisons, geographically distant soils could, hypothetically, be matched if they have environments encouraging growth of micro-organisms with similar phenotypes. Also, the difference between soil communities when examining culturable organisms alone can be much lower than if total bacterial populations are analysed (Ovreas and Torsvik 1998). Both phenotypic and PLFA profiles of soils are not truly representative of the genotypic structure of soil communities and comparison of samples may be limited by time as metabolic states and composition of cell lipids can be extremely dynamic. The metabolic state of soil microbes and their lipid composition can produce significantly different profiles within hours of tillage or disturbance (Calderon*, et al.* 2000; Calderon*, et al.* 2001). More commonly DNA based techniques are used as they are able to give a comprehensive analysis of community structure and by definition an indication of genotypic composition. Comparison and matching of genotypes rather than phenotypic traits is ideal, as coincidental matches are reduced. Organisms may share many phenotypic traits, but in reality may not be genetically similar, leading to false positive results.

1.2.1 DNA Based Analysis of Soil Community DNA

Some of the techniques applied to soil DNA in the past and currently in use include G+C composition of community DNA, DNA reassociation kinetics studies (Torsvik, Goksoyr, *et al.* 1990; Torsvik, Salte, *et al.* 1990; Torsvik, *et al.* 1994), community DNA hybridisation (Ritz and Griffiths 1994; Griffiths, *et al.* 1996), restriction fragment length polymorphism (RFLP) (Moyer, *et al.* 1994) and terminal restriction fragment length polymorphism (TRFLP) (Liu, *et al.* 1997) analysis, randomly amplified polymorphic DNA analysis (RAPD) (Xia, *et al.* 1995), cloning and sequencing variable regions of the genome (Yap, *et al.* 1996; Zhou, *et al.* 1997), ribosomal intergenic spacer analysis (RISA) (Ranjard, *et al.* 2001), denaturant gradient or temperature gradient gel electrophoresis (DGGE/TGGE) (Duineveld, *et al.* 1998; Kozdroj and van Elsas 2000) and various methods of microarray analysis (for a review see Zhou 2003 and Chpt 1.3).

1.2.1.1 G/C Composition of Soil Community DNA

Analysis of soil DNA based on the percentage of the nucleotide bases guanine (G) and cytosine (C) compared to adenine (A) and thymine (T) is an established method of looking at genotype composition. This technique is the most simple, single, low resolution measure of global genotypic structure. This technique, as it applies to soil DNA, is primarily ecological in nature. When used it can determine change in a single community structure over time. However, as a comparative technique, G+C composition is unsuitable as varying soil types containing distinct biological make-ups may give similar or even identical G+C values. Another trait of G+C analysis is the requirement for large amounts of highly purified DNA. These properties do not make this particular technique suitable for forensic applications due to the small size of many forensic samples and the need for suitable individualisation of samples.

1.2.1.2 Reassociation Kinetics Studies of Soil Community DNA

DNA reassociation kinetics experiments measure the rate at which denatured single stranded DNA reforms double stranded DNA. Simple DNA samples lacking diversity will reassociate rapidly in comparison with highly

diverse DNA populations due to the relative frequencies of complementary sequences. These experiments are performed to determine the relative diversity of the DNA population. Soil community DNA has been analysed in this way to determine the diversity of soil bacteria within a single sample (Torsvik, Goksoyr, *et al.* 1990; Torsvik, Salte, *et al.* 1990; Torsvik, *et al.* 1994). Equal measures of soil bacterial diversity cannot be used as proof of matching soil populations, as it is possible that two distinct soils may coincidentally support populations with the same diversity. However, this technique does have the potential to be altered so that a mixture of soil DNA from two samples could be analysed and the reassociation rate determined, relative to the pure, individual samples. If the reassociation rate of the mixed DNA and both individual samples were comparable, similar DNA composition could be inferred. Regardless, as with G+C comparisons, the large amount of highly pure DNA required for such experiments do not make this technique amenable to the forensic sciences.

1.2.1.3 Community DNA Hybridisation Analysis

A technique that is similar in concept to the reassociation technique described above is DNA community hybridisation. These experiments determine the relative similarity between two populations of DNA by measuring the extent to which the DNA cross-hybridises. Two samples are said to be similar or identical when the DNA from one, cross-hybridises to the other as well as it does to itself. Community DNA hybridisation was first used to assess species compositions of natural bacterioplankton assemblages (Lee and Fuhrman 1990) but was later used for soil microbial community structure analysis (Ritz and Griffiths 1994; Griffiths, *et al.* 1996). In theory, this method of comparison lends itself to the forensic sciences as cross-hybridisation depends on DNA sequence similarity and not similar properties of the DNA. Another theoretical advantage to this technique is that unlike many other comparative techniques outlined below. It does not select a subset of sequences for comparison, where biases can occur, but compares total DNA content. This method was used to objectively quantify the similarity of four soil types, with similarity percentages ranging from 25 to 74% (Griffiths, *et al.* 1996). The downside to this technology is that very large amounts of highly pure, whole community DNA and long hybridisation times are

required to handle the enormous diversity of soil communities, making this technology unsuitable for forensic soil comparison.

1.2.2 Profiling the Soil Community Metagenome

The DNA technologies discussed thus far have focused on broad-scale, analysis of whole soil communities, known as the soil "metagenome". Limitations due to the enormous diversity of soil communities or the low resolution of the analysis tends to make these technologies less than ideal for forensic analysis of soil evidence. The analytical techniques discussed below have attempted to overcome the difficulties associated with the large diversity of soil communities by focusing on a subsection of the soil metagenome. Subsections analysed include specific genes and loci or an arbitrary, randomly selected component. This reproducible selection of a subsection of the soil community DNA is referred to in this thesis as profiling.

1.2.2.1 DNA Sequence Data of Variable Alleles

Amplification of highly variable regions of genomes, cloning then sequencing these regions, is a method used for extremely high-resolution analysis of soil communities (Yap *et al*. 1996; and Zhou *et al*. 1997). This method is in principle highly reliable but may take a prohibitively long time per comparison, even when focused on a small subset of organisms within the soil. The cost and time required to amplify, clone and sequence enough DNA to be statistically representative of the general soil community is impracticably large and certainly does not lend itself to large scale, high throughput, time critical analysis that is demanded of forensic analysis.

1.2.2.2 Length Polymorphism Based Analysis of Amplification Products from a Single Locus.

The methods outlined in this section have been grouped together as they share many of the same principles, concepts, advantages and shortcomings. The most useful rely on PCR to amplify specific loci in the genomes of soil organisms

and then distinguishing samples on the basis of differences between the sequences amplified. Commonly the locus amplified with PCR is within the 16S rDNA gene but the principle of the technique is more important than the particular loci chosen. Oligonucleotide primers are designed to prime highly conserved regions of loci then amplify the regions between containing sequences unique to each genotype. The amplification products can then be analysed in various ways, outlined below.

1.2.2.2.1 Restriction Fragment Length Polymorphism (RFLP) Analysis

Historically, the first method to analyse variants of an individual locus was RFLP. Amplification products were cloned into plasmid vectors then treated with restriction endonucleases that cut the DNA into specific lengths depending on the sequence between the two primers. Lengths of the DNA fragments were compared and sequences giving rise to the same pattern were grouped as a genotype.

1.2.2.2.2 Terminal Restriction Fragment Length Polymorphism (TRFLP) Analysis

The financial obstacles and time needed to clone and restrict enough sequences to be representative of the soil using RFLP was overcome by utilising the analytical technique known as terminal restriction fragment length polymorphism (TRFLP) (Liu *et al*. 1997). TRFLP was used to identify close to all restriction based genotypes simultaneously. One of the primers used to amplify DNA is tagged with a label. After PCR amplification, the entire mixture of products is restricted using an endonuclease and the fragments separated by electrophoresis on polyacrylamide gels or through capillaries. The labelled fragments are then located by standard imaging techniques and the pattern of labelled bands is analysed. Identical or similar banding patterns are assumed to indicate identical or similar starting material.

1.2.2.2.3 Denaturant or Temperature Gradient Gel Electrophoresis (DGGE or TGGE) Analysis

The financial obstacles and lengthy periods of analysis using RFLP were also overcome with the introduction of DGGE and TGGE. These analytical

techniques do not require primer labelling or endonuclease restriction, instead the entire amplification mixture is separated on gels with a gradient of denaturant (chemical for DGGE or temperature for TGGE) (Muyzer, *et al.* 1993). When sequences reach a point in the gel at which their unique sequence no longer holds the double stranded DNA together, it separates into single strands and is prevented from travelling any further in the gel. The result is separate bands where unique sequences are sequestered. Like TRFLP, the pattern of bands is representative of the DNA make-up of the community.

These single locus based banding analyses seem to be very powerful tools. However, on closer inspection several drawbacks present themselves. It has been suggested that such systems "will generally detect a limited number of dominant, ubiquitous and ecologically recalcitrant bacterial types in a given soil." (Gelsomino, *et al.* 1999), which must raise questions about the usefulness of such a technique for distinguishing between soils. Even though oligonucleotide primers are designed to prime the most conserved regions, they will not prime these regions from all bacteria let alone all organisms that reside in the soil including actinomycetes, fungi, plants, nematodes and insects. Kozdroj and van Elsas (2000) found that dilution of the initial template substantially alters the pattern of bands observed when 16S rRNA genes were amplified and separated by DGGE. The competitive nature of the PCR reaction can exclude many sequences from being amplified in proportion to the actual number present (Mathieu-Daude, *et al.* 1996; Lueders and Friedrich 2003). Varying levels of PCR inhibitors commonly found in soils, such as humic acid, may also bias the amplification differently in different samples.

It should then come as no surprise that replicate reactions produce DGGE banding patterns with similarities of only 90% (Kozdroj and van Elsas., 2000) and in a study with nine replicates, 85% of TRFLP bands were not reproducible in all reactions (Dunbar *et al*., 2001). There also remains the philosophical problem that a specific band location may not always contain the same sequence. These loci based, banding analytical tools hold great promise for soil ecology investigations. However, the low reproducibility of replicates raises issues of doubt regarding the suitability of forensic application.

1.2.2.3 Length Polymorphism Based Analysis of Arbitrary or Random Loci.

There are other banding based techniques that do not rely on any particular loci but profile entire genomes or genomic mixtures. Alternate fragment length polymorphism (AFLP) (Vos, *et al.* 1995), arbitrarily primed polymerase chain reaction (AP-PCR) (Welsh and McClelland 1990), random amplified polymorphic DNA (RAPD) (Williams, *et al.* 1990), Repetitive Extragenic Palindromic (REP) PCR (Versalovic, *et al.* 1991) and the recently developed profiling system, sequence mediated inter-primer selectivity (SMIPS) (Rogers 2002). These techniques have not been used extensively for soil community DNA analysis.

1.2.2.3.1 Alternate fragment length polymorphism (AFLP) Analysis

AFLP works by digesting DNA with a restriction endonuclease, attaching oligonucleotide linkers to the ends of the restricted fragments and amplifying fragments using PCR and oligonucleotides that prime to the linkers. The result is a mixture of amplified DNA of differing lengths giving a representative pattern of bands when separated by gel or capillary electrophoresis. Only one recent study has used AFLP to analyse soil microbial communities (Franklin and Mills 2003), providing relatively little evidence on the repeatability and reliability of the technique regarding soil community analysis. What was determined by Franklin and Mills (2003) is that analysis of profiles generated from replicate soil samples had relatively low similarities (30-35% similarity) and samples taken less than a metre apart showed further reductions in AFLP profiles similarities (~20% similar). No tests were performed to assess the level of similarity between soils separated by larger distances or distinct types of soil. This level of replication and spatial variation is not acceptable for forensic investigations, which require high similarities between replicate profiles and smaller spatial variation over practical distances.

1.2.2.3.2 Arbitrary Primed PCR (AP-PCR) and Random Amplified Polymorphic DNA (RAPD) Analysis

Arbitrary Primed PCR (AP-PCR) and Random Amplified Polymorphic DNA (RAPD) both amplify regions of DNA without any *a priori* knowledge of

the sequence of the template DNA. AP-PCR uses a single arbitrarily chosen oligonucleotide primer and a PCR with two initial low stringency annealing cycles followed by many rounds of normal amplification. Like AFLP and the majority of soil community analytical techniques, the result is a representative pattern of band lengths separated on a gel. AP-PCR has not been used for microbial community analysis. However, the very similar technique RAPD has previously been used for the analysis of creek water (Franklin, *et al.* 1999), termite mounds (Harry, *et al.* 2001), plant root zones (Kang and Mills 2004) and soils (Xia, *et al.* 1995; Sudarshana, *et al.* 2000; Yang, *et al.* 2000). RAPD utilises a short, random oligonucleotide that primes compatible genomic DNA sequences and regions of DNA with correctly orientated oligonucleotides close enough to each other to allow amplification. A number of PCR amplifications are performed with a battery of primers and a representative fingerprint of DNA lengths can be analysed for each primer used.

1.2.2.3.3 Repetitive Extragenic Palindromic (REP) PCR Analysis

Repetitive Extragenic Palindromic (REP) PCR is based on the use of oligonucleotide primers designed to bind conserved interspersed repetitive extragenic sequences in bacteria. REP primers will extend outwards from the repetitive sequences to amplify DNA between adjacent repetitive elements, resulting in an assortment of amplified sequences of characteristic lengths which can be used as a genomic fingerprint. REP-PCR has primarily been used for the identification of individual species isolates (McManus and Jones 1995; Snelling, *et al.* 1996; van der Zee, *et al.* 1999; Cherif, *et al.* 2003). Recently, REP-PCR has been used to observe differences in soil microbial communities (Parham, *et al.* 2003), however no analysis on the reproducibility or discriminatory power of REP-PCR for soil microbial communities was performed.

1.2.2.3.4 Structurally Mediated Inter-Primer Selection (SMIPS) Profiling.

SMIPS profiling is a PCR based method of amplifying a fraction of the total DNA sequence of a sample (Rogers 2002). SMIPS involves a single primer PCR amplification without primer-template specificity to initiate a relatively unselective, broad-scale amplification. With further amplification the annealing processes of the PCR reaction, specifically the self annealing of the single-

stranded amplicons between the primer sites, create structures that either facilitate or impede primer loading resulting in a selective amplification of a relatively small number of sequences.

The concept of PCR amplification occurring without any (or minimal) primer-template specificity may be foreign to many molecular biologists but this proposed priming of non-complementary sequences is supported by empirical evidence. Human sequences were SMIPS amplified from reverse transcribed RNA (cDNA) and DNA isolated from sera, cloned and sequenced. From the 15 amplified sequences that had good (>99% confidence) homology with known loci on the human genome, the mean number of consecutive bases from the 3' end of the primer matching the database sequence was a mere 2.7bp before a mismatch intervened and fifty percent of sequences had no primer-to-database homology whatsoever, not even a match at the primer's 3' terminus nucleotide (Rogers 2002).

The second half of the SMIPS amplification system is the selective amplification of DNA with particular secondary structures that facilitate primer loading. This proposition was supported by *in silico* analysis of SMIPS amplified products using the MFOLD software available on the BioManager website (http://biomanager.angis.org.au) to predict single stranded secondary structure. The majority of SMIPS amplified products analysed by MFOLD displayed secondary structures with free ends available for the loading of additional primers, while secondary structures involving the ends of the PCR products (hindering subsequent primer loading) were rarely observed (Rogers 2002).

Although the single arbitrary primer amplification of SMIPS is similar to AP-PCR and RAPD, it has several distinct differences. Firstly, oligonucleotide primers used in SMIPS profiling do not rely on sequence compatibility to prime the template. Secondly, selective pressure is applied within the SMIPS PCR to yield higher amplification of DNA with particular sequence structures *in vitro*. It is this selection that is utilised to profile the DNA.

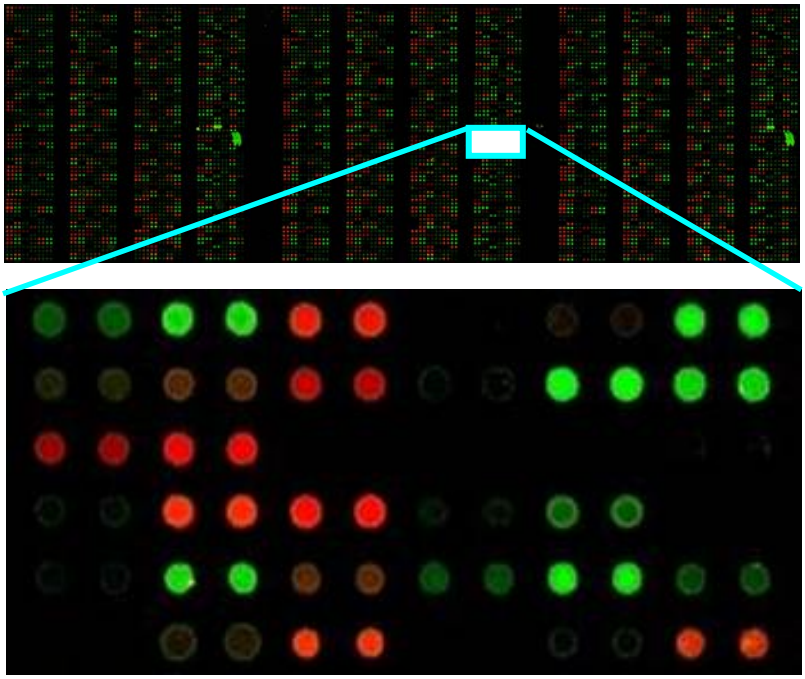
1.3 Microarrays

Microarray (or Gene chip) technology is a recently developed, powerful genomics tool allowing high-throughput analysis of up to tens of thousands of DNA or oligonucleotide sequences simultaneously. Microarrays have allowed

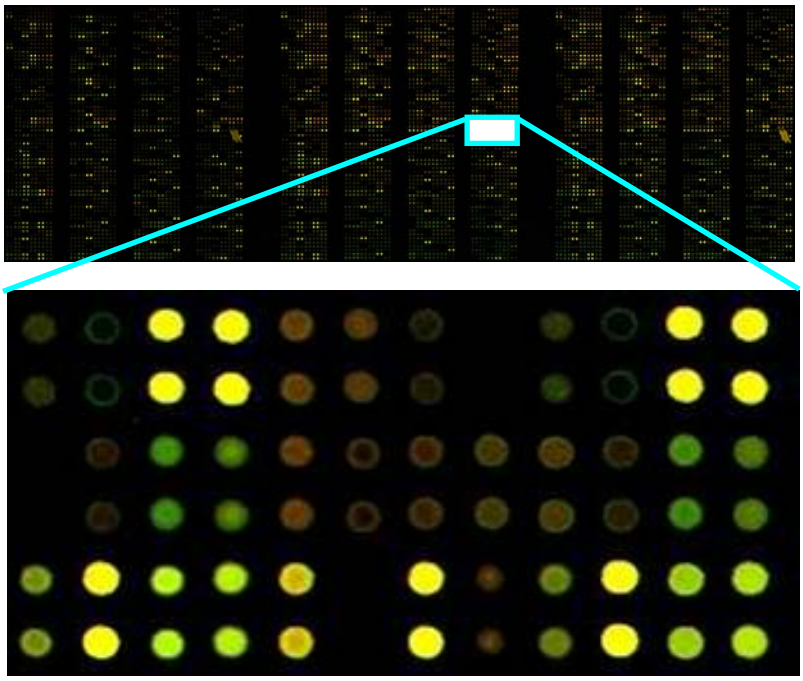
investigation of the composition and working physiology of genetic components of living cells and tissues more comprehensively than ever before. The key advantage of microarray technology is due to the miniaturisation of the DNA amounts deposited for hybridisation, allowing a high density of DNA sequences available for parallel analysis. Robotics and automation allow tens of thousands of sequences to be deposited either as spots on a non-porous glass-slide-based microarray system or within immobilised gel pads, fixed to a solid matrix. These array bound spots are referred to as "probes" and can be hybridised with labelled DNA called "target DNA". This nomenclature seems contrary to traditional methods of DNA hybridisation, where labelled probes are hybridised to membrane bound target DNA. The reason for this reversal is that the amount of target DNA (usually a complex mixture of genomic fragments or reverse transcribed RNA molecules) hybridizing to many individual probes can be assessed as with traditional methods. The key difference with microarrays is the ability to assess large numbers of probes with a single target DNA simultaneously. Labelled target DNA is hybridised to the array and where complementary sequences exist between a probe and target molecule, hybridisation occurs and the labelled target DNA is sequestered at the probe's location. After hybridisation and subsequent processing steps, the fluorescence of the label at each probe on the array can be determined using imaging equipment and software. This fluorescence corresponds to the amount of target DNA at each probe on the array, giving an indication of the relative amount or similarity of target DNA with homology to each probe sequence. Frequently, two distinct target DNA samples are hybridised to a single array in order to give a measure of relative abundance. This can be achieved by the use of different labels. Commonly, the two labels used are the fluorochromes Cy3 and Cy5 (Yu, *et al.* 1994), which after excitation (peak absorption at 550nm and 649nm, respectively), emit different wavelengths as the excited electrons return to their ground state (peak emission at 570nm and 670nm, respectively). Software used to analyse the intensity of each label, generally display the presence of the fluorochromes as green and red for Cy3 and Cy5, respectively. The co-hybridisation of the two labels to a specific probe indicates that they both share homology to that probe. This shared homology is indicated visually by the appearance of yellow (equal amounts of green and red) spots. Figure 1.1 (pg21)

Figure 1.1: Examples of a small section of a microarray hybridised with two differentially labelled profiles which have different (A) or similar (B) DNA sequence. Where labelled profiles have distinct DNA sequences (A), hybridisation of one or the other labelled profile (red or green) occurs at some spots. When labelled profiles share DNA sequences (B), hybridisation of both labelled profiles occurs at some spots (red and green) resulting in a yellow fluorescence.

A



B



shows two examples of small microarrays sections after hybridisation with two profiles labelled with fluorochromes Cy3 or Cy5. The first microarray, section (A), shows the hybridisation of two labelled profiles, comprised of different sequences. The effect is a separation of the two fluorescent labels resulting in predominantly red or green probes. The second microarray, section (B), shows the hybridisation of two labelled profiles containing similar sequences. The appearance of yellow array elements reflects the ability of both labelled profiles to hybridise to the same probe sequence. Relative levels of bound label can be measured to give an indication of the degree of correspondence of sequence between target DNA samples at each position in the array.

Microarray technology has been used to answer many scientific questions. Some common uses for microarrays include comparing global gene expression in cells or tissues under different conditions or between organisms (Costouros and Libutti 2002; Crow and Wohlgemuth 2003; Koizumi and Yamada 2003), determining the genetic divergence between pairs or groups of organisms (Schoolnik 2002; Daran-Lapujade, *et al.* 2003; Watanabe, *et al.* 2004), detecting specific mutations (Gerry, *et al.* 1999; Shi 2001) and single nucleotide polymorphism (SNP) mapping (Wang, *et al.* 1998; Cronin, *et al.* 1999; Hacia, *et al.* 1999; Ji, *et al.* 2004). Most microarray experiments have used pure cultures of specific model organisms. However, the use of microarrays for environmental samples, containing complex mixtures of mostly unknown and uncharacterized organisms has received some recent attention. A variety of microarray formats has been suggested and developed for the analysis and comparison of complex microbial communities including Phylogenetic Arrays (PAs), Functional Gene Arrays (FGAs), Community Genome Arrays (CGAs) and Random Oligonucleotide Arrays (ROAs)

1.3.1 Phylogenetic Arrays (PAs)

Phylogenetic Arrays are used to study phylogenetic structures of microbial communities (Small, *et al.* 2001; Koizumi, *et al.* 2002; Valinsky, *et al.* 2002). Typically, a large variety of 16S rDNA genes are amplified, or oligonucleotides that hybridise 16S rDNA sequences are designed and deposited onto array slides. 16S rDNA or rRNA extracted from samples of interest is then labelled (with

reverse transcription in the case of 16S rRNA) with or without PCR amplification and hybridised to the PA. Resulting fluorescence fingerprints give an indication of which ribotypes are present in the environmental sample and in what proportions. Ribosomal DNA genes are present in all micro-organisms and contain regions which are highly conserved as well as hyper-variable regions, making specific detection of rDNA genes difficult, particularly with PAs containing full length or large rDNA sequences as significant cross-hybridisation between different but closely related sequences is frequent. The use of oligonucleotide probes designed to hybridise short, variable regions of rDNA sequences can help reduce the amount of cross-hybridisation between probes and targets of different ribotypes. However, cross-hybridisation of mismatched DNA does still occur with oligonucleotide PAs. Probe specificity of oligonucleotide microarrays depends on a number of factors including probe length, with a trade off existing between sensitivity and specificity (Guschin, *et al.* 1997). Longer probes were found to be less discriminatory (low specificity), while short probes caused a decrease in fluorescence intensity (low sensitivity) (Guschin, *et al.* 1997). Studies have been performed that elucidate the extent to which mismatched sequences contribute to fluorescence readings after hybridisation. Zhou and Thompson (2002) reveal that hybridisation signal intensities of target DNA bound to 19mer oligonucleotide probes with a single base mismatch was reduced to 10-30% of perfectly matched probes, depending on the type of mismatched nucleotide base. Additional increases in mismatched bases further reduced the hybridisation signal to 5-25% (2 mismatched bases) and <5% (three or four mismatched bases) of perfectly matched probes. Other researchers have also found difficulties with complete discrimination of 16S rDNA genes (Bavykin, *et al.* 2001; Small, *et al.* 2001; Loy, *et al.* 2002; Urakawa, *et al.* 2002). This indicates that for the purposes of identifying and quantifying particular ribotypes in environmental samples, errors may be prevalent when interpreting microarray data due to cross-hybridisation. Cross-hybridisation is less concerning for forensic investigations where a distinguishable profile of an environmental sample is all that is required, rather than extensive phylogenetic data. However, the potential for errors in interpretation to occur is still present. Two distinct but similar ribotypes may exist in separate samples that both cross-hybridise equally well to a probe with distinct sequence to both ribotypes. This would falsely

contribute to the similarity of the profiles generated from these two samples. Investigations into this particular source of error would need to be performed to determine its influence on sample matching. It may be that the large number of probes on microarrays would dilute the effect of such coincidental errors if the majority of the target DNA extracted from both samples of interest will not cross-hybridise in this way.

1.3.2 Functional Gene Arrays (FGAs)

An alternative to phylogenetic arrays are functional gene arrays (FGAs). FGAs contain microarray-bound probes (either cloned genes or oligonucleotides) that will bind genes encoding functional enzymes, usually involved in biogeochemical cycling processes, from environmental organisms. These arrays are used to determine the physiological status and functional activity of microbial communities (Wu, *et al.* 2001; Bodrossy, *et al.* 2003; Taroncher-Oldenburg, *et al.* 2003; Rhee, *et al.* 2004; Stralis-Pavese, *et al.* 2004). These arrays also have limitations of specificity. An FGA constructed to assess the activity of genes involved in nitrogen cycling (nitrite reductases *nirS* and *nirK*, ammonia mono-oxygenase *amoA* and methane mono-oxygenase *pmoA*), containing full-length genes amplified from pure cultures and marine sediments, displayed cross-hybridisation fluorescence signals from probes containing >80% sequence identity and in some cases, weak hybridisation signals were observed for some sequences with <75% sequence identity, even under highly-stringent hybridisation conditions (65°C) (Wu, *et al.* 2001). Taroncher-Oldenburg *et al.* (2003) found similar results using a 70mer oligonucleotide based FGA (containing the nitrogen cycle genes *amoA, nifH, nirK* and *nirS*), with 87% sequence identity the threshold for hybridisation, with these sequences retaining 20% fluorescence of perfectly matched sequences under highly-stringent conditions (65°C). A 50mer FGA (containing oligonucleotides designed from biodegradation and metal resistance genes) also shows similar levels of cross-hybridisation with 88% sequence identity the threshold for hybridisation (Rhee, *et al.* 2004). The use of oligonucleotides slightly increased the specificity compared to arrays containing full-length sequences. The potential for non-matching sequences from different profiles to hybridise to the same probes may erroneously

increase the apparent similarity of those profiles. As previously mentioned, if the overwhelming majority of sequences did not cross-hybridise in a manner that would alter the apparent similarity of the profiles, the vast number of probes contained on the array may dilute this potential source of error. Stralis-Pavese *et al.* (2004) found that 98.9% of the 3660 individual target-to-probe hybridisations, performed on their 50mer oligonucleotide methanotroph community (methane mono-oxygenase *pmoA*) FGA gave the result predicted by the weighted-mismatches-based software CalcOligo version 2.03 (http://www.diagnostic-arrays.com/calcoligo/index.htm). This indicates that the potential exists for accurate prediction of cross-hybridisation levels of known probes and targets. However, the number and diversity of organisms within environmental samples and the DNA sequences they contain, will be mostly unknown. This still leaves open the potential for cross-hybridisation to lead to errors in determining sample similarity.

1.3.3 Community Genome Arrays (CGAs)

Another microarray formats exist that does not assess the variations of a small number of genes but rather whole genomes. Community genome arrays (CGAs) are a very recent development that comprise an array of whole genomic DNA samples, isolated from pure cultures. These genomes can be hybridised with whole DNA extracted directly from soil or other environmental samples, or from the culturable or other fractions of biota. Currently, no data have been published regarding the use of CGAs for microbial community analysis but, it has been reported that differential hybridisation intensity at the species-level has been observed (Zhou and Thompson 2002; Zhou 2003). Although differentiation of individual species may be possible with CGAs, the analysis of entire populations of microbial communities will introduce many difficulties. The possibility of cross-hybridisation with such a microarray format is high due to the number of potential probes (genes) at each position on the array. Cross-hybridisation will lead to dubious interpretation of data. Also, only the small culturable fraction of organisms will be available for comparison as the construction of CGAs requires DNA extracted from pure isolates.

1.3.4 Random Oligonucleotide Arrays (ROAs)

A microarray based fingerprinting technology has been recently developed that utilizes short (9mer) random oligonucleotides as probes (Kingsley, *et al.* 2002). The 47 probe prototype was capable of distinguishing 14 closely related *Xanthomonas* strains after hybridisation with REP-PCR amplified fingerprints. Despite its current use to distinguish between specific strains, the principal of the technology could be expanded to accommodate microbial community fingerprints. The random selection of oligonucleotide probes allows easy construction, optimisation and standardisation of ROAs and is capable of generating fingerprints from any organism or environmental sample. The method generates a hybridisation-based fingerprint of the sample but does not provide any detailed phylogenetic, physiological or genetic information about the sample, making this technique of little use to molecular ecologists, although it may be ideal for forensic investigations where only a distinguishing fingerprint is required.

1.4 Requirements for an Ideal Analytical Technique for Forensic Soil Examination

The many biological techniques for soil analysis must be assessed in the context of their application. The primary context for the development of many of the techniques described above is ecological not forensic which demands a different set of requirements. For application in forensic analysis, appropriate techniques will possess a high level of repeatability and discrimination between genuinely different samples (exclusions) and be able to match samples taken from practical spatial and temporal scales that will help with investigations (inclusions). An effective technique for soil examination will also need to cope with practicalities related to forensic science, including environmental variables that could potentially alter the profile between sampling times, the small sample sizes often associated with forensic cases and contamination of samples with other soils, biota or anthropomorphic materials.

1.4.1 Reproducibility of Analysis

A primary concern for all scientific analysis of soils, particularly in forensic investigations, is the reproducibility of the information gathered. Reproducibility is a foundation of empirical science and needs to be achieved to lend weight to conclusions drawn from comparison experiments. Obviously, any test developed for forensic analysis of soil communities will need to be reproducible. The reproducibility of the tests performed in this thesis has been determined.

1.4.2 Discrimination Between Soils

The practical purpose of analyses for forensic comparisons of soils is to distinguish soil samples to include or exclude the possibility of samples originating from the same location. If an analytical technique was incapable of excluding the possibility of common origin, even when in reality the samples originated from different locations (giving a high number of false positive results) such a technique would contribute little evidential value. Alternatively, techniques that are too discriminatory, excluding the possibility of common origin when sample are actually from the same location (yielding many false negative results) will not be practically useful. The possibility of obtaining differences between repeated samples from the same site is a practical reality of soils. The inherent variation of soils means that no two samples will be exactly the same, even sub-samples of soil will differ in some manner. This variation can often be enhanced during transfer to a suspect's person or property by preferential transfer and/or persistence of particular fractions of the soil. Any method for forensic analysis of soils must be tolerant of these realities.

1.4.3 Dealing with Environmental Variables

It is inevitable, if put to use, that any method for comparing soil samples will need to surmount hurdles such as temporal and spatial variation of soil populations. The dynamic nature of soil micro-organism composition means that samples taken a period long enough apart may no longer match. The effect of time on the ability of tests to successfully match soils can be used to determine

preliminary timeframes of usefulness for any comparative technique developed. Samples taken large periods apart may not be useful in forensic investigations as matches may not be achieved despite originating from the same location but the value of successful matches may be strengthened by dating samples within a finite period of time. Likewise, micro-organism populations are expected to change with space. Subtle differences in environmental conditions will drive alterations in the proportions of the microbial species in a population or even the loss or gain of species. These practical realities of soil communities may be present in all cases making them of high importance. The variables of time and space will need to be assessed to determine the level of influence they will have, allowing informed decisions to be made about the value of soil analysis results.

1.4.4 Suitable for Minute Sample Sizes

The nature of the application of a soil comparison technique for forensic science means that it must be amenable to minute samples of soil as commonly all that can be retrieved from a forensic scene is a few milligrams. Therefore, when contemplating using DNA from soil organisms it would be ideal if all of the DNA contained within the sample is recovered for use. In practice, this is not possible. However, the maximisation of high molecular weight DNA extracted should be strived for. Potential approaches to achieve high yields from minute samples include scaling down of existing methods or the development of new miniature extraction methods. Extraction and purification of DNA from soil samples is discussed in detail below in Chpts 1.5 and 1.6.

One question that will need serious attention regards the degree of uniformity of soil. It is reasonable to expect that small soil samples will be less representative of the original site than larger samples, leading to potentially erroneous assumptions being made due to differences between DNA profiles of small recovered samples and the parent soil. This reality of the heterogeneous nature of soils will need to be tolerated by any technique to be used for forensic analysis of soil evidence.

1.4.5 Practical Considerations Regarding Forensic Soil Samples

Less critical but genuine, practical concerns regarding the comparison of soil communities from forensic samples, crime-scenes and control sites include how site disturbance, removal of samples from original locations, storage of samples *ex situ*, interactions with clothing, footwear, entrenching tools or skin, and contamination of samples affect the profiles generated. The transfer of soil from the ground to a person or property involves some level of disruption to the native soil. The degree of disruption may vary from mild, e.g. transfer to footwear, to severe e.g. excavation for burial purposes. The potential for any disruption to alter the soil community profile means the affect of such disturbances should be assessed to determine if profiles do indeed change in a way that would render soil community profiling techniques ineffective. The nature of all transferred evidence is that it is removed from the scene. As micro-organisms are adapted to their native environment, being removed to a new and potentially different environment may alter the composition of the soil community in the transferred sample. Information on the *ex situ* conditions, if any, that significantly alter soil profiles will be valuable to forensic investigators. Also, any interactions with clothing, footwear, tools or even skin may potentially alter soil community profiles. The extent of such an affect should be determined to allow informed decisions by forensic investigators. It is also likely that any soil evidence recovered from a suspect's person or property may be contaminated by other materials including soils from a variety of locations. An ideal technique for forensic analysis of soil communities will be flexible enough to allow comparisons involving contaminated samples.

Before a comparison between soil DNA populations can be made, the DNA must be extracted and purified to a useable standard. The following is a review of work performed previously in the areas of extracting DNA from soil and its purification from inhibitors of downstream applications.

1.5 DNA Extraction Methods

Vigdis L. Torsvik (Torsvik 1980) described the first procedure for the isolation of DNA from soil bacteria. This pioneering effort initiated the field of

soil molecular microbial ecology. In the years since, there have been numerous modifications to the Torsvik protocol and new methods for extracting DNA from soil. However, all procedures can be classified into two main categories, direct or indirect extraction.

1.5.1 Indirect Extraction of DNA from Soil

The indirect methods of extracting DNA from soil, including that of Torsvik (1980), work on the basis of separating the bacterial fraction from the remainder of the soil, then lysing the cells and isolating DNA by standard methods. Traditionally, bacteria have been isolated by fractional centrifugation however cation exchange resin has been introduced to separate bacteria from soil (Jacobsen and Rasmussen 1992).

1.5.2 Direct Extraction of DNA from Soil

Direct methods of extracting DNA from soil break open cells whilst still in the soil and then purify the liberated DNA from soil contaminants. This method of DNA extraction was pioneered by Ogram *et al*. (Ogram*, et al.* 1987). Alterations and modifications of this method have been plentiful, including the use of various combinations of: lysozyme, high salt buffers, freeze-thaw cycles, extended heating, CTAB, proteinase K, ammonium acetate, SDS, bead-mill homogenisation and phenol-chloroform extractions (Tsai and Olson 1991; Porteous and Armstrong 1993; Smalla*, et al.* 1993; More*, et al.* 1994; Zhou*, et al.* 1996; Porteous*, et al.* 1997; Kuske*, et al.* 1998; Heinz and Platt 2000).

1.5.3 Comparing Direct and Indirect Methods of DNA Extraction from Soils

The various methods of direct and indirect extraction of DNA have been subject to rigorous comparisons, not only by their inventors but also by independent reviewers (Steffan*, et al.* 1988; Trevors 1992; Miller*, et al.* 1999; Martin-Laurent*, et al.* 2001). Interestingly, there is no one method that stands alone as superior to all others. There are advantages and disadvantages to all. Generally, direct methods tend to yield higher amounts of DNA than indirect

methods. Bacterial cells associated with particulate matter that do not separate with free bacteria is partially overcome in the direct method. Also, all cells including actinomycetes, fungi, plant matter and animal cells are subject to lysis and any extracellular DNA is retained, expanding contributions to the DNA extracted from bacterial cells to all organisms and cell debris present in the sample. There is also variation of yields between different direct extraction methods. More physically aggressive methods such as bead-mill homogenisation tend to yield more DNA than the chemical extraction methods because of the lysis of spores and other recalcitrant cells that are difficult to break open. However, bead-mill homogenisation can lead to shearing of DNA into smaller fragments. This may be undesirable when the objective is to extract large fragments of DNA intact or perform PCR where sub-optimal product amounts or PCR generated chimeras may be formed as a result (Liesack, *et al.* 1991). Direct methods are commonly more contaminated with humic substances (Zhou, *et al.* 1996), heavy metals and other pollutants that interfere with downstream uses such as restriction digestion, hybridisation and the polymerase chain reaction (PCR). This means that DNA extracted directly from soil needs to be purified before use. The ability of many direct extraction methods to yield higher amounts of DNA from a wider source of micro-organisms has spawned much interest and research into purifying DNA from high levels of common soil contaminants, enabling direct extraction methods to yield DNA of a quality required for DNA analysis.

1.6 DNA Purification Methods

DNA extracted from soil commonly contains contaminants such as humic substances and heavy metals such as iron that interfere with downstream processes. Such contaminants are known to inhibit polymerases used in PCR (Tsai and Olson 1992b; Tsai and Olson 1992a; Tebbe and Vahjen 1993; Watson and Blackwell 2000), restriction endonucleases, transformation of competent bacterial hosts and DNA-DNA hybridisation (Tebbe and Vahjen 1993) making their removal essential in soil DNA analysis.

Several methods have been utilised to remove soil contaminants, primarily humic substances, from DNA. These include the use of the chemical reagents caesium chloride and potassium acetate for precipitation and spermine-HCl or glassmilk for further purification (Smalla, *et al.* 1993), sodium hydroxide

(Bourke, *et al.* 1999), Sephadex, Chelex or polyvinylpyrrolidone (Tsai and Olson 1992b; Straub, *et al.* 1994; Berthelet, *et al.* 1996) which separate DNA and humic substance by differences in affinity, or size exclusion chromatography, polyacrylamide and agarose gels which separate large DNA molecules from smaller contaminants (Tsai and Olson 1992b; Zhou, *et al.* 1996; Harry, *et al.* 1999), or a combination of both chemical and size exclusion methods (Young, *et al.* 1993; Harry, *et al.* 1999). There are also more specific methods of purification, which use magnetic beads covered with DNA complementary to specific desired sequences that are purified, recovered, then used for PCR (Jacobsen 1995). However, this method is not useful for broad-scale community analysis and only single genes or a specific sequence of interest can be isolated.

Although all of the above purification methods have been proven to increase the effectiveness of downstream processes, there has been little research undertaken to determine the method with the best yield and purity of high molecular weight DNA. Tsai and Olson (1992b) have reported that Sephadex G-200 removes more contaminants than polyacrylamide gel columns (Bio-Gel P-6 and P-30) and Sephadex G-50. However, it is unlikely that any one method will be superior to others for all soil types, as differing levels of various contaminants are found in all. As is expected, it has been shown that combining two or more of the above methods will greatly reduce contaminant levels (Harry, *et al.* 1999).

1.6.1 Blocking the Influence of Soil-Borne Inhibitors on Downstream Processes

An approach that could be used in concert with methods to remove contaminants from DNA, involves removing or reducing the influence of soil contaminants on downstream analysis of soil DNA. The activity of PCR inhibitors can be reduced or blocked by the addition of BSA, greater quantities of Taq polymerase, or the single stranded DNA stabilising T4 gene 32 protein (Vahjen and Tebbe 1994). Combinations of purification techniques and additives could be used to achieve DNA of a purity that is usable for forensic analysis of soil evidence.

It may be possible to discard purification stages when samples are in such low amounts as the concentration of soil contaminants will be low, potentially having minimal effect on downstream applications. Additives such as BSA or T4

gene product 32 (Vahjen and Tebbe 1994) may reduce contaminant based inhibition without the need for purification. If purification of DNA was found to be unnecessary for small samples, losses during processing would not occur and greater amounts of DNA would be recovered enabling more accurate and/or numerous tests.

1.7 Introduction to the Research and Findings Encompassed in this Thesis

In this thesis, an investigation of the potential of two DNA based profiling techniques for the analysis of forensic soil evidence is presented. These profiling techniques are Terminal Restriction Fragment Length Polymorphism Analysis (TRFLP) of 16S DNA (Chpt 3.1) and an Arbitrarily Amplified DNA (AAD) profiling method (Chpts 3.2 and 3.3). The AAD method was based on the arbitrary priming PCR methods of AP-PCR, RAPD and SMIPS. AAD profiles were analysed by length polymorphism of DNA fingerprints (AADLP) (Chpt 3.2.2) and DNA sequence similarity (AADSS). AADSS was investigated using Southern hybridisation (Chpt 3.2.1) and Microarray technology (Chpt 3.3).

It will be shown that these methods were successful at distinguishing samples of soil to varying degrees. TRFLP analysis was capable of generating low but significant differences in similarity statistics between replicate and distinct soil profiles, while both AAD analyses (AADLP and AADSS) generated large and significant differences in similarity statistics between replicate and distinct soil profiles.

The affect of several technical, environmental and practical variables on the biological profiles generated using these techniques was also investigated. These variables included sampling and processing of soils (assessed with TRFLP, AADLP and AADSS), time *in situ* (TRFLP and AADSS), as well as time *ex situ* under a number of storage conditions, and spatial variations of microbial communities over small distances (AADSS only).

The molecular mechanism of the arbitrary amplification system was also investigated (Chpt 3.4), allowing insight into the way these profiles are generated and potential ways to control the process in order to generate profiles suited for various applications. Many potential improvements and developments are

suggested which may further enhance the utility of the techniques presented in this thesis.

The findings presented in this thesis demonstrate the potential for biological profiling of soil communities as a relatively simple, high resolution, objective tool that permits stringent statistical analysis, is not reliant on expert interpretation and is complementary to existing strategies for the forensic examination of soil evidence. This thesis has also highlighted some limitations of the technologies regarding geographical ranges of selectivity (distance), time frames of collection and storage conditions, allowing informed decisions to be made regarding the reliability of any results.

2 MATERIALS AND METHODS

2.1 Chemicals

General laboratory chemicals, solutions and reagents were all biotechnology, analytical, certified molecular biology or ultra pure grade and supplied by Sigma-Aldrich, Amresco, Merck or Chem-Supply unless stated otherwise.

2.2 Enzymes

Restriction endonucleases were purchased from New England Biolabs unless otherwise stated. DNA polymerases were obtained from multiple sources; Taq DNA polymerase from Promega, RedHot DNA polymerase from ABgene and BioTaq from Bioline.

2.3 Oligonucleotide Primers

All oligonucleotide primers were purchased from Geneworks and were of Sequencing/PCR purity. A complete list of primers used in this study is presented in Table 1 (pg36)

Table 1: Primers used for PCR amplification.

Primer Name	Sequence 5'-3'	Use
FAM-8f	FAM-AGAGTTTGATCCTGGCTCAG	16S-rDNA loci amplification
926r	CCGTCAATTCCTTTRAGTTT	16S-rDNA loci amplification
T7	TAATACGACTCACTATAGGGAGA	Sequencing
Sp6	CATACGATTTAGGTGACACTATAG	Sequencing
Seq005	CCCTCGAACACCACCTCC	Arbitrary Amplification
FAM-Seq005	FAM-CCCTCGAACACCACCTCC	Arbitrary Amplification
antiSeq005	GGAGGTGGTGTTCGAGGG	Arbitrary Amplification
Seq5A	CCTCCAACACCACCTCC	Arbitrary Amplification
antiSeq5A	GGAGGTGGTGTTGGAGG	Arbitrary Amplification
Seq5B	CACCCTCCAACACCACCTCC	Arbitrary Amplification
Seq5C	CTCCACCCTCCAACACCACCTCC	Arbitrary Amplification
Seq5D	CCTCCAACACCACCTCG	Arbitrary Amplification
Seq5E	CCTCCAACACCCCTCAC	Arbitrary Amplification
Seq5F	CCTCCAACACCCCACTC	Arbitrary Amplification
Seq5G	CCTCCAACACCCAGTCG	Arbitrary Amplification
Seq5H	CCTCCAACAACACCACC	Arbitrary Amplification
Seq5I	CCACCAACAACACCTCC	Arbitrary Amplification
Seq5J	CCTCCTTCTTCTCCTCC	Arbitrary Amplification
Seq5K	CCACCAACAACACCACC	Arbitrary Amplification
Seq5L	CCACCTTCTTCTCCACC	Arbitrary Amplification
Seq5M	CCTCCAACAACACCTGC	Arbitrary Amplification
Seq5N	CCTCCAACAACACGTCC	Arbitrary Amplification
Seq5O	CCTCCAACAACAGCTCC	Arbitrary Amplification
Seq5P	CCTCCACCACAACCTCC	Arbitrary Amplification
Seq5Q	CCTCCCAACAACCCTCC	Arbitrary Amplification
Seq5R	CCTCCAACCCAACCTCC	Arbitrary Amplification
Seq5S	CCTCCCACCAAACCTCC	Arbitrary Amplification
Seq5T	CCTCCAAACCACCCTCC	Arbitrary Amplification
Seq5U	CCTCCAACACCACCACCTCC	Arbitrary Amplification
Seq5V	CCTCCAACACCACCACCACCTCC	Arbitrary Amplification
Seq5W	CCTCCAACAACACCACCTCC	Arbitrary Amplification
Seq5X	CCTCCAACAACAACACCACCTCC	Arbitrary Amplification
Seq5Y	CCTCCAACAACACCTCC	Arbitrary Amplification
Seq5Z	CCTCCAACACCACCTGC	Arbitrary Amplification
Seq5AA	CCTCCAACACCACCACC	Arbitrary Amplification
Seq5AB	CCTCCAACACCACGTCC	Arbitrary Amplification
Seq5AC	CCTCCAACACCAGCTCC	Arbitrary Amplification
Seq5AD	CCACCAACACCACCTCC	Arbitrary Amplification
Seq5AE	CTCCTCCACCACCACCACCTACC	Arbitrary Amplification
Seq5AF	CCACCACCACACCTCCTCCTACC	Arbitrary Amplification
Seq5AG	CACCACCACCACCTCTCACCTCC	Arbitrary Amplification
Seq5AH	CCACACCTCCACTCCACACCTCC	Arbitrary Amplification
Seq5AI	CCATCCACCTCCACCACACCTCC	Arbitrary Amplification
Seq5AJ	TCCACCTCCACCACACCTCCACC	Arbitrary Amplification
Seq5AK	CTCCACCACCACCACTCTCCACC	Arbitrary Amplification
Seq5AL	CACCACCTCTCCACCACTCCACC	Arbitrary Amplification
Seq5AM	CCTCCAACACCACCGCC	Arbitrary Amplification
Seq5AN	CCTCCAACACCACCTTC	Arbitrary Amplification
Seq5AO	CCTCCAACACCACCTAC	Arbitrary Amplification
Seq5AP	CCTCCAACACCACCTCA	Arbitrary Amplification
Seq5AQ	CCTCCAACACCACCTCT	Arbitrary Amplification

2.4 Collection of Soils and Extraction of DNA

Up to 50g of soil was collected from the soil surface (0 to 5cm depth) from sites within South Australia (Figure 2.1, pg38).

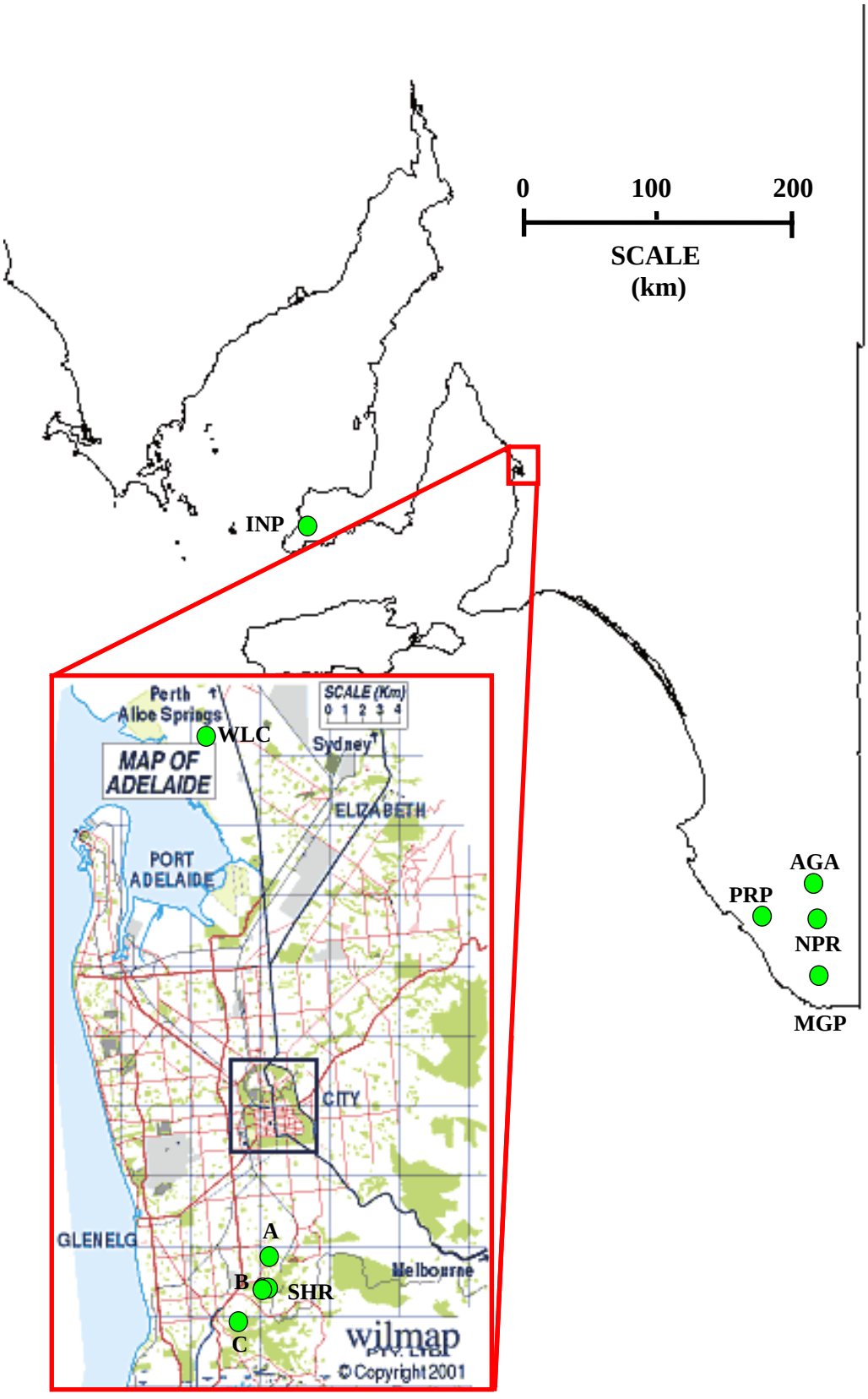
Site A (34.99058° S, 138.59355° E) soil was a dark loam rich in organic matter taken from Grant Jacob Reserve, Site B (35.00251° S, 138.59329° E) was a silty clay taken from Shepherds Hill Reserve and Site C (35.02614° S, 138.56720° E) was a sandy loam taken from the Flinders University grounds. All three sites were within 5km of each other and repeat sampling was performed up to twelve-months after the initial sample, from the original locations.

In addition to these three sites were seven other locations from South Australia including an agricultural soil from the outskirts of Penola in the south-east of the state (AGA, 37.36334° S, 140.82957° E), one sandy soil from the coastal Innes National Park (INP, 35.26499° S, 136.93759° E) on the Yorke Peninsula, three pine forest plantation soils within 100km of each other in the south-east of the state (MGP, 35°50.888' S, 140°46.199'; NPR, 37.54182° S, 140.81443° E; and PRP 37°46.403' S, 140°41.148' E) and two other soils from greater Adelaide, (SHR, 35°00'52.9" S, 138°34'51.4" E, also from Shepherds Hill Reserve, within 200m of Soil B) and the suburb Waterloo Corner (WLC 34°43.521' S, 138°34.690' E).

2.4.1 Lysis and Extraction of DNA from Soil Organisms.

A number of DNA extraction and purification protocols were trialled (see below for detailed protocols), including methods based on lysozyme and freeze/thaw treatments (Tsai and Olson 1991) (Chpt 2.4.1.1), SDS lysis and Proteinase K treatment (Zhou, *et al.* 1996) (Chpt 2.4.1.2), bead-mill homogenisation and SDS lysis (Tsai and Olson 1991; Zhou, *et al.* 1996; Miller, *et al.* 1999) (Chpt 2.4.1.3) and a commercially available bead-mill homogenisation method UltraClean™ Soil DNA Kit (MoBio Laboratories, Solana Beach CA) (Chpt 2.4.1.4). Due to the increased yield of DNA from the bead-mill homogenisation techniques (data not shown) and the quick and easy operation of the commercially available kit, UltraClean™ Soil DNA Kit (MoBio Laboratories, Solana Beach CA) was used for DNA extractions from all soils presented in this thesis.

Figure 2.1: A map of South Eastern South Australia and Adelaide City insert (reproduced with the consent of Wilmap Pty. Ltd. who own the copyright to the insert of Adelaide City), with the locations (Green Spots) of the ten soils used in the studies presented in this thesis.



2.4.1.1 Lysozyme and Repeated Freeze-Thaw Lysis of Soil Organisms.

This protocol is based on the method described by Tsai and Olson (1991). One gram of soil was mixed into 2mL of 120mM Na_2HPO_4 (pH 8) and shaken on an orbital rotor at 150rpm for 15 minutes. The slurry was then centrifuged at 6000g for 10 minutes, the supernatant discarded and the pellet washed in the above phosphate buffer before centrifugation again at 6000g. The pellet was resuspended in 2mL of lysis solution (150mM NaCl, 100mM EDTA, 15mg/mL lysozyme) and incubated at 37°C for 2hrs, mixing every 20-30 minutes. Two millilitres of 100mM NaCl, 500mM Tris HCl (pH 8), 10% (w/v) SDS was added and the slurry subjected to 3 rounds of freezing at -70°C and thawing in a 65°C water bath. Two millilitres of 100mM Tris-HCl (pH 8), saturated phenol was added and briefly vortexed before centrifugation at 6000g for 10 minutes. The supernatant was transferred to a new tube and one volume of Phenol:Chloroform:Isoamylalcohol (25:24:1) was added, mixed then centrifuged at 6000g for 10 minutes. Retrieval of the aqueous layer was followed by the addition of one volume of Chloroform:Isoamylalcohol (24:1) and centrifugation at 6000g for 10 minutes. The aqueous layer was transferred to a fresh tube and one volume of cold isopropanol was added to precipitate the DNA overnight at -20°C. Centrifugation at 10,000g for 10 minutes was performed to pellet the DNA which was then air dried and resuspended in 100µl of 10mM Tris, 1mM EDTA (pH 8).

2.4.1.2 SDS Lysis and Proteinase K Treatment of Soil Organisms.

This protocol is based on the method described by Zhou *et al.* (1996). One gram of soil is mixed with 2.7ml of DNA extraction buffer (100mM Tris-HCl pH 8, 100mM EDTA, 100mM Na_2HPO_4, 1.5M NaCl and 1% (w/v) CTAB) and 20µl of Proteinase K (10mg/ml) and the slurry shaken on an orbital shaker at 225rpm for 30 minutes at 37°C. 300µl of 20% (w/v) SDS was added and incubated at 65°C for 2hrs, mixing every 30 minutes. The suspension was then centrifuged at 6000g for 10 minutes the supernatant recovered and kept. 900µl of the DNA extraction buffer above and 100µl of 20% SDS was added to the soil pellet, which was resuspended, incubated at 65°C for 10 minutes then pelleted again by

centrifugation and the supernatant kept. This step was repeated once more before the supernatants were pooled, and added to one volume of Chloroform:Isoamylalcohol (24:1), mixed, centrifuged and the aqueous phase recovered. The DNA in the aqueous phase was then precipitated with 0.6 volumes of isopropanol for 1hr before centrifugation at 16000g for 20 minutes. The pelleted DNA was washed with 70% ethanol before it was air dyed and resuspended in 100µl of 10mM Tris-HCl, 1mM EDTA.

2.4.1.3 Bead-Mill Homogenisation and SDS Lysis of Soil Organisms.

This protocol is based on the method described by Miller *et al*. (1999). 100mg of soil was added to 2ml tubes containing 2g of acid washed, sterile glass beads (equal parts <106µ, 425-600µ and 710-1180µ glass beads, Sigma) along with 300µl phosphate buffer (100mM NaH_2PO_4, pH 8), 300µl SDS lysis buffer (100mM NaCl, 500mM Tris-HCl, 10% (w/v) SDS, pH 8) and 300µl chloroform:isoamlyalcohol (24:1). The soil suspension was vortexed for 5 minutes before centrifugation at 10,000g for 30 seconds. The supernatant was transferred to a new tube and the DNA precipitated with one volume of isopropanol at -20°C for 16hrs. The precipitated DNA was pelleted by centrifugation at 10,000g for 10 minutes and the pellet washed with 70% ethanol, air dried and resuspended in 100µl of 10mM Tris-HCl, 1mM EDTA (pH 8).

2.4.1.4 Commercially Available Bead-Mill Lysis of Soil Organisms and Column Purification of Liberated DNA.

DNA was extracted and purified from 500mg of whole soil using an UltraClean™ Soil DNA Kit (MoBio Laboratories, Solana Beach CA). This extraction method encompassed a bead-mill homogenisation and detergent based lysis technique coupled with micro-column purification. Due to the commercial nature of the extraction method, the exact details regarding the nature and concentration of reagents, beads and micro-columns are unavailable. However, the protocol involved adding up to 1g of soil to the Bead Solution tubes provided, along with 60µl of Solution 1 (containing SDS) and 200µl of Inhibitor Removal Solution (required if DNA is to be used for PCR). The tubes were then secured

horizontally and vortexed for ten minutes. The tubes were centrifuged for 30 seconds at 10,000g and the supernatant transferred to a clean tube. 250µl of Solution 2 was added, mixed, incubated at 4°C for 5 minutes and centrifuged for 1 minute at 10,000g. 450µl of the supernatant was transferred to another clean tube and 900µl of Solution 3 was added, mixed well, loaded into the spin filter and passed through by centrifugation in two batches. 300µl of Solution 4 (containing ethanol) was added to the spin filter and passed through by centrifugation to wash the filter bound DNA. The dry spin filter was then transferred to a clean collection tube and 50µl of water was added and centrifuged to elute the DNA.

2.4.2 Lysis and Extraction of Human DNA from Whole Blood

2.4.2.1 Extraction and Purification of Human DNA from Fresh Whole Blood.

10ml of whole blood was pelleted by centrifugation at 4000g and the plasma discarded. The pelleted cells were washed in 5ml of digest buffer (10mM Tris-HCl, 10mM EDTA, 100mM NaCl, pH 8.0) for 2 minutes before being pelleted by centrifugation at 4000g. Five millilitres of fresh digest buffer containing 1% SDS and 100µg/ml Proteinase K was used to resuspend the pelleted cells and incubated at 37°C for 16hrs. with mild agitation. One volume (5ml) of phenol was added, mixed well then centrifuged for 5min at 4000g. The aqueous phase was recovered to a fresh tube and one volume (5ml) of phenol was added, mixed well and centrifuged for 5min at 4000g. The aqueous phase was recovered to a fresh tube and one volume (5ml) of chloroform was added, mixed well and centrifuged for 5min at 4000g. The aqueous phase was recovered to a fresh tube and two volumes (10ml) of absolute ethanol added to precipitate the DNA. Ethanol precipitation was allowed to occur at -20°C for 16hrs. The DNA was pelleted by centrifugation at 10,000g for 20min, the supernatant discarded and the DNA allowed to air dry. DNA was solubilised in 400µl of 10mM Tris-HCl, 1mM EDTA (pH 8.0) before quantification spectrophotometrically (DNA conc. µg/ml = 50 x Optical Density at 260nm).

2.4.2.2 Purification of DNA from FTA® Paper for use as PCR Template DNA.

FTA® paper is a solid matrix for the long term storage of blood and biological samples that is optimised for bio-safety and DNA technology, in particular, automated PCR. The composition of FTA® is described as the preferred composition of a patent (Burgoyne 1996). Whole blood was dropped onto the FTA® paper, allowed to dry and stored in the dark at room temperature or -80°C. The preservative constituents of the FTA® paper were removed by washing before the DNA was used for PCR. Small discs (1mm diameter) were punched from the blood soaked FTA® paper and placed into 1.5ml tubes containing 200µl phenol. FTA® discs were washed for 10min. with gentle agitation. The phenol was removed and replaced with 200µl of fresh phenol and washed a second time for 10min. with gentle agitation. Phenol was removed by two 5min washes with 200µl of 95% isopropanol, 5% 1M Tris-HCl and gentle agitation. The paper samples were dried in a 50°C oven and used immediately for amplification of arbitrary DNA sequences (Chpt 2.5.3 PCR1, Cycle1).

2.5 Polymerase Chain Reaction

2.5.1 Hardware and Devices

Thermal cycling was performed using GeneAmp PCR System 2400 (Applied Biosystems) for up to 24 samples, FTS 960 (Corbett Research) for up to 96 samples or a RotorGene 2000 (Corbett Research) for real-time PCR. Either individual 0.2mL UltraFlux PCR tubes (Research Products International), 0.2ml ThermoFast skirted 96 well PCR plates (ABgene) or clear 0.2ml thin walled PCR tubes (Axygen Scientific) for real-time PCR were used.

2.5.2 Amplification of 16S-rDNA Loci from Soil-Borne Bacterial DNA.

An approximately 900bp region of the 16S-rDNA loci was amplified from soil DNA using conditions referred to herein as **PCR0**; 200µM of each dNTP, 400nM (20 pmol in 50µl) of 926r primer, 400nM (20 pmol in 50µl) FAM labelled 8f primer, 2.5mM $MgCl_2$, 10mM Tris-HCl, 50mM KCl, 0.1% Triton X-100, 400µg/ml BSA, 1.25U of Taq DNA polymerase (Promega, Madison WI) and 5ng of template DNA in a total volume of 50µl.

This reaction was then subjected to the following thermal cycling regime **Cycle0:**

94˚C 5min;

30X 94˚C 30sec., 50˚C 30sec., 72˚C 90sec;

72˚C 8min.

2.5.3 Amplification of Arbitrary DNA Sequence

The final conditions, cycling and use of primers for AAD profiles were chosen after small pilot experiments which identified appropriate parameters for successful amplification (results not shown).

Arbitrary DNA sequences were amplified in two stages. Products were initially amplified with low stringency under the following conditions, referred to herein as **PCR1**: 200µM of each dNTP, 400nM of a single arbitrary primer (Table 1, pg36), 2mM $MgCl_2$, 1X reaction buffer (20mM $(NH_4)_2SO_4$, 75mM Tris-HCl pH 9 at 25°C, 0.01% Tween), 400µg/ml BSA, 1.25U of thermostable DNA polymerase and 5ng or a 1mm FTA® disc (Chpt 2.4.2.2) of DNA template in a total volume of 50µl.

This reaction was then subjected to the following thermal cycling regime **Cycle1:**

94 °C 5min;

2X 94°C 30sec, 30 °C 3min, 72 °C 3min;

35X 94 °C 30sec, 62 °C 30sec, 72 °C 3min;

72 °C 7min.

A second, high-stringency amplification (**PCR2**) was then performed, with reaction conditions identical to the first stage (PCR1) but with 5µl of the first round amplification products used as template and the thermal cycling regime **Cycle2**:

94 °C 5min;

35X 94 °C 30sec, 62 °C 30sec, 72 °C 3min;

72 °C 7min.

Products of the reaction were then stored at 4 °C or –20 °C until used.

2.5.4 Real-Time Amplification of Arbitrary DNA Sequence

DNA was PCR amplified using Taq DNA polymerase (Promega) under the following conditions, referred to herein as PCR3: 200μM of each dNTP, 400nM of a single arbitrary primer (Table 1, pg36), 2mM $MgCl_2$, 10mM Tris-HCl, 50mM KCl, 0.1% Triton X-100, 400ug/ml BSA, 1X SYBR Green I (Molecular Probes), 1.25U of Taq polymerase and unless otherwise stated 5ng of template DNA in a total volume of 50ul.

This reaction was then subjected to the following thermal cycling regime **Cycle3**:

	95°C 5min;
2X	95°C 30sec., 30°C 3min., 72°C 3min;
15X	95°C 30sec, 62°C 30sec, 72°C 3min (not acquiring);
65-100X	95°C 30sec, 62°C 30sec, 72°C 3min (acquiring to channel 2);
	72°C 7min.

The cycle at which the fluorescence rose above an arbitrary 10% fluorescence threshold was recorded and used to determine the efficiency of amplification of various primers with various sources of template DNA.

2.6 Generation and Analysis of Length Polymorphism Fingerprints

There are a variety of methods (multiple primers, multiple restriction enzymes) and analytical procedures (eg. those that take into account peak heights or simple presence/absence) for band based fingerprint analyses. The method used in this thesis (Chpt 2.6.1) was chosen for being a simple version with an analytical technique (Chpt 2.6.3) that is better suited for discrimination of samples.

2.6.1 Terminal Restriction Fragment Length Polymorphism Fingerprinting

Amplification products from PCR0/Cycle0 (Chpt 2.5.2) reactions were digested with 10 units of the endonuclease *Msp*I (New England Biolabs, Beverly MA) in the PCR buffer, incubated at 37˚C for 16hrs. The digestion was terminated by heat inactivation at 70°C for 1hr.

The digestion products were separated by polyacrylamide gel electrophoresis and fluorescently labelled fragments of 16S sequences detected on an ABI377 genetic analyser (Applied Biosystems). TRFLP profiles were analysed

using GENESCAN™ software (Applied Biosystems) and the size of fragments determined with the use of an internal 1000bp standard. Fragments with peak heights less than 50 fluorescence units were excluded from analyses.

2.6.2 Arbitrarily Amplified DNA Length Polymorphism (AADLP) Fingerprinting

DNA samples to be fingerprinted by band length polymorphism were amplified using a single FAM labelled primer (FAM-Seq005). Amplified products were separated by polyacrylamide gel electrophoresis and fluorescently labelled fragments detected on an ABI377 genetic analyser (Applied Biosystems). Profiles were analysed using GENESCAN™ software (Applied Biosystems) and the size of fragments determined with the use of an internal 1000bp standard. Fragments with peak heights less than 50 fluorescence units were excluded from analyses.

2.6.3 Analysis of Length Polymorphism Data

The tabulated electrophoretic data containing fragment lengths, peak heights and peak areas for each TRFLP and AADLP sample were used to compare profiles using the T-RFLP Analysis program from the Ribosomal Database Project II website http://rdp.cme.msu.edu/html/ (Cole, *et al.* 2003). The output data are similarity indices (SI) based on the Jaccard coefficient of similarity, calculated as twice the number of fragments with common lengths from two samples, divided by the total number of fragments in the two samples.

The similarity index of TRFLP and arbitrarily amplified DNA profiles were calculated using the Jaccard coefficient, which utilises only the presence or absence of peaks and not another method that takes the peak height or peak area into account due to the variation associated with these parameters. Previous studies have shown variation in peak heights of replicate TRFLP profiles electrophoresed on the same gel can average ~10% (range of 4.2% -18.3%) of the peak height (Osborn, *et al.* 2000) or 7% after rigorous standardisation of the data (Dunbar, *et al.* 2001). When compounded over a large number of peaks within each profile, such variation can lead to an appreciably reduced similarity between replicate profiles. The first paper describing the TRFLP technique (Liu, *et al.*

1997) utilized both peak-area sensitive analysis and the Jaccard coefficient to analyse TRFLP profiles generated from four microbial environments. The Jaccard coefficient analysis discriminated the four microbial communities with similarity indices ranging from 38%-48% and grouped two replicate control samples (size references) with 98% similarity. The area sensitive analysis was also capable of discriminating the microbial communities, but the similarity indices were higher (50%-60% similarity) and the control samples showed reduced similarity (95%). Obviously for a forensic application, high SI's for replicates and low similarity indices for distinct microbial communities are desired. Thus, the presence or absence of peaks without taking into account peak height or area was used to determine similarity of profiles.

The significance of similarity index differences between group means was determined by independent measures *t* tests.

2.7 Low Resolution, Conventional Agarose Gel Electrophoresis, Southern Transfer and Hybridisation

5µl of product generated in the second amplification (PCR2, Cycle2, Chpt 2.5.3) and 1µl of 6X loading buffer was electrophoresed through a 1X TAE, agarose (BIORAD) gel in 1X TAE buffer. DNA markers, either 100bp ladder (Figure 2.2, pg47) or *Eco*RI digested SPP-1 phage DNA (Figure 2.3, pg47) were electrophoresed in each gel to provide size and mass references. After electrophoresis, gels were stained in dilute ethidium bromide solution, destained in distilled water, observed and recorded under UV transillumination.

When Southern transfer of DNA to membranes was required, Agarose gels were soaked in transfer buffer (1M NaCl, 0.4N NaOH) for 15min, then again for 20min in fresh buffer. The positively charged nylon membrane, ZetaProbe (BIORAD), was soaked in transfer buffer for 15min prior to capillary transfer for 16hrs using a reservoir of transfer buffer. Membranes were neutralised in Neutralisation Buffer (1M NaCl, 0.5M Tris-HCl, pH 7.2) for 15 min and air-dried.

A second stage, high stringency amplification (PCR2, Cycle2, Chpt 2.5.3) with the addition of 50µCi ^{35}S-dATP was undertaken to produce a radioisotope labelled probe.

Figure 2.2: 100bp ladder supplied by NEB displaying all fragments with corresponding size and mass of each band, assuming 500ng total load.

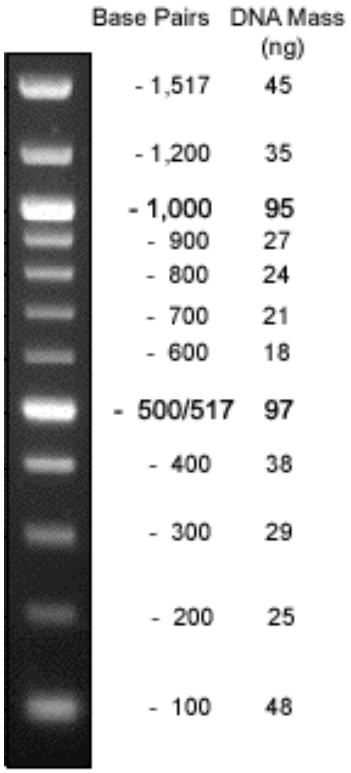
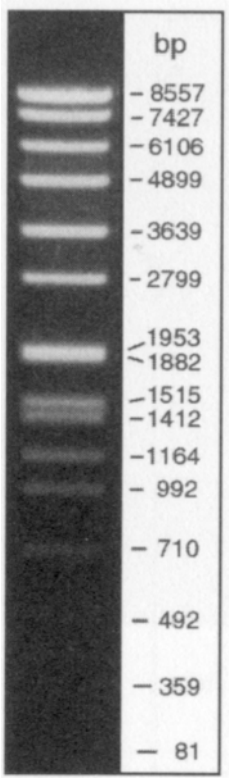


Figure 2.3: SPP-1 Phage (GenBank Accession Number: BSPP1GENM) ***Eco*RI restricted DNA marker** supplied by Geneworks displaying all fragments with corresponding sizes.



Membranes were pre-hybridised in 25ml hybridisation buffer (0.5M sodium phosphate buffer, 2mM EDTA, 7% w/v SDS, 1% w/v BSA fraction V, pH 7.2) rotating at 65°C for at least 1hr. The hybridisation buffer was replaced with 5ml of new pre-warmed hybridisation buffer containing freshly heat-denatured probe.

Hybridisation was carried out at 65°C, rotating for 16hrs. The membrane was washed at 65°C twice in Wash Buffer I (40mM sodium phosphate buffer, 2mM EDTA, 5% w/v SDS, 0.5% w/v BSA fraction V, pH 7.2), then eight times in Wash Buffer II (40mM sodium phosphate buffer, 1mM EDTA, 1% w/v SDS, pH 7.2) each for five minute periods. Membranes were air-dried and exposed to X-ray film X-Omat K (Kodak), developed and fixed after an appropriate exposure time.

2.8 Microarray Techniques

2.8.1 Microarray Preparation and Printing

2.8.1.1 The Pilot Microarray

A pilot microarray was constructed containing 12 replicates of 73 AAD profiles from various soils and 47 control spots, giving 120 different DNA spots, 1440 spots in total on each array. The profiles generated from soil samples were amplified from various amounts of template DNA ranging from 250pg to 25ng in order to best identify the optimal amount of template DNA for AADSS analysis.

Second round arbitrarily amplified profiles, generated from soil DNA and control DNA samples (PCR2, Cycle2, Chpt 2.5.3), were purified using UltraCleanTM PCR purification DNA clean up kits (MoBio Laboratories) according to the manufacturer's instructions, quantified using Sybr Green I (Molecular Probes) with a standard curve made from a dilution series of DNA of known concentrations. Equal amounts of each profile and control samples (2µg) were dried and provided to the Adelaide Microarray Facility to be printed to microarray slides. DNA was spotted in a solution of 50% dimethyl sulfoxide (DMSO) to 15 slides (GAPS II amino-silane coated slides, Corning).

2.8.1.2 The Final, Prototype Microarray

The final, prototype microarray was constructed with 725 AAD profiles and 43 controls, spotted 12 times on each microarray slide, giving a total of 9,216 elements per array.

Second round arbitrarily amplified profiles, generated from soil DNA and control DNA samples (PCR2, Cycle2, Chpt 2.5.3), were purified using NucleoFast 96 PCR plates (Machery-Nagel), quantified using Sybr Green I (Molecular Probes) with a standard curve made from a dilution series of DNA of known concentrations. Equal amounts of each profile and control samples (2µg) were dried and provided to the Adelaide Microarray Facility to be printed to microarray slides. DNA was spotted in a solution of 50% dimethyl sulfoxide (DMSO) to 20 slides (GAPS II amino-silane coated slides, Corning).

2.8.2 Quality Control of Microarray Printing

Two slides, the first and last of the print run, were stained with the Sybr Green II (Molecular Probes) to assess the quality and reproducibility of the array spots. Mean fluorescence minus background fluorescence values were used to determine the reproducibility of the quantity of DNA deposited at each spot.

2.8.3 Labelling and Hybridisation of DNA profiles

Arrays were probed with labelled profiles. To label a profile, the second round amplification (PCR2, Cycle2, Chpt 2.5.3) was performed in the presence of 2nmole Cy3-dCTP or Cy5-dCTP (Amersham Biosciences) and only 80µM unlabelled dCTP. Unbound Cy3-dCTP or Cy5dCTP, dNTPs and primer were removed using UltraCleanTM PCR purification DNA clean up kit (MoBio Laboratories) according to the manufacturer's instructions. Unlabelled primer (10pmol) was added to the labelled profile to block the labelled primers from cross-hybridising to primer sequences of profiles amplified with the same primer. Hybridisation buffer was added (a final concentration of 25% formamide, 2.5X SSC and 0.1% SDS) and the labelled profile denatured by heating to 100°C for 10min followed by 2min on ice. After 45min. incubation with pre-hybridisation buffer (10mg/ml BSA, 25% formamide, 5X SSC and 0.1% SDS) at 45°C, the

array was hybridised with denatured probe at 45°C for 6-8hrs. The array was then washed at 45°C for 5min in 1X SSC, 0.2% SDS, 5min in 0.1X SSC, 0.2% SDS and twice for two min. each time in 0.1X SSC, dried and scanned in a GenePix 4000A (Axon Instruments).

2.8.4 Analysis of Microarray Data

All spots deemed "bad elements" by the software package GenePix™ Pro 3.0 were removed from further analysis. Mean fluorescence minus the background fluorescence of the remaining replicate spots were averaged, then values were normalised, relative to the most fluorescent profile on the array image, to give relative fluorescence units (RFU) before further data analysis. The significance of differences between two group means was determined using non-parametric Mann-Whitney *U* tests.

2.9 Cloning of Arbitrarily Amplified DNA Profile Sequences

2.9.1 Ligation Reactions

Ligation of arbitrarily amplified products into pGEM-T Easy vectors (Promega) was performed according to the suggested protocol. 50ng of vector and approximately 50ng of PCR product (~1:3 molar ratio of vector:insert) were ligated using 3 Weiss units of T4 DNA ligase in a final volume of 10μl at 4°C overnight.

2.9.2 Transformation of Ligated Vectors into Competent E. coli JM109.

2.9.2.1 Preparation of Competent Cells

2.5ml of Luria Broth (LB) (10g NaCl, 5g Bacto Yeast Extract, 10g Bacto Tryptone in 1L H_2O) media was inoculated with a single *E.coli* JM109 colony (the genotype of JM109 is *rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17(r_{K-},m_{K-}), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*I^qZΔM15]) (Messing, *et al.* 1981) from an LB plate (1.5% w/v agar) and incubated overnight at 37°C, shaking at

approximately 225rpm. The entire 2.5ml culture was then used to inoculate 250ml of LB medium containing 20mM $MgSO_4$. The cells were grown in a 1L flask until the A_{600} was between 0.4-0.6 above the LB media's absorbance. The cells were pelleted by centrifugation at 4500g for 5 min at 4°C. Cells were then gently resuspended in 100ml of filter sterilised TFB1 (10mM $CaCl_2$, 50mM $MnCl_2$, 100mM RbCl, 30mM potassium acetate and 15% glycerol, pH 5.8) and kept on ice for 5 min. The cells were pelleted by centrifugation at 4500g for 5 min at 4°C. Cells were gently resuspended in 10ml of filter sterilised TFB2 (10mM MOPS pH 6.5, 75mM $CaCl_2$, 10mM RbCl, 15%glycerol, pH 6.5) and incubated on ice for 15-60min. 200µl aliquots were then dispensed, quickly frozen in liquid nitrogen and stored at –80°C.

2.9.2.2 Transformation of Competent Cells

3-5µl of ligation products (Chpt 2.9.1) was added to 100-200µl of competent *E. coli* strain JM109 in 1.5ml tubes on ice which were gently mixed then incubated for 20 min on ice. The transformation mixture was heat shocked for 90 sec in a 42˚C water bath and returned to ice for 2 min. 1ml of LB or SOC media (2g Bacto Tryptone, 0.5g Bacto Yeast Extract, 1ml 1M NaCl, 0.25ml 1M KCl, 97ml deionised water, autoclaved and cooled to room temperature before addition of 1ml of filter sterilised 2M Mg^{++} stock [1M $MgCl_2$, 1M $MgSO_4$] and 1ml filter sterilised 2M glucose, pH7.0) was added and incubated at 37˚C for 60 min shaking at 150rpm. 150µl, 300µl and a rest of mixture aliquots were plated onto selective LB plates (1.5% w/v agar) containing 100µg/ml ampicillin, 0.5mM IPTG and 80µg/ml X-Gal for blue/white selection. Plates were incubated for 16-24 hrs at 37˚C.

2.9.3 Isolation of Plasmid DNA from Recombinant Hosts.

Plasmid DNA was extracted and purified from the recombinant host *E.coli* cells using UltraClean™ Mini Plasmid Prep Kits (MO BIO Laboratories) as per the manufacturer's instructions. Briefly, 5ml of LB media containing 100µg/ml ampicillin was inoculated with a single colony from the transformation plates and

incubated at 37˚C for 16-24 hrs. or until the absorbency was approximately 0.6 at 600nm. 1.5ml of culture was transferred to a 1.5ml tube and centrifuged at 10,000g for 30 sec. The supernatant was discarded and the pellet of cells resuspended in 50µl of Solution 1 (a resuspension buffer containing RNase A) by vortexing. 100µl of Solution 2 (an alkaline lysis buffer [pH12] containing the detergent SDS) was then added and gently mixed by inversion. 325µl of Solution 3 (neutralisation buffer containing potassium acetate and salt) was added and gently mixed by inversion. The tubes were centrifuged at 10,000g for one minute and the supernatant transferred to the spin filter tube provided with the kit. These spin filter tubes were centrifuged at 10,000g for 30 seconds and the eluate discarded. 300µl of Solution 4 (washing solution containing approximately 50% ethanol) was added to the spin filter and centrifuged for 30 seconds at 10,000g, the eluate discarded and the dry spin filter containing the DNA was transferred to a new collection tube. 50µl of Solution 5 (10mM Tris) was added directly to the spin filter, which was then centrifuged at 10,000g for 30 seconds to elute the plasmid DNA.

2.9.4 Sequencing Reactions.

All sequencing reactions were prepared by adding 800ng of plasmid template to 6.4pmol of a single primer in water to a total volume of 16µl. Primers T7 and Sp6 (see Table 1, pg36) were used to sequence from the pGEM-T Easy vector into the arbitrarily amplified DNA insert. Sequencing reactions and gel separation was performed by the Australian Genome Research Facility (AGRF), Brisbane. Sequence data were viewed and edited using the program Sequencher™ version 4.0.5 (GeneCodes Corporation).

3 RESULTS AND DISCUSSION

3.1 Terminal Restriction Fragment Length Polymorphism (TRFLP) Analysis

3.1.1 Soil Profile Comparisons

DNA derived from the microbial communities in morphologically distinct soils (A, B and C), sourced from three geographically separate sites 2-5km apart, were collected over a six-month period and used to generate TRFLP profiles (PCR0, Cycle0, Chpt 2.5.2). An example of an electrophoretogram of a TRFLP profile generated from Soil A template is shown in Figure 3.1 (pg54). The profiles were compared in pair-wise combinations and a similarity index (SI) calculated (Chpt 2.6.3 for SI computation, see Appendix CD\TRFLP\TRFLP SI Matrix.xls for all Day0 profile comparisons and SI's). Pairs of TRFLP profiles generated from the same site and time had a mean SI of 0.843 while profiles generated from DNA's of different origins had a mean similarity index of 0.693 (Table 2, pg55). The mean SI values of replicates of the same soil sample were 0.761 for Site A, 0.850 for Site B and 0.867 for Site C (Table 2, pg55). Despite the low difference in SI values between replicate (0.843) and different (0.693) site comparisons, the profiles of all three replicate groups were significantly different to distinct soils when means were compared using an Independent measures *t* test (Table 2, pg55). The SI values presented in Table 2 (pg55) are in accord with the published data of Tiquia *et al.* (2002) who found SI's of 0.81-0.91 among triplicate samples and Horswell *et al.* (2002) who obtained a SI of 0.91 for a single comparison of soil from the sole of a shoe, to soil from the shoe-print it left behind and SI values averaging 0.612 for comparisons of profiles from the five soil samples used in the Horswell e*t al*. (2002) study. However, it must be noted that the use of alternate oligonucleotide primers and restriction endonucleases between these studies will alter the profiles generated and this must be kept in mind when comparing results from other studies.

Figure 3.1: An electrophoretogram of a TRFLP profile (Chpt 2.6.1) generated from site A soil DNA.

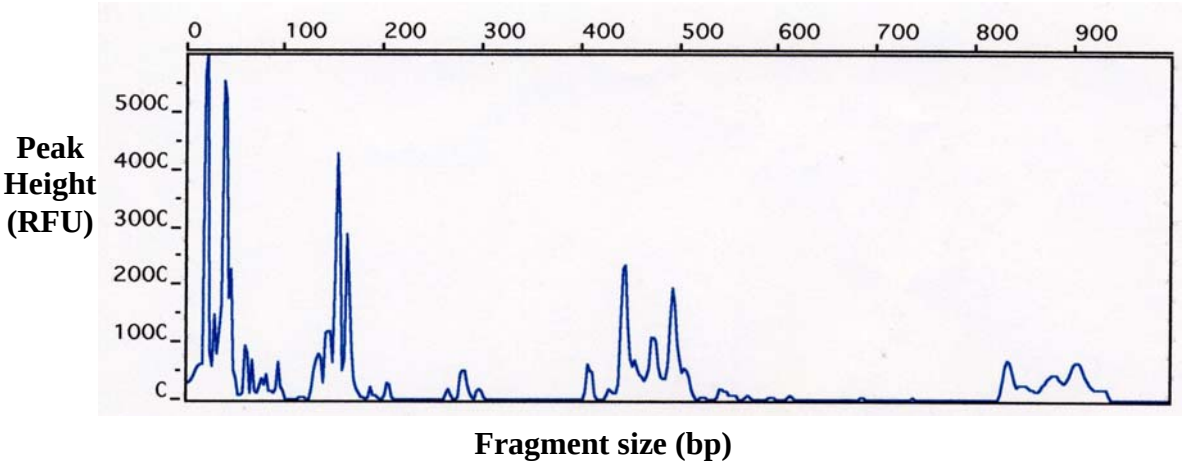


Table 2: Distinguishing soils from each other by TRFLP. Mean similarity index (SI) (Chpt 2.6.3) of replicate and inter-soil comparisons of profiles from three soils (A, B and C) generated with amplification conditions PCR0 and Cycle0 (Chpt 2.5.2).

I	II	III	IV
Profile Comparisons	**Mean Similarity Index (SI) ± S.D.**	**Number of pair-wise comparisons (*n*)**	**Significance of mean SI difference to different soils (*P*)**
Soil A replicates	0.761 ± 0.073	6	$1.33x10^{-3}$
Soil B replicates	0.850 ± 0.026	10	$4.79x10^{-17}$
Soil C replicates	0.867 ± 0.051	18	$1.10x10^{-24}$
Combined Same Soil Comparisons	**0.843 ± 0.062**	**34**	**$4.93x10^{-26}$**
Soil A vs. Soil B	0.675 ± 0.044	20	
Soil A vs. Soil C	0.690 ± 0.043	24	
Soil B vs. Soil C	0.713 ± 0.041	30	
Combined Different Soil Comparisons	**0.693 ± 0.045**	**74**	**-**

I The nature of the profiles compared are listed. Soil A, Soil B and Soil C indicate the comparison of replicate profiles of individual soils. Soil A vs. Soil B, Soil A vs. Soil C and Soil B vs. Soil C indicate the comparison of profiles of two distinct soils. Combined Same Soil Comparisons and Combined Different Soil Comparisons are the comparison of all profiles from replicate or distinct soils respectively.

II Mean Similarity Index (SI) values of each comparison category are listed with the standard deviation (S.D.).

III The number (*n*) of paired profile comparisons performed for each comparison category.

IV *P* values, determined by independent measures *t* tests (α = 0.05, 2 tailed), when comparing the difference of SI values generated from replicate profiles of individual or combined soils, to Combined Different Soil Comparisons SI values.

At first sight, it appears that TRFLP profiles are poor discriminators of soil communities as the similarity statistic (SI) for replicate samples of the same soil is not greatly different to the SI of different soils. However, the statistical significance of differences between mean SI values of replicate samples and mean

SI values of soils of different origins has been determined in this thesis and were found to be significant, with a low probability of error (P=4.93x10^{-26}), largely due to the degree of replicate analyses performed. The PCR based nature of the technique allows multiple replicate analyses to strengthen the power of the statistical tests even from trace samples, as many profiles can be generated from a small amount of soil.

The small difference between SI's of identical and different soils is the main shortcoming of the TRFLP technology and may arise from technical variations including PCR biases, slight variation in template concentration (Kozdroj and van Elsas 2000) humic acid contamination (LaMontagne*, et al.* 2002) and potentially, pseudo-TRF's generated by secondary structure (Egert and Friedrich 2003). These may all lower the similarity index of identical soil samples. Oligonucleotides designed to prime amplification of conserved sequences, may yield amplicons from a similar subsection of organisms, reducing the actual diversity of soils observed using this technology. Moreover, the assumption that TRF's of identical length are identical between samples, regardless of the sequence between the primer and restriction site, misleadingly increases the similarity of different soils as polymorphisms exposed by a single restriction endonuclease are likely to underestimate total polymorphism within the sequence.

The discrimination power of TRFLP between three distinct soils was investigated and presented in this thesis. Whilst TRFLP analysis was able to distinguish the three soils, proving its potential applicability for discriminating between soils, TRFLP will need to be trialled on a wider range of soils of similar type or usage (e.g. multiple agricultural soils under identical crops). It may be that similar soil types or soils with similar vegetation harbour similar microbial communities, yielding highly similar TRFLP profiles. This possibility is supported by a previous study which concluded from physiological and 16S rDNA (DGGE and TRFLP) data that soil type was the key factor in determining bacterial community composition, while cropping regimens and management strategies played minor roles in determining bacterial community composition (Girvan*, et al.* 2003). Alternatively, even though such soil communities may share functional or phenotypic similarity, their genotypic make-up may be distinguishable by TRFLP. In a previous study (Smalla*, et al.* 2001) it has been shown by DGGE analysis of 16S DNA that sowing two lots of a uniform soil with different plant species can

rapidly alter the microbial community make-up, particularly in the rhizosphere. Different agricultural practices such as tillage (compared to zero tillage) and crop rotation (continuous wheat, wheat/summer fallow, wheat/red clover, wheat/field peas) have also been shown to cause divergence of microbial communities within a small trial field (Lupwayi, *et al.* 1998). The influence of plant species and management strategies over the microbial make-up of soils indicates that even highly similar soils may contain different microbial communities. This gives great power to biological methods of soil comparison, which may be capable of narrowing the range of matching samples beyond the capabilities of physical and chemical methods.

3.1.2 Technical Variation of TRFLP Profiles

Some sources of technical variation including i) DNA extraction and purification from soil and ii) PCR replication were evaluated. The reduction in similarity indices of TRFLP profiles generated from replicate DNA extractions (0.839), from those generated using the same DNA extract (0.861), was not significant (Table 3 pg58, see Appendix CD\TRFLP\TRFLP SI Matrix.xls for all Day0 profile comparisons and SI's).

Table 3: Effect of technical sources of variation on similarity indices and significance values for each source of variation.

I	II	III	IV
Technical Variation	**Mean Similarity Index (SI) ± S.D.**	**No. of pair-wise comparisons (*n*)**	**Significance of mean SI difference to (1) (*P*)**
(1) Same DNA Extraction Same PCR Reagent Batch	0.861 ± 0.036	6	-
(2) Different DNA Extraction Same PCR Reagent Batch	0.839 ± 0.066	28	0.449
(3) Same DNA Extraction Different PCR Reagent Batch	0.727 ± 0.050	6	0.018
(4) Different DNA Extraction Different PCR Reagent Batch	0.727 ± 0.038	12	0.002

I The source of technical variation. Profiles were generated using either the same or different DNA extraction replicates and PCR reagent batches.

II Mean Similarity Index (SI) values of profiles from the same Soil DNA for each technical variation source are listed with the standard deviation (S.D.).

III The number (*n*) of paired profile comparisons performed for each source of technical variation.

IV *P* values, determined by independent measures *t* tests (α = 0.05, 2 tailed), when comparing the difference of SI values generated with differing conditions (rows 2-4) to profiles generated with identical conditions (row 1).

However, profiles obtained from amplifications using a different batch of DNA polymerase and PCR reagents differed significantly from replicates using the same batch of reagents, although the SI values for profiles generated using different reagent batches were the same regardless of DNA extraction (0.727). The statistical differences, determined by an independent measures *t* test, were greater when replicate DNA extracts were combined with different reagent batches (row 4, Table 3, pg58). The difference in mean SI of the technical variants in the case of varying DNA extraction and PCR batches was the same as the difference between that technical variation and a completely different soil. This makes allocating such a result to a group of either ‘same’ or ’different’

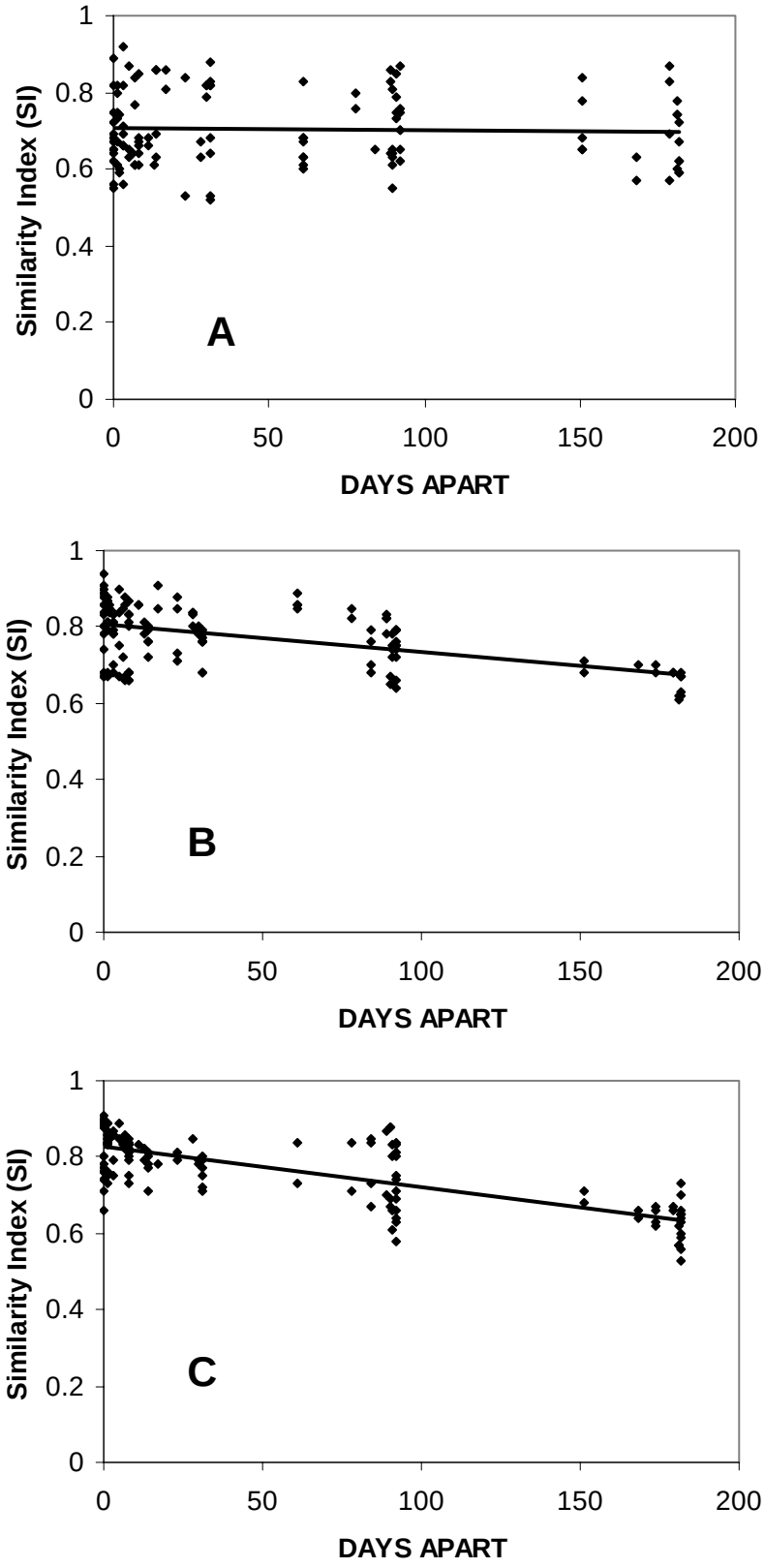
impossible as it is significantly different to both groups in equal amounts (P=0.002, independent measures t test). In this study, the sample size of this particular source of variation is small, only one soil (Site C) was tested on two occasions. A previous study determined that potential sources of variation such as electrophoresis (intra-gel and inter-gel), restriction digests, PCR and DNA isolation only introduced very small changes in TRFLP profiles, but the use of different brands of *Taq* polymerase dramatically altered the number of fragments generated (Osborn, *et al.* 2000). This will need to be further evaluated and if found to be a persistent source of variation, stringent standardisation employed.

3.1.3 Temporal Variation of TRFLP Profiles

Comparison of profiles obtained from samples from site B and also from site C showed increasing divergence between samples taken over a six-month period (Figure 3.2B and 3.2C, pg60). This supports previously published conclusions that time allows populations of some soil communities to change appreciably (Lukow, *et al.* 2000; Horswell, *et al.* 2002).

Similarity indices generated from samples taken 179, 181 and 182 days apart were grouped into a six-month category. Comparison of six-month SI values at each of sites B (0.651) and C (0.625) differed significantly from Day 0 replicates of the corresponding soil (P=1.97x10^{-11} and 6.93x10^{-16} respectively, independent measures t test). In contrast, over the six-month period, profiles generated from site A (Figure 3.2A, pg60) did not differ significantly (P= 0.154, independent measures t test), despite a mean SI value of 0.688 for samples taken six months apart, lower than the distinct soil SI (0.693). Site A may indeed harbour a relatively temporally stable community, which has been observed on at least one other occasion (Mummey and Stahl 2003). However, it is difficult to reach this conclusion for Site A as the SI of samples taken six-months apart, is lower than soils taken from distinct locations. Analysis of additional fingerprints may increase the strength of the independent measures t test performed, allowing significant differences to be observed.

Figure 3.2: Trends of TRFLP similarity indices with temporal distance. TRFLP profiles were generated from three soils (A, B and C) separated by 2-5km). All profiles generated from soils sampled over six-months from three individual sites were compared and the SI plotted against days apart.



The likelihood that Site A profiles actually do change over time is reflected in the *P* value determined when comparing SI values from distinct soils, sampled on the same day, and Site A samples taken six-months apart (P=0.560, independent measures *t* test). Thus, the SI values generated from site A soils taken six-months apart are not significantly different to SI values generated from distinct soils. When SI values from samples taken six-months apart from Sites B and C were compared to SI values of distinct soil types, sampled on the same day, significant differences were observed (P= 0.003 and 5.03×10^{-8}, respectively, independent measures *t* test). Thus, samples taken six-months apart from either Site B or C show a greater difference than samples taken from different locations, on the same day. This result suggests that despite being separated by distance (2-5km), Sites A, B and C may share seasonally dependant bacterial population subsets, increasing the similarity of TRFLP profiles from the three sites. However, bacterial populations change and it appears that TRFLP profiles reflect more intense bacterial population changes over six-months within one site than between sites 2-5km apart at one time.

This thesis has not examined the question of seasonal cycling of microbial communities using TRFLP, as samples taken more than six-months apart were amplified with a different batch of PCR reagents which was found to affect the similarity index (Table 3, pg58). The effect of seasonal cycling on TRFLP profiles should be experimentally investigated as it is possible that although the microbial communities may be appreciably different after six months, they may regain similarity after twelve months due to season dependent selection of micro-organisms. However, if some soils exhibit temporally stable microbial communities such as site A appears to do, then each site may need to be assessed on a case by case basis until more information is known about which environments harbour such communities. For those soil microbial communities that change over time, TRFLP technology will be limited to cases where the two samples for testing are collected within a timeframe that allows valid comparison. The data presented in this thesis suggest that the allowable time frame can be less than six months. This potential limitation could also be an advantage. If two soils have a similar TRFLP profile, it can be suggested that not only do the soils come from the same location but it also within a specifiable time frame.

3.1.4 Conclusions for the TRFLP Technique

The data presented in this thesis suggest that TRFLP analysis is capable of discriminating soil samples based on the DNA of the microbial community residing within and that the make-up of these communities can change over time. Perhaps the most crucial area for improvement is the reproducibility of replicates between batches of reagents. The data presented here are not encouraging in that respect. However, with a sample size of only two reagent sets, it is by no means a comprehensive study of this particular source of variation. The SI's of different soils differed less than is ideal for discrimination between soils, making discrimination of soils less obvious or self-evident, necessitating statistical analysis for the distinction of soils. It has been suggested (Liu, *et al.* 1997; Clement, *et al.* 1998) that the discriminatory power of TRFLP can be improved by separately digesting amplified products with several restriction endonucleases. This would reduce the number of identical TRF's, by enabling amplicons that give identical TRF's with one enzyme to be distinguished by the use of another**.** Also, the use of distinct labels at both ends of the PCR product would give an additional level of information allowing better discrimination. These suggested improvements will add a significant cost to the process of soil comparison but will not require collection of larger amounts of soil, a limiting factor in some forensic investigations. They will not however, escape the problem of temporal changes

Unlike many of the other chemical and physical methods of soil analysis, TRFLP does not require a high degree of expertise to perform, is highly amenable to automation and does not have any subjective component in its analysis. TRFLP shows moderate promise as an additional tool to complement existing methods of soil analysis for the forensic sciences.

3.1.5 A Soil TRFLP Profile Database is Not Currently Feasible.

A major strength of TRFLP is that raw data of profiles can be easily stored in a database for comparison to new profiles generated, reducing labour and progressively increasing the power of matches and exclusions as more profiles are added to the database and available for comparison. This database would be of immense value to forensic investigators as such a tool could be available to State and National centres. However, two major fundamental issues must be satisfied

before the value of such databases can be realised. Reproducibility and the temporal stability of profiles must be high in order to justify the creation of a soil profile database. A high degree of reproducibility is essential to accurately include or exclude profiles as matching and temporally transient profiles will not allow "mapping" of areas as the maps will be continually changing. The effort to maintain a database where reference soil fingerprints need to be continually updated, to provide current profiles required for comparison, will be prohibitively costly and laborious. This thesis has identified problems with both reproducibility and temporal stability of soil TRFLP profiles. However, these issues have little impact on the ability of TRFLP to discriminate soils, but are only problematic in regard to the ability to database profiles.

3.2 Arbitrarily Amplified DNA Profiles

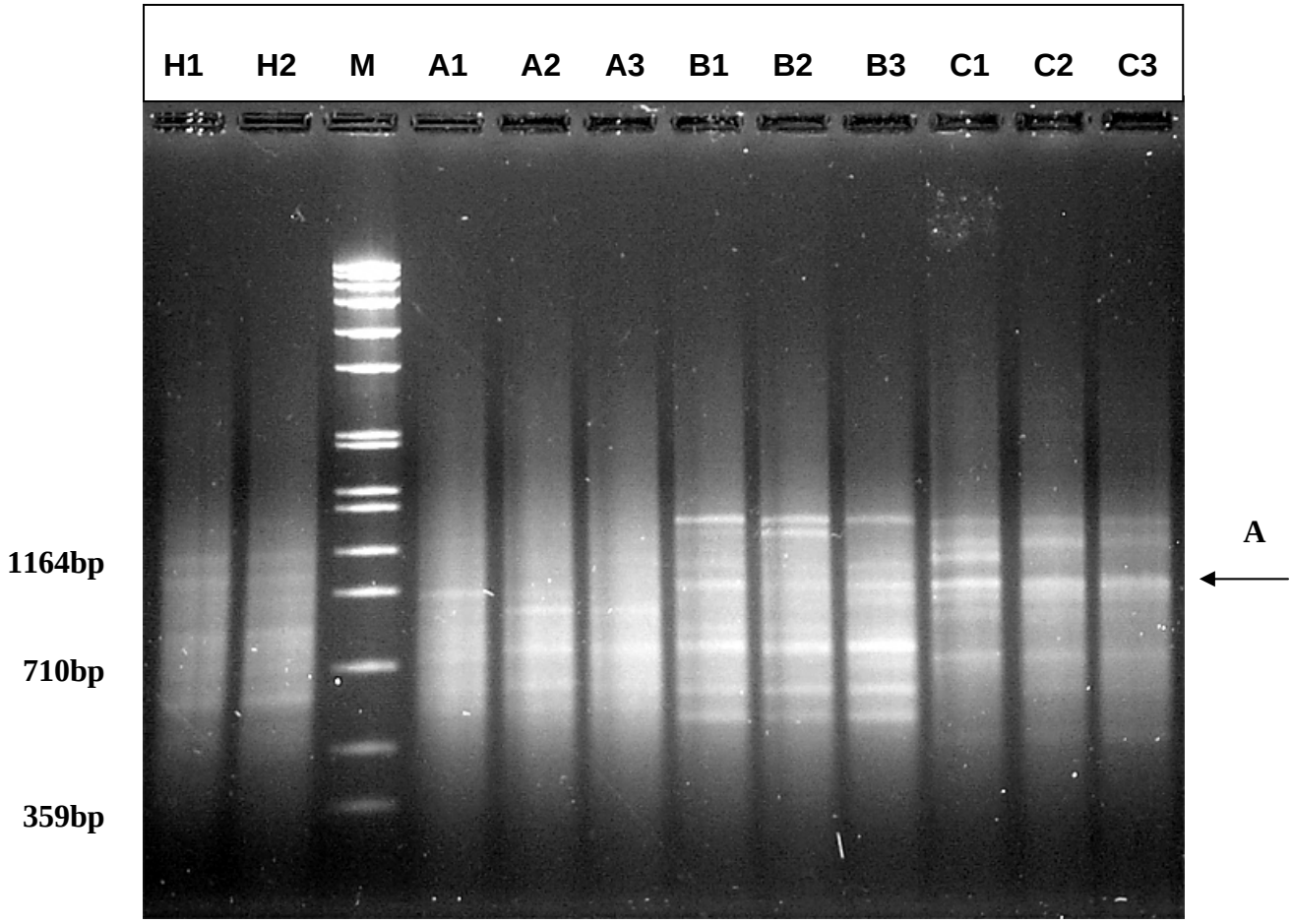
DNA isolated from soils, human, SPP-1 phage as well as "No Template" controls were amplified to produce AAD profiles (Chpt 2.5.3). Some of the amplicons were cloned using pGEM T-Easy vectors and sequenced (Chpt 2.9) to examine the types of sequences that were being amplified. A selection of these sequences are compiled and presented in the Appendix CD (Sequences.doc).

Two contrasting methods of analysis were tested for arbitrarily amplified DNA profiles. The first was a standard length polymorphism method similar to the analytical method used for TRFLP analysis. The second was to compare the sequences amplified, by cross-hybridisation.

3.2.1 Electrophoresis and Southern Analysis

Profiles generated from primer Seq005 (Table 1, pg36) and the DNA of three soils and one human (for a control) were resolved on an agarose gel (Figure 3.3, pg64). A similar pattern of bands or "fingerprint" was generated for each DNA source. These fingerprints were used to distinguish different and identical soil DNA communities (Chpt 3.2.2). This thesis has not relied on length polymorphisms alone to imply similarity, but has extended the technique to look at sequence similarity between samples to determine the similarity of biological communities.

Figure 3.3: Length polymorphisms of arbitrarily amplified DNA from replicate Human genomic DNA (H1 and H2), triplicate DNA extractions from Soil A (A1-3), Soil B (B1-3) and Soil C (C1-3) using primer Seq005 and the amplification conditions PCR2 with Cycle2 (Chpt 2.5.3). *Eco*RI digested SPP-1 is the size-marker (M) (see Figure 2.3 for all band lengths). This gel was subsequently probed with ^{35}S-labelled B1 (see Figure 3.4 overleaf). The arrow identified with A identifies bands of similar length between samples B and C and is discussed further on page 65.



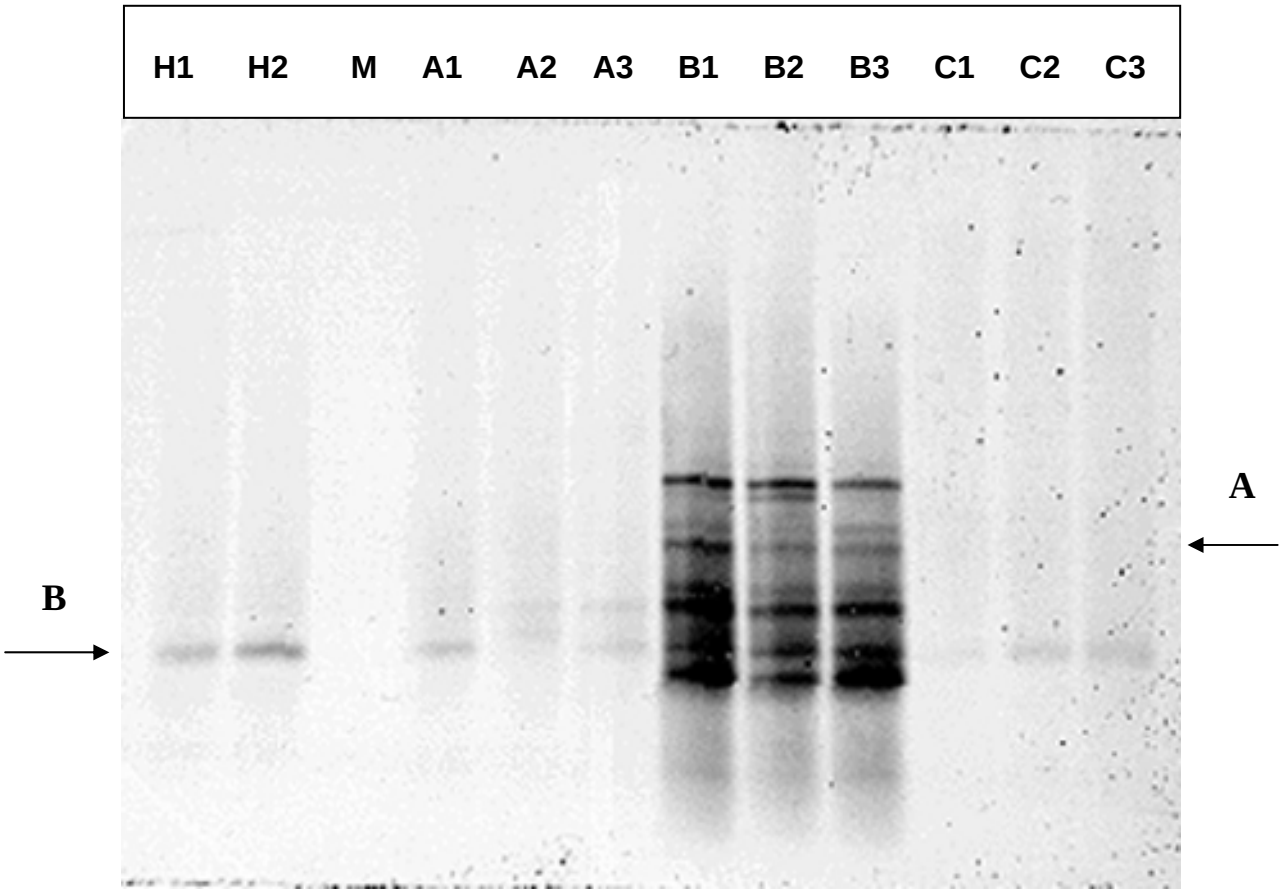
This is because length polymorphism analyses do not utilise the bulk of the information within sequences, only the sequence of the restriction site, while sequence similarity, at least in theory, utilises all of the information available.

DNA profiles were Southern blotted to nylon membrane and probed with a ^{35}S-labelled profile (see Chpt 2.7 for methods) generated from Soil B, extraction 1 (B1). The autoradiograph (Figure 3.4, pg66) shows that signal is predominantly sequestered at profiles from Soil B, indicating DNA profiles from soil B share a high degree of sequence similarity, as expected, while DNA profiles from other soils and the human sample share very little sequence similarity to the B1 profile with the exception of one fragment ~600bp in length (arrow B, Figure 3.4, pg66) that is present in AAD profiles from all DNA sources. This fragment may be common to all DNA sources or could have been amplified from DNA contaminating the PCR reagents (discussed in more detail below in Chpt 3.4.2).

Sequence based comparisons do provide better data in the way expected because amplicons of equal or similar lengths have been generated that do not appear to share any real sequence similarity (see arrow A, Figures 3.3, pg64 and 3.4, pg66). The presence of such bands that would be "matched" by length polymorphism analysis alone, will contribute to the calculated similarity of the two samples if length polymorphism is the only consideration, but by utilising hybridisation techniques, such inaccuracies may be avoided. Also, sequence that is amplified in low amounts may be missed when looking at length polymorphisms, but the sensitivity of hybridisation techniques allow amplicons with sequence similarity to be observed where otherwise they may not have been recognised (arrow B, Figure 3.4, pg66).

Profiles were also deposited directly onto membranes to produce a macroarray of soil DNA profiles. These simple macroarrays were then hybridised with ^{35}S labelled profiles of Site A, B or C DNA under the same conditions as the Southern transferred DNA membranes (Chpt 2.7). Macroarrays were also used to determine the relative amounts of label that binds to each profile. In all three macroarray hybridisations, the profiles with the same soil origin as the labelled profile have the strongest signal (Appendix CD\Macroarray.doc), however, increased levels of background signal relative to the Southern membrane experiments were observed.

Figure 3.4: Sequence similarity of arbitrarily amplified DNA profiles as determined by Southern analysis of the gel shown in Figure 3.3. Replicate Human genomic DNA (H1 and H2), triplicate DNA extractions from Soil A (A1-3), Soil B (B1-3) and Soil C (C1-3), amplified using primer Seq005 and the amplification conditions PCR2 with Cycle2 (Chpt 2.5.3), and unamplified *Eco*RI digested SPP-1 size-marker (M), were probed with ^{35}S labelled B1. See Figure 3.3 for entire DNA profiles. The arrow identified with A identifies bands of similar length between samples B and C and is discussed further on page 65. The arrow identified with B identifies bands of similar length and sequence between all samples and is discussed further on page 65.



3.2.2 Length Polymorphism Comparisons of Arbitrarily Amplified Soil DNA Profiles

DNA derived from the biological communities in the same soils (Sites A, B and C, see Figure 2.1 pg38) used for TRFLP analysis (Chpt 3.1) were used to amplify a subset of arbitrary sequences with primer Seq005 (Table 1, pg36) to generate a profile representative of the existing biota.

The length polymorphisms between pairs of profiles generated from the same site gave a mean similarity index of 0.778, while profiles generated from DNA samples of different origins had a mean similarity index of 0.147 (Table 4, pg68). Appendix CD\AADLP\AADLP SI Matrix.xls contains all Day0 profile comparisons and SI values. The mean SI values of replicates of the same soil sample were 0.789 for Site A, 0.836 for Site B and 0.707 for Site C. The profiles of each of the three soils were significantly different to profiles from distinct soils when compared using an independent measures *t* test (Table 4, pg68) indicating the potential utility of this general method.

Previous studies using Random Amplified Polymorphic DNA (RAPD) for soil microbial community analysis have not investigated the similarity indices of replicate profiles. However, Yang *et al*. (Yang, *et al.* 2000) found that profiles generated using 12 short (10mer) random primers, from four different soils with varying levels of pesticide and chemical fertilizer contamination, gave similarity indices between 0.455 and 0.655, well above the similarity indices between the three soils of this study. This suggests that despite the use of only one primer, for the purpose of excluding distinct soils from a match, the particular arbitrary amplification presented in this thesis (PCR 2 and 3, Cycle 2 and 3) may be superior. However, the use of different soils means that these studies are not directly comparable, as the approach presented in this thesis may indeed yield different results when applied to the soils actually used by Yang *et al.* (2000).

Table 4: Distinguishing soils from each other by Arbitrarily Amplified DNA profiles. Mean similarity index (SI) (Chpt 2.6.3) and relative fluorescence units (RFU) (Chpt 2.8.4) of replicate and inter-soil comparisons of profiles from three soils (A, B and C) generated with primer Seq005 and amplification conditions PCR2 and Cycle2 (Chpt 2.5.3).

I	II	III	IV	V	III	VI
Profile Comparisons	Mean Similarity Index **SI** ± (S.D.).	Number of pair-wise comparisons (*n*)	Significance of mean SI difference to different soils (*P*, *t* test)	Mean Relative Fluorescence **RFU** ± (S.D.).	Number of replicates (*n*)	Significance of mean RFU difference between different soils (*P*, Mann-Whitney *U*)
Soil A replicates	**0.789** (0.095)	66	5.4×10^{-174}	**0.706** (0.163)	46	2.21×10^{-25}
Soil B replicates	**0.836** (0.079)	66	5.7×10^{-189}	**0.663** (0.298)	18	3.41×10^{-12}
Soil C replicates	**0.707** (0.086)	66	2.1×10^{-152}	**0.710** (0.156)	22	2.35×10^{-14}
Combined Same Soil Comparisons	**0.778** (0.102)	198	10^{-289}	**0.697** (0.195)	86	4.05×10^{-39}
Soil A vs. Soil B	**0.248** (0.088)	144		**0.031** (0.029)	64	
Soil A vs. Soil C	**0.055** (0.064)	144		**0.072** (0.034)	68	
Soil B vs. Soil C	**0.136** (0.086)	144		**0.028** (0.026)	40	
Combined Different Soil Comparisons	**0.142** (0.112)	432	-	**0.047** (0.037)	172	-

I The nature of the profiles compared are listed. Soil A, Soil B and Soil C indicate the comparison of replicate profiles of individual soils. Combined Same Soil Comparisons and Combined Different Soil Comparisons are the comparison of all profiles from replicate or distinct soil profiles respectively.

II Mean Similarity Index (SI) and standard deviation (S.D.) of each comparison category is listed.

III The number (*n*) of paired profile comparisons performed for each comparison category.

IV *P* values, determined by independent measures *t* tests ($\alpha = 0.05$, 2 tailed), when comparing the difference of SI values generated from replicate profiles of individual or combined soils, to Combined Different Soil Comparisons SI values.

V Mean Relative Fluorescence Units (RFU) and S.D. of each comparison category is listed.

VI *P* values, determined by Mann-Whitney *U* tests when comparing the difference of RFU values generated from replicate profiles of individual or combined soils, to Combined Different Soil Comparisons RFU values.

The similarity between microbial community RAPD profiles from other environments, such as termite mounds constructed by different termite species, was between 0.21 and 0.46 (Harry, *et al.* 2001). Franklin *et al*. (Franklin, *et al.* 1999) found replicate comparisons of RAPD profiles generated from creek water communities gave similarity indices ranging from 0.83 to 0.97, while profiles from different creeks gave similarity indices of 0.30. These statistics are similar to those determined in this study, although are not directly comparable due to the different types of microbial communities and the use of different primers and amplification regimes. This thesis has for what appears to be the first time, demonstrated the repeatability and discriminatory power of length polymorphism fingerprints generated by arbitrary amplified DNA from whole soil communities.

All previous RAPD analyses utilised multiple amplifications with a set of primers to generate many more 'bands' per sample than this study. Despite the simple, single primer amplification and analysis of only ten bands per sample, the length polymorphism analysis employed in this thesis was reproducible and capable of discriminating the different soil community profiles.

A major potential advantage of length polymorphism method is that raw data of profiles can be easily stored in an electronic database as tables of peaks with corresponding sizes, for comparison to new profiles generated. The application of such databases will depend on a number of factors including the repeatability of the arbitrary amplified DNA profiles and the stability of profiles generated from sites over time. The discriminatory power (Table 4, pg68) and repeatability (discussed in Chpt 3.3.4 and Table 5, pg80 along with AADSS repeatability) of the profiles generated has been determined in this thesis and found to be potentially useful, but the question of AADLP profile stability from soils over time will need to be investigated further.

3.3 Microarray Analysis

3.3.1 General Principles of the Array

The hybridisation of AAD profiles was expanded using microarray technology. Entire AAD profiles, printed as miniaturised spots on slides, were hybridised with fluorescently labelled AAD profiles from individual soil (Chpt 2.8.3) and where sequence similarity existed between them, the labelled profile was bound to the position of the printed profile, allowing the relative quantification of total sequence similarity for many profiles simultaneously. Relative levels of bound label can be measured to give an indication of the amount of corresponding sequence in the DNA at each array position.

3.3.2 Monitoring the Quality of Microarray Probes

The first (#1) and last (#22) slide printed during the production of a batch of microarrays were tested for quality control purposes. The dye SybrGreenII, which binds single strand DNA, was used to determine the shape of the spots and relative amounts of DNA deposited at each position of the array (see Appendix CD\AADSS-Microarray\Prototype Microarray\Microarray Images\Quality Control slide #1.jpeg and Appendix CD\AADSS-Microarray\Prototype Microarray\Microarray Images\Quality Control slide #22.jpeg). Relative fluorescence at each spot on the array (see Appendix CD\AADSS-Microarray\Prototype Microarray\Fluorescence data\Quality Control.xls) was used to determine a coefficient of variance (CV) of the DNA deposition, calculated by dividing the standard deviation of the fluorescence values by the mean fluorescence value to express the variation as a percentage of the mean. The CV was calculated for DNA deposited between arrays, of different and replicate profiles within arrays and of replicate spots between arrays. The CV of mean fluorescence of identical spot positions between the two arrays (#1 and #22) averaged 20.5%. CV of mean fluorescence for the 12 replicate spots of each profile within arrays averaged 50.1% and 34.6% for array #1 and #22 respectively. The average CV of mean fluorescence of replicate spots on arrays #1 and #22 combined was 36.0%. The CV of mean fluorescence of all spots within arrays averaged 68.6% and after all twelve replicate spots were averaged, giving

single values for each profile spotted on the array, the CV between profiles was 55.0%.

Thus, despite spotting with solutions adjusted to equal DNA content, there is a considerable degree of variation in the amount of DNA fixed to each spot. Variation in the amount of DNA that is deposited on the slide could arise from many areas throughout the microarray manufacturing process including, varying amounts of DNA acquired by different printing pins or individual pins between samples, and varying levels of DNA deposited by the printing pins either by different pins, the same pin between samples, or whilst printing a single sample.

This variation in microarray production has been the subject of a previous study into the reproducibility of cDNA microarrays. Rickman *et al.* (2003) found replicate DNA spotted with all solutions tested (including 50% DMSO as used in this study) showed variable levels of fluorescence after hybridisation with a labelled probe. Mean CV of overall signal intensity of replicate spots from two array slides, printed in 50% DMSO, was almost 50%, similar to the 50.1% found on array #1 but much higher than the 36.0% of the #1 and #22 combined slide average. Thus although the array preparation technology used here seems to have a high level of variation of fixation of DNA to the glass, it is in the order of the experience of other labs (Rickman, *et al.* 2003)

Ideally, a stringent correction for the effects of deposition efficiency, requiring a preview of each array before probing, to generate a correction for each spot should be performed. In practice however, the most significant problem was the complete loss of spots when printing of a spot almost failed completely leading to a loss of data for those profiles (see Appendix CD\AADSS-Microarray\Prototype Microarray\Microarray Images\Quality Control slide #1.jpeg and Appendix CD\AADSS-Microarray\Prototype Microarray\Microarray Images\Quality Control slide #22.jpeg). Except for these, the actual hybridisation comparisons below showed little or no sign of bias due to the admittedly highly variable levels of DNA deposited on the arrays. After an experiment using one soil profile to probe itself and the others on the array, the amount of DNA printed onto the arrays at each spot had little influence over the expected hybridisation pattern, unless a spot had failed to print any DNA. This reality is a major factor in making the current array technology useful although advances in printing

technology would still be of great value if only in reducing the numbers of failed spots that represent lost data.

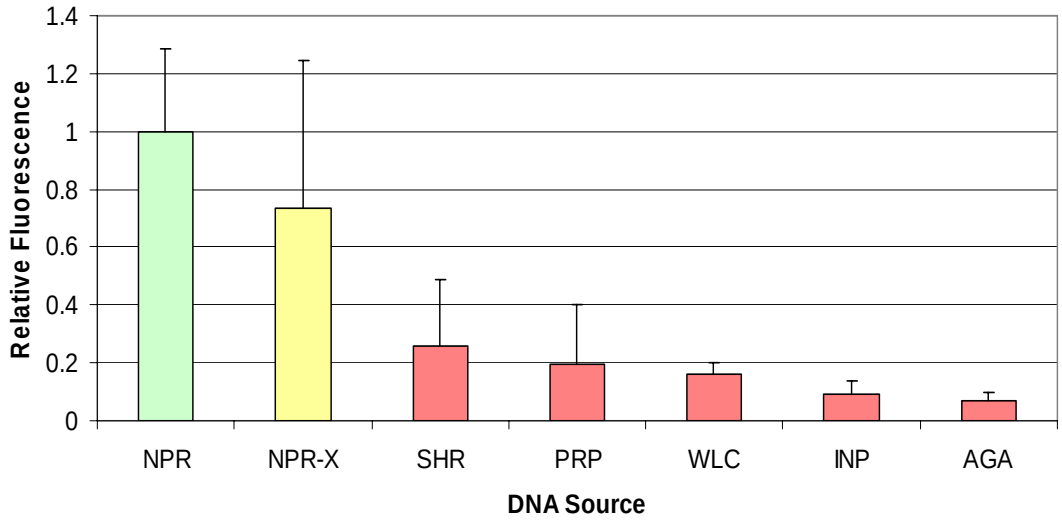
3.3.3 Soil Discrimination via Sequence Similarity of DNA profiles

3.3.3.1 A Pilot Microarray Study

The general principles of the microarray technology were initially demonstrated using the pilot microarray (Chpt 2.8.1.1). The pilot microarray was hybridised with only one Cy3-labelled profile under the same hybridisation conditions (outlined in Chpt. 2.8.3) as the final, prototype microarray. Images and fluorescence data of the pilot microarrays were generated (see Appendix CD\AADSS-Microarray\Pilot Microarray) after hybridisation with Cy3 labelled AAD profiles. The labelled profiles of microbial community DNA hybridised to profiles from different soils disproportionately. Labelled profiles of the *Pinus radiata* plantation soil NPR hybridised best to other replicate NPR profiles. Profiles generated from the same plantation soil up to 100 metres away from NPR (NPR-X, Figure 3.5, pg73) also hybridised well to labelled NPR profiles. Profiles generated from five distant soils hybridised much less of the labelled NPR profile indicating the sequence similarity of the profiles were low.

Detailed analyses of the pilot microarray data do not appear in this thesis for the sake of clarity, as all experiments on the pilot array were superseded by the construction and use of the final, prototype microarray, which contained 725 profiles and 43 controls (spotted 12 times on each microarray slide), giving a total of 9,216 elements per microarray (Chpt 2.8.1.2). The remaining microarray data presented in this thesis was generated using the final, prototype microarray but in many instances were corroborated by the results obtained from the pilot microarray study (data not shown).

Figure 3.5: Cross-hybridisation of arbitrarily amplified DNA profiles from various soils on the <u>Pilot Microarray</u>. Relative fluorescence units (RFU) (+1SD) of profiles generated from soils NPR (Green column), soils taken within 100m of NPR (NPR-X, Yellow column) and AGA, PRP, WLC, INP, SHR (Red columns) after hybridisation with Cy3-labelled NPR profile.



3.3.3.2 The Prototype Microarray Studies

Images and fluorescence data of the prototype microarray were generated (see Appendix CD\AADSS-Microarray\Prototype Microarray) after hybridisation with Cy3 or Cy5 labelled AAD profiles.

Mean relative fluorescence of the prototype microarray profiles generated from soils A, B and C after hybridisation with labelled profile were calculated (Chpt 2.8.4) and taken as a representation of total DNA sequence similarity. Where profiles were replicates of the labelled profile, mean relative fluorescence averaged 0.697, while profiles generated from DNA of different origin to the labelled profile had mean RFU values of 0.047 (Table 4, pg68). This is a convincingly large 14.8 fold difference. The mean RFU values of replicates of the same soil sample were 0.706 within Site A, 0.569 within Site B and 0.710 within Site C (Table 4, pg68). The profiles of all three soils were significantly different from each other as determined by independent measures *t* tests (Table 4, pg68). This result is in accord with those observed in much more detail using Southern hybridisation (Figure 3.4, pg66). That is to say, both Southern and microarray analysis gave high levels of signal between profiles of the same soil and very low levels of signal from different soil profiles. This difference was far greater than the difference in SI values for TRFLP comparisons of same soil and different soil profiles.

Relative fluorescence values (normalised so the labelled profile's RFU value is put to 1) of individual soils probed with each of the ten soil profiles and one human profile can be seen in Figures 3.6 through 3.8 (pgs75-77). It can be seen that a number of soil profiles share some real sequence similarity to profiles of other soils (eg. A, AGA and WLC in Figure 3.6, PRP and NPR in Figure 3.7 and B and SHR in Figure 3.8), but other soil profiles appear to share very little sequence similarity to any other soil profile (C and INP in Figure 3.8). These very clear mismatches, or exclusions, increase the confidence that the small sequence similarity between some soil profiles reflect a biological reality.

Figure 3.6: Cross-hybridisation of arbitrarily amplified profiles generated from different soil DNA templates. Relative fluorescence (+/- 1SD) of profiles from ten soils and one Human control, generated using primer Seq005, PCR2 and Cycle2 (Chpt 2.5.3) after hybridisation with labelled profiles generated using Seq005 and DNA from Soil A (A), Soil AGA (AGA) or Soil WLC (WLC). Profiles are arranged in rank order along the X-axis.

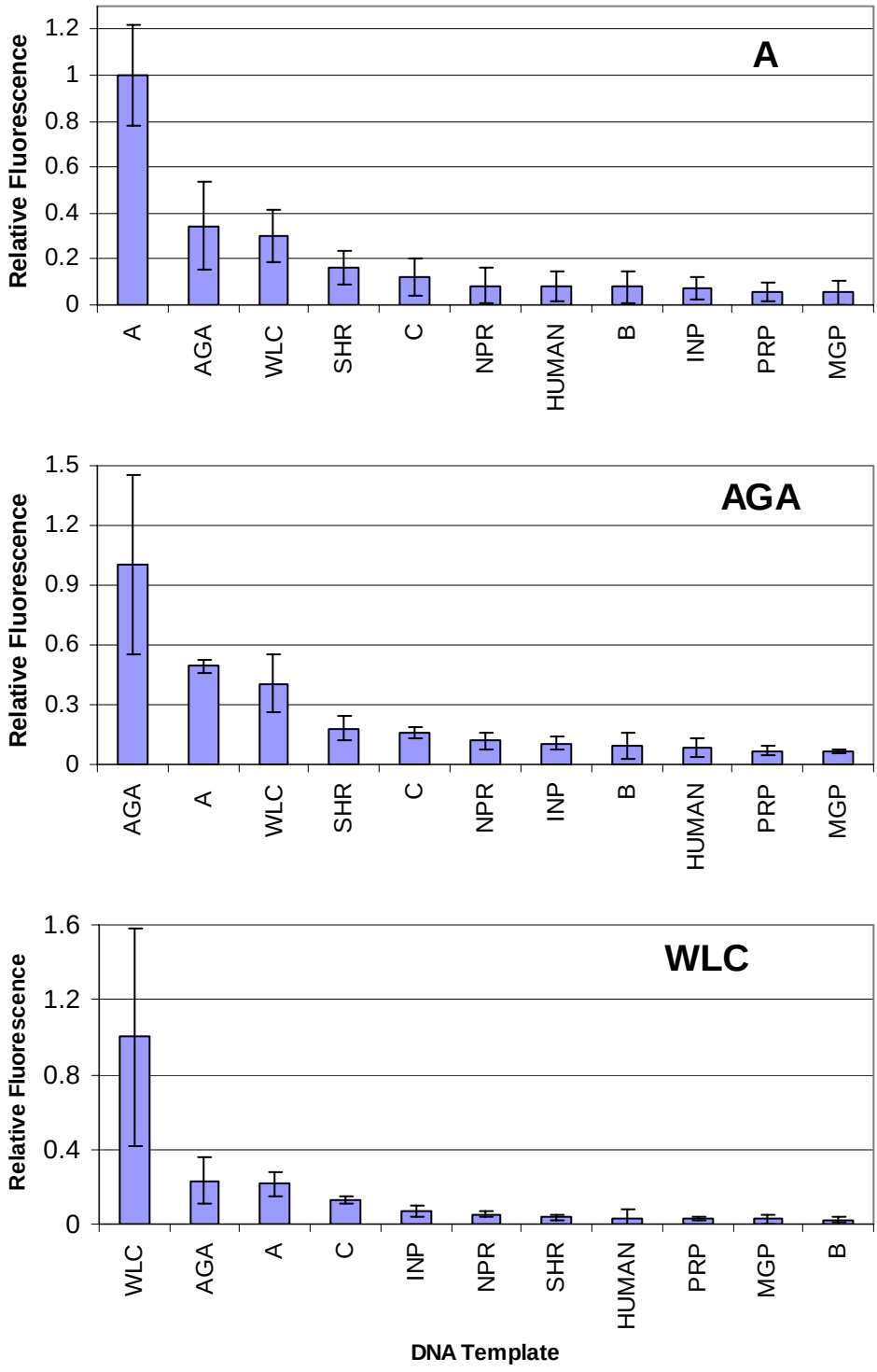


Figure 3.7: Cross-hybridisation of arbitrarily amplified profiles generated from different soil DNA templates. Relative fluorescence (+/- 1SD) of profiles from ten soils and one Human control, generated using primer Seq005, PCR2 and Cycle2 (Chpt 2.5.3) after hybridisation with labelled profiles generated using Seq005 and DNA from Soil PRP (PRP), Soil NPR (NPR) or Soil MGP (MGP). Profiles are arranged in rank order along the X-axis. (***** - only one profile was available for analysis thus no standard deviation was determined)

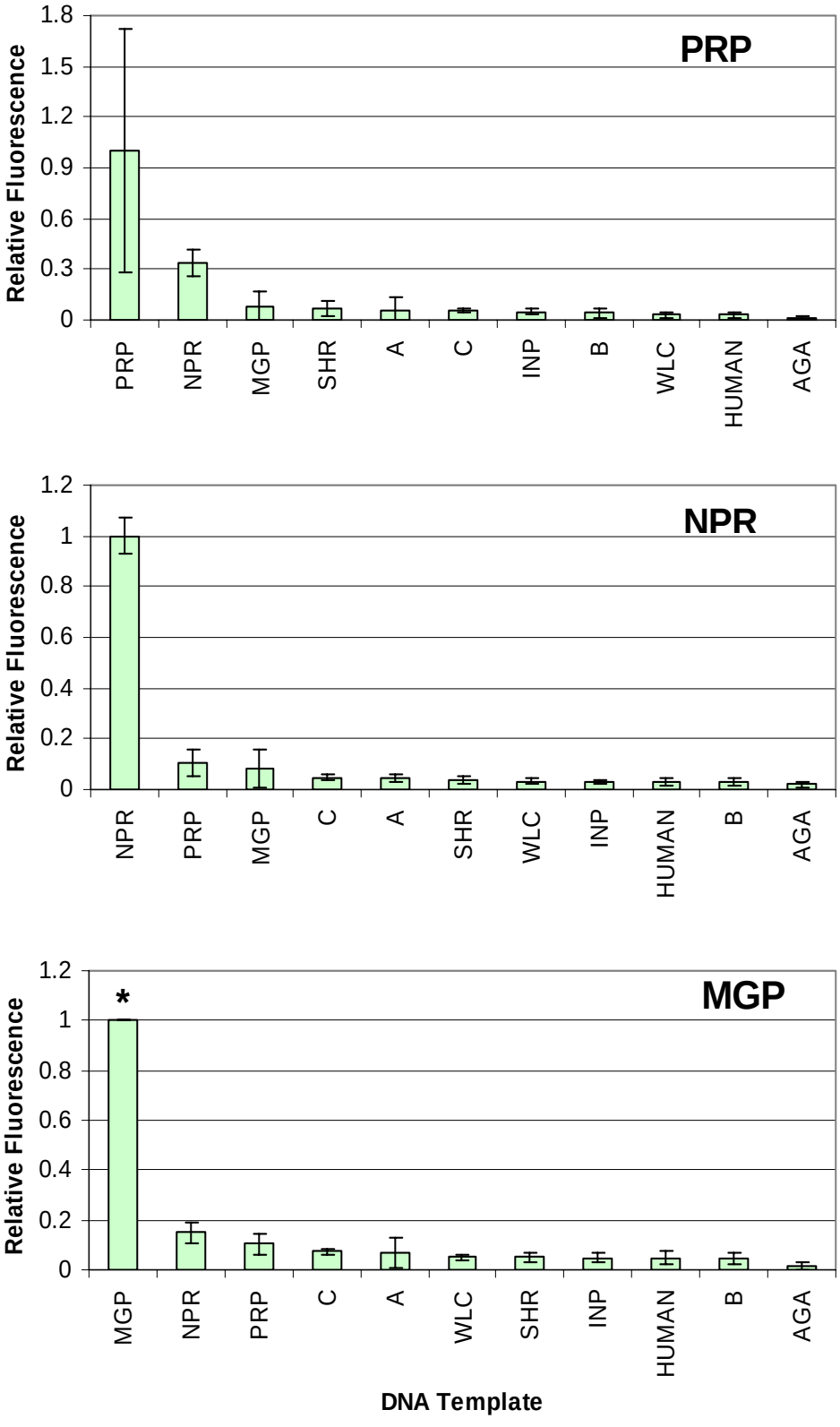
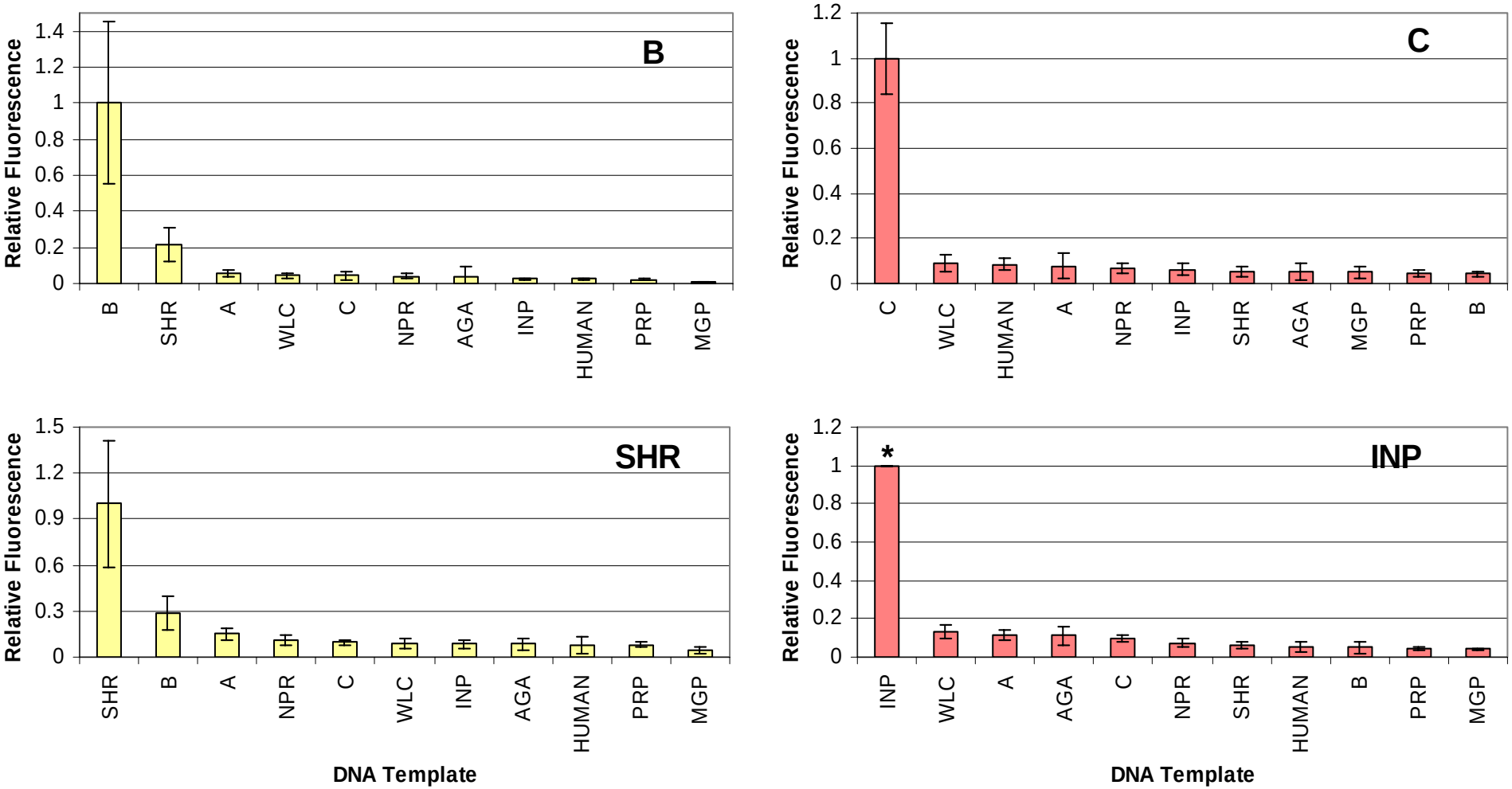


Figure 3.8: Cross-hybridisation of arbitrarily amplified profiles generated from different soil DNA templates. Relative fluorescence (+/- 1 SD) of profiles from ten soils and one Human control, generated using primer Seq005, PCR2 and Cycle2 (Chpt 2.5.3) after hybridisation with labelled profiles generated using Seq005 and DNA from Soil B (A), Soil SHR (B), or Soil C (C) or Soil INP (D). Profiles are arranged in rank order along the X-axis. (***** - only one profile was available for analysis thus no standard deviation was determined)



When profiles from these seven additional soils were added to the statistical analysis of soils A, B and C, the mean RFU of same soil comparisons dropped slightly to 0.662 and different soil comparisons rose to 0.062. However, the ten-fold difference between these groups of mean RFU remains significant (P=1.47x10^{-65}, independent measures t test).

Despite the minor changes in mean RFU's due to the addition of seven soils, the discrimination power of the method was very strong. These extra soils included three pine forest plantation sites (PRP, NPR and MGP Figure 3.7) from sites within 100km and a soil within 200m of soil B (SHR) (see Figure 2.1, pg38). These soils displayed sequence similarity when hybridised with other pine forest soils or soil B respectively, above the level of other distinct soils for which there was no *a priori* evidence to expect any similarity. This real similarity has contributed to the raise in mean RFU value to 0.062 for supposedly dissimilar soils, confirming that in reality the method is actually be more powerful than even the ten-fold difference cited above and possibly nearer to the 14.8 fold first estimated. The shared sequence similarity in the case of the monoculture pine forest soils may be due to either the presence of similar microbial communities between sites (particularly microbes associated with the rhizosphere), the presence of *Pinus radiata* DNA itself from degrading plant debris, or a combination of both. Unsurprisingly, the sample close to site B (SHR) yielded an above normal RFU when hybridised with labelled profile from site B DNA indicating some level of shared sequence similarity and biological community. However, there were other soils that appeared to share no common link but nonetheless showed varying levels of sequence similarity above the norm for distinct soils. This suggests that classifying soils for the purpose of forensic investigations may not always be as decisive as for soils A, B and C. Soils with comparable plant cover, usage (such as cropping or grazing) or in near vicinity may share sequence similarity between profiles. However, in many cases this may be advantageous, as soil typing may be able to provide information about a crime scene when its location is unknown. The proposition that soil type can influence the biological community that resides within it was advanced by Girvan *et al.* (2003). It was concluded from physiological and 16S rDNA (DGGE and TRFLP) data that soil type was the key factor in determining bacterial community composition, while cropping regimens and management strategies played minor

roles in determining bacterial community composition. The three sites used in the study by Girvan et al. (2003) were all within 90km and the two similar soil types were ~46km apart.

As well as potentially providing information about soil type, these secondary cross-hybridisations may be useful in their own right as a profiling system. The ability to distinguish soil profiles by their preferences of cross-hybridising to other soil profiles may provide higher levels of information about the DNA profile, above that of the degree of cross-hybridisation between two soil profiles alone. While it remains to be seen if this has any practical value this is an important prospect for the future of this field.

3.3.4 Technical Sources of Arbitrarily Amplified DNA Profile Variation

Potential sources of technical variation including i) the extraction and purification of DNA from soil and ii) PCR replication were evaluated. DNA profiles created using template from different DNA extractions of the same soil were not significantly different to replicate analyses of the same extract for either length polymorphism or sequence similarity analysis (Table 5, pg80, see Appendix CD\AADLP\AADLP SI Matrix.xls for all Day0 profile length polymorphism comparisons and SI values). Likewise, profiles generated using different batches of PCR enzyme and reagents, and profiles generated from both different DNA extractions and PCR enzymes and reagents did not differ significantly from replicate analyses of the same extract (Table 5, pg80). Overall, this technique seems robust and repeatable.

However, the use of different sources of thermostable DNA polymerases and corresponding buffers when amplifying DNA does noticeably alter both the quantity and type of DNA amplified (Figure 3.9, pg81). Hybridisation studies also suggest that the total sequence similarity of AAD profiles is reduced between samples amplified with different thermostable polymerases (Figures 3.10 through 3.12, pgs82-84) although rarely to levels of different soils or control profiles. All other hybridisation comparisons within this thesis are between profiles generated with the same thermostable DNA polymerase (Chpt 2.5.3), so are not prejudiced by this probable source of variation.

Table 5: Effect of technical sources of variation on mean relative fluorescence and significance values for each level of variation.

I	II	III	IV	V	III	VI
Experimental Variation	Mean Similarity Index **SI** ± (S.D).	Number of pair-wise comparisons (*n*)	Significance of mean SI difference to (1) (*P*, *t* test)	Mean Relative Fluorescence **RFU** ± (S.D).	Number of replicates (*n*)	Significance of mean RFU difference to (1) (*P*, Mann-Whitney *U*)
(1) Same Extraction Same PCR reagents	**0.790** (0.096)	36	-	**0.760** (0.266)	4	-
(2) Different Extraction Same PCR reagents	**0.776** (0.099)	54	*0.498*	**0.780** (0.162)	16	*0.850*
(3) Same Extraction Different PCR reagents	**0.779** (0.113)	54	*0.643*	**0.610** (0.158)	30	*0.109*
(4) Different Extraction Different PCR reagents	**0.769** (0.098)	54	*0.301*	**0.671** (0.176)	28	*0.297*

I The source of technical variation. Profiles were generated using either the same or different DNA extraction replicates and PCR reagent batches.

II Mean Similarity Index (SI) values of profiles from the same Soil DNA for each technical variation source are listed with the standard deviation (S.D.).

III The number (*n*) of paired profile comparisons performed for each source of technical variation.

IV *P* values, determined by independent measures *t* tests (α = 0.05, 2 tailed), when comparing the difference of SI values generated with differing conditions (rows 2-4) to profiles generated with identical conditions (row 1).

V Mean Similarity Index (SI) values of profiles from the same Soil DNA for each technical variation source are listed with the standard deviation (S.D.).

VI *P* values, determined by Mann-Whitney *U* tests when comparing the difference in RFU values of profiles generated with differing conditions (rows 2-4) to profiles generated with identical conditions (row 1).

Figure 3.9: Amplified profiles of triplicate No template controls (1-3) and replicate Human DNA (4 and 5) using four commercially available thermostable DNA polymerases; Promega Taq in Buffer A (A), Promega Taq in Buffer B (B), Austral Scientific BIOTAQ (C) and ABgene RedHot polymerase (D). Phage SPP-1/*Eco*RI digest loaded as marker (M) (see Figure 2.3 for all band sizes). Amplified products from contaminating DNA are highly evident in all commercial sources; superficially they are most evident in enzyme-C but close examination indicates manufacturers impurities tend to follow the total ability of a product to produce long molecules. This may not mean that enzyme-C has the most contaminating DNA, but the high activity of the enzyme results in more being amplified.

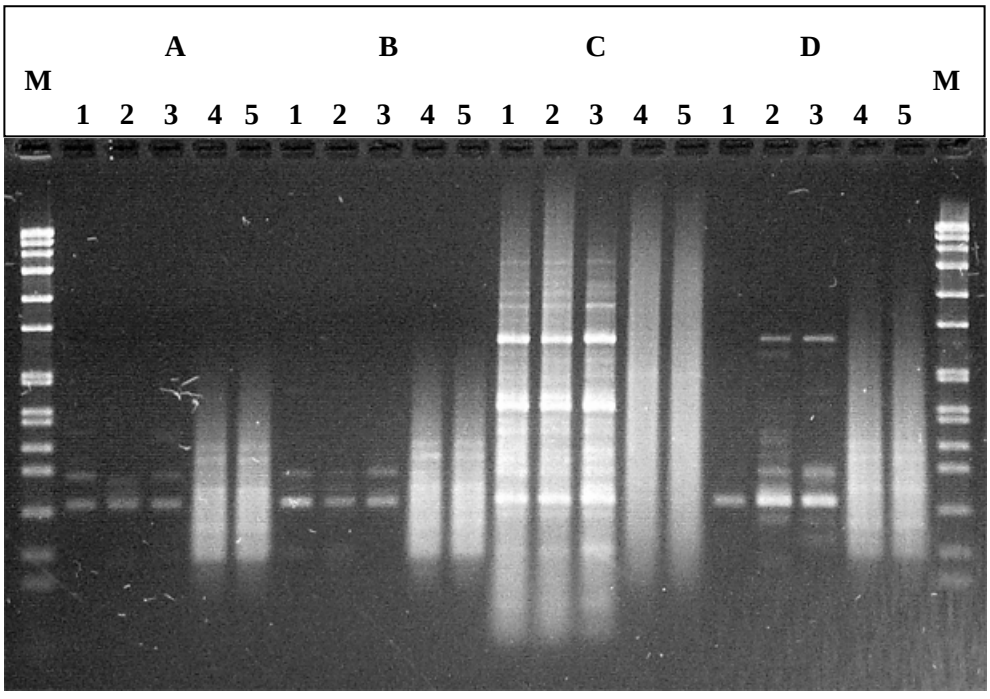


Figure 3.10: Cross-hybridisation of arbitrarily amplified profiles amplified with various DNA polymerases. Relative fluorescence (+ 1SD) of profiles amplified using either Promega Taq in buffer A (Blue), Promega Taq in buffer B (Green), RedHot polymerase (Yellow) or BioTaq (Red) after hybridization with labelled profile generated using **Promega Taq in buffer B** and DNA from soils A, B or C.

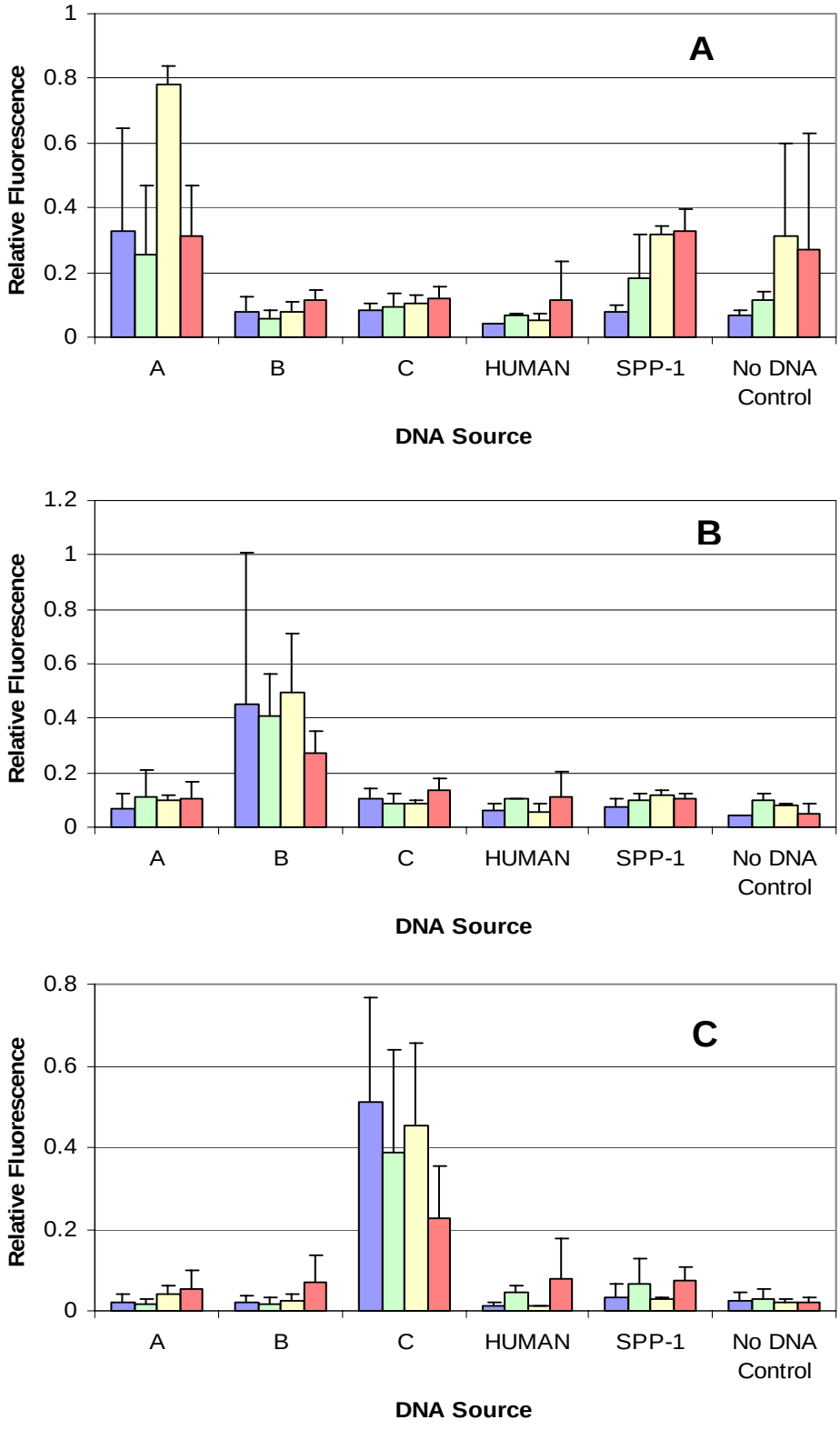


Figure 3.11: Cross-hybridisation of arbitrarily amplified profiles amplified with various DNA polymerases. Relative fluorescence (+ 1SD) of profiles amplified using either Promega Taq in buffer A (Blue), Promega Taq in buffer B (Green), RedHot polymerase (Yellow) or BioTaq (Red) after hybridization with labelled profile generated using **RedHot Polymerase** and DNA from soils A, B or C.

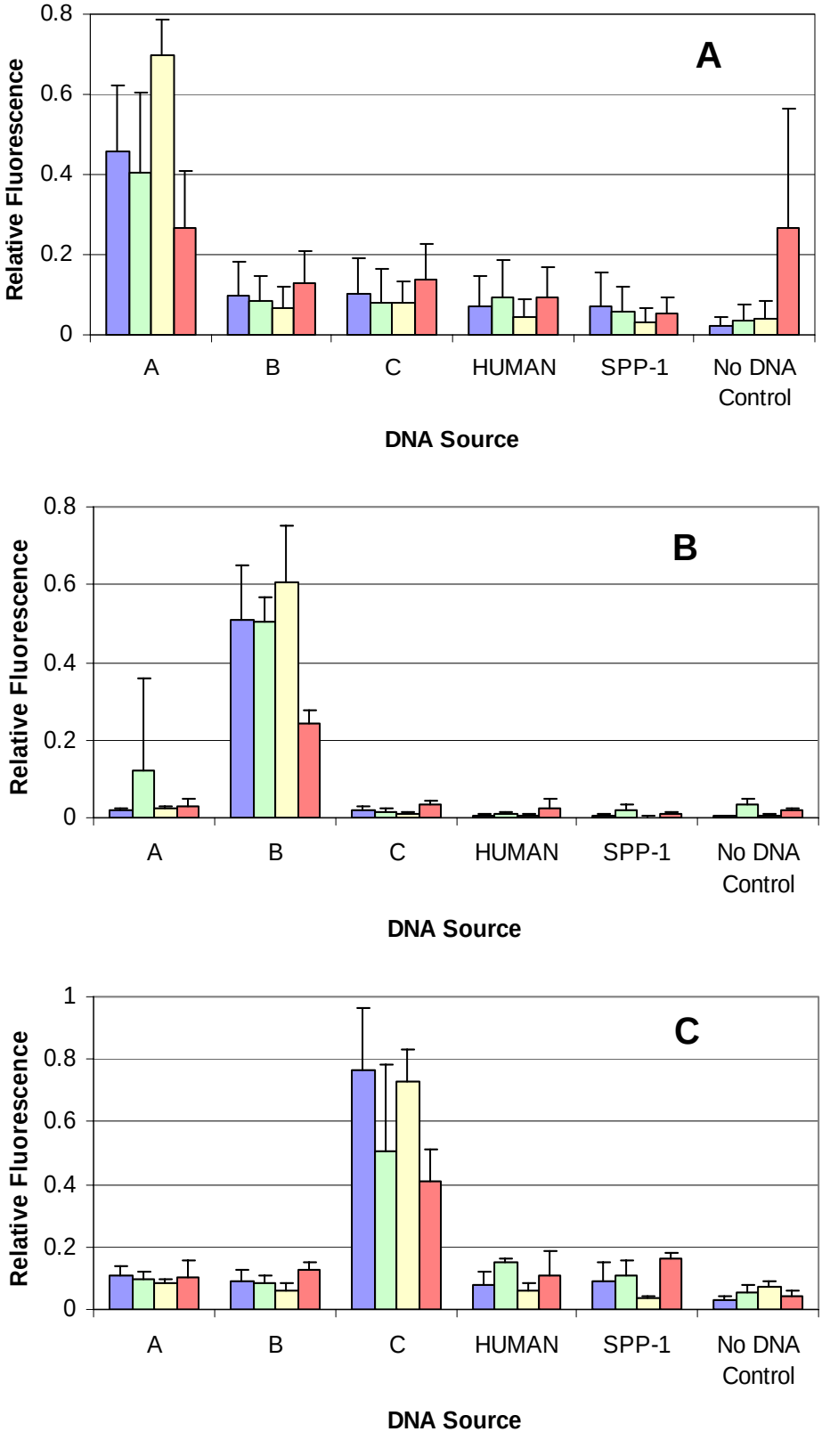
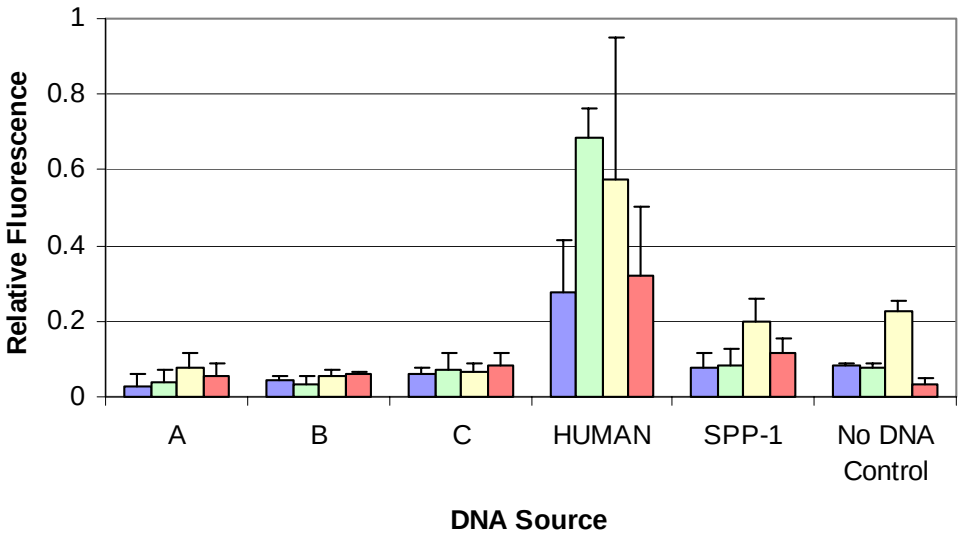


Figure 3.12: Cross-hybridisation of arbitrarily amplified profiles amplified with various DNA polymerases. Relative fluorescence (+ 1SD) of profiles amplified using either Promega Taq in buffer A (Blue), Promega Taq in buffer B (Green), RedHot polymerase (Yellow) or BioTaq (Red) after hybridization with labelled profile generated using **RedHot Polymerase** and Human DNA.



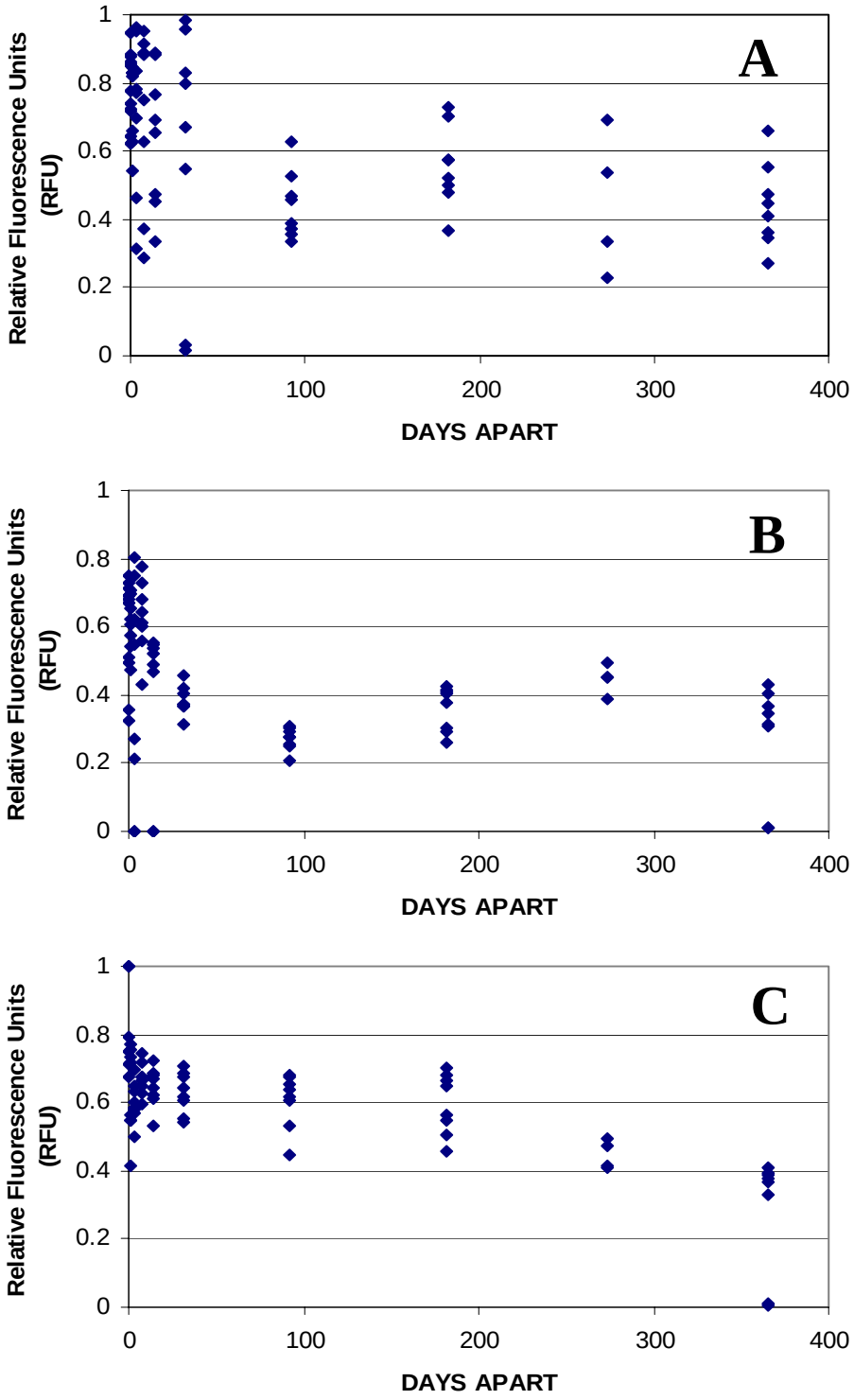
3.3.5 Environmental Variables that may affect Profiles.

3.3.5.1 Temporal Variation of DNA Profiles

Hybridisation of profiles generated from various times over a twelve-month period showed that profiles changed over time (Figure 3.13, pg86). RFU values from day 0, day1 and day3 profiles of the same soil were grouped and compared to RFU values of other profiles using the Mann-Whitney *U* test. Soil A profiles from day92, day182, day275 and day365 had different ($p<0.01$) RFU values to the day0-3 group, Soil B profiles from day92 and day365 had different ($p<0.05$) RFU values to the day0-3 group and Soil C profiles from day275 and day365 had different ($p<0.01$) RFU values to the day0-3 group. This comes as no surprise as the make-up of soil microbial communities is known to change over time (Duineveld*, et al.* 1998; Lukow*, et al.* 2000; Mummey and Stahl 2003; Kang and Mills 2004). For sites A and B the changes are non-linear, while for site C the RFU value between profiles does not differ dramatically until more than six months have passed.

Despite the reduction in RFU values between samples taken at different times, generally the total similarity of the profile sequence is above the level of soils at quite different sites, indicating the level of sequence similarity between soils separated by time (up to twelve months) (Figure 3.13, pg86), is well above the sequence similarity of soils separated by distance (2-5km) (Table 4, pg68). The reduction in overall sequence similarity with time may be due to a number of factors including fluctuations in relative levels of organisms, departure or arrival of transient species and perhaps to a small extent genetic drift and modification of organisms. However, there appears to be a level of similarity maintained over time (Figure 3.13, pg86), above the level of similarity to soils from other sites (Table 4, pg68). This similarity may be due to persistent micro-organisms, or the constant presence of plant matter in the soil. The effect of time on the profile generated may have a seasonal component, whereby samples taken from the same time of year will have more similarity than those taken at different times due to seasonal selection of micro-organisms. This study has sampled soils from three sites (A, B and C) at a one year interval, but found little increase in sequence similarity above levels generated from samples taken different seasons throughout

Figure 3.13: Trends of Arbitrarily amplified DNA profile sequence similarity with time. Arbitrarily amplified DNA profiles were generated from three soils (A, B and C) separated by 2-5km. Microarray analysis was performed on amplified profiles generated from soils sampled over twelve-months from the three sites after hybridisation with labelled Day0 profiles. The RFU of profiles are plotted against the number of days between labelled target and probe profile samples.



the year (Figure 3.13, pg86). However, more extensive, long term sampling will be required to conclusively detect such an influence.

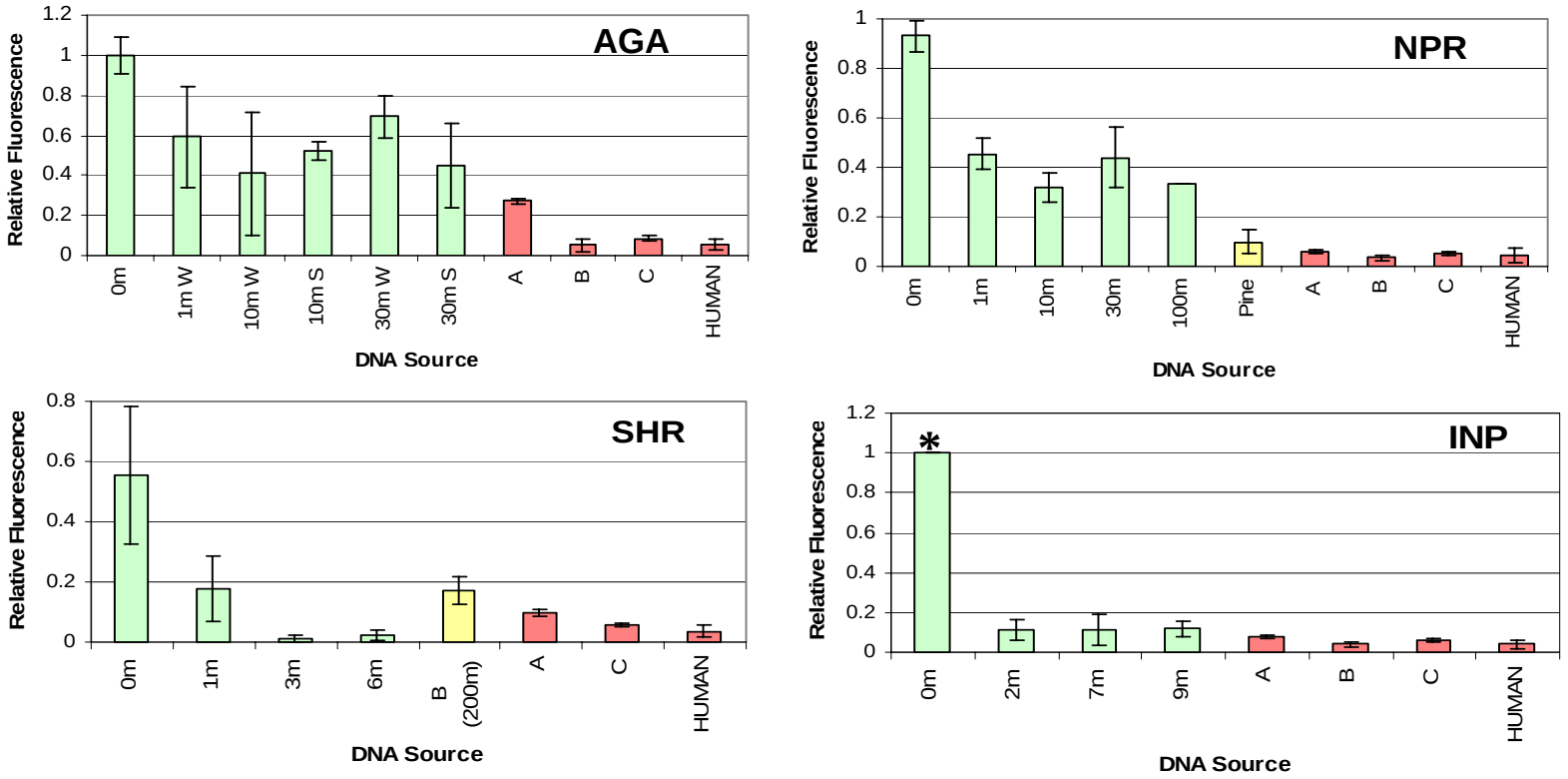
The data presented in this thesis suggest that the sequence similarity of profiles does indeed change with time, suggesting that the length polymorphism patterns may also change. However, like the RFU values for sequence hybridisation analysis, the length polymorphism analysis may yield similarity indices above the level of distinct soils, even when samples are separated by large amounts of time, still providing beneficial information. The current as well as previous studies (Lukow*, et al.* 2000; Horswell*, et al.* 2002) have shown that length polymorphism patterns generated by TRFLP do differ over time. If temporally transient profiles also occur with length polymorphism analysis of arbitrarily amplified DNA profiles, the potential of a soil profile database will be limited as the cost and logistics of maintaining a database with current soil profiles may be prohibitive. However, any potential limitation of transient profiles can also be an advantage. If two samples have similar profiles, it can be suggested that not only do the samples come from the same location but also within a specifiable time frame.

Regardless of the potential for length polymorphism profiles to be databased, the sequence hybridisation method can be databased both physically and electronically. A standard array containing whole profiles or individual sequences is a compact, physical database of DNA profiles, while fluorescence patterns after hybridisation can be stored on electronic databases and these patterns, although still of unexplored utility, almost certainly contain much more information over and above the very simple information of simple pair-wise comparisons. It is quite reasonable to expect that this higher level of information will have considerable utility. These databases will be subject to the same restrictions as length polymorphism databases regarding temporally transient profiles, as the sequence similarity of profiles do change over time. However, the arbitrarily amplified profiles appear to maintain a level of sequence similarity above that of soils from different sites, even after a one year interval. The temporal stability of arbitrarily amplified profiles is much higher than TRFLP profiles which differ significantly after less than six months

3.3.5.2 Small Scale Spatial Variation of DNA profiles

Profiles generated from soil samples taken various distances, only metres apart, were compared by microarray hybridisation. A decline in profile sequence similarity between soils separated by distance was expected and observed in all four cases (Figure 3.14, pg89). However, the nature and severity of the decline could be divided into two distinct categories that corresponded to groups of soils, those under monoculture crops and those under wild, native flora. Both AGA (agricultural) and NPR (*Pinus radiata* plantation) soils showed a limited decrease in profile similarity of soils taken 30m and 100m apart, respectively. However, INP (National Park) and SHR (Reserve) both show dramatic reductions in profile similarity, to near background levels, between soils taken within metres of each other. It is unsurprising that soils under monoculture crops show the least heterogeneity as plant roots, leaves and other debris could contribute a significant proportion of DNA isolated from soils. The predominance of a single plant species may promote or foster the growth of similar micro-organisms and these farmed environments generally undergo mechanical homogenisation by ploughing or other tending implements. The reduction of profile sequence similarity between these crop growing soils may be due to micro-fluctuations of community structures across the sample area. Conversely, native and relatively undisturbed soils generally appear to support a more diverse community on the scales tested although more studies of this expected effect would be most valuable as it suggests that native soils may be much more discriminatory in a forensic environment than cultivated soils. The issues are probably the converse of those for monoculture soils in that the diversity of above ground species influences the apparent diversity of organisms in and on the ground. The lack of any mechanical ploughing or other artificial homogenisation is also likely to facilitate the heterogeneity of soil communities observed over short distances. This spatial variability has been observed previously with TRFLP profiles of homogeneous grassland soils, which change very little over distances up to 100m, and soil from a shrubland containing heterogeneous plant cover, which changed dramatically within metres (Mummey and Stahl 2003).

Figure 3.14: Cross-hybridisation of arbitrarily amplified DNA profiles from soils taken from varying distances apart. Relative fluorescence units (RFU) (+/- 1SD) of profiles generated from soils AGA, NPR, INP, SHR and soils taken from small distances away (Green columns), similar soils from further away (Yellow columns) and the controls; Human DNA and soils A, B and C for reference (Red columns). (***** - only one profile was available for analysis thus no standard deviation was determined)

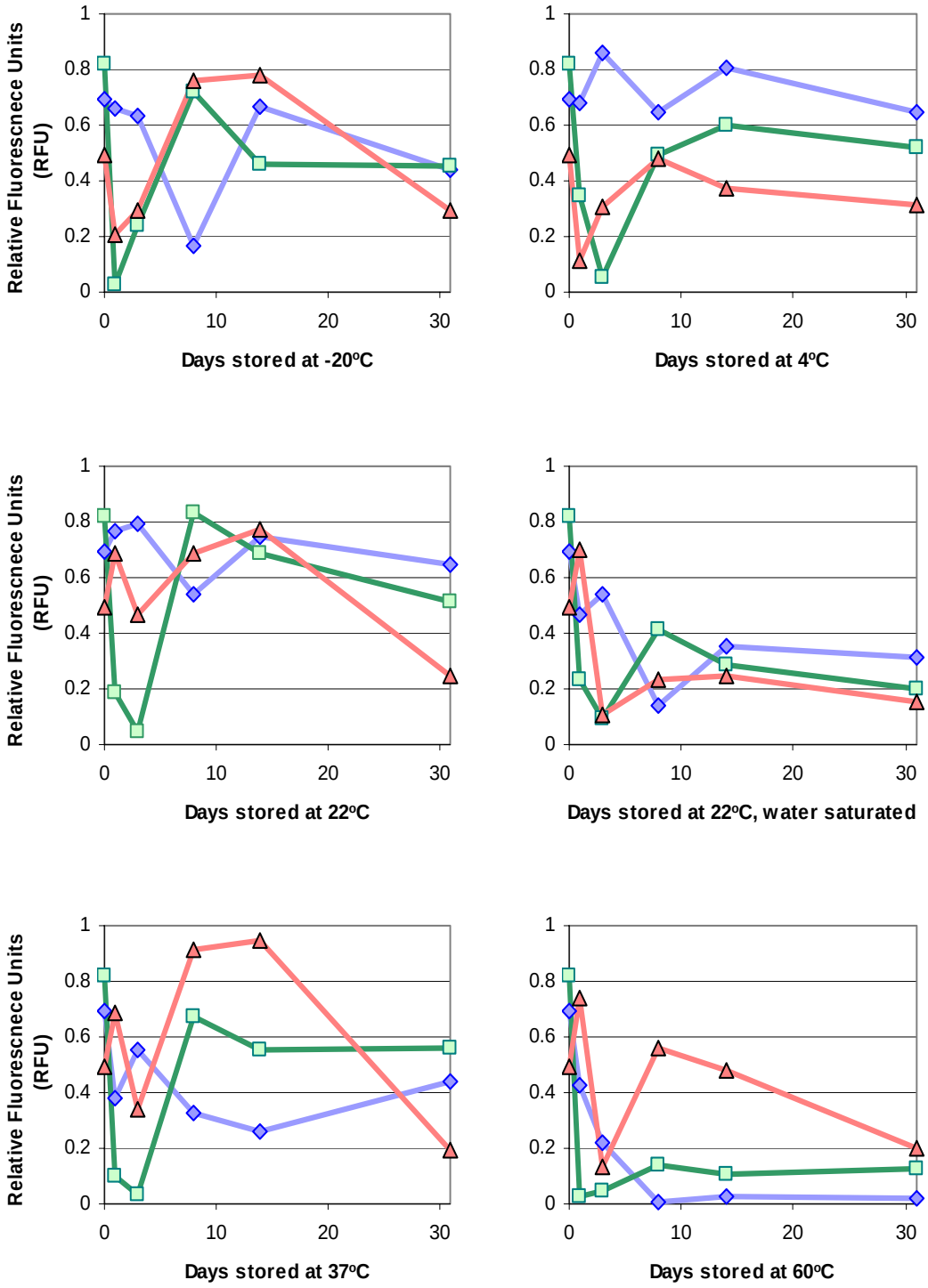


The implications of such a finding are that extensive sampling of an area of interest may be necessary, particularly from native or undisturbed areas, in order to identify exact locations where samples have matching profiles. In some cases, this inconvenience may be outweighed by the ability to pinpoint matching samples to a highly defined position. Alternatively, slightly different methods of processing the samples, for example the use of alternate primers during amplification, could lead to profiles containing sequences from more spatially stable organisms, that are representative of larger areas. This could be an initial approach with the higher-resolution approach to follow if required.

3.3.5.3 Variation of DNA Profiles after Storage of Soil at Various Temperatures.

Samples of soils A, B and C freshly taken from their place of origin, were stored at either -20°C, 4°C, 22°C, 22°C saturated with water, 37°C or 60°C and DNA extracted in duplicate over a period of one month to determine how various storage conditions alter the profiles generated. The relative fluorescence of profiles after hybridisation with a labelled profile generated from DNA extracted on initial sampling (Day 0) is shown in Figure 3.15 (pg91). The relative fluorescence of soil B profiles and all but one (22°C) soil C profile decrease dramatically under all temperature conditions after 1-3 days. However, in most cases this dramatic decrease in profile similarity is followed by a restoration of similarity by 8 days post sampling. In support of this, a previous study looking at the effects of freeze-thaw stress on microbial communities found that soil frozen at -20°C underwent a change in the amount of eucaryal (18S) and archaeal (16S) rDNA PCR products 0-3 days after thawing, but also recovered to pre-frozen levels within 10 days (Pesaro and Widmer 2002). Although these studies and this thesis are not directly comparable, the results presented in this thesis also show rapid changes in community structure after adjustments in water content or temperature and the subsequent recovery of communities toward pre-stress levels. There are two conditions tested here that appear to dramatically alter the profile generated in less than one month, for all three soils tested; saturating soil at 22°C with water and storing dry at 60°C.

Figure 3.15: Cross-hybridisation of arbitrarily amplified DNA profiles from soils stored at various temperature conditions for up to one month. Relative fluorescence of profiles generated with primer Seq005 and DNA isolated from Soil A (Blue Diamonds), Soil B (Green Squares) or Soil C (Red Triangles) after hybridisation with labelled profile generated from Seq005 and DNA isolated from the corresponding soil prior to storage (Day0).



The addition of water to soil severely alters the profile generated and is likely to be due to a dramatic but persistent change in proportions of organisms, either flourishing water loving species and/or the decline of others. A previous study investigating the affect of a single drying-rewetting cycle found dramatic changes in TRFLP profiles of bacterial community structures after rewetting dried soil (Pesaro*, et al.* 2004). In this thesis, dry storage at 60°C notably decreased the profile similarities (Figure 3.15, pg91), which correlated with a fall in the amount of DNA extracted from the soil (data not shown). Despite the same quantity of DNA being used as template in the PCR amplifications, relative levels of surviving genomes appear to change. This is presumably due to the autolysis of DNA by thermolabile species, until only the most thermo-tolerant species remain. It is the change in relative levels of organisms that cause the reduction of profile similarity, which can cause differences like those of distinct soil community profiles. However, dry storage at temperatures ranging from -20°C to 37°C for up to one month, seems to only moderately affect the profiles of these three soils, with RFU values above the levels of distinct soil comparisons. In order of influence over changing the profile generated, 4°C and 22°C have a minimal affect, -20°C and 37°C have a moderate affect and 60°C and saturation with water have the largest affect.

In practical terms these results indicate that soil removed from its native site, stored in a dry place without exposure to extreme temperatures is likely to maintain similar microbial community structure to soil *in situ*, except during the few days after removal where stress-induced fluctuations of community structure may occur.

3.3.6 Conclusions for Arbitrarily Amplified DNA Profiles.

This study has found that both length polymorphism and hybridisation analysis of arbitrarily amplified DNA are capable of discriminating soil samples based on the DNA of the biological community residing within (Chpts 3.2.2 and 3.3.3). The differences in both mean SI and RFU values between soil replicates and soils of different origins were significant, with a very low probability of error (Table 4, pg68).

However, the profiles were dependant on the source of DNA polymerase used and the number of rounds of amplification although the affect of DNA extractions and PCR batches were low, highlighting the robust nature of the method once a defined protocol has been established and adhered to (Chpt 3.3.4, Table 5 pg80).

The make-up of these communities was shown to change over time by the use of hybridisation analysis of arbitrarily amplified DNA (Chpt 3.3.5.1). Community make-up was also found to change spatially, to varying degrees depending on the type of soil (Chpt 3.3.5.2).

The affect of storage conditions on AAD profiles was investigated (Chpt 3.3.5.3). Storage of soils was found to significantly alter the profile generated in less than one month when exposed to 60°C or saturated with water at 22°C. Conditions affecting profiles the least were 4°C and 22°C. Two of the three soils underwent a stress related change in profiles between 1 and 3 days post-sampling, but recovered to pre-stress levels in less than 10 days post-sampling.

Unlike many of the other methods of soil analysis, this technology does not require a high degree of expertise to perform and does not have any subjective component in its analysis. This technology shows promise in its own right and as a complementary tool for the forensic examination of soil.

3.4 The Molecular Mechanisms of the Arbitrary Amplification

If this profiling strategy is to be utilised, the underlying principles of the arbitrary amplification process need to be understood in order to maximise the usefulness of such a technique and anticipate its shortcomings. Issues regarding the sensitivity and amplification efficiency of various primers are examined in the following section along with issues regarding the amplification of profiles from trace DNA samples, the selection processes that occur during the amplification resulting in the production of a minute subset of the DNA analysed, the degree of specificity or resolution the system is operating at and the degree of conservation of both length polymorphisms and sequence similarity between profiles amplified with different but related primers.

3.4.1 Amplification Efficiency and Sensitivity of Various Primers

Trials in previous studies indicated that primers with little or no G residues and a high melting point may have peculiar but ill understood advantages for amplifying DNA from extremely low amounts of template (Rogers 2002). In this thesis, a series of 46 primers were directly tested for total amplification efficiency using real-time PCR with either Soil A, Human DNA or no-template controls. An example of fluorescence traces of real-time PCR generated using various primers and Human DNA are shown in Figure 3.16 (pg95). The number of cycles each primer requires to raise fluorescence, and by inference total dsDNA, above an arbitrary threshold varies with template and primer (Table 6, pg96). The applicability of real time PCR in assessing the efficiency of primers during standard PCR is uncertain as the presence of Sybr Green I can adversely affect the activity of thermostable polymerases (Nath, *et al.* 2000). The different thermostable polymerase used that contains no coloured dye (PCR3, Chpt 2.5.4) and the different thermocycling regime (Cycle3, Chpt 2.5.4) may also alter the conditions of amplification so that real-time PCR may not reflect the actual behaviour of the standard PCR. A sample of profiles (using human DNA as template) generated using real-time PCR were electrophoresed on an agarose gel (Figure 3.17, pg97). Despite different profiles being generated by real-time PCR (with Sybr Green I) as to standard PCR (without Sybr Green I), the quantity of amplified DNA from real-time PCR was consistent with profiles generated using standard PCR (Figure 3.18, pg98) with the exception of the Seq5K primed profile which successfully amplified during real-time PCR but not during standard PCR. Some primers fail to amplify (Figure 3.18, pg98, lanes G and K) or poorly amplify (Figure 3.18, pg98, lane D) particular sources of DNA and some primers seem to be so efficient at amplifying DNA that even trace amounts in PCR reagents can be amplified in no-template controls (Figures 3.9, pg81 and 3.19, pg99). The ability to amplify DNA well or inadequately did not appear to result from any single feature of the primers and may depend on a number of factors including the composition of template DNA sequence facilitating appropriate priming and allow amplification, the properties of the primer itself, or a combination of both.

Figure 3.16: Illustration of the appearance of the data from normalised fluorescence readings of Real-Time PCR amplified Human DNA using various primers.

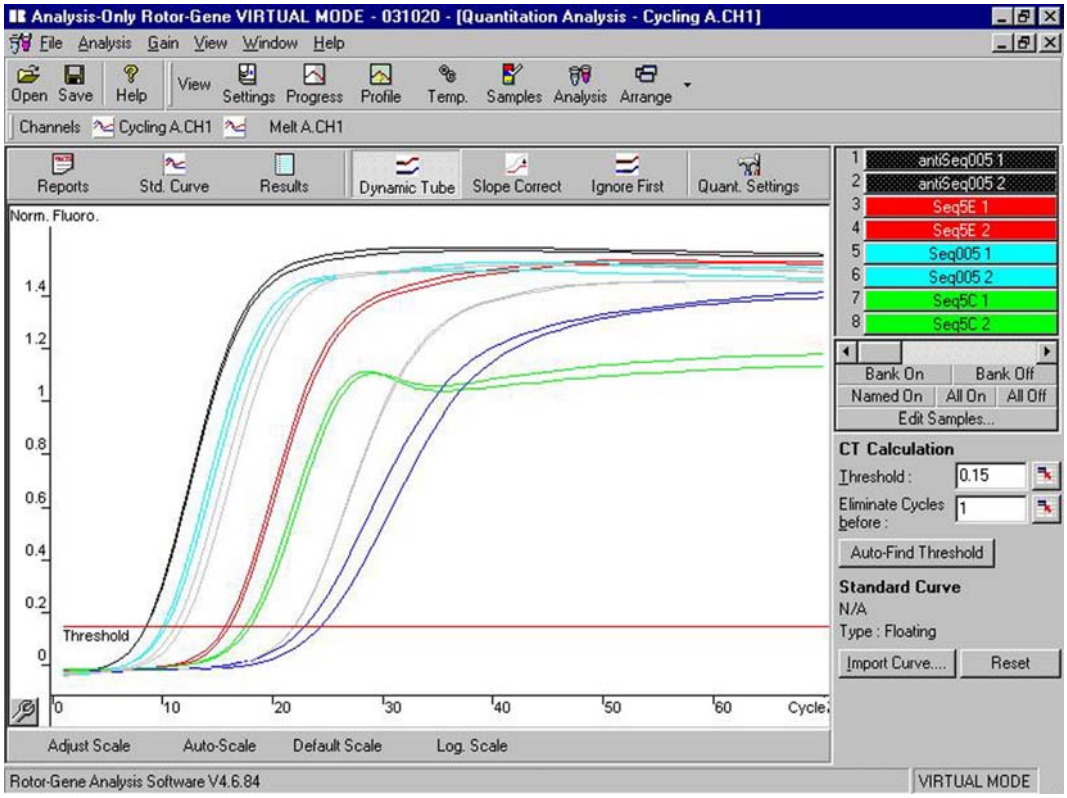


Table 6: Real-Time PCR efficiency of primers when amplifying Human, Soil or No DNA template. Listed is the mean number of cycles taken to increase fluorescence above the threshold and the coefficient of variation (CV) for each primer (Sequences listed in Table 1, pg36). If neither or one of duplicate PCR reactions increased the fluorescence above the threshold in 80 cycles for Human and Soil DNA or 120 cycles for No DNA controls, * is indicated for Cycle No. and CV(%) respectively.

Primer	Soil A DNA Template		Human DNA Template		No DNA Template	
	Cycle No.	CV (%)	Cycle No.	CV (%)	Cycle No.	CV (%)
Seq005	28.13	0.651	31.81	0.840	46.12	3.285
antiSeq005	23.91	0.933	27.89	0.858	46.25	1.739
Seq5A	27.90	0.105	45.15	1.535	47.10	2.632
antiSeq5A	24.20	0.682	45.15	1.081	49.04	0.492
Seq5B	25.58	0.057	29.45	0.218	46.67	3.218
Seq5C	30.84	0.842	49.67	25.829	*50.15*	*
Seq5D	32.79	3.920	*	*	*	*
Seq5E	39.09	3.888	45.22	1.212	*	*
Seq5F	37.25	2.929	40.34	1.284	*	*
Seq5G	36.04	1.131	*	*	*	*
Seq5H	28.32	0.259	32.72	2.313	*	*
Seq5I	29.71	0.770	34.29	2.706	*	*
Seq5J	*	*	63.60	2.649	*	*
Seq5K	27.40	0.507	32.06	10.776	*47.65*	*
Seq5L	39.36	0.597	41.02	0.300	*	*
Seq5M	33.14	0.045	*46.00*	*	*	*
Seq5N	35.64	1.044	*	*	*	*
Seq5O	35.03	1.319	38.67	3.045	*	*
Seq5P	31.59	0.543	33.16	2.850	*	*
Seq5Q	*38.80*	*	*	*	*	*
Seq5R	34.98	1.819	36.88	7.101	*	*
Seq5S	32.83	1.462	31.97	2.617	*	*
Seq5T	33.87	1.224	*	*	*	*
Seq5U	28.78	0.587	31.96	1.384	50.46	24.645
Seq5V	25.84	0.139	30.85	1.800	47.84	10.721
Seq5W	28.07	0.208	*40.43*	*	*47.25*	*
Seq5X	26.97	0.054	26.69	0.276	50.59	54.449
Seq5Y	52.43	1.518	59.55	0.115	*50.20*	*
Seq5Z	36.63	0.146	46.09	0.735	*49.20*	*
Seq5AA	28.11	0.692	33.98	6.432	*51.02*	*
Seq5AB	35.66	1.026	48.79	1.917	*	*
Seq5AC	34.71	0.000	44.21	0.620	50.29	*
Seq5AD	29.26	0.320	33.24	4.061	47.72	*
Seq5AE	30.57	0.821	37.49	0.356	*	*
Seq5AF	34.53	0.574	37.16	1.843	*	*
Seq5AG	*	*	*	*	*	*
Seq5AH	43.21	2.223	59.32	14.519	*	*
Seq5AI	29.20	0.294	30.17	0.144	*47.14*	*
Seq5AJ	28.93	0.108	29.49	1.776	48.80	52.830
Seq5AK	28.26	0.110	30.74	0.228	*50.17*	*
Seq5AL	33.42	0.663	*34.87*	*	*	*
Seq5AM	36.55	1.287	53.21	1.605	*	*
Seq5AN	40.47	14.313	55.80	1.928	*	*
Seq5AO	*40.84*	*	60.57	2.397	*	*
Seq5AP	*43.94*	*	50.76	2.727	*	*
Seq5AQ	*	*	57.46	2.719	*	*

Figure 3.17: Arbitrarily Amplified profiles of Human DNA generated using Real-Time PCR and several Seq5 primers (antiSeq005, antiSeq5A, Seq5A, Seq5B, Seq5C, Seq5D, Seq5E, Seq5F, Seq5H, Seq5I, Seq5K, Seq5AF and Seq5AH). Phage SPP-1/*Eco*RI digest loaded as marker (M) and corresponding fragment sizes shown. Note how the sizes of products generated with Real-Time PCR contrast with products of standard PCR (Figure 3.18)

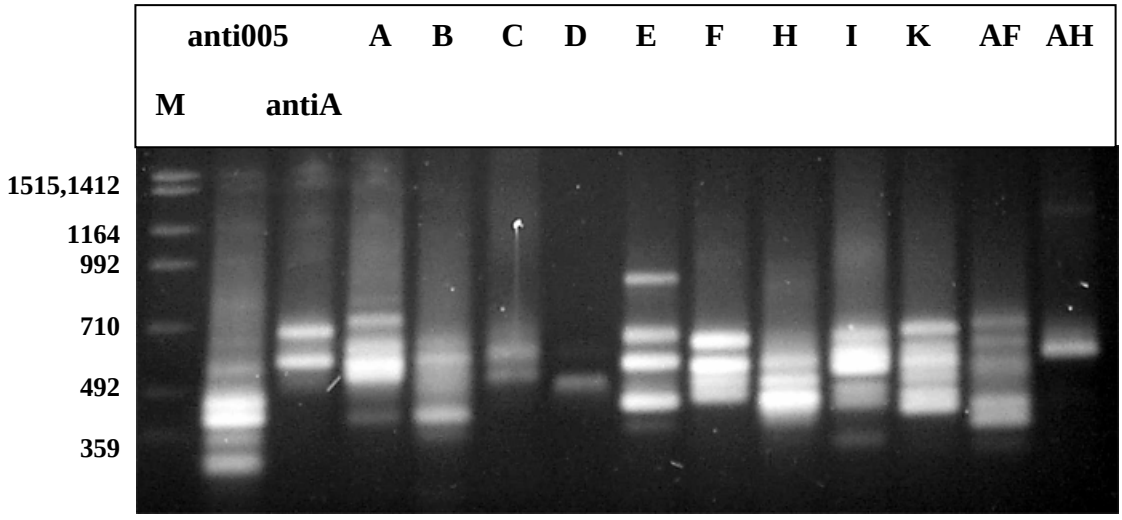


Figure 3.18: Amplified profiles of Human DNA using primers Seq5A through Seq5O (A-O) (See Table 1 for a complete list of all primer sequences). Phage SPP-1/*Eco*RI digest loaded as marker (Mk) (see Figure 2.3 for all band lengths).

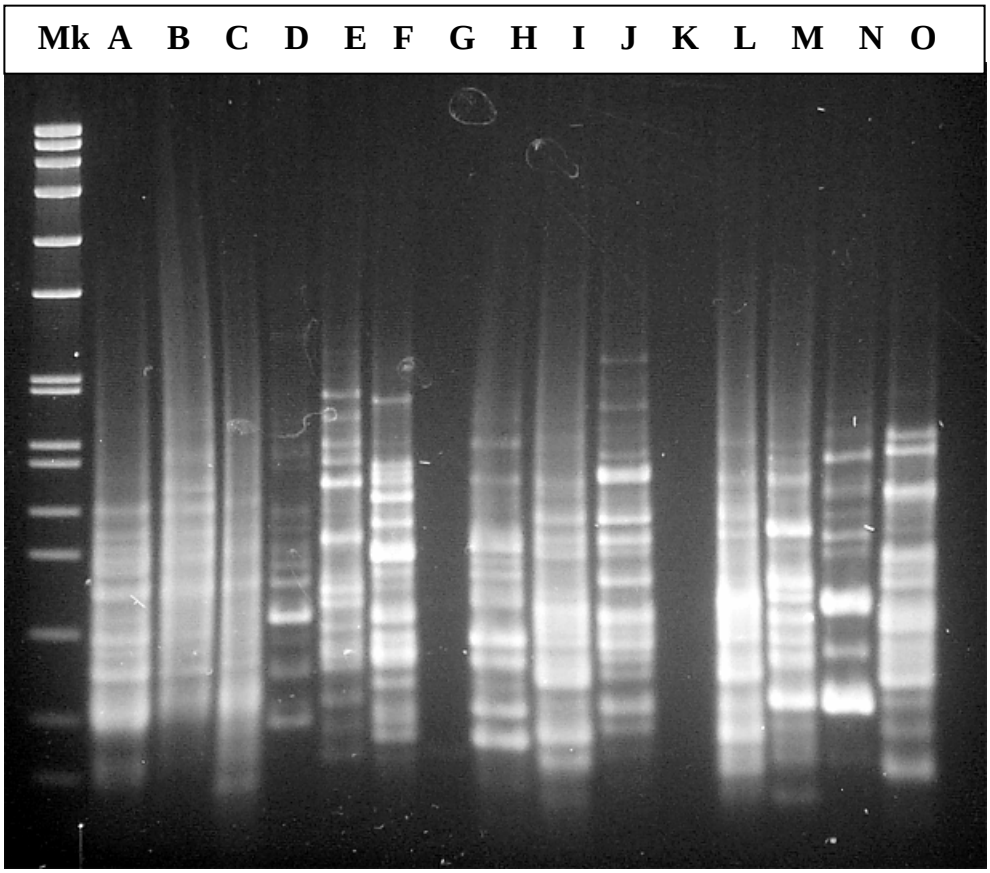
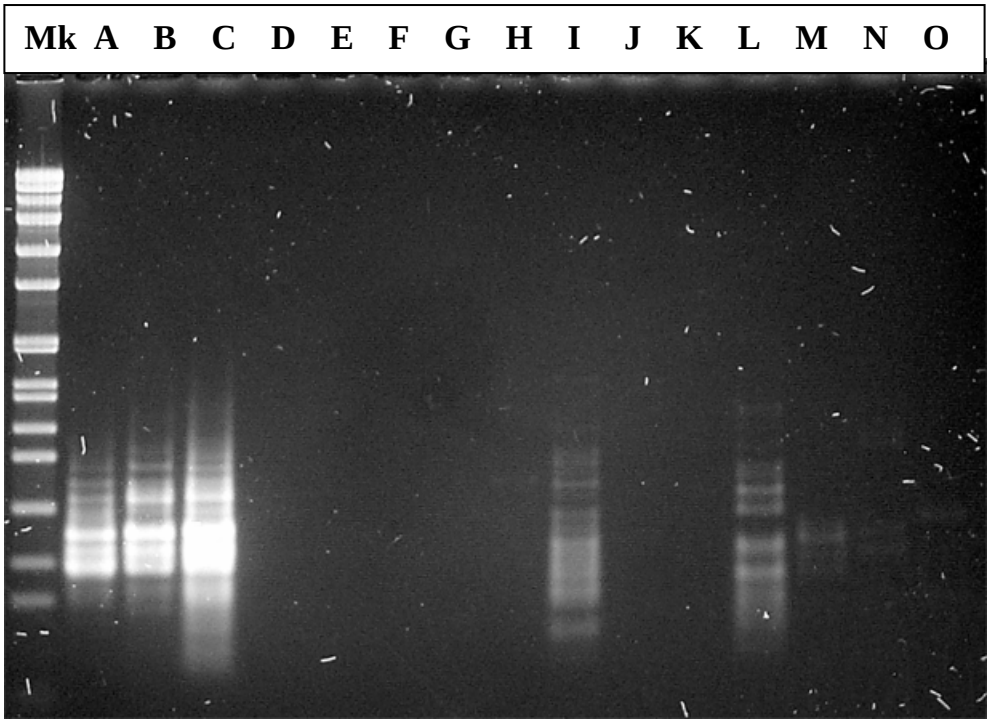
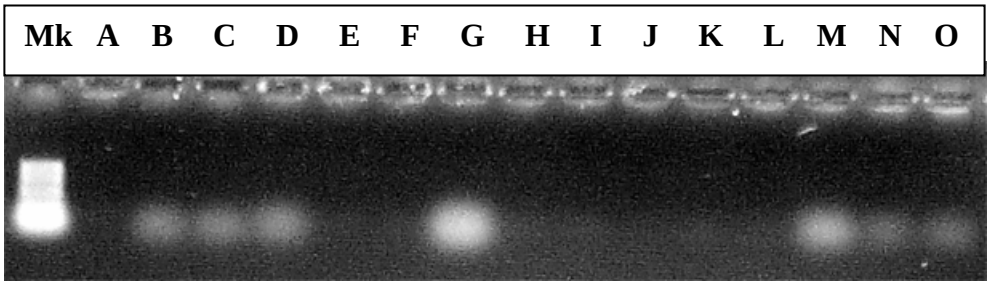


Figure 3.19: Arbitrarily Amplified DNA profiles of No Template Controls using primers Seq5A through Seq5O (A-O) (See Table 1 for a complete list of primer sequences). Phage SPP-1/*Eco*RI digest loaded as marker (Mk) (see Figure 2.3 for all band lengths). This figure illustrates the sensitivity of some of the primers used in this study. Data from Real-Time PCR trials on No Template Controls using all primers is summarised in Table 6.



The failure of primers Seq5G and Seq5K to amplify human DNA and low number of fragments generated by Seq5D and Seq5N primed amplifications may be attributed to the presence of CpG nucleotides, as Human and other mammalian genomes are known have low CpG frequencies (Josse*, et al.* 1961; Swartz*, et al.* 1962), reducing the number of potential priming sites for these four primers. To assess whether any other structural properties of the primers were causing reduced amplification efficiency, primers Seq5A to Seq5O were loaded on an agarose gel, electrophoresed and stained with ethidium bromide, which binds to double stranded DNA with a much higher affinity than single stranded DNA (Figure 3.20, pg101). The amount of ethidium bromide bound to the primer is an indication of secondary structure in the primer itself. The presence of ethidium bromide stained DNA in lanes loaded with Seq5B, C, D, G, M, N and O shows little correlation between double stranded secondary structure of primers and amplification efficiency. This lack of correlation may be due to the electrophoresis and PCR being performed under different magnesium, salt and temperature conditions. Some secondary structures of primers may exist under one set of conditions and dissociate under another. Although, the primer with the most intense ethidium bromide stain, Seq5G failed to amplify human DNA, indicating either intra-primer or inter-primer secondary structures may have contributed to the failure of Seq5G to amplify. However, some primers successful at amplifying human DNA also show signs of secondary structures and Seq5K, which failed to amplify shown no sign of secondary structure. Seq5K may be an exceptional example as it consists entirely of cytosine and adenosine residues. Such a primer is likely be relatively soluble, seeking highly insoluble guanine and thymine rich regions of DNA and may not easily form secondary structures by itself, but have difficulties priming template DNA during the PCR. Also, it has been speculated previously that $d(CA/TG)_n$ repeats (Kladde*, et al.* 1994) and poly(CCA)·poly(TGG) sequences (Packer*, et al.* 2000) may have an unusual (non-B-DNA) structure. If this is true then template DNA that would otherwise be targeted by Seq5K may be inaccessible or have structures that inhibit primer loading and/or amplification. Thus, ethidium bromide staining to identify secondary structure was not a useful predictor of successful amplification.

Figure 3.20: Levels of Secondary Structure demonstrated by ethidium binding of Primers Seq5A through Seq5O (A-O) used to amplify profiles (See Table 1 for a complete list of primer sequences). PCR Marker used as size standard (Mk). Sizes of bands are equal to 1000,750, 500, 300, 150 and 50bp (see Figure 2.2).



3.4.2 Amplification of Trace Amounts of DNA from PCR Reagents

Amplification of DNA from "no-template" negative controls is of major concern to the reliability of arbitrary amplification of DNA from minute samples, particularly if the application of such a technology is for forensic investigations. To investigate the nature of this undesirable DNA and potential ways to overcome it, several products from no-template negative controls were ligated into T-vectors, transformed into competent *E.coli*, cultured and the resulting plasmid DNA sequenced (detailed methods in Chpt 2.9). Sequence data were used to search the Main GenBank DNA database and SwisProt and SpTrEMBL protein databases for potential matches using the blastn and blastx programs (Altschul, *et al.* 1997) respectively, available on the BioManager website (http://biomanager.angis.org.au). Many close matches were found from bacterial sources (see Appendix CD\Sequences.doc). The possibility exists that common reagent-borne sequences may be amplified in two separate samples and when compared, the contaminating DNA will artificially increase the apparent similarity of the two samples. Conversely, distinct reagent borne sequences may be amplified in separate samples and when compared may artificially decrease the similarity of the two DNA samples by introducing new DNA fragment lengths (analysed by AADLP) or by lowering the total sequence similarity between profiles (AADSS analysis).

3.4.2.1 The Possibility of Primer Concatenates Contributing to Cross-Hybridisation.

Interestingly, one cloned sequence was found containing a series of intact and corrupted primer concatenates at one end of the sequence (see Appendix CD\Sequences.doc, insert code 215+primer concatenate tail). It may be possible that primer concatenates contribute a substantial proportion of the amplified product from "no template" controls, but may not have been cloned with as much ease as other sequences. This theory is supported by the result represented in Figure 3.10, pg82 (panel A) and Figure 3.11, pg83 (panel A). Both Figures 3.10 (panel A) and 3.11 (panel A) show a high level of fluorescence associated with

"no template" controls amplified with reagents different to those of the labelled profile. There are two potential explanations for this result, both of which may occur, the first is the possibility of all enzyme sources containing similar contaminating DNA sequences and the second is the possibility that the amplified sequence contains a large degree of primer-derived sequence (possibly in the form of primer concatenates). It is possible that different enzyme batches may contain similar contaminating sequences as similar organisms may be used to produce the thermostable enzymes between manufacturers. Any contaminating sequences shared by reagent sets could lead to high levels of cross-hybridisation between a profile amplified from soil DNA and a "no template" profile amplified with a different enzyme. However, the absence of any real fluorescence in the "no template" control using the same reagents as the labelled profile suggests identical reagent-borne sequences are not frequent. The presence of cross-hybridising sequence in the "no template" controls of RedHot and BioTaq polymerases may be due to their increased activity (Figure 3.9, pg81), which in the absence of any real template may result in the amplification of primer concatenates. Countervailing this result is the absence of any cross-hybridisation between labelled profiles of soils B or C and "no template" controls (Figures 3.10, pg82 and 3.11, pg83 panels B and C), which would be expected if labelled profiles were cross-hybridising to primer concatenate sequence in the "no template" controls. It may be the case that legitimately amplified sequence will not cross-hybridise to primer concatenate but soil A profile may have also amplified an amount of primer concatenate sequence which cross-hybridised to the controls. Regardless of the possibility of primer concatenate sequence interfering with profile comparison, the observed reality is that when comparing profiles generated from an adequate amount of template, the level of cross-hybridisation is small indicating any influence of primer concatenate sequences is minimal. However, it is important to be aware of this effect and it should always be monitored by observing cross-hybridisation to a small suite of no-template controls on the array.

3.4.2.2 Removing the influence of reagent-borne DNA.

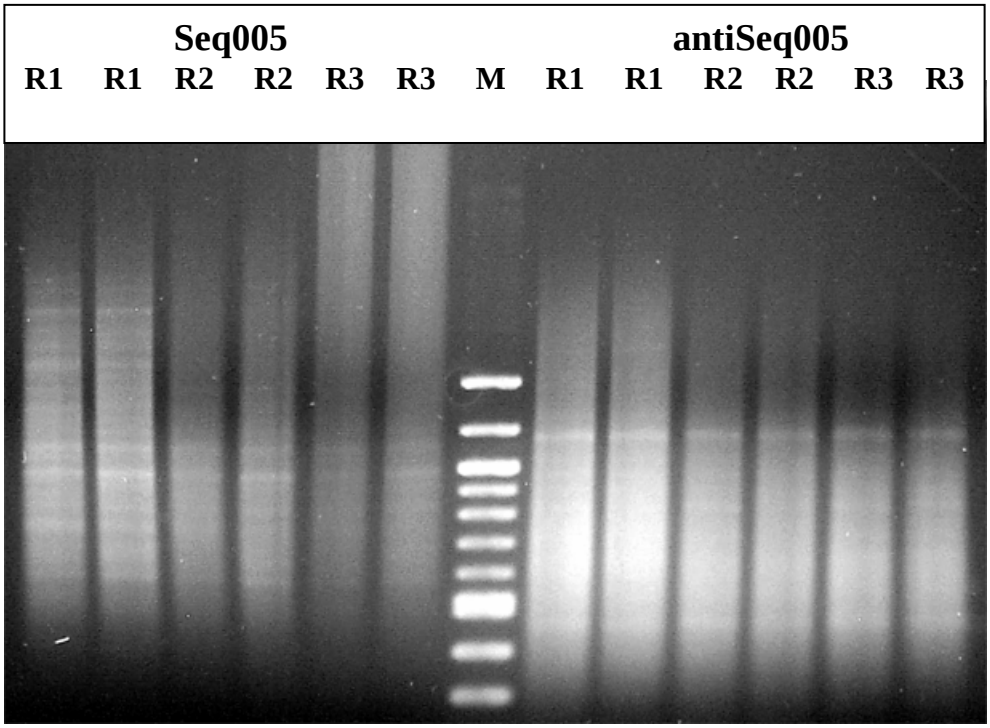
There are a number of ways to combat such an important issue. Care should be taken when selecting both PCR reagents and oligonucleotide primers. The relative purity of polymerases and buffers differs between products and manufacturers. The difference in the range and amount of sequences amplified from no-template control reactions between four commercially available thermostable DNA polymerases is large (Figure 3.9, pg81). Careful selection of PCR enzymes and reagents can limit the amount of contaminating DNA carried over, while maintaining a high level of amplification for legitimate samples. Optimisation of the chosen enzyme(s) with a variety of primers will also allow control over which primers to use for analysis. For example, many primers failed to amplify any DNA from no-template controls while maintaining high amplification efficiencies for soil A and human DNA (e.g. Seq5H, I, O, P, R, S, AE, and AF; Table 6, pg96).

More aggressive means such as UV irradiation and DNaseI treatment of PCR reagents were also investigated. PCR reagents were exposed to varying levels of UV irradiation before adding template and primers and thermal cycling. However, a high dose of UV irradiation (250mJ) caused the profile of human DNA to change before the complete removal of contaminating DNA in no-template controls (data not shown). DNaseI was also used in an attempt to remove contaminating DNA by incubating PCR reagents with varying amounts of the nuclease before heat denaturation of the DNaseI, addition of template and primer and thermal cycling. Low doses of nuclease did not remove contaminating DNA sufficiently from no-template controls, while higher doses caused the amplification to fail even when DNA was added after treatment (data not shown).

3.4.3 The Effect of Successive Rounds of PCR on Profile Composition

The number of rounds of PCR a DNA sample is subjected to affects the profile generated. Fragment lengths and fingerprints (Figure 3.21, pg105) and total sequence similarity (Figure 3.22, pg106) of profiles generated from soil DNA change after successive rounds of amplification, as expected.

Figure 3.21: Arbitrarily Amplified DNA profiles of Soil A (A) and Soil B (B) DNA after one, two or three rounds of amplification (R1, R2 or R3 respectively) using primers Seq005 or antiSeq005. R1 samples underwent one round of low stringency amplification (PCR1, Cycle1 Chpt 2.5.3). R2 samples underwent one low stringency and one high stringency amplification (PCR2, Cycle2 Chpt 2.5.3). R3 samples underwent one low stringency and two rounds of high stringency amplification (PCR2, Cycle2 then PCR2 and Cycle2 again, Chpt 2.5.3). 100bp DNA ladder (Figure 2.2) used as marker DNA (M).



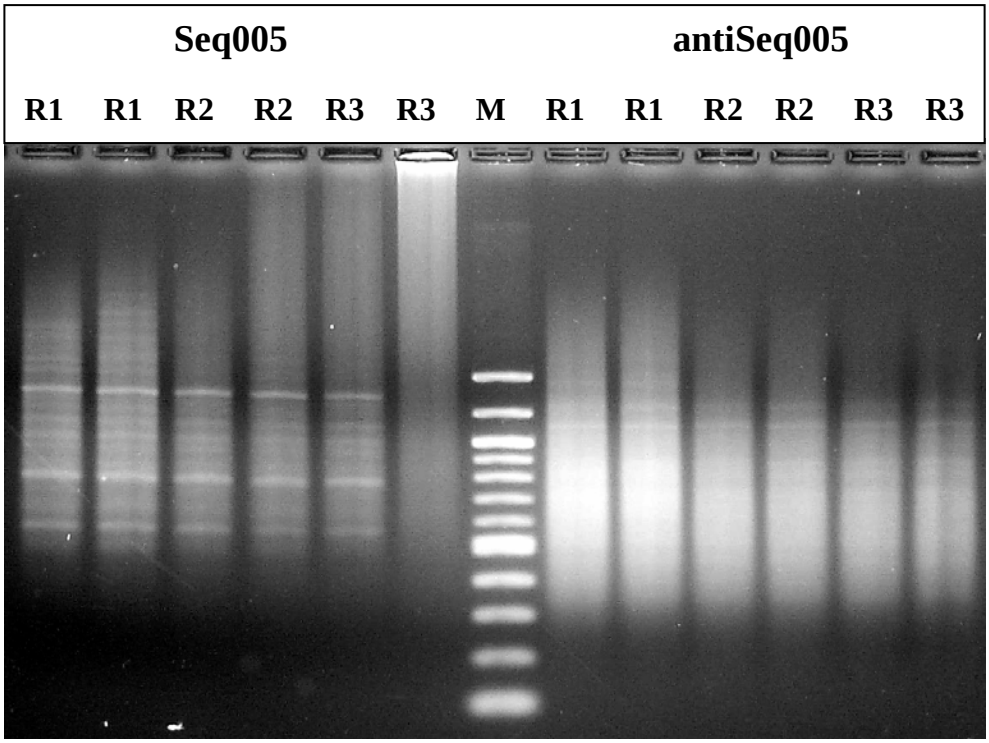
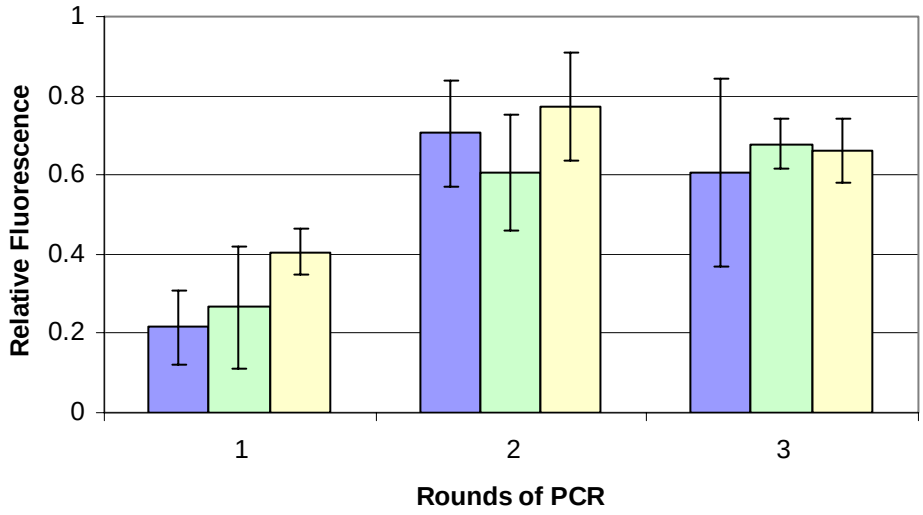


Figure 3.22: Cross-hybridisation of arbitrarily amplified profiles amplified for 1, 2 or 3 rounds of amplification. Relative fluorescence (+/- 1SD) of profiles amplified from soil A (Blue), B (Green) or C (Yellow) DNA using one, two or three 35 cycle rounds of PCR after hybridisation with the labelled profiles of corresponding soils, amplified with **two 35 cycle rounds** of PCR.



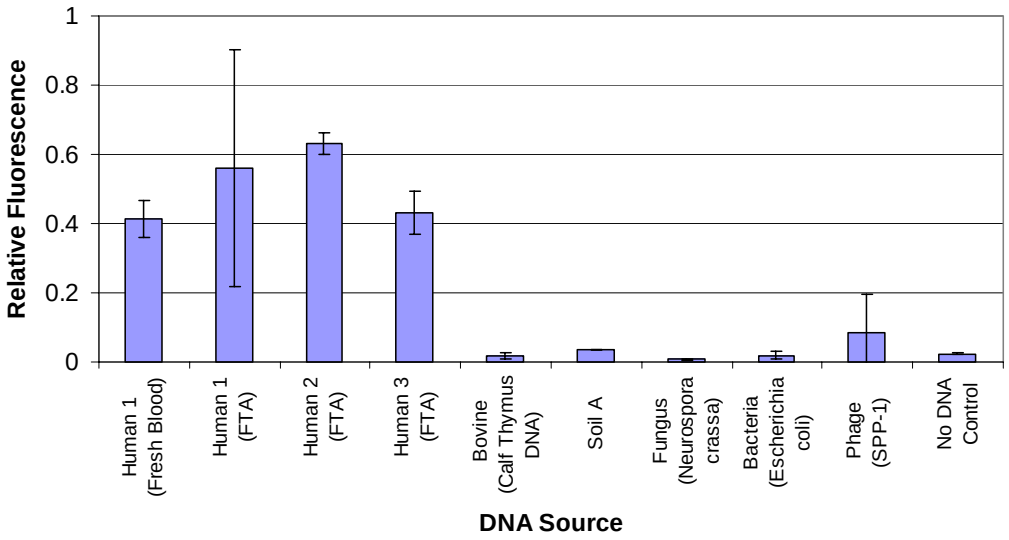
It is interesting to note that while profiles generated using one round of PCR have reduced sequence similarity to profiles generated using two rounds of PCR, third round profiles display high levels of sequence similarity to second round profiles. This differs from length polymorphism patterns where round one and two profiles look alike and round three profiles show signs of high molecular weight DNA (possibly concatemers or chimeras of individual sequences, joined together by excessive amplification) replacing the distinct fragment lengths used for fingerprint analysis. Regardless, once standard conditions and reagents have been established, this technology is reproducible, robust and discriminatory.

3.4.4 DNA Profiles of Different Humans and Organisms

DNA isolated from blood stored on FTA®, of three people (2 Caucasian [LAB, JMW] and 1 Asian [JC]), Calf Thymus, *Neurospora crassa*, *Escherichia coli*, *SPP-1* phage (GenBank Accession Number: BSPP1GENM) or soil A was arbitrarily amplified and subsequently compared to a Caucasian human DNA profile (LAB) by microarray analysis (Figure 3.23, pg108). The relative fluorescence of all human profiles from FTA®-stored blood, hybridised with a labelled human profile generated from fresh blood, were as high as or higher than the relative fluorescence of fresh blood profiles (positive control), indicating highly similar sequences in all human profiles. Despite the Asian background of JC, this profile gave the highest RFU value of all human profiles when probed with labelled profile of Caucasian origin (LAB), indicating ethnic differences did not affect the sequences amplified. However, the amplification of DNA from FTA® paper means that an accurate determination of the amount of template DNA used in the amplification was not possible. Different amounts of DNA used to generate profiles, does affect the resulting profile to some degree (data not shown). Thus, the profiles generated from FTA® stored DNA may not be entirely standardised and repeatable.

Profiles generated from other organisms, even another mammal (Bovidae), showed no significant sequence similarity to the human profile. This lack of cross-hybridisation gives an indication of the level of discrimination at which the system is operating, which appears to be at the species level for Humans and possibly other eukaryotes.

Figure 3.23: Cross-hybridisation of arbitrarily amplified DNA profiles from DNA of three humans isolated from FTA storage paper as well as fresh blood of Human 1 and genomic DNA from other organisms. Relative fluorescence (+/- 1SD) of profiles generated using Seq005 and various genomic DNA sources were hybridised with a labelled profile of Human 1 DNA isolated from fresh blood (Human 1 Blood).



3.4.5 The Degree of Overlap of Amplification Products using Different but Related Primers.

3.4.5.1 Length Polymorphism Differences.

Unsurprisingly, arbitrarily amplified DNA profiles generated from a particular source of DNA were altered by the use of different but related oligonucleotide primers. Fragment lengths of PCR products, used to analyse samples based on fingerprints of length polymorphisms, were dramatically altered by the use of alternate primers (Figure 3.18, pg98). However, similar banding patterns were observed when amplifying Human DNA with primers Seq5A, Seq5B and Seq5C in accord with conventional theory, as these primers share the same 3′ sequence and thus, should bind to the same template DNA sequence. The difference between Seq5A, Seq5B and Seq5C is the additional 3 and 6 residues at the 5′ end of Seq5B and Seq5C, respectively (Table 1, pg36). Conventional expectations would be that any sequence that Seq5C primed would also be primed by Seq5B and Seq5A, and any sequence bound by Seq5B would also be primed by Seq5A due to the overlap of similarity between these primers. It was expected that Seq5A, Seq5B and Seq5C would all prime the same sites at approximately the same efficiency. During the amplification performed in this study, Seq5C amplified Human DNA (Figure 3.18, pg98) and "no template" controls (Figure 3.19, pg99) contained a more diverse population of products than Seq5B amplified DNA, which in turn contained more diverse products than Seq5A amplified DNA. The increased diversity of Seq5C primed products, relative to Seq5B and Seq5A primed products, is likely to be because of the additional bases on the primers, which may endow these longer primers with increased stability when binding to mismatched sequences. Thus, the short Seq5A primer may not be afforded this binding stability and prime less frequently than its longer counterparts. The affect of primer length on sensitivity and selectivity may be used to advantage by designing primers of different lengths to perform amplifications of varying sensitivity and selectivity, without the need for altering other PCR parameters.

It is also interesting to note that primers containing a CpG dinucleotide, particularly if located towards the 3' end, amplified Human DNA poorly. This is probably due to the relative scarcity of CpG dinucleotides in the Human genome (Josse, *et al.* 1961; Swartz, *et al.* 1962).

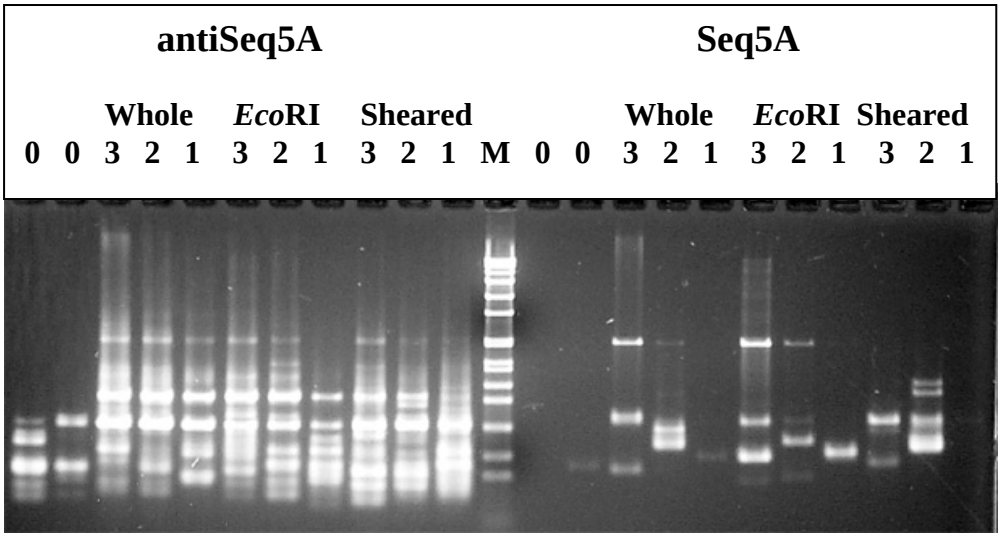
Primers with a high level of sequence similarity may not always amplify similar sequences. Seq5H (CCTCCAACAACACCACC) and Seq5K (CCACCAACAACACCACC) have very similar sequences but did not amplify Human DNA in a similar way (Figure 3.18, pg98). Despite only one changed residue near the 5′ end (T vs. A at position 3 [5′-3′]), Seq5K failed to amplify any DNA while Seq5H amplified many alleles. This example may not represent all instances of this amplification discrepancy as the Seq5K primer, containing only C and A residues, may have special properties. As discussed above, it has been speculated previously that $d(CA/TG)_n$ repeats (Kladde, *et al.* 1994) and poly(CCA)·poly(TGG) sequences (Packer, *et al.* 2000) may have an unusual (non-B-DNA) structure. If this is true then template DNA that would otherwise be targeted by Seq5K may be inaccessible or have structures that inhibit primer loading and/or amplification. This suggests that factors other than Watson-Crick base pairing influence the priming and/or amplification of DNA under the conditions used in this study.

Oligonucleotide solubility may contribute to primer loading during amplification, especially during the initial low stringency round. Each deoxynucleotide residue has a different solubility. The most soluble deoxynucleotide is cytosine (C) with a saturation point of 69.2 nmol/L, followed by thymine (T, 31.7 nmol/L), adenine (A, 3.7 nmol/L) and finally guanine (G) which is virtually insoluble (Budavari 1989). Highly soluble oligonucleotides (containing a large proportion of C and T residues) may prime less readily than highly insoluble primers (containing a large proportion of G and A residues) as these insoluble oligonucleotides will bind to DNA in preference to remaining in solution. This tendency to bind to DNA rather than remain in solution may increase the level of mismatched priming during the low stringency annealing for high G/A primers, leading to a greater diversity of amplicons. Another contributing factor to the greater diversity of amplicons generated by oligonucleotides high in G and A residues is the nature of the sequences they prime. These target sequences are highly soluble, C and T rich sequences and are

generally freely soluble in the PCR solution. This means that they will be less likely to be involved in secondary structures that prohibit the binding of primers and prevent subsequent amplification. Thus, G and A rich oligonucleotides may be expected to amplify more broadly and more efficiently than their C/T rich counterparts. This disparate amplification between G/A rich and C/T rich primers has been observed during the amplification of two soil DNA samples (Figure 3.21, pg105). For both DNA templates (Soil A and Soil B) the amount of product generated and the relative diversity of fragment lengths is higher for amplifications using the G/A rich primer antiSeq005 as opposed to the C/T rich primer Seq005. For Seq005 primed amplifications, a number of distinct bands are observed whereas a smear of DNA is observed in the lanes of antiSeq005 amplified DNA, indicating that a much larger number of sequences have been primed and successfully amplified using the G/A rich primer as opposed to the C/T rich primer. This result is corroborated by real-time PCR (Table 6, pg96). Soil A and Human DNA are both amplified faster using antiSeq005 as opposed to Seq005. It is particularly notable that antiSeq005 amplified Human DNA more efficiently than Seq005 as the CpG dinucleotide (present in both primers) is closer to the critical 3` end in antiSeq005, which would be expected to reduce the number of successful priming events due to the under-representation of CpG dinucleotides in the Human genome (Josse, *et al.* 1961; Swartz, *et al.* 1962). This amplification bias may be due to a disparity in target sequences that provides antiSeq005 with more opportunities to prime and amplify. Even though antiSeq005 and Seq005 are the reverse complement of each other and theoretically should be in equal amounts in the template DNA, the proximity and orientation of target sites and low stringency annealing steps, facilitating imperfect priming, may allow such biased amplification to occur. However, the likelihood of priming sites being orientated and located to favour antiSeq005 amplifications occurring in both soil DNA templates as well as Human DNA is low.

Priming disparity between two complementary primers was also observed with Seq5A and antiSeq5A amplified SPP-1 phage DNA (Figure 3.24, pg112) and real-time PCR amplified Soil A but not Human DNA (Table 6, pg96).

Figure 3.24: Arbitrarily amplified SPP-1 phage DNA using complementary antiSeq5A and Seq5A primers, which behave differently. 100ng (3), 10ng (2), 1ng (1) of SPP-1 phage DNA that was untreated (Whole), restriction digested with *Eco*RI (*Eco*RI) or sheared by sonication to fragments below 3000bp (Sheared) were amplified with antiSeq5A or Seq5A primers, along with no template controls (0). Phage SPP-1/*Eco*RI digest loaded as marker (M) (see Figure 2.3 for all band lengths).



Varying amounts of the small SPP-1 genome (44 Kbp) was amplified with either Seq5A or antiSeq5A and electrophoresed. It is obvious from Figure 3.24 (pg112) that products of the antiSeq5A primed amplification are more plentiful than their Seq5A amplified counterparts. The target sequences of Seq5A and antiSeq5A are very similar as these primers are almost palindromic, but slight differences in the target sequences exist that could account for the biased amplification. SPP-1 does not contain any sequences exactly complementary to the entire Seq5A or antiSeq5A primers, but does contain approximately equal numbers of consecutive complementary bases of various lengths, which should bind the 3′ end of each primer (Table 7, pg114). Assuming an equal distribution and orientation of such potential priming sites, if base pairing at the 3′ region of the primer was the only consideration regarding initiation of amplification, it would be expected that both Seq5A and antiSeq5A would amplify a near equal number of alleles. Equal amplification of the SPP-1 genome was not observed, indicating factors other than Watson-Crick base-pairing may influence priming and/or subsequent amplification of template DNA. Admittedly, the nature of primer-template interactions are highly complex and cannot be fully accounted for by simply measuring the number consecutive bases complementary to the 3′ end of a primer. Other sequences within SPP-1 may facilitate base-pairing by involving the middle or 5′ end of primers, stabilising them to template DNA, despite potential mismatches in the primer's 3′ end.

Table 7: The frequency of target sequences for primers Seq5A and antiSeq5A in both strands (arbitrarily assigned sense and antisense) of the double stranded phage SPP-1. Target sequences are consecutive complementary bases that will bind the 3´ end of each primer, listed in decreasing numbers, from eight to four consecutive matches.

antiSeq5A (5´ - 3´) GGAGGATGGTGTTGGAGG			**Seq5A (5´ - 3´) CCTCCAACACCACCTCC**		
Target sequence (5'-3')	Frequency in SPP-1 Genome		Target sequence (5'-3')	Frequency in SPP-1 Genome	
	Sense	Anti-Sense		Sense	Anti-Sense
CCTCCAAT	0	0	GGAGGTGG	1	0
CCTCCAA	0	7	GGAGGTG	2	0
CCTCCA	3	19	GGAGGT	16	2
CCTCC	12	77	GGAGG	77	12
CCTC	49	244	GGAG	267	23

Thus, it cannot be claimed conclusively that factors other than Watson-Crick base-pairing are involved in priming and amplification of template DNA during arbitrary amplification but it seems highly likely. Priming disparity was observed between Seq005 and antiSeq005 on one Human and two soil DNA samples, as well as between Seq5A and antiSeq5A on Soil A DNA and the defined SPP-1 phage DNA, thus the likelihood of this disparity being observed by chance is low.

Priming disparity may be exploited to suit various applications. High C/T primers may be used to quickly simplify the enormous diversity of a soil metagenome and generate band-based fingerprints, while high G/A primers may be used to broadly amplify the soil DNA in order to give a more representative profile for the purpose of sequence hybridisation.

3.4.5.2 Differences in Sequence Similarity

Not only does the length polymorphism fingerprint change between profiles amplified with different primers but so does the sequence similarity. DNA isolated from Soil A was amplified with a range of primers and hybridised with labelled profiles amplified using primers Seq005, antiSeq005, Seq5A or

antiSeq5A (Figure 3.25, pg116), as well as primers Seq5C, Seq5AC or Seq5D (Figure 3.26, pg117). The resulting fluorescence of the profiles indicates the relative level of sequence similarity between the profiles on the array and the labelled profiles. Despite the large difference in primer sequence between Seq005 and antiSeq005 (the reverse-complements of each other), similarities often exist in the way they hybridise to the other profiles, which was unexpected. The labelled Seq005 profile did not hybridise well to antiSeq005 profiles but after its own profile, the labelled antiSeq005 profile hybridised best to Seq005 profiles (Figure 3.25, antiSeq005, pg116).

The sequence similarity of profiles generated using Seq005 and Seq5A is high, indicated by the high fluorescence of profiles generated by one when probed with the other (Figure 3.25, Seq005 and Seq5A). Subsequent rankings of profiles generated by other primers when probed with either Seq005 or Seq5A are also very similar indicating that both Seq005 and Seq5A prime and amplify similar regions of the metagenome of soil A DNA. An overlap of amplification between Seq005 and Seq5A is not surprising as both these primers share high degrees of sequence similarity, only differing by one mismatched base and one insertion/deletion at the 5'end. Unlike the similarity of profiles generated by Seq005 and antiSeq005, Seq5A and antiSeq5A seem to generate very distinct profiles of soil A DNA (Figure 3.25, Seq5A and antiSeq5A). Almost no sequence similarity appears to exist between profiles generated by antiSeq5A and any other primer. Seq5C amplified profiles contained sequences similar to profiles amplified with many other primers including Seq5A (Figure 3.26, Seq5C, pg117). However, due to the 17 base identical 3' sequence between Seq5A and Seq5C, if sequence similarity of the primer to template is the only factor for PCR priming and amplification, it is expected that profiles generated using these two primers would cross-hybridise more than profiles generated using Seq5AI or Seq5V which share only 7 residues and 10 residues consecutive 3' sequence similarity, respectively to Seq5A and Seq5C. However, profiles generated using Seq5V or Seq5AI bind more of labelled Seq5C profiles than Seq5A profiles. Profiles generated with Seq5D or Seq5AC primers, which differ to Seq5A by a C to G substitution at the first or fourth (3' to 5' orientation) base respectively, do not show any strong sequence similarity to other profiles generated using alternate primers (Figure 3.26, Seq5AC and Seq5D).

Figure 3.25: Cross-hybridisation of arbitrarily amplified profiles of Soil A DNA, amplified with various primers. Relative fluorescence (+/- 1SD) of profiles generated using Soil A DNA as template and various primers (for sequences see Table 1) after hybridisation with labelled profiles of Soil A using primers Seq005 (Blue), antiSeq005 (Yellow), Seq5A (Green) or antiSeq5A (Red).

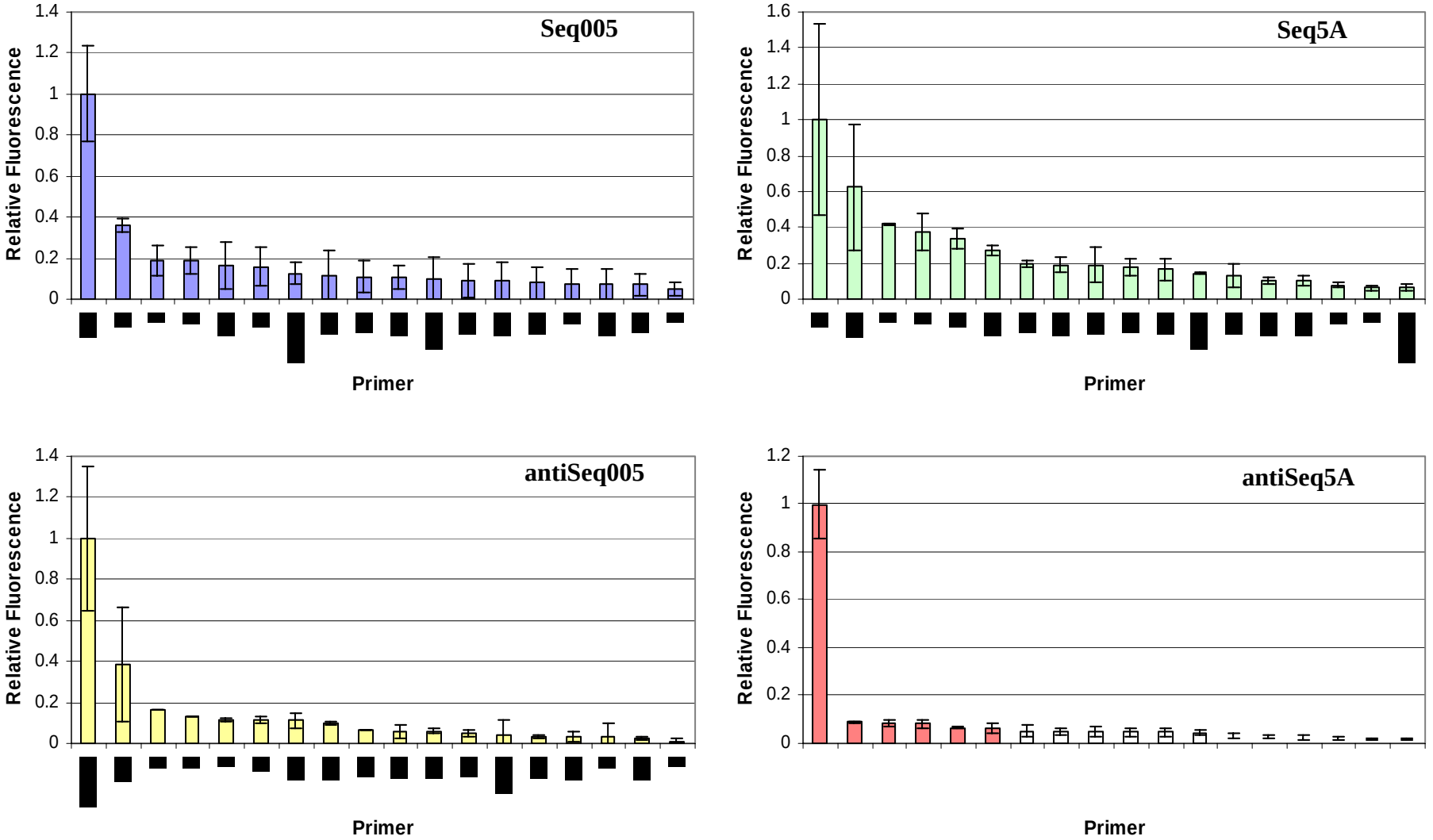
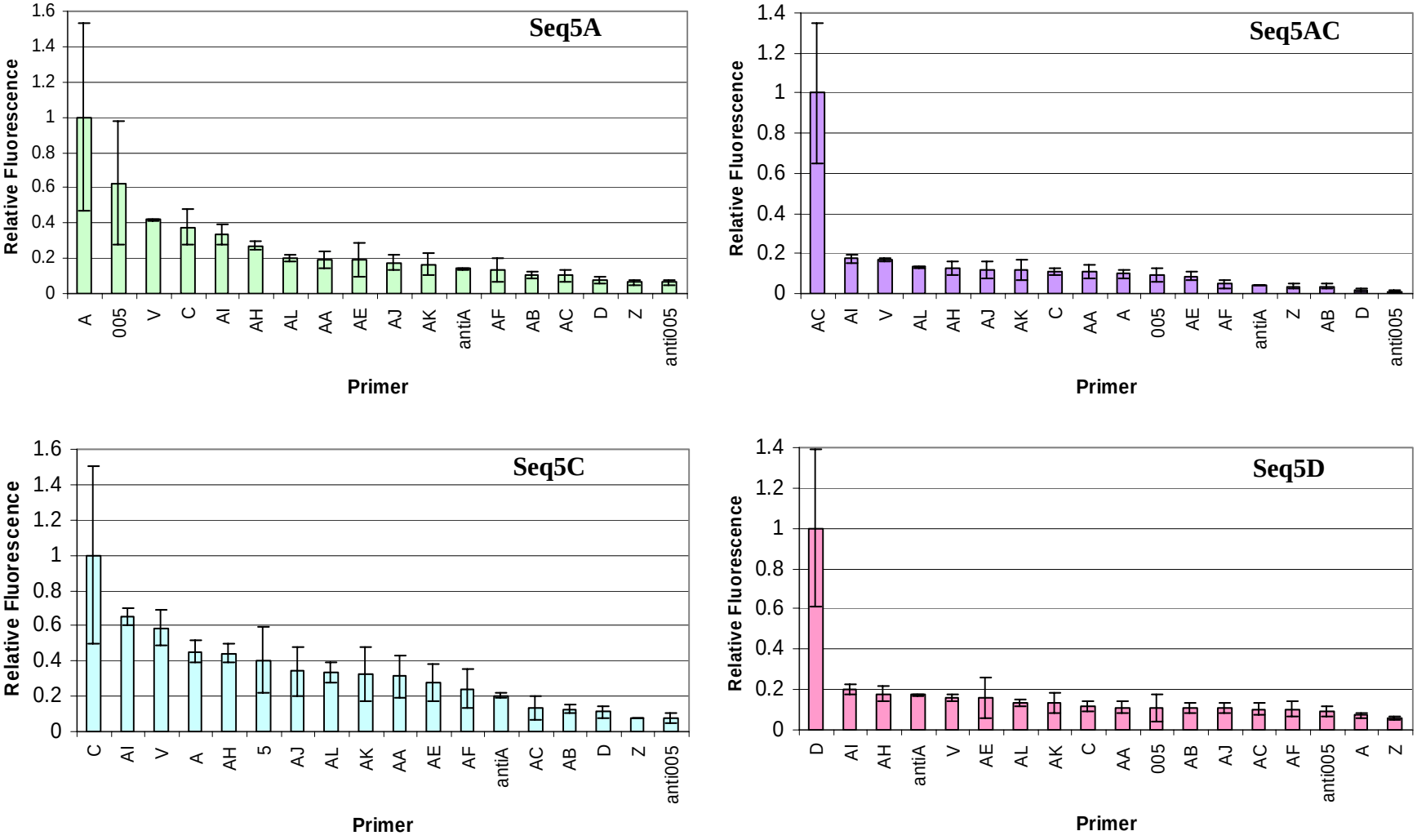


Figure 3.26: Cross-hybridisation of arbitrarily amplified profiles of Soil A DNA, amplified with various primers. Relative fluorescence (+/- 1SD) of profiles generated using Soil A DNA as template and various primers (for sequences see Table 1) after hybridisation with labelled profiles of soil A using primers Seq5A (Green – as Figure 3.25), Seq5C (Light Blue), Seq5AC (Purple) or antiSeq5D (Pink).



DNA isolated from human blood was also amplified with the same set of primers (Figure 3.27, pg119). Human DNA profiles hybridised with labelled Seq005 amplified profiles showed a sequence similarity hierarchy similar but not identical to profiles of soil A DNA (Figure 3.25, pg116). When the Human DNA profile was labelled using Seq005 primer and human ***genomic*** DNA as template (PCR1, Cycle1, Chpt 2.5.3), rather than PCR products from a first round amplification (PCR2, Cycle2, Chpt 2.5.3), the level of sequence similarity between profiles generated with different primers increased relative to Seq005 profiles. However, not all primers underwent the increase in sequence similarity equally, as some profiles increased more, relative to others.

Unlike the profiles of soil A DNA hybridised with labelled antiSeq005 profiles, human DNA profiles show little sequence similarity to labelled antiSeq005 profiles (Figure 3.27, antiSeq005, pg119). Like Seq005, labelled profiles generated with Seq5A shared sequence similarity to profiles amplified using other primers in much the same manner as soil A DNA profiles.

In many cases the level of similarity between profiles can be correlated to the similarity of the oligonucleotide sequence i.e. similar oligonucleotide sequences will bind similar target sequences leading to the amplification of similar or identical alleles. However, in some cases this correlation does not appear to fully explain the level of sequence similarity between profiles. For example, profiles generated from Soil A DNA with primers Seq005, Seq5A, Seq5C and Seq5V all share some sequence similarity (Figures 3.25, pg116 and 3.26, pg117), from 21% to 62 % RFU, relative to the positive control profiles (replicate profiles of the labelled profile). Some degree of sequence similarity between amplification products of Seq005, Seq5A, Seq5C and Seq5V is unsurprising as all these primers have similar sequences (Table 1, pg36). However, the rank order of the degree of similarity does not appear to correlate with primer sequence similarity. The similarity of Soil A profiles of primers Seq005, Seq5A, Seq5C and Seq5V (Figures 3.25 and 3.26) is represented by the dendrogram in Figure 3.28 (pg120). Labelled profiles generated using Seq5C are more similar to Seq5V profiles (59% of Seq5C profile RFU values) than Seq5A profiles (45% of Seq5C profile RFU values). However, Seq5A has its entire 17mer sequence in common to the 3` end of the 23mer Seq5C, while Seq5V has 7 of its 23 residues (30%) mismatching Seq5C (see below).

Figure 3.27: Cross-hybridisation of arbitrarily amplified profiles of Human DNA, amplified with various primers. Relative fluorescence (+/- 1SD) of profiles generated using Human DNA as template and various primers (for sequences see Table 1) after hybridisation with labelled profiles of Human DNA generated using primers Seq005 and genomic DNA (Blue – First Round Profile), Seq005 and first round PCR as template (Green – Second Round Profile), antiSeq005 (Yellow) or Seq5A (Red).

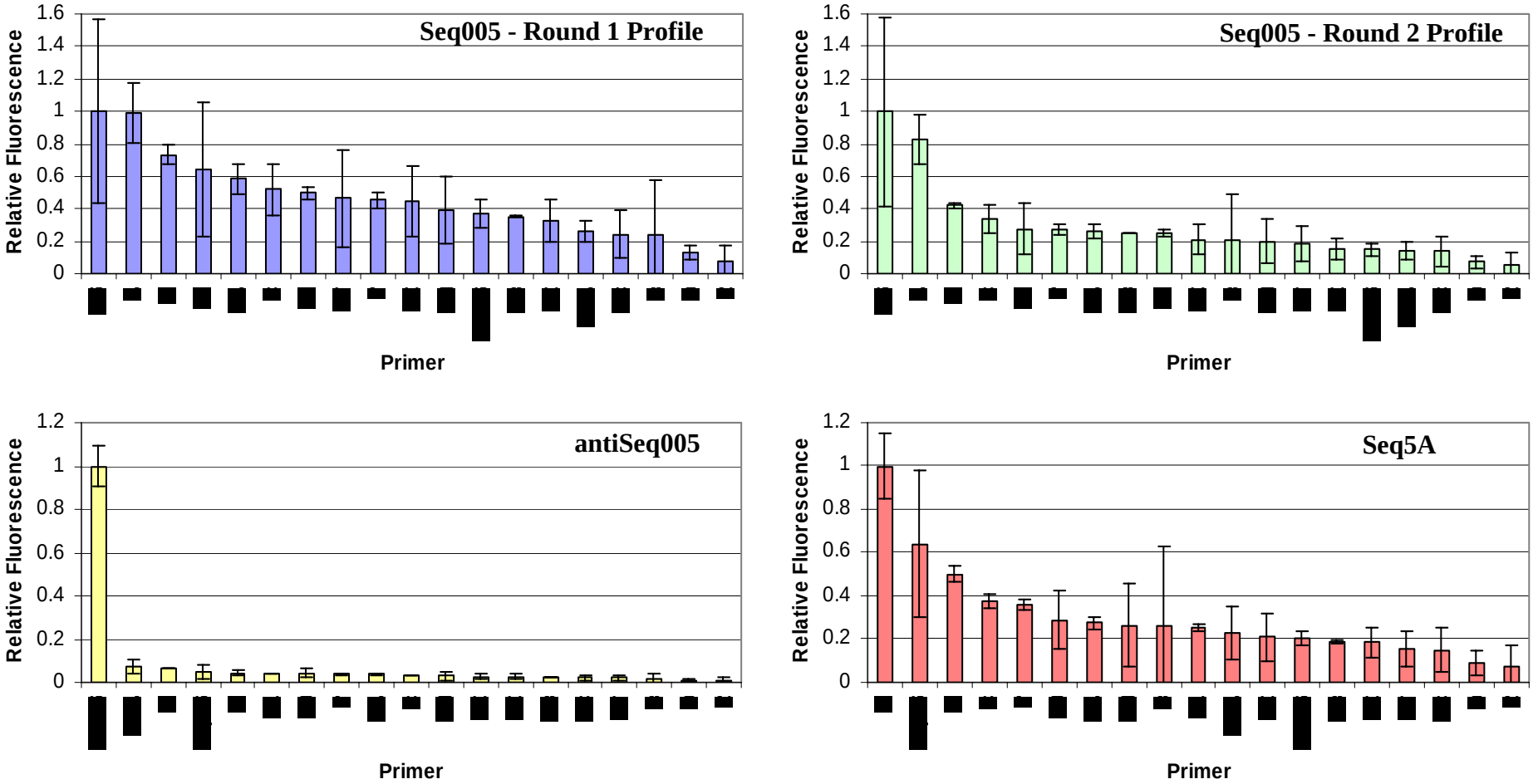
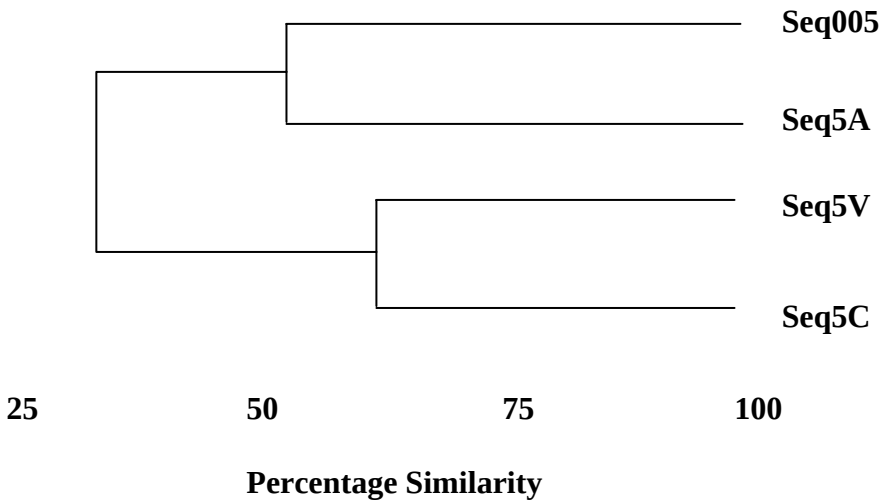


Figure 3.28: A representative dendrogram of the sequence similarity similarity of arbitrarily amplified DNA profiles of Soil A DNA generated using four separate primers (Seq005, Seq5A, Seq5C and Seq5V). Similarity is estimated as the mean RFU of profiles when hybridised to each other, relative to positive control profiles (replicate profiles of the labelled profile).



Seq5C (5` to 3`) **CTCCACCCTCCAACACCACCTCC**

Seq5A (5` to 3`) **CCTCCAACACCACCTCC**

Seq5C (5` to 3`) **CTCCACCCTCCAACACCACCTCC**

Seq5V (5` to 3`) **CCTCCAACACCACCACCACCTCC**

Theoretically, Seq5A primers should bind at all sequences Seq5C binds. However, the additional 6 residues at the 5` end of Seq5C may confer extra stability of primer-template duplexes during the low stringency annealing, allowing additional target sequences to be primed for amplification. As discussed above (Chpt 3.4.5.1), the diversity of banding patterns from Seq5A and Seq5C derived profiles (Figures 3.18, pg98 and 3.19, pg99) suggests additional alleles were amplified using Seq5C primers. These additional sequences amplified by Seq5C may account for the lowering of profile sequence similarities. Despite less than 70% sequence identity, Seq5C and Seq5V derived profiles have a high level of sequence similarity. It is likely due to the absence of mismatches in the 3` end of Seq5C and Seq5V (see above) that these primers amplify homologous sequences, confirming the importance of the 3` end of primers as a major factor for primer-template binding. However, the high degree of mismatched bases in the 5' end of Seq5C and Seq5V highlights the relative unimportance of the 5` sequence in template DNA identification. It may be the case that the 3` end of primers are involved in sequence recognition and binding, while the 5` ends are less important in binding specific sequences but may bestow a degree of primer-template stability, particularly with longer primers, allowing successful priming and amplification of a greater number of alleles to occur. Amplifications with longer primers may produce profiles containing a more diverse population of sequences and may be more representative of the entire DNA community.

The greater diversity of sequences amplified using the longer Seq5C as opposed to Seq5A may also be due to differences in selective pressures during the PCR amplification. Biases can occur during PCR reactions that cause unequal amplification rates for different sequences. Such biases can occur when sequences are amplified to abundant levels and rehybridisation occurs before primers can hybridise to the single stranded product (Mathieu-Daude, *et al.* 1996). This "C_ot

effect" means that rare sequences are amplified at a faster rate than abundant sequences. Also, DNA sequences interact with themselves and one another, forming base pairing between complementary sequences, providing the basis on which primers bind to template and PCR can occur. However, in some cases primer-binding sites may be bound to other sequences forming secondary structures that impede the binding of new primers for amplification. Such sequences will be amplified much slower than those which do not form secondary structures involving primer binding sites. Amplification biases between sequences will cause a selection pressure whereby the "fittest" sequences that do not impede primer loading or extension are amplified in greater numbers than "unfit" sequences, which inhibit primer loading and/or extension by the formation of secondary structures. This is the basis of the SMIPS profiling method of Rogers (2002) (Chpt 1.2.2.3.4)

A potential observation of this corrupt reassociation of amplicons may be seen in the real time PCR experiment involving Human DNA as template and Seq5C primer (Green curve, Figure 3.16, pg 95). As the amplification progresses, the fluorescence increases as normal until it peaks (after 27 cycles) then decreases sharply before slowly increasing again. The decrease in fluorescence indicates a decrease in the amount of double stranded DNA, which may be due to corrupted reassociation of amplicons. It is interesting to note that this decrease in fluorescence occurs late in the amplification when the majority of the primers have been incorporated into product. The abundance of PCR products and deficiency of primers may allow corrupted reassociation of amplicons to occur, thus any PCR biases may be expected to have most influence late in the amplification process.

This selection pressure can be observed by looking at profiles amplified from successive rounds of arbitrary amplification. The banding pattern of Soil A and B profiles amplified with Seq005 and antiSeq005 from three successive rounds (35 cycles each) of PCR show a decreasing diversity of bands as the amplification progresses (Figure 3.21, pg105). Transferring a sub-sample of product from one round as template for the next will not account for the loss of diversity as the amounts transferred (5µl) is the amount loaded onto agarose gels. Thus any band observable on the gel has an appreciable amount transferred to subsequent amplifications. Differences in sequence homologies between profiles

generated with different primers are also exaggerated by PCR biases. By hybridising the microarray with labelled profiles generated using Seq005 primers and either Human ***genomic*** DNA or a first round amplification as template DNA, differences were observed between Human DNA profiles generated using different primers (Figure 3.27, pg119). Different primer profiles were more similar to the Seq005 profile generated using genomic DNA template (a first round profile) rather than a first round amplification as template (a second round profile), suggesting that the longer an amplification proceeds the more PCR biases influence the profiles, which become increasingly divergent. It is also possible that errors introduced by polymerases will cause profiles to diverge, but this is likely to have a lower impact than varying quantities of amplified sequences arising from prolonged exposure to PCR biases.

3.4.5.3 Practical Outcomes of using Different but Related Primers for Arbitrary Amplification of DNA.

Amplification differences between primers are also of practical use. The overlap of profile sequence similarity generated by the use of some primers (Seq005, antiSeq005 and Seq5A in Figure 3.25, pg116; Seq5C in Figure 3.26, pg117; Seq005 and Seq5A in Figure 3.27, pg119) and the absence of sequence similarity overlap between others (antiSeq5A in Figure 3.14, pg89; Seq5AC and Seq5D in Figure 3.26, pg117; antiSeq005 in Figure 3.27, pg119) may be exploited. Multiple profiles of the same template DNA with very little redundancy may be possible with careful selection of primers. This will allow more information to be generated for each sample and a more accurate comparison of the biological communities that reside within each sample.

Careful selection of primers may also provide highly sensitive amplification with very little amplification of reagent-borne sequences. Primers Seq5H, Seq5I, Seq5P, and Seq5S, all amplify soil and Human DNA above the arbitrary threshold within 35 cycles while failing to amplify “no template” control samples. Optimisation of primers for each DNA sample may not be readily performed in all cases but could be tested on a number of typical samples and appropriate primers applied to real cases. Preventing amplification of reagent-borne DNA by prudent primer selection will be a large advantage. Although

experiments in this study showed that the use of primers that did amplify reagent borne DNA did not seem to have any adverse affect, such an issue will undoubtedly negatively affect the weight this evidence would carry and is best eliminated to remove any potential criticism.

4 DISCUSSION

4.1 Comparison of TRFLP and Arbitrarily Amplified DNA Length Polymorphism (AADLP) Analyses.

4.1.1 Similarity Index Comparisons

Similarity indices of TRFLP and AADLP profiles generated from identical samples were similar, averaging 0.843 and 0.778 respectively, with TRFLP comparisons slightly superior to AADLP. However, the SI values for comparisons of profiles from different soils were dissimilar for the two methods. TRFLP analysis yielded a mean SI of 0.693, much higher than the 0.147 average for profiles generated from different soils by AADLP. The higher SI of TRFLP comparisons may reflect authentic similarity of the selected bacterial populations within the three soils. This may not be observed for AAD profiles as DNA from all organisms is amplified, which may dilute the effect of bacterial population similarity with the diversity of non-bacterial DNA. Alternatively, the higher number of bands analysed by TRFLP (~100) as opposed to AADLP (10) may have increased the occurrence of amplified fragments being matched coincidentally. Regardless, the higher SI value for non-matching samples generated by TRFLP makes categorisation difficult, as there is only a small difference between SI values of "same" and "different" soils. The superior discrimination of AADLP is also apparent when looking at the ratios of SI values for profiles generated from the same sample against those for different samples. TRFLP "matching" SI values are only 1.22 times the "non-matching" SI value. In contrast, matching AADLP profile SI values are 5.29 times the non-matching values, allowing easy and reliable discrimination between matching and non-matching categories.

4.1.2 Reproducibility Comparisons

Both TRFLP and AADLP length polymorphism techniques are reproducible under standardized conditions. However, in this study TRFLP

profiles were affected by the use of different batches of PCR enzymes and reagents, while AADLP profiles were not. Under standardized conditions, profiles generated from the same sample have high SI values and low standard deviation. The low standard deviation indicates that although SI values are below the ideal value of 1.0, the degree to which they fall below this value is generally consistent. This consistency and replicate analyses (leading to high degrees of freedom) has contributed to the high significance of the difference between matching and non-matching SI values. The statistical analysis of SI values in this study is novel and will allow more objective comparisons of soil samples in future analyses.

4.1.3 Significance of Discriminatory Power

Although both TRFLP and arbitrarily amplified DNA profiles produce significantly different SI values for replicate and distinct soil comparisons, it is not surprising that the probability of SI values from matching and non-matching soils being different by chance (determined by independent measures *t* tests) is far less likely for arbitrarily amplified DNA fingerprints than TRFLP fingerprints (Tables 3 and 1, respectively). The lower probability for AADLP is mostly due to the greater actual difference in SI values between the groups and greater number of pair-wise comparisons performed with arbitrarily amplified DNA profiles. It is difficult to directly compare probability statistics between TRFLP and arbitrarily amplified DNA analyses due to the varying degrees of freedom. However, as they stand, arbitrarily amplified DNA profiles appear to distinguish the three soils samples better than TRFLP.

4.1.4 Fundamental Differences of Analysis

TRFLP and AADLP analyses both allow soils to be matched or distinguished but by different principles. 16S-TRFLP amplifies bacterial ribosomal loci from a subsection of organism to which “universal” primers will bind and facilitate amplification, while AADLP amplifies arbitrary sequences from genomes of potentially any organisms DNA extracted from the soil sample. The question of which profiling strategy is theoretically the best is unable to be answered conclusively. TRFLP examines a subsection of prokaryotes, reducing

the scope of organisms that could potentially contribute to the profile. However, AADLP also examines a subsection of organisms but in a more random manner. Sequences that are amplified using arbitrary primers represent a subsection or profile of all organisms that inhabit the soil but, by using a battery of primers large enough, one could potentially represent almost all organisms in the sample. However, the number of primers to do this would be large with no actual guarantee that all organisms will be represented. Regardless, total species representation is not required for forensic analysis, which only needs a distinguishing profile of the sample. Therefore, either strategy is acceptable as they both generate repeatable and distinctive profiles.

4.1.5 Potential for Improvement

A number of suggestions have been made in this study (Chpt 3.1.4) and elsewhere regarding potential improvements of TRFLP analysis. Utilising these suggested improvements may lead to better discrimination of distinct samples and a larger difference between matching and non-matching SI values. The arbitrarily amplified profiling method and length polymorphism analysis presented here also has the potential for enhanced discrimination powers. By using more than one primer to generate multiple fingerprints of the same sample, profiles that are more comprehensive can be produced from each sample, allowing better discrimination. However, as the technologies stand in this study, the match:non-match ratio and the significance of the difference between mean SI values suggest that AADLP is superior to TRFLP for distinguishing the three soils tested in this study.

4.1.6 Other Banding Based DNA Technologies that may be suited to Forensic Soil Evidence

There are other technologies, similar to the two banding technologies presented in this study that may be suitable for forensic comparison of soil samples. Ribosomal intergenic spacer analysis (RISA) of soil communities (Ranjard*, et al.* 2001; Casamayor*, et al.* 2002; Sigler and Zeyer 2002; Hery*, et al.* 2003) was adapted from a tool developed to identify species of Enterococci in

clinical samples (Tyrrell, *et al.* 1997). This technology amplifies the DNA between the small (16S) and large (23S) ribosomal subunits in the bacterial rDNA operon, using primers that bind to conserved regions of the small and large subunit sequence. The length heterogeneity of this intergenic region between species allows the diversity and composition of bacterial communities to be elucidated. RISA length polymorphisms of fragments amplified from bacterial communities of five distinct soils have allowed investigators to distinguish all five soils using principal component analysis (PCA) (Ranjard, *et al.* 2001). PCA of Length Heterogeneity PCR (LH-PCR) fragments, amplified from the 5' region of 16S rDNA, has also been used to distinguish soil microbial communities from three sites (Ritchie, *et al.* 2000).

Possibly the most common fingerprinting technology after TRFLP is denaturant or temperature gradient gel electrophoresis (DGGE/TGGE). Typically, the 16S rDNA is amplified by PCR as with TRFLP, but the analysis of the amplified genotypes differs. Full-length amplicons are electrophoresed through an acrylamide matrix, with a gradient of increasing temperature or denaturant concentrations, in order to separate distinct genotypes based on the melting point of the sequences they contain. DGGE and TGGE have been used extensively by soil ecologists and microbiologists to study the structure and changes of microbial communities from microbial mats and bacterial biofilms (Muyzer, *et al.* 1993), deep sea hydrothermal vents (Muyzer, *et al.* 1995), hot springs (Ferris, *et al.* 1996; Ferris, *et al.* 1997; Ferris and Ward 1997), estuaries (Murray, *et al.* 1996), lakes, sand dunes (Kowalchuk, *et al.* 1997), dust (Bruns and Scow 2000), rhizospheres (Duineveld, *et al.* 1998), sediments and soils (Kozdroj and van Elsas 2000; McCaig, *et al.* 2001; Peixoto, *et al.* 2002; Girvan, *et al.* 2003).

Some studies have determined the reproducibility of the profiles generated by DGGE separation of soil derived 16S rDNA and most find SI values of replicate analyses to be 0.90 or above (Duineveld, *et al.* 1998; Kozdroj and van Elsas 2000; Girvan, *et al.* 2003). However, one study found within-group similarity to be predominantly comparable to between-group similarity for three distinct grassland soils (McCaig, *et al.* 2001). The SI values obtained for replicate fingerprints of one of these grassland soils was 0.58, only slightly higher than comparisons of the most distant profiles (0.53; P=0.12 as determined by Student's t test) and the three soils could not be distinguished by principal component

analysis (PCA) (McCaig, *et al.* 2001). However, canonical variate analysis (CVA), a subjective technique that looks at the overall pattern of variation across the entire data set and maximizes between-group differences, did distinguish the three grassland soils.

Despite the predominantly repeatable and discriminating profiles generated by DGGE, the band-based nature of the technique is subject to the same limitations as other band-based fingerprinting methods. These shortcomings include the limited resolution of the gels used to separate the bands, under-representing the number of genotypes. McCaig *et al*. (McCaig, *et al.* 2001) found only 77 bands in DGGE fingerprints when clone libraries of the same soil DNA contained ~130 distinct operational taxonomic units (OTU's). This under-representation may result from the inability to detect minor components of the profile on gels or the comigration of distinct OTU's resulting in single bands containing multiple sequences. Alternatively, overestimation of the diversity of genotypes present in microbial communities is possible due to other limitations of PCR amplification coupled with band-based fingerprint analysis. During PCR amplification, particularly when amplifying a single locus like 16S rDNA, the formation of heteroduplex and chimeric molecules is possible (Speksnijder, *et al.* 2001). These full-length fragments, comprised of sequences originating from multiple sources, introduce erroneous bands into subsequent analyses and alter the resulting profile. Another limitation of single locus based amplification is the possibility of multiple different alleles being amplified from a single organism due to various operon copy numbers (Crosby and Criddle 2003). 16S rDNA amplified from a pure culture of *Paenibacillus polymyxa* generated ten different bands when analysed by TGGE (Nubel, *et al.* 1996), highlighting the potential of single organisms to cause misinterpretations of data regarding the diversity of micro-organism populations. The ramifications of this particular potential for error is only of importance to molecular ecologists who wish to discover diversity and structure of bacterial populations and is of little concern for forensic investigators who wish only to generate a repeatable and distinctive profile of those populations.

The limitations outlined above are present in all band-based fingerprint technologies, but do not necessarily need to be overcome, as many band-based methods are capable of distinguishing soil samples despite any limitations they

may have. However, each technology has its own set of unique problems that limit its utility. DGGE is no exception. A significant drawback of the DGGE technology is the requirement to optimize the gradient and duration of the electrophoresis in order to maximize the separation of individual sequences. Although previous studies have been able to separate sequences differing by only one base pair (Nubel, *et al.* 1996; Kowalchuk, *et al.* 1997) others have found fragments from different species fail to resolve under the conditions used (Buchholz-Cleven, *et al.* 1997; Vallaeys, *et al.* 1997). As each soil sample is likely to contain a unique population of genotypes, optimisation for each profile may be required. A set of standard parameters capable of producing high degrees of separation for a majority of profiles may be established, however such generic conditions may result in some 16S rDNA fragments failing to be resolved leading to an underestimate of diversity and potentially incorrect assumptions of matching bands between profiles that may contain distinct sequences. As well as resolution difficulties with the gel itself, errors introduced during PCR present a special concern for DGGE analysis. Heteroduplex DNA can be formed during PCR, consisting of two strands, amplified from different templates that have annealed after denaturation. The melting temperature (and as a result the mobility through denaturant gradient gels) of heteroduplex DNA is lower than homoduplex molecules. Additional bands are formed as a result of heteroduplex molecules being analysed by DGGE, leading to an overestimate of microbial diversity and possibly artificial increases or decreases in the similarity of profiles. Although potentially problematic, Murray *et al*. (1996) concluded that the issue of heteroduplex DNA in DGGE analysis of microbial communities was not likely to be significant. Limitations will be associated with soil analysis regardless of the technology used to generate and analyse profiles. However, the solution is to choose the technology best suited for its intended application. It has been suggested previously that due to the increased sensitivity, ease of operation and more objective comparisons associated with the automation of its analysis, TRFLP may be better suited to highly diverse communities such as soil than DGGE or TGGE, which may be favoured for communities with low or moderate complexity such as marine or freshwater samples (Lukow, *et al.* 2000). This claim is supported by difficulties (including mostly smears and low resolution of bands)

encountered by other researchers when applying DGGE to bulk soil communities (Ovreas and Torsvik 1998; Alvey*, et al.* 2003).

4.2 Comparison of Microarray and Southern Membrane Sequence similarity Analysis of Arbitrarily Amplified DNA Profiles.

As it is possible that nylon-membrane spots may provide the basis of a simple initial version in the transfer of this hybridisation-based technology to forensic laboratories it is worth noting that there were differences observed between nylon membranes and glass arrays, mainly in the increased levels of non-specific, background binding to spots on the nylon membranes under the conditions used. However, overall, the RFU of soil profiles, hybridised using microarray technology, correlates well with the signal generated from Southern transfer (Figure 3.4, pg66) and direct spotting of DNA to nylon membrane (data not shown) after hybridisation with ^{35}S labelled profiles.

The ability of both Southern membrane and microarray technologies to analyse the level of sequence similarity between AAD profiles allows the option of either a cheap, low-tech method capable of a relatively low number of comparisons (Southern) or a more expensive, high-tech, high throughput technology (microarray). Although Southern analysis of AAD profiles is capable of distinguishing soils, it is a dead-end technology with very little potential for development. In contrast, microarray analysis is capable of providing more information from individual profiles and has the potential for a large number of technological developments, including the creation of standardised arrays, incorporation of random oligonucleotide arrays (ROA's) (Chpt 1.3.4), establishment of electronic databases and a high level of automation.

The choice of which technology to pursue (Southern membrane or Microarrays) should be based on a number of factors including the quality of data produced, the longevity of the technology, ease of operation and cost effectiveness. Southern membrane and microarray analysis are two very different approaches to answering the same question. It is my opinion that microarray technology should be pursued in preference to Southern analysis as the relatively high costs associated with microarrays will inevitably diminish as the technology matures and the higher quality and quantity of data, the ability to archive data and future technological developments will overshadow the financial costs.

4.3 Comparison of Length Polymorphism and Sequence similarity Analysis of Arbitrarily Amplified DNA Profiles.

4.3.1 Comparison of Similarity Index and Relative Fluorescence Unit Values

The similarity index of two profiles generated from the same soil sample using length polymorphism analysis of AAD was 0.778 while the SI of profiles generated from two distinct soils averaged 0.147. The value of "matching" samples was 5.3 times the value of "non-matching" samples when analysed using length polymorphisms of arbitrarily amplified DNA. Although quite high, this match:mismatch ratio is still less than that obtained by sequence hybridisation of the same arbitrarily amplified DNA, which yields RFU values of 0.697 for matching samples, 14.8 times the value of non-matching samples (0.047).

Interestingly, the order of similarity between pairs of soil samples was different for both techniques. Length polymorphism analysis determined soils A and B to be most similar (SI of 0.248), followed by soil B and C (SI of 0.136) with the most dissimilar soils being A and C (SI of 0.055). However, when sequence hybridisation was used to assess profile homology A and C were the most similar (RFU of 0.072), followed by A and B (RFU of 0.031) with the most dissimilar soils being B and C (RFU of 0.028). Admittedly, both A vs. B and B vs. C RFU values are very low with high standard deviations. This low level of relative fluorescence may actually be representative of the baseline level of homology, associated with no real sequence similarity. Regardless, soil A and C profiles do appear to share some small degree of sequence similarity above this baseline level despite sharing negligible fingerprint length homology suggesting that the two analytical methods of arbitrarily amplified DNA profiles i.e. that predominating fragment lengths in profile fingerprints and the total profile sequence, distinguish soils differently. Generally, the length polymorphism based SI values for differing soils are larger than corresponding hybridisation based RFU values, even as a percentage of same soil (positive control) values, suggesting that there may be some degree of falsely matching bands when length polymorphism analysis is used. The theoretical alternative (reduced hybridisation of matching sequences, leading to artificially low RFU values) is less likely as the

stringent hybridisation conditions required for such errors would result in the concurrent reduction in same soil (positive controls) RFU values, which was not observed.

4.3.2 Reproducibility Comparisons

Despite the better exclusion ability of the AADSS method, that is to say, its apparent freedom from false inclusions, the reproducibility of AADLP is better than the AADSS method presented in this study, with the standard deviation of SI values remaining relatively low. When analysing AAD profiles by sequence hybridisation, the reproducibility is not as high as length polymorphism analysis, with standard deviations 2-3 times length polymorphism levels (Table 4, pg68). The variability between standard deviation values is not reflective of reproducibility of the profiles but rather the differing methods of analysis. Analysing the ten most abundant bands for fingerprint analysis, means disregarding many of the other minor components in the profile. These minor sequences may indeed be more variable between profiles and since they are not ignored during sequence hybridisation analysis, they may lead to more variable levels of fluorescence after hybridisation. Also, the use of microarray technology with variable amounts of DNA printed onto the slides (Chpt 3.3.2) may contribute to the higher variation of RFU values compared to the SI values of the length polymorphism analysis.

Reproducibility of arbitrarily amplified DNA profiles when generated using replicate DNA extractions and/or different PCR batches was high for both length polymorphism and sequence similarity analyses, with no significant difference in SI or RFU values, respectively.

4.3.3 Significance of Discriminatory Power

Differences between matching and non-matching soil comparisons, for both SI and RFU values, were found to be significant. However, the probability of SI values from matching and non-matching soils being different by chance (determined by independent measures *t* tests) is far less likely than RFU values (determined by Mann-Whitney *U* tests) (Table 4, pg68). It is problematic to

compare these probability statistics as they are both extraordinarily low and were determined using different analytical methods, prohibiting any meaningful comparison. The reason a non-parametric analysis (Mann-Whitney *U* test) was performed on the microarray data was the lack of normally distributed RFU values and the high difference of variation between groups. Therefore, the probability statistics (Table 4, columns IV and VI, pg68) should not be used as a comparative statistic for the discriminatory power of length polymorphism and sequence similarity analysis. The high number of comparisons performed for both AADLP and AADSS are not economical for a real forensic situation but the high significances achieved (Table 4, pg68) indicate that much smaller numbers of comparisons will allow categorisation of soil samples to be made with economically attainable levels of confidence.

The high inclusion:exclusion value ratios for both methods of arbitrarily amplified DNA analysis allow decisions regarding the origin of soil samples to be made easily. The statistical analysis of the data resulting from arbitrarily amplified DNA profiles also shows the significant difference between soil profiles from the same sample and soil profiles from different samples for both total sequence similarity and length polymorphism analysis.

4.3.4 Fundamental Differences of Analysis

Analysing the similarity of profiles by sequence hybridisation may give a more accurate indication of community similarity as bands that coincidentally have the same length but different sequence will not falsely contribute to the determined similarity. Using sequence similarity may also reduce the error introduced by determining whether fragments separated by electrophoresis have "identical" lengths or not. Some degree of error is assumed when comparing profiles so that bands within a small range of sizes (that may result from electrophoresis variability) will be considered identical. This flexibility leaves open the possibility that two fragments, genuinely different in length by a small amount, will be classed as identical if they fall within the range specified by the analytical software. This type of error is absent when comparing profiles by sequence hybridisation. It is true that slightly different sequences may cross-hybridise, depending on the stringency of the hybridisation, resulting in an

increased fluorescence. However, such inexactness is not radically misrepresenting the similarity of the sequence, in contrast to length polymorphism errors.

The relative proportion of each sequence in the profile is also taken into account with sequence hybridisation, including poorly amplified sequences that would not be analysed using the length polymorphism method. There are, however, analytical methods that consider band intensity by weighting individual bands when analysing length polymorphisms, which could be employed to overcome any inaccuracies introduced by ignoring the relative abundance of each fragment length.

AADLP and AADSS are both very competent at discriminating soil samples and have unique properties, giving advantages to each under different circumstances. There is no reason why one should be favoured exclusively over the other for forensic investigation. It may be possible to utilise either, or in some cases combine both methods of analysis during the course of investigations, in concert with other traditional analyses.

In some instances a contributing factor in deciding which analysis to perform may be the infrastructure available to each laboratory. Genetic analysers like the one used for AADLP analysis in this study are more readily available to forensic laboratories than microarray scanning equipment. Logistical considerations like access to hardware and cost may favour the use of AADLP over AADSS, despite the better discrimination of the AADSS technique.

4.3.5 Other Sequence Hybridisation Based Methods that may be suited to Forensic Soil Evidence.

Other technologies of soil analysis exist that utilize DNA hybridisation. Community DNA hybridisation measures the extent to which two microbial communities are similar by measuring the relative amount of DNA cross-hybridisation between them. Community DNA hybridisation was first used to assess species compositions of natural bacterioplankton assemblages (Lee and Fuhrman 1990), but was later used for soil microbial community structure analysis (Ritz and Griffiths 1994; Griffiths*, et al.* 1996). Although this method was capable of measuring the similarity of four soil types, with percent similarity

ranging from 25 to 74% (Griffiths, *et al.* 1996), several factors make this technology inappropriate for forensic investigation. Community DNA hybridisation will not be suitable for forensic soil comparison due to the requirements for large quantities of community DNA of high purity, for targets of hybridisation and probe labelling, as well as very long hybridisation times to allow adequate hybridisation of probe to matrix bound template. Another similar technique is the reassociation of DNA for the determination of genetic diversity of soil communities (Torsvik, Goksoyr, *et al.* 1990; Torsvik, Salte, *et al.* 1990; Torsvik, *et al.* 1994). This procedure measures the rate of change in absorbance associated with DNA reassociation, giving an indication of diversity, as diverse populations will take longer to anneal than simple ones. This procedure could be modified to incorporate equal amounts of two samples for comparison and determine the relative reassociation rate compared to the pure samples individually. Although much quicker and easier to perform than community DNA hybridisation, this technique also requires large amounts of pure DNA making it unsuitable for forensic investigations. The requirements of large amounts of pure DNA and long hybridisation times is due to the enormous complexity of soil community DNA and the difficulties of hybridisation arising from this complexity. Some researchers have tried to overcome this by reducing the complexity of the DNA being hybridized. The relative amount of specific genes (Guo, *et al.* 1997), diversity of 16S-rRNA genes by hybridisation with defined oligonucleotide probes (Liesack and Stackebrandt 1992) and even microarray analysis of 16S rDNA and rRNA sequences (Small, *et al.* 2001; Koizumi, *et al.* 2002; Valinsky, *et al.* 2002) or variants of specific functional genes (Wu, *et al.* 2001; Bodrossy, *et al.* 2003; Taroncher-Oldenburg, *et al.* 2003; Stralis-Pavese, *et al.* 2004) have been investigated as potential tools for soil community analysis. By focusing on specific genes, hybridisation problems associated with the complexity of the entire microbial community are circumvented. Evading complexity related problems has also been achieved with the hybridisation method presented in this thesis. Entire soil metagenomes have been reduced to a number of sequences manageable during hybridisation. The only difference between the study in this thesis and existing studies is that the mode of simplification has been relatively arbitrary in this thesis, rather than focusing on particular genes as has been performed in other studies. Another potential

advantage of the amplification method outlined in this study is the attachment of primers flanking amplified sequences which will facilitate the alignment of sequences during hybridisation, allowing quicker hybridisation of homologous amplified sequences. A distinct difference between this study and other studies is the hybridisation of many different loci as opposed to variants of a single gene or locus. This distinction means the stringency and optimisation of the current method will be far simpler, as trying to find conditions where highly similar but distinct 16S rDNA will not cross-hybridise may not be possible. The limitations of distinguishing highly similar sequences by microarray analysis are outlined in the introduction of this thesis (Chpt 1.3). Since the amplification of random loci greatly reduces the probability of highly conserved sequences being amplified, the level of loosely similar DNA sequences cross-hybridising should be negligible or at least, much reduced.

Hybridisation based methodologies generate more accurate results due to absences of purely coincidental matching of sequences as with length polymorphism analysis.

4.4 Comparison of TRFLP and Sequence Hybridisation Analysis of Soil Samples Taken at Different Times from the Same Location.

A practical consideration for any DNA analysis of soil communities is the ability to match samples taken from an area at different times. The temporal stability of biological community profiles from three soils was determined for TRFLP (Chpt 3.1.3) and AADSS (Chpt 3.3.5) profiles. Both profiling techniques showed changing DNA profiles between samples taken at different times from the same location. However, differences in the effects of time, relative to distance (2-5km), are apparent for the two profiling strategies. TRFLP profiles changed to the extent that similarity index (SI) values of within-site samples taken six months apart, for all three soils, were lower than SI values between sites. In two of these soils the SI values were significantly lower than SI values of between-site samples, collected on the same day (P= 0.003 and 5.03×10^{-8} for Soils B and C respectively, independent measures t test), suggesting that despite being separated by distance (2-5km), soils may share seasonally dependant bacterial population subsets, increasing the similarity of TRFLP profiles from these three sites at a given time. It appears that TRFLP profiles reflect more intense bacterial

population changes over six-months within one site than between sites 2-5km apart at one time.

Although sizeable reductions in RFU values were observed between profiles separated for a period of one year, AADSS profiles maintained a level of sequence similarity above the level of soils separated by distance.

Although the time sensitivity of the TRFLP method is interesting, the focus of this research project has always been to develop a viable test for soil evidence. Thus, the temporal stability and greater discrimination of the AADSS technology will probably be better suited for practical applications.

5 CONCLUSIONS

This study has examined the potential of certain DNA based analytical techniques for the examination of soil evidence for forensic investigations. Although a simple TRFLP analysis (one labelled primer, one restriction enzyme) of the microbial content of soils was capable of generating significant differences between mean similarity indices of replicate and different soil comparisons, the low relative difference in similarity indices between these types of comparisons makes this simple TRFLP analysis of doubtful use in a forensic context. The level of change in TRFLP profiles over time *in situ* would be problematic when applying this version of TRFLP analysis to real cases.

The arbitrary amplification of DNA outlined in this thesis allowed two types of analyses to be performed. AADLP (length polymorphism) analysis of arbitrarily amplified DNA profiles was capable of discriminating replicate soil profiles from profiles generated from different soils at a level sufficient for evidential purposes. This study has for the first time assessed the reproducibility and discriminatory power of AADLP from soil microbial communities and found them to be useful for forensic purposes. This study has, also for the first time, hybridised arbitrarily or randomly amplified DNA profiles to assess the sequence similarity of the profiles generated (the AADSS method). It was found that cross-hybridisation, using either membrane bound profiles with radioisotope labelled DNA or microarray technology, was capable of distinguishing replicate profiles of the same soil from profiles of different soils.

It appears that both of the arbitrary amplification methods were superior to the TRFLP analysis presented in this thesis for distinguishing soils for the purposes of forensic investigations. It is less obvious whether sequence similarity or length polymorphism analysis is superior for the analysis of arbitrarily amplified DNA profiles so either of these analytical methods can be used individually or together.

A number of variables were investigated, principally but not exclusively using sequence similarity (AADSS) analysis, to determine how these variables would influence the arbitrarily amplified DNA profiles. It was found that time *in situ*, distance between sample sites and time under various storage conditions all

caused changes in the sequence similarity of arbitrarily amplified DNA profiles. This knowledge will be valuable when applying AADSS to cases, as the conditions allowing successful "matching" of soils will be established.

Importantly, the molecular mechanism of the arbitrary amplification has been investigated allowing insight into the way AAD profiles are generated and potential ways to control the process in order to generate different profiles suited for various analyses, including rapid simplification of the metagenome for banding analysis and broad amplification for metagenome sequence similarity comparisons. Potential improvements and developments including the use of multiple primers for more extensive profiles and standard microarrays for sequence similarity analyses may further enhance the utility of arbitrarily amplified DNA profiles.

The work reported in this thesis has shown that DNA technologies are capable of contributing to forensic investigations of soil evidence. The methods presented in this thesis have demonstrated some principles that allow comparisons of soil evidence, which could be further refined into a generally applicable tool for routine forensic use. Overall, a variety of DNA technologies are capable of contributing to forensic investigations of soil evidence, of which AADSS performed best here.

5.1 Future Directions for Biological Profiling of Soil Evidence

5.1.1 Improving the Profiling Techniques

A large trial with many soils, including physically and chemically similar soils, is the next appropriate step. The degree to which physically and chemically similar soils can be distinguished by the microbial community profiles will determine the ultimate value of these methods of soil analysis. The degree of independence of DNA profiling and physicochemical profiling methods will determine the combinational value of these methods of soil examination.

The profiling of soil microbial communities by arbitrary amplification of DNA with multiple primers to generate several profiles of the same soil should also be performed to investigate any potential increase in the information and

discrimination achieved by the technique when multiple, distinct profiles are generated from the same sample.

A long term developmental pathway for DNA profiling technologies may include the routine sequencing of amplified DNA. Although laborious to begin with, amplified DNA could be cloned, the predominating sequences ascertained and databased to provide highly detailed profiles of biological communities. These databased sequences can then be easily used in subsequent *in silico* analyses. This highly desirable technology will depend on a reduction in price and labour associated with sequencing high numbers of PCR products. This may be achieved in part with automation, however the cost of such a venture is likely to be prohibitive for this particular end, at this time.

5.1.2 Practical Issues that may Arise in Forensic Casework Need Addressing

A practical issue of forensic investigation of soil evidence that was not directly addressed in this thesis was the ability to extract DNA and generate adequate profiles from minute samples of soil. The methods in this study extracted DNA from 500mg of soil, taken from a larger sample collected from sites (Chpt 2.4.1.4). The amount of DNA used in the generation of each DNA profile was low (5ng) compared to the amount of DNA recovered from the 500mg samples (averaging ~10µg). This implies that much smaller samples of soil can be used to recover sufficient DNA to generate a profile, possibly as low as 1mg. However, the critical issue will be the representativeness of such a small soil sample, of the original site. This issue is not a drawback of the DNA technologies used but a reality of sampling biases and will likely be of major concern in any analytical technique dealing with such minute samples. Regardless, this issue should be investigated in order to provide some guidelines regarding the reliability and accuracy of analyses between trace soil samples and the parent soil.

Another practical issue that must be addressed is the ability of these profiling techniques to cope with mixtures of soils. It is likely that some cases will arise where a mixture of soils or sediments will be recovered and need to be compared to an individual site. The ability of any soil analytical technique to determine the provenance of such samples is highly desirable. It may be expected that length polymorphism analysis will fail to provide sufficient evidence for

matching a mixture of soils to either of the original sites as a mixture of band lengths will cause profiles to differ from both. However, sequence similarity analysis of arbitrarily amplified DNA profiles may not fail to the extent of the length polymorphism analysis, as sequence similarity between sample profiles may exist above the baseline level of distinct soils indicating some degree of shared sequence.

Once the above issues regarding mixtures, trace samples and the additional value of DNA profiling techniques have been addressed, mock cases may be performed with soil recovered from clothing, footwear, the hair, skin or fingernails of a person before being analysed and compared to the suspected crime scene, nearby soils and reference soils. Many blind trials, presented as simulations of real casework will be required to establish the validity of the technique before beginning in real cases.

The ability of DNA techniques to match or distinguish soil samples based on the microbial content that resides within is undoubted and further demonstrated in this thesis. The application of such technologies to real forensic investigations may be valuable.

REFERENCES

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 25: 3389-3402.

Alvey, S., C.-H. Yang, A. Buerkert and D. E. Crowley. (2003). Cereal/legume rotation effects on rhizoshpere bacterial community structure in west african soils. *Biology and Fertility of Soils*. 37 (2): 73-82.

Bavykin, S. G., J. P. Akowski, V. M. Zakhariev, V. E. Barsky, P. A. N. and A. D. Mirzabekov. (2001). Portable system for microbial sample preparation and oligonucleotide microarray analysis. *Applied and Environmental Microbiology*. 67: 922-928.

Berthelet, M., L. G. Whyte and C. W. Greer. (1996). Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiology Letters*. 138 (1): 17-22.

Bodrossy, L., N. Stralis-Pavese, J. C. Murrell, S. Radajewski, A. Weilharter and A. Sessitsch. (2003). Development and validation of a diagnostic microbial microarray for methanotrophs. *Environmental Microbiology*. 5 (7): 566-582.

Bourke, M. T., C. A. Scherczinger, C. Ladd and H. C. Lee. (1999). NaOH treatment to neutralize inhibitors of Taq polymerase. *Journal of Forensic Sciences*. 44 (5): 1046-1050.

Bruns, M. A. and K. M. Scow. (2000) DNA Fingerprinting as a Means to Identify Sources of Soil-Derived Dust: Problems and Potential. In: *Integrated Assesment of Ecosystem Health* (eds Scow, K., G. E. Fogg, D. E. Hinton and M. L. Johnson) Lewis Publishers, Inc.,

Buchholz-Cleven, B. E. E., B. Rattunde and K. L. Straub. (1997). Screening for genetic diversity of isolates of anaerobic Fe(II)-oxidizing bacteria using DGGE and whole-cell hybridization. *Systematic and Applied Microbiology*. 20: 301-309.

Budavari, S. (1989). The Merck Index. *11th Edition*. Merck & Co., Inc., Rahway,N.J., USA.

Burgoyne, L. A. (1996). Solid medium and method for DNA storage. U.S. Patent 5,496,562.

Calderon, F. J., L. E. Jackson, K. M. Scow and D. E. Rolston. (2000). Microbial responses to simulated tillage in cultivated and uncultivated soils. *Soil Biology & Biochemistry*. 32 (11-12): 1547-1559.

Calderon, F. J., L. E. Jackson, K. M. Scow and D. E. Rolston. (2001). Short-term dynamics of nitrogen, microbial activity, and phospholipid fatty acids after tillage. *Soil Science Society of America Journal*. 65 (1): 118-126.

Casamayor, E. O., R. Massana, S. Benlloch, L. Ovreas, B. Diez, V. J. Goddard, J. M. Gasol, I. Joint, F. Rodriguez-Valera and C. Pedros-Alio. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environmental Microbiology June*. 4 (6): 338-348.

Cherif, A., L. Brusetti, S. Borin, A. Rizzi, A. Boudabous, H. Khyami-Horani and D. Daffonchio. (2003). Genetic relationship in the 'Bacillus cereus group' by rep-PCR fingerprinting and sequencing of a Bacillus anthracis-specific rep-PCR fragment. *Journal of Applied Microbiology*. 94 (6): 1108-1119.

Clement, B. G., L. E. Kehl, K. L. DeBord and C. L. Kitts. (1998). Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *Journal of Microbiological Methods*. 31: 135-142.

Cole, J. R., B. Chai, T. L. Marsh, R. J. Farris, Q. Wang, S. A. Kulam, S. Chandra, D. M. McGarrell, T. M. Schmidt, G. M. Garrity and J. M. Tiedje. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research January*. 31 (1): 442-443.

Costouros, N. G. and S. K. Libutti. (2002). Microarray technology and gene expression analysis for the study of angiogenesis. *Expert Opinion on Biological Therapy*. 2 (5): 545-556.

Croft, D. J. and K. Pye. (2003). The potential use of continuous-flow isotope-ratio mass spectrometry as a tool in forensic soil analysis: a preliminary report. *Rapid Communications in Mass Spectrometry*. 17: 2581-2584.

Cronin, M. T., F. Frueh, M. Pho, D. Dutta and T. M. Brennan. (1999). Applying rapid DNA microarray optimization capability to SNP screening and genotyping. *American Journal of Human Genetics*. 65 (4): A224-A224.

Crosby, L. D. and C. S. Criddle. (2003). Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques*. 34 (4): 790-+.

Crow, M. K. and J. Wohlgemuth. (2003). Microarray analysis of gene expression in lupus. *Arthritis Research & Therapy*. 5 (6): 279-287.

Daran-Lapujade, P., J. M. Daran, P. Kotter, T. Petit, M. D. W. Piper and J. T. Pronk. (2003). Comparative genotyping of the Saccharomyces cerevisiae laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays. *Fems Yeast Research*. 4 (3): 259-269.

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 (12): 4950-4957.

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 (1): 190-197.

Egert, M. and M. W. Friedrich. (2003). Formation of Pseudo-Terminal Restriction Fragments, a PCR-Related Bias Affecting Terminal Restriction Fragment Length Polymorphism Analysis of Microbial Community Structure. *Applied and Environmental Microbiology*. 69 (5): 2555-2562.

Faegri, A., V. Torsvik and J. Goksoyr. (1977). Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated centrifugation technique. *Soil Biology & Biochemistry*. 9: 105-112.

Ferris, M. J., G. Muyzer and D. M. Ward. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot springmicrobial mat community. *Applied and Environmental Microbiology*. 62: 340-346.

Ferris, M. J., S. C. Nold, N. P. Revsbech and D. M. Ward. (1997). Population structure and physiological changes within a hot spring microbial may community following disturbance. *Applied and Environmental Microbiology*. 63: 1367-1374.

Ferris, M. J. and D. M. Ward. (1997). Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*. 63: 1375-1381.

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 (3): 335-346.

Franklin, R. B., D. R. Taylor and A. L. Mills. (1999). Characterization of microbial communities using randomly amplified polymorphic DNA (RAPD). *Journal of Microbiological Methods*. 35 (3): 225-235.

Garland, J. L. and M. S. Mills. (1994) A community-level physiological approach for studying microbial communities. In: *Beyond the Biomass* (eds Ritz, K., J. Dighton and K. E. Giller) John Wiley & Sons, Chichester.

Gelsomino, A., W. A. C. Keijzer, G. Cacco and E. J. D. van. (1999). Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *Journal of Microbiological Methods*. 38 (1-2): 1-15.

Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany and F. Barany. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *Journal of Molecular Biology*. 292 (2): 251-262.

Girvan, M. S., J. Bullimore, J. N. Pretty, A. M. Osborn and A. S. Ball. (2003). Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Applied and Environmental Microbiology*. 69 (3): 1800-1809.

Green, C. T. and K. M. Scow. (2000). Analysis of phospholipid fatty acids (PLFA) to characterize microbial communities in aquifers. *Hydrogeology Journal*. 8 (1): 126-141.

Griffiths, B. S., K. Ritz and L. A. Glover. (1996). Broad-scale approaches to the determination of soil microbial community structure: Application of the community DNA hybridization technique. *Microbial Ecology*. 31 (3): 269-280.

Guo, C., W. Sun, J. B. Harsh and A. Ogram. (1997). Hybridization analysis of microbial DNA from fuel oil-contaminated and noncontaminated soil. *Microbial Ecology*. 34 (3): 178-187.

Guschin, D. Y., B. K. Mobarry, D. Proudnikov, D. A. Stahl, B. E. Rittmann and A. D. Mirzabekov. (1997). Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Applied and Environmental Microbiology*. 63 (6): 2397-2402.

Haack, S. K., H. Garchow, M. J. Klug and L. J. Forney. (1995). Analysis of factors affecting the accuracy, reproducibility, and interpretation of fatty acid methyl ester profiles of model bacterial communities. *Applied and Environmental Microbiology*. 61: 1458-1468.

Hacia, J. G., J. B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R. A. Mayer, B. Sun, L. Hsie, C. M. Robbins, L. C. Brody, D. Wang, E. S. Lander, R. Lipshutz, S. P. A. Fodor and F. S. Collins. (1999). Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nature Genetics*. 22 (2): 164-167.

Harry, M., B. Gambier, Y. Bourezgui and S. E. Garnier. (1999). Evaluation of purification procedures for DNA extracted from organic rich samples: Interference with humic substances. *Analusis*. 27 (5): 439-442.

Harry, M., N. Jusseaume, B. Gambier and E. Garnier-Sillam. (2001). Use of RAPD markers for the study of microbial community similarity from termite mounds and tropical soils. *Soil Biology & Biochemistry*. 33 (4-5): 417-427.

Heinz, R. A. and H. W. Platt. (2000). Improved DNA extraction method for Verticillium detection and quantification in large-scale studies using PCR-based techniques. *Canadian Journal of Plant Pathology*. 22 (2): 117-121.

Hery, M., S. Nazaret, T. Jaffre, P. Normand and E. Navarro. (2003). Adaptation to nickel spiking of bacterial communities in neocaledonian soils. *Environmental Microbiology January*. 5 (1): 3-12.

Hiraoka, Y. (1994). A possible approach to soil discrimination using X-ray fluorescence analysis. *Journal of Forensic Sciences*. 39 (6): 1381-1392.

Hopkins, D. W., S. J. MacNaughton and A. G. O'Donnell. (1993). A dispersion and differential centrifugation technique for representatively sampling micro-organisms from soil. *Soil Biology & Biochemistry*. 23: 217-225.

Horswell, J., S. J. Cordiner, E. W. Maas, T. M. Martin, B. W. Sutherland, T. W. Speir, B. Nogales and A. M. Osborn. (2002). Forensic comparison of soils by bacterial community DNA profiling. *Journal of Forensic Sciences*. 47 (2): 350-353.

Jacobsen, C. S. (1995). Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. *Applied and Environmental Microbiology*. 61 (9): 3347-3352.

Jacobsen, C. S. and O. F. Rasmussen. (1992). Development and Application of a New Method to Extract Bacterial-DNA from Soil Based on Separation of Bacteria from Soil with Cation-Exchange Resin. *Applied and Environmental Microbiology*. 58 (8): 2458-2462.

Ji, M. J., P. Hou, S. Li, N. Y. He and Z. H. Lu. (2004). Microarray-based method for genotyping of functional single nucleotide polymorphisms using dual-color fluorescence hybridization. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*. 548 (1-2): 97-105.

Josse, J., A. D. Kaiser and A. Kornberg. (1961). Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbour base sequences in deoxyribonucleic acid. *Journal of Biological Chemistry*. 236: 864-875.

Kang, S. H. and A. L. Mills. (2004). Soil bacterial community structure changes following disturbance of the overlying plant community. *Soil Science*. 169 (1): 55-65.

Kingsley, M. T., T. M. Straub, D. R. Call, D. S. Daly, S. C. Wunschel and D. P. Chandler. (2002). Fingerprinting closely related *Xanthomonas* pathovars with random nonamer oligonucleotide microarrays. *Applied and Environmental Microbiology*. 68 (12): 6361-6370.

Kladde, M. P., Y. Kohwi, T. Kohwishigematsu and J. Gorski. (1994). The Non-B-DNA Structure of D(Ca/Tg)N Differs from That of Z-DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (5): 1898-1902.

Koizumi, S. and H. Yamada. (2003). DNA microarray analysis of altered gene expression in cadmium-exposed human cells. *Journal of Occupational Health*. 45 (6): 331-334.

Koizumi, Y., J. J. Kelly, T. Nakagawa, H. Urakawa, S. El-Fantroussi, S. Al-Muzaini, M. Fukui, Y. Urushigawa and D. A. Stahl. (2002). Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology. *Applied and Environmental Microbiology*. 68 (7): 3215-3225.

Kowalchuk, G. A., J. R. Stephen, W. de Boer, J. I. Prosser, T. M. Embley and J. W. Woldendorp. (1997). Analysis of ammonia-oxidising bacteria of the ß-subdivision of the class *Proteobacteria* in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Applied and Environmental Microbiology*. 63: 1489-1497.

Kozdroj, J. and J. D. van Elsas. (2000). Application of polymerase chain reaction-denaturing gradient gel electrophoresis for comparison of direct and indirect extraction methods of

soil DNA used for microbial community fingerprinting. *Biology and Fertility of Soils*. 31 (5): 372-378.

Kuske, C. R., K. L. Banton, D. L. Adorada, P. C. Stark, K. K. Hill and P. J. Jackson. (1998). Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Applied and Environmental Microbiology*. 64 (7): 2463-2472.

LaMontagne, M. G., F. C. Michel, P. A. Holden and C. A. Reddy. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *Journal of Microbiological Methods*. 49 (3): 255-264.

Lee, S. and J. A. Fuhrman. (1990). DNA hybridization to compare species compositions of natural bacterioplankton assemblages. *Applied and Environmental Microbiology*. 56: 739-746.

Liesack, W. and E. Stackebrandt. (1992). Occurrence of Novel Groups of the Domain Bacteria as Revealed by Analysis of Genetic Material Isolated from an Australian Terrestrial Environment. *Journal of Bacteriology*. 174 (15): 5072-5078.

Liesack, W., H. Weyland and E. Stackebrandt. (1991). Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microbial Ecology*. 21: 191-198.

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 (11): 4516-4522.

Loy, A., A. Lahner, N. Lee, J. Adamczyk, H. Meier, J. Ernst and K. H. Schleifer. (2002). Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Applied and Environmental Microbiology*. 68: 5064-5081.

Lueders, T. and M. W. Friedrich. (2003). Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Applied and Environmental Microbiology*. 69 (1): 320-326.

Lukow, T., P. F. Dunfield and W. Liesack. (2000). Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. *FEMS Microbiology Ecology*. 32 (3): 241-247.

Lupwayi, N. Z., W. Z. Rice and G. W. Clayton. (1998). Soil microbial diversity and community structure under wheat as influenced by tillage and crop rotation. *Soil Biology & Biochemistry*. 30 (13): 1733-1741.

Martin-Laurent, F., L. Philippot, S. Hallet, R. Chaussod, J. C. Germon, G. Soulas and G. Catroux. (2001). DNA extraction from soils: Old bias for new microbial diversity analysis methods. *Applied and Environmental Microbiology*. 67 (5): 2354-2359.

Mathieu-Daude, F., J. Welsh, T. Vogt and M. McClelland. (1996). DNA rehybridization during PCR: The 'C(O)t effect' and its consequences. *Nucleic Acids Research*. 24 (11): 2080-2086.

McCaig, A. E., L. A. Glover and J. I. Prosser. (2001). Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns. *Applied and Environmental Microbiology*. 67 (10): 4554-4559.

McManus, P. S. and A. L. Jones. (1995). Genetic fingerprinting of Erwinia amylovora strains isolated from tree-fruit crops and Rubus spp. *Phytopathology*. 85 (12): 1547-1553.

Messing, J., R. Crea and P. H. Seeburg. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research*. 9 (2): 309-321.

Miller, D. N., J. E. Bryant, E. L. Madsen and W. C. Ghiorse. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology*. 65 (11): 4715-4724.

More, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse and E. L. Madsen. (1994). Quantitative Cell-Lysis of Indigenous Microorganisms and Rapid Extraction of Microbial DNA from Sediment. *Applied and Environmental Microbiology*. 60 (5): 1572-1580.

Moyer, C. L., F. C. Dobbs and D. M. Karl. (1994). Estimation of Diversity and Community Structure through Restriction-Fragment-Length-Polymorphism Distribution Analysis of Bacterial 16s Ribosomal-RNA Genes from a Microbial Mat at an Active, Hydrothermal Vent System, Loihi Seamount, Hawaii. *Applied and Environmental Microbiology*. 60 (3): 871-879.

Mummey, D. L. and D. A. Stahl. (2003). Spatial and temporal variability of bacterial 16S rDNA-based T-RFLP patterns derived from soil of two Wyoming grassland ecosystems. *FEMS Microbiology Ecology*. 46: 113-120.

Murray, A. E., J. T. Hollibaugh and C. Orrego. (1996). Phylogenetic compositions of bacterioplankton from two California esturies compared by denaturing gradient gel electrophoresis of 16S rDNA frogments. *Applied and Environmental Microbiology*. 62: 2676-2680.

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 Ribosomal-Rna. *Applied and Environmental Microbiology*. 59 (3): 695-700.

Muyzer, G., A. Teske, C. O. Wirsen and J. H. W. (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch. Microbiol.* 164: 165-171.

Nath, K., J. W. Sarosy, J. Hahn and C. J. Di Como. (2000). Effects of ethidium bromide and SYBR Green I on different polymerase chain reaction systems. *Journal of Biochemical and Biophysical Methods*. 42 (1): 15-29.

Nubel, U., B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R. I. Amann, W. Ludwig and H. Backhaus. (1996). Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *Journal of Bacteriology*. 178: 5636-5643.

Nute, H. D. (1975). An improved density gradient system for forensic science soil studies. *Journal of Forensic Sciences*. 20 (4): 668-73.

Ogram, A., G. S. Sayler and T. Barkay. (1987). The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*. 7 (1): 57-66.

Osborn, A. M., E. R. B. Moore and K. N. Timmis. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environmental Microbiology*. 2 (1): 39-50.

Ovreas, L. and V. Torsvik. (1998). Microbial diversity and community structure in two different agricultural soil communities. *Microbial Ecology*. 36 (3): 303-315.

Packer, M. J., M. P. Dauncey and C. A. Hunter. (2000). Sequence-dependent DNA structure: Tetranucleotide conformational maps. *Journal of Molecular Biology*. 295 (1): 85-103.

Parham, J. A., S. P. Deng, H. N. Da, H. Y. Sun and W. R. Raun. (2003). Long-term cattle manure application in soil. II. Effect on soil microbial populations and community structure. *Biology and Fertility of Soils*. 38 (4): 209-215.

Peixoto, R. S., C. C. H. L. da, N. G. Rumjanek, A. Macrae and A. S. Rosado. (2002). Use of *rpoB* and 16S rRNA genes to analyse bacterial diversity of a tropical soil using PCR and DGGE. *Letters in Applied Microbiology*. 35 (4): 316-320.

Pesaro, M., G. Nicollier, J. Zeyer and F. Widmer. (2004). Impact of soil drying-rewetting stress on microbial communities and activities and on degradation of two crop protection products. *Applied and Environmental Microbiology*. 70 (5): 2577-2587.

Pesaro, M. and F. Widmer. (2002). Identification of novel Crenarchaeota and Euryarchaeota clusters associated with different depth layers of a forest soil. *FEMS Microbiology Ecology*. 42 (1): 89-98.

Petraco, N. (1994a). Microscopic Examination of Mineral Grains in Forensic Soil Analysis .1. *American Laboratory*. 26 (6): 35-37.

Petraco, N. (1994b). Microscopic Examination of Mineral Grains in Forensic Soil Analysis .2. *American Laboratory*. 26 (14): 33-35.

Porteous, L. A. and J. L. Armstrong. (1993). A simple mini-method to extract DNA directly from soil for use with polymerase chain reaction amplification. *Current Microbiology*. 27: 115-118.

Porteous, L. A., R. J. Seidler and L. S. Watrud. (1997). An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. *Molecular Ecology*. 6 (8): 787-791.

Ranjard, L., F. Poly, J. C. Lata, C. Mougel, J. Thioulouse and S. Nazaret. (2001). Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. *Applied and Environmental Microbiology*. 67 (10): 4479-4487.

Rhee, S.-K., X. Liu, L. Wu, S. C. Chong, X. Wan and J. Zhou. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Applied and Environmental Microbiology*. 70 (7): 4303-4317.

Rickman, D. S., C. J. Herbert and L. P. Aggerbeck. (2003). Optimizing spotting solutions for increased reproducibility of cDNA microarrays. *Nucleic Acids Research*. 31 (18): e109.

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 (4): 1668-1675.

Ritz, K. and B. S. Griffiths. (1994). Potential Application of a Community Hybridization Technique for Assessing Changes in the Population-Structure of Soil Microbial Communities. *Soil Biology & Biochemistry*. 26 (8): 963-971.

Rogers, C. D. G. (2002) Principles of processing and Analysis of Clinically important genomes from nucleic acid protection media. *Doctor of Philosophy* School of Biological Sciences, Flinders University of South Australia.

Schoolnik, G. K. (2002). Functional and comparative genomics of pathogenic bacteria. *Current Opinion in Microbiology*. 5 (1): 20-26.

Shi, M. M. (2001). Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clinical Chemistry*. 47 (2): 164-172.

Sigler, W. V. and J. Zeyer. (2002). Microbial diversity and activity along the forefields of two receding glaciers. *Microbial Ecology*. 43 (4): 397-407.

Small, J., D. R. Call, F. J. Brockman, T. M. Straub and D. P. Chandler. (2001). Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Applied and Environmental Microbiology*. 67 (10): 4708-4716.

Smalla, K., N. Cresswell, L. C. Mendoncahagler, A. Wolters and J. D. Vanelsas. (1993). Rapid DNA Extraction Protocol from Soil for Polymerase Chain Reaction-Mediated Amplification. *Journal of Applied Bacteriology*. 74 (1): 78-85.

Smalla, K., G. Wieland, A. Buchner, A. Zock, J. Parzy, S. Kaiser, N. Roskot, H. Heuer and G. Berg. (2001). Bulk and rhizoshpere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Applied and Environmental Microbiology*. 67 (10): 4742-4751.

Snelling, A. M., P. GernerSmidt, P. M. Hawkey, J. Heritage, P. Parnell, C. Porter, A. R. Bodenham and T. Inglis. (1996). Validation of use of whole-cell repetitive extragenic palindromic sequence-based PCR (REP-PCR) for typing strains belonging to the Acinetobacter calcoaceticus Acinetobacter baumannii complex and application of the method to the investigation of a hospital outbreak. *Journal of Clinical Microbiology*. 34 (5): 1193-1202.

Song, X. H., P. K. Hopke, M. A. Bruns, K. Graham and K. Scow. (1999). Pattern recognition of soil samples based on the microbial fatty acid contents. *Environmental Science & Technology*. 33 (20): 3524-3530.

Speksnijder, A., G. A. Kowalchuk, S. De Jong, E. Kline, J. R. Stephen and H. J. Laanbroek. (2001). Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Applied and Environmental Microbiology*. 67 (1): 469-472.

Steffan, R. J., J. Goksoyr, A. K. Bej and R. M. Atlas. (1988). Recovery of DNA from Soils and Sediments. *Applied and Environmental Microbiology*. 54 (12): 2908-2915.

Stralis-Pavese, N., A. Sessitsch, A. Weilharter, T. Reichenauer, J. Reisling, J. Csontos, J. C. Murrell and L. Bodrossy. (2004). Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environmental Microbiology*. 6 (4): 347-363.

Straub, T. M., I. L. Pepper, M. Abbaszadegan and C. P. Gerba. (1994). A Method to Detect Enteroviruses in Sewage Sludge-Amended Soil Using the Pcr. *Applied and Environmental Microbiology*. 60 (3): 1014-1017.

Sudarshana, P., J. R. Hanson and K. M. Scow. (2000) Application of a Random Amplified Polymorphic DNA (RAPD) Method for Characterization of Microbial Communities in Agricultural Soils. In: *Integrated Assessment of Ecosystem Health* (eds Scow, K., G. E. Fogg, D. E. Hinton and M. L. Johnson) Lewis Publishers, Inc.,

Sugita, R. and Y. Marumo. (1996). Validity of color examination for forensic soil identification. *Forensic Science International*. 83 (3): 201-210.

Swartz, M. N., T. A. Trautner and A. Kornberg. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbour base sequences in deoxyribonucleic acids. *Journal of Biological Chemistry*. 237: 1961-1967.

Taroncher-Oldenburg, G., E. M. Griner, C. A. Francis and B. B. Ward. (2003). Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Applied and Environmental Microbiology*. 69 (2): 1159-1171.

Tebbe, C. C. and W. Vahjen. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and Environmental Microbiology*. 59 (8): 2657-2665.

Tiquia, S. M., J. Lloyd, D. A. Herms, H. A. J. Hoitink and F. C. J. Michael. (2002). Effects of mulching and fertilization on soil nutrients, microbial activity and rhizoshpere bacterial community structure determined by analysis of TRFLPs of PCR-amplified 16S rRNA. *Applied Soil Ecology*. 21: 31-48.

Torsvik, V. (1980). Isolation of Bacterial DNA From Soil. *Soil Biology & Biochemistry*. 12: 15-21.

Torsvik, V., J. Goksoyr and F. L. Daae. (1990). High Diversity in DNA of Soil Bacteria. *Applied and Environmental Microbiology*. 56 (3): 782-787.

Torsvik, V., J. Goksoyr, F. L. Daae, R. Sorheim, J. Michalsen and K. Salte. (1994) Use of DNA analysis to determine the diversity of microbial communities. In: *Beyond the Biomass* (eds Ritz, K. D., J. and Giller, KE.) Wiley-Sayce, British Society of Soil Science (BSSS).

Torsvik, V., K. Salte, R. Sorheim and J. Goksoyr. (1990). Comparison of Phenotypic Diversity and DNA Heterogeneity in a Population of Soil Bacteria. *Applied and Environmental Microbiology*. 56 (3): 776-781.

Trevors, J. T. (1992). DNA extraction from soil. *microb releases*. 1: 3-9.

Tsai, Y. L. and B. H. Olson. (1991). Rapid Method for Direct Extraction of DNA from Soil and Sediments. *Applied and Environmental Microbiology*. 57 (4): 1070-1074.

Tsai, Y. L. and B. H. Olson. (1992a). Detection of Low Numbers of Bacterial-Cells in Soils and Sediments by Polymerase Chain-Reaction. *Applied and Environmental Microbiology*. 58 (2): 754-757.

Tsai, Y. L. and B. H. Olson. (1992b). Rapid Method for Separation of Bacterial-DNA from Humic Substances in Sediments for Polymerase Chain-Reaction. *Applied and Environmental Microbiology*. 58 (7): 2292-2295.

Tyrrell, G. J., R. N. Bethune, B. Willey and D. E. Low. (1997). Species identification of enterococci via intergenic ribosomal PCR. *Journal of Clinical Microbiology*. 35 (5): 1054-1060.

Urakawa, H., P. A. Noble, S. El Fantroussi, J. J. Kelly and D. A. Stahl. (2002). Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Applied and Environmental Microbiology*. 68: 235-244.

Vahjen, W. and C. C. Tebbe. (1994). Enhanced detection of genetically engineered Corynebacterium glutamicum pUN1 in directly extracted DNA from soil, using the T4 gene 32 protein in the polymerase chain reaction. *European Journal of Soil Biology*. 30 (2): 93-98.

Valinsky, L., G. Della Vedova, A. J. Scupham, S. Alvey, A. Figueroa, B. Yin, J. Hartin, M. Chrobak, D. E. Crowley, T. Jiang and J. Borneman. (2002). Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. *Applied and Environmental Microbiology*. 68 (7): 3243-3250.

Vallaeys, T., E. Topp, G. Muyzer, V. Macheret, G. Laguerre and G. Soulas. (1997). Evaluation of denaturing gradient gel electrophoresis in the detection of 16S rDNA sequence variation in rhizobia and methanotrophs. *FEMS Microbiology Ecology*. 24: 279-285.

van der Zee, A., H. Verbakel, J. C. van Zon, I. Frenay, A. van Belkum, M. Peeters, A. Buiting and A. Bergmans. (1999). Molecular genotyping of Staphylococcus aureus strains:

Comparison of repetitive element sequence-based PCR with various typing methods and isolation of a novel epidemicity marker. *Journal of Clinical Microbiology*. 37 (2): 342-349.

Versalovic, J., T. Koeuth and J. R. Lupski. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Research*. 19 (24): 6823-6831.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. (1995). Aflp - a New Technique for DNA-Fingerprinting. *Nucleic Acids Research*. 23 (21): 4407-4414.

Wang, D. G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M. S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. J. Hudson, R. Lipshutz, M. Chee and E. S. Lander. (1998). Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science*. 280 (5366): 1077-1082.

Watanabe, T., Y. Murata, S. Oka and H. Iwahashi. (2004). A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. *Yeast*. 21 (4): 351-365.

Watson, R. J. and B. Blackwell. (2000). Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Canadian Journal of Microbiology*. 46 (7): 633-642.

Welsh, J. and M. McClelland. (1990). Fingerprinting Genomes Using Pcr with Arbitrary Primers. *Nucleic Acids Research*. 18 (24): 7213-7218.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey. (1990). DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic-Markers. *Nucleic Acids Research*. 18 (22): 6531-6535.

Winding. (1994) Fingerprinting bacterial soil communities using Biolog microtitre plates. In: *Beyond the Biomass* (eds Ritz, K., J. Dighton and K. E. Giller) Wiley, London.

Wu, L. Y., D. K. Thompson, G. S. Li, R. A. Hurt, J. M. Tiedje and J. Z. Zhou. (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Applied and Environmental Microbiology*. 67 (12): 5780-5790.

Xia, X. Q., J. Bollinger and A. Ogram. (1995). Molecular-Genetic Analysis of the Response of 3 Soil Microbial Communities to the Application of 2,4-D. *Molecular Ecology*. 4 (1): 17-28.

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 (1): 72-79.

Yap, W. H., X. Li, T. W. Soong and J. E. Davies. (1996). Genetic diversity of soil microorganisms assessed by analysis of hsp70 (dnaK) sequences. *Journal of Industrial Microbiology*. 17 (3-4): 179-184.

Young, C. C., R. L. Burghoff, L. G. Keim, V. Minakbernero, J. R. Lute and S. M. Hinton. (1993). Polyvinylpyrrolidone-Agarose Gel-Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. *Applied and Environmental Microbiology*. 59 (6): 1972-1974.

Yu, H., J. Chao, D. Patek, R. Mujumdar, S. Mujumdar and A. S. Waggoner. (1994). Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Research*. 22 (15): 3226-3232.

Zhou, J., M. E. Davey, J. B. Figueras, E. Rivkina, D. Gilichinsky and J. M. Tiedje. (1997). Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. *Microbiology Reading*. 143 (12): 3913-3919.

Zhou, J. H. (2003). Microarrays for bacterial detection and microbial community analysis. *Current Opinion in Microbiology*. 6 (3): 288-294.

Zhou, J. H. and D. K. Thompson. (2002). Challenges in appying microarrays to environmental studies. *Current Opinion in Biotechnology*. 13: 204-207.

Zhou, J. Z., M. A. Bruns and J. M. Tiedje. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*. 62 (2): 316-322.