ADAPTATION OF BIOLOGICAL FINGERPRINTING METHODS FOR FUGITIVE DUST MONITORING

FINAL REPORT
CONTRACT NO. 97-321

Prepared for:
California Air Resources Board
1102 Q Street
P.O. Box 2815
Sacramento, CA 95812

Submitted by:
Department of Land, Air, and Water Resources
One Shields Avenue
University of California, Davis
Davis, CA 95616

Prepared by:

Kate M. Scow	**Principal Investigator**
Mara J. Johnson	**Post-doctoral researcher**
Geoff N. Elliott	**Post-doctoral researcher**
Kenneth J. Graham	**Post-graduate researcher**
Philip K. Hopke	**R. A. Plane Professor**

The statements and conclusions in this Report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

ACKNOWLEDGMENTS

This Report was submitted in fulfillment of Air Resources Board Contract No. 97-321: "ADAPTATION OF BIOLOGICAL FINGERPRINTING METHODS FOR FUGITIVE DUST MONITORING" by the University of California at Davis under the sponsorship of the California Air Resources Board. Work was completed as of March 2001.

TABLE OF CONTENTS

LIST OF FIGURES vi

LIST OF TABLES viii

ABSTRACT ix

EXECUTIVE SUMMARY x
Background x
Methods x
Results xi
Conclusions xi

1.0 INTRODUCTION 1
1.1 Background 1
1.2 Objectives 2

2.0 APPROACH AND METHODS 3
2.1 Collection of Samples and Soil Property Analysis 3
2.1.1 San Joaquin Valley Sources (TSS-12) 3
2.1.2 Sustainable Agriculture Farming System (SAFS) soils and dust 5
2.1.3 Central Valley soils 5
2.1.4 Small-scale dust generation and collection 6
2.1.5 Large-scale dust ceneration and collection 8
2.2 DNA-Based Methods 9
2.2.1 Extraction and purification of DNA from sources and dust 9
2.2.2 PCR amplification of DNA 10
2.2.3 DNA fingerprinting and analysis 11
2.3 Lipid-Based Methods 11
2.3.1 PLFA extraction and analysis 11
2.3.2 SFAME extraction and analysis 12
2.4 Statistical Analyses 12
2.4.1 Similarity Matrices and Cluster Analysis 12
2.4.2 Principal Components, Redundancy and Correspondence Analyses 13
2.4.3 Classification Models 14

3.0 RESULTS AND DISCUSSION 15
3.1 Construction of Dust Generation Chambers 15
3.2 DNA-Based Approaches 16
3.2.1 Reproducibility of the PCR Method for source and dust samples 16
3.2.2 Analysis of source material from the San Joaquin Valley and comparison to laboratory-generated dust 19
3.2.2.1 Reproducibility on the field scale: source samples 19
3.2.2.2 Reproducibility on the field scale: laboratory-generated dust 23
3.2.2.3 Differentiation of sources 26

3.2.2.4 Comparison of laboratory-generated dust to source samples: San Joaquin Valley soils 33
3.2.2.5 Comparison of field-generated dust to source samples: SAFS soils 35
3.2.3 Differentiation of dust samples and potential biomarkers 36
3.2.4 Detection limits of DNA fingerprinting 40
3.3 Lipid-Based Approaches 42
3.3.1 Analysis of source material: Comparison of PLFA and SFAME methods 42
3.3.1.1 Fatty acid yields and the mass of source material required for analysis: soil set 1 44
3.3.1.2 Variability 46
3.3.2 Comparison of laboratory-generated dust to source samples: Comparison of PLFA and SFAME Methods 47
3.3.3 Detection limits of PLFA and SFAME 49
3.3.4 Summary 50
3.4 Statistical Approaches for Classifying Source and Dust Samples 51
3.4.1 Principal Components, Redundancy, and Correspondence Analyses 51
3.4.2 Classification Models 54

4.0 COMPARISON OF DIFFERENT BIOLOGICAL APPROACHES FOR CLASSIFYING SOURCES AND DUST 63

5.0 SUMMARY AND CONCLUSIONS 64

6.0 FUTURE WORK AND RECOMMENDATIONS 65

7.0 REFERENCES 66

8.0 PUBLICATIONS PRODUCED 69

9.0 GLOSSARY OF TERMS, ABBREVIATIONS, AND SYMBOLS 70

10.0 APPENDICIES 71
10.1 Technical Support Study 12 Source Sampling Protocol 71
10.2 Design and Construction of Small-Scale Dust Generation and Collection Chamber 72
10.3 Image Processing Points for DNA Fingerprint Analysis 74
10.4 PLFA Extraction and Purification Protocol 75
10.5 SFAME Extraction and Purification Protocol 78
10.6 Comparison of Laboratory-Generated Dust to Source Samples: San Joaquin Valley Soils 80

LIST OF FIGURES

Figure 1. California and detail maps of TSS-12 sample locations. 4

Figure 2. Small-scale respirable dust generation and collection chamber. 7

Figure 3. Large-scale dust generation and collection chamber 8

Figure 4. Reproducibility of laboratory replicates of filter-collected dust from small-scale chamber. 18

Figure 5. Reproducibility on the field scale: sources. Bacterial DNA fingerprints of field replicated samples from TSS-12. 20

Figure 6. Reproducibility on the field scale: sources. Eucaryotic DNA fingerprints of field replicated samples from TSS-12.. 22

Figure 7. Reproducibility on the field scale: dust. Bacterial DNA fingerprints of field replicated laboratory-generated dust from TSS-12 samples.. 24

Figure 8. Reproducibility on the field scale: dust. Eucaryotic DNA fingerprints of field replicated laboratory-generated dust from TSS-12 samples.. 25

Figure 9. Cluster diagram of bacterial DNA fingerprints for crop samples.. 27

Figure 10. Cluster diagram of eucaryotic DNA fingerprints for crop samples.. 28

Figure 11. Cluster diagram of bacterial DNA fingerprints for non-crop samples..... 30

Figure 12. Cluster diagram of eucaryotic DNA fingerprints for non-crop samples..... 31

Figure 13. Bacterial and eucaryotic DNA fingerprints of soil and respirable dust from almond samples. 34

Figure 14. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from samples collected during a corn cultivation operation at the Sustainable Agriculture Farming Systems Project.. 35

Figure 15. Bacterial DNA fingerprints of respirable dust from all source samples.. 37

Figure 16. Eucaryotic DNA fingerprints of respirable dust from all source samples. 38

Figure 17. Bacterial DNA fingerprints of sources and respirable dust from all grape samples, and potential biomarkers G1 and G2. 39

Figure 18. Correspondence analysis of Vin3 soil using either PLFA or SFAME lipid extraction on varying amounts of soil. 45

Figure 19. Correspondence analysis of six samples in soil set 1, each with two to four replicates, using either PLFA or SFAME lipid extraction 47

Figure 20. Principal components analysis of TSS-12 crop samples. 52

Figure 21. Redundancy analysis of TSS-12 crop samples. 53

Figure 22. Measured versus predicted organic carbon content of test samples using Partial Least Squares 58

Figure 23. Measured versus predicted organic carbon content of test samples using Artificial Neural Net (ANN).. 59

Figure 24. Measured versus predicted organic carbon content of test samples using Principal Components Analysis and Artificial Neural Net (ANN). Root Mean Square Error (RSME) = 1.664. 60

Figure A1. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from cotton samples. 81

Figure A2. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from grape and safflower samples. 82

Figure A3. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from tomato samples.. 83

Figure A4. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from cattle dairy and cattle feedlot samples. 84

Figure A5. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from construction and disturbed land (irrigation ditch and dry irrigation drainage pond) 85

Figure A6. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from roadway samples. 86

LIST OF TABLES

Table 1. The variability of estimates of fragment sizes and their relative abundances in DNA fingerprints. 17

Table 2. Selected properties of TSS-12 fugitive dust source samples. 21

Table 3. The estimated amounts of laboratory-generated dust required to produce a DNA fingerprint..... 41

Table 4. Descriptions of samples in soil set 1 and soil set 2: lipid analysis..... 43

Table 5. Summary of the total amounts of fatty acids detected in each sample of soil set 1, and the lower limits of the mass required to extract reliable profiles 44

Table 6. Summary of the mean total amounts of fatty acids detected across all samples in soil set 1, and the mean lower weight limits for analysis.. 46

Table 7. Comparison of the total fatty acid content and the lower weight limits for analysis of source material or PM10 from soil set 1..... 48

Table 8. Summary of the mean total amounts of fatty acids detected across all PM_{10} samples generated from in soil set 1, and the mean lower weight limits for reliable analysis... 49

Table 9. Number of samples of different source types used in DNA-based classification models. 54

Table 10. Source types used in training and test sets of classification models. 55

Table 11. PLS and Neural Net classification results for full-variable data. 55

Table 12. Classification of PLS and BP-ANN model results after variable selection using D-PLS. 56

Table 13. Classification of PLS and BP-NN model results after variable selection using Genetic Algorithms. 56

Table 14. Source sample classes used for training and test sets in PLFA-based classification models. 61

Table 15. D-PLS and RDA classification results for full-variable data. 62

Table 16. D-PLS and RDA classification results for selected-variable data. 62

ABSTRACT

We developed and applied two types of biological fingerprinting methods to characterize sources of fugitive dust. The first type was DNA based [Intergenic Transcribed Spacer (ITS) analysis), and the second type was fatty acid based (Phospholipid Fatty Acid (PLFA) and Soil Fatty Acid Methyl Ester (SFAME) analyses]. Major goals included overcoming detection limit problems associated with small samples of dust, determining relationships between sources and dust, and classifying source materials. Two dust generation/collection chambers were constructed to enable source and dust sample comparisons under controlled conditions. Source and dust comparisons were also performed on samples collected during an agricultural operation. Detection limits were lower for DNA-based than fatty acid-based methods. Both methods yielded unique biological signatures from Central Valley fugitive dust sources. The DNA-based method revealed strong similarities between source and dust fingerprints, indicating its promise for source characterization and apportionment. Classification models including artificial neural networks were optimized to analyze the large data sets generated by both types of biological fingerprinting. Appropriately applied, they classified source and dust samples with 99% accuracy. Continuing advances in molecular biology technologies will increase the ability to rapidly characterize large numbers of samples and streamline the biological fingerprinting methods currently used.

EXECUTIVE SUMMARY

Background

Levels of PM_{10} (particulate matter 10 μm aerodynamic diameter or less) in California's Central Valley regularly exceed both State and Federal air quality standards. Particularly in late summer and early fall, soil-derived fugitive dust constitutes the dominant fraction of PM_{10} and may be generated by agricultural operations, vehicular traffic on paved and unpaved roads, and construction activities. These principle sources of fugitive dust are indistinguishable by conventional analytical methods, prompting basic research for alternate methods of source characterization, including the use of biological markers derived from soil microorganisms (bacteria, fungi, and protozoa). We have developed fingerprinting methods based on the direct extraction and analysis of two classes of biological markers, nucleic acids (DNA) and fatty acids. DNA analysis focuses on the genetic code in the nucleus of cells. The principle advantages of DNA analysis include low detection limits and the potential use of highly specific gene sequences. Fatty acid analysis includes analysis of phospholipid fatty acids (PLFAs), found in cell membranes of living organisms, or total fatty acids from whole cells and nonliving biological material in the case of SFAME (soil fatty acid methyl ester) analysis. The principle advantage of fatty acid analysis is that it is currently more quantitative than DNA analysis. In our previous contract, lipid analysis distinguished source soils from one another, yet a major limitation was large sample size requirements. The objectives of the current contract were to continue the development and application of biological tools for characterizing fugitive dust sources, overcome detection limit problems associated with the inherently small sample sizes of dust, address issues of applying biological methods to source apportionment, and to use multivariate statistics to investigate the relationships between biological profiles of fugitive dust and its sources.

Methods

Both DNA and fatty acid analyses use standardized procedures that extract biochemical information from biological material in source or dust samples. DNA analysis is performed by the production of multiple, identical copies of targeted DNA (gene) fragments using specific primer sequences with the PCR (Polymerase Chain Reaction). The fragments (bands) are separated based on their length via migration in a gel matrix and are stained to visualize and analyze the resulting band pattern, or DNA fingerprint. Fingerprinting of the Intergenic Transcribed Spacer (ITS) portion of the bacterial or eucaryotic genome was conducted on forty-seven San Joaquin Valley sources and on dust samples (generated in the laboratory in a customized small-scale chamber, or collected in the field during agricultural operations). Fatty acid analysis is performed by separating fatty acids based on size and chemical properties with gas chromatography. Fatty acid fingerprints consist of percentages of fatty acids detected as peaks. PLFA analysis was conducted on nearly 500 source samples. In addition, PLFA and SFAME analyses were conducted on six dust samples generated in the laboratory in a customized, large-scale chamber. Both types of fingerprinting methods generate multivariate data (DNA band identities and densities or fatty acid types and percentages) that are used in statistical analyses and predictive mathematical models.

Results

We were successful in obtaining DNA fingerprints for source materials and dust that were reproducible among replicate extractions, different PCR conditions, and across multiple gels. Bacterial DNA fingerprints are more complex (contain 25-30 bands) than eucaryotic DNA fingerprints, which contain 8-15 bands. Each DNA fingerprint was very specific for each of the samples tested, but in some cases shared common features when from specific agricultural crops and/or geographic location. Reproducibility was also high for agricultural samples collected from the same field, and for replicate samples of laboratory-generated dust. Paired comparisons of source-dust DNA fingerprints revealed that the majority of DNA sequences detected in source samples are also detected in filter-collected dust under laboratory conditions. The amounts of filter-collected dust required for adequate DNA fingerprinting ranged from 20-200 μg (most of source samples) to 1-6 mg (unpaved road samples).

The estimated amounts of source sample required for fatty acid fingerprinting ranged from 0.5 g (SFAME) to 5 g (PLFA). The estimated amounts of dust required for fatty acid fingerprinting ranged from 0.1 g (SFAME) to 0.8 g (PLFA). The PLFA method was more reproducible than SFAME. Relationships between dusts and their source soils were evident in only two of the five soil-dust pairs tested by PLFA analysis, and in none of the sample pairs by SFAME analysis.

Artificial neural net-based and other classification methods were successful in classifying soil and dust samples by their sources greater than 90% of the time, in particular when variable selection strategies were used. These approaches were optimized for use in rapid analysis of large sets of data generated by both the DNA- and fatty acid-based methods.

Conclusions

The DNA fingerprinting method (ITS) used in this study was far more sensitive and more successful in relating dust (generated in the lab) to source samples than were fatty-acid based methods (PLFA and SFAME). DNA fingerprints were highly reproducible in laboratory and field replicates. The ITS method was in fact so specific for an individual sample, that it was possible to differentiate each of the forty-seven source samples analyzed in this study. Further work is needed to evaluate whether conclusions from our study can be extrapolated to field conditions. Continuing advances in molecular biology technologies will increase the ability to rapidly analyze large numbers of samples and streamline the fingerprinting methods currently used.

1.0 INTRODUCTION

1.1 Background

The purpose of this project was to continue progress in the development of biological tools for characterizing sources of fugitive dust. Fugitive dust contributes to non-attainment of State and Federally mandated PM_{10} (particulate matter of 10 μm aerodynamic diameter or less) levels in the Central Valley, particularly in the fall season (Chow et al., 1993). Developing protocols to differentiate suspected sources of fugitive dust, such as agricultural soils, paved or unpaved roadways, cattle industry sites, and construction sites, will contribute to the eventual application of appropriately focused abatement measures. As fugitive dust is derived from soil, and soils harbor specialized microbial communities as a function of specific environmental influences (such as moisture and nutrient availability), biological tools represent a potential avenue of source characterization. Specifically, biological fingerprinting rests on the principle that soils contain complex communities of bacteria (prokaryotes), fungi (eucaryotes), protozoa (eucaryotes), and other microorganisms, all of which contain biochemical material that can be extracted and analyzed. The types and amounts of extracted biochemical material not only comprise a fingerprint of the microbial community in a fugitive dust source, but also constitute a rich set of multivariate data applicable in mathematical predictive modeling. The scientific basis for the microbiological characterization of source soils is well established. By extension, characterizing microorganisms in fugitive dust and linking characteristics of dust and source soils should be possible, forming the major hypothesis of this research.

Previous work with the Air Resources Board (Contract No. 94-321) set the foundation for using biological tools to characterize fugitive dust, with a focus on source characterization. We developed fingerprinting methods based on two classes of biological markers, lipids (fatty acids) and, to a lesser extent, nucleic acids (DNA). Phospholipid fatty acids (PLFAs) are major components of the cell membrane and vary in their composition between eucaryotes and prokaryotes, as well as among many bacterial groups. These compounds are apparently rapidly degraded upon cell death, making them indicators of living organisms (White et al., 1979). PLFAs extracted directly from soil provide a “fingerprint” of the microbial community present and have been used to study changes in soil communities in agricultural soils during the growing season (Bossio et al., 1998), in soils subjected to heavy metal contamination (Baath et al., 1995), among others. Soil fatty acid methyl esters (SFAMEs) are derived from storage compounds in cells, as well as membranes, and can be extracted from living and dead microbial and animal cells, as well as from plant tissues in various stages of decomposition. The SFAME method is more rapid than PLFA analysis, and has been used to describe microbial communities in agricultural soils (Buyer and Drinkwater, 1997; Cavigelli et al., 1995; Ibekwe and Kennedy, 1999).

Previous analysis of biochemical information contained within source soils and a small number of dust samples indicated that both lipid-based approaches were promising for PM_{10} source characterization. However, as relatively large sample sizes were required for the PLFA method, one of the conclusions of our previous contract was that gaining additional sensitivity using nucleic acid-based methods should be pursued. Nucleic acid analysis can target specific portions of the genetic code held in DNA (deoxyribonucleic acid), making use of the PCR

(polymerase chain reaction). The primary advantage gained with the PCR is an ability to amplify the signal of microbial DNA, greatly decreasing detection limits. The main assumption in the use of PCR is that the amplified signal of targeted DNA accurately represents the types and amounts of DNA from the original microbial community. The limitations of the use of PCR for environmental samples have been reviewed (Wintzingerode et al., 1997). The final component of our previous contract was to develop statistical approaches for classifying source material (in collaboration with Dr. Philip K. Hopke at Clarkson University).

Our current contract further refined the use of biological tools for source characterization, and extended their application to the analysis of dust samples. A major goal was to identify links between the biological material in soil and dust under laboratory-controlled conditions. To accomplish this we modified methods developed in our previous contract, and adopted new methods with greater potential for successful application in the characterization of fugitive dust. In particular, we standardized and advanced a method of DNA fingerprinting. This method targets a portion of DNA that is present in all microorganisms, but differs in length and sequence among species (Intergenic Transcribed Spacer or ITS region) (Garcia-Martinez et al., 1999; Jensen et al., 1993). The technique involves a step of DNA amplification with PCR, and then the amplified sequences are separated by length to yield a DNA fingerprint of the microbial community within each sample. DNA- and lipid-based methods were performed on source samples and on dust generated and collected in two chambers designed and constructed in our laboratory.

1.2 Objectives

The objectives of this contract were to:

1) continue the development and application of biological tools for the characterization of fugitive dust sources,

2) overcome detection limit problems associated with inherently small sample sizes of dust by improving existing methods or utilizing new methods,

3) provide technical support for field studies on agricultural and urban sources of dust, and

4) use classification models, such as artificial neural networks, with data derived from DNA- and lipid-based techniques to classify source material and investigate the relationships between sources and dusts.

5) An important and unanticipated objective of this contract was the design and construction of two chambers for the generation and collection of dust from source samples.

2.0 APPROACH AND METHODS

The overall approach to achieve the objectives was to: i) adapt a DNA fingerprinting method for application to sources and dust and, analyze a comprehensive set of paired source and dust samples (source samples collected from the San Joaquin Valley), ii) construct two dust-generation chambers to produce dust from potential fugitive dust sources, iii) compare two lipid-based methods with respect to reliability, sensitivity and detection limits, as well as overall feasibility of their application to dust samples, and iv) develop and apply statistical and classification methods in the analysis of DNA and lipid data.

2.1 Collection of Samples and Soil Property Analysis

2.1.1 San Joaquin Valley Sources from Technical Support Study 12 (TSS-12)

Forty-seven samples representing suspected fugitive dust sources were collected in the San Joaquin Valley (SJV) during fall, 1997/1998 as part of the California Regional $PM_{10}/PM_{2.5}$ Air Quality Study (Figure 1). Thirty-two samples composed the agricultural sources and 15 samples represented construction sites, cattle dairy and feedlot sites, the bare surfaces of drainage ditches and basins, public and residential unpaved roads, and urban and rural paved roads. The sampling plan for agricultural sources was designed to compile same-crop samples from multiple locations to determine the variability in crop source profiles. In addition, replicates in the same location were collected from five agricultural fields (one almond, three cotton, and one tomato) to determine location-specific variability and analytical reproducibility (field-replicated samples). The sampling plan for other source categories was designed to compile a set samples from diverse locations throughout the SJV to determine the variability in profiles that represent a source type. In this study, the agricultural sources are termed crop samples, and all other sources are termed non-crop samples. Details of the sampling regime are presented in Appendix 10.1.

Information recorded for each sample site included coordinates for latitude and longitude, a five-year previous crop history for agricultural samples, and detailed site descriptions of soil conditions and management practices. At the time of sampling (fall), all agricultural crops had been harvested and annual crop residues had been incorporated into the soil by tillage. Air-dried samples were sieved (2 mm mesh), and stored at 25° C until analysis. All samples were analyzed for percent sand, silt, and clay with the pipette method (United States Department of Agriculture, 1996), for total organic carbon and nitrogen using an elemental analyzer (NA 1500 Series 2, Fisons Instruments, Beverly MA), for inorganic carbon (carbonate) with a CO_2 analyzer, and pH and electrical conductivity (1:1 soil-water paste). Relationships among soil properties were evaluated using analysis of variance and the Pearson correlation (significance determined by the t-test).

In addition, several samples were collected in the spring, 2000 from 0-5 cm of agricultural soils cropped with cotton or grape, and from fallow fields, all located in the SJV. These samples are designated as SPF 1-12.

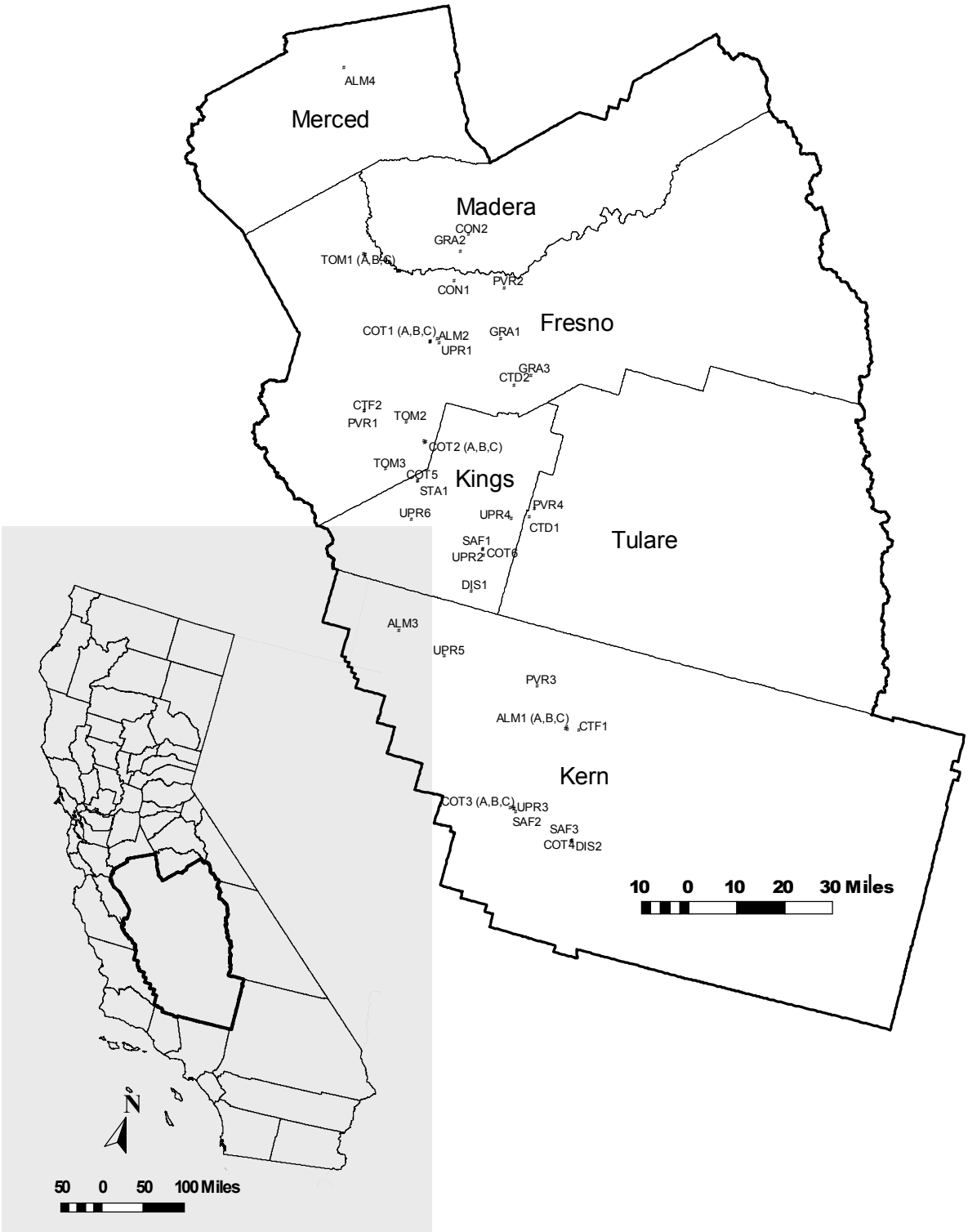


Figure 1. California and detail maps of TSS-12 sample locations. ALM = almond, COT = cotton, GRA = grape, TOM= tomato, SAF = safflower, CTD = cattle dairy, CTF = cattle feedlot, PVR = paved road, UPR = unpaved road, STA = staging area, CON = construction, DIS = "disturbed" land (DIS1 = dry irrigation water drainage area, DIS2 = irrigation ditch).

2.1.2 Sustainable Agriculture Farming System (SAFS) soils and respirable dust

The long term Sustainable Agriculture Farming Systems (SAFS) project is located in Davis, CA. The 56-plot experiment has a randomized complete block design, with crop rotations as split plots within each main plot for each farming system, with four replications, as described by Gunalpala and Scow (1998). The farming systems include organic, low input, and conventional two-year rotations. The organic system relies on organic sources of nutrients obtained from a vetch winter cover crop, manure, seaweed, and fish powder. No pesticides are used and the plots are managed according to California Certified Organic Farmers requirements. The low-input system, which is intermediate between the organic and conventional systems, relies on vetch cover crops as a partial source of nitrogen but is supplemented with mineral fertilizers and limited amounts of pesticides. The conventional system uses only mineral fertilizers, some pesticides, and the only organic matter inputs are in the form of stubble and roots from the previous cash crop.

For DNA analysis, source soil and respirable (50% cutoff 4 μm aerodynamic diameter) dust samples were collected during June 1999. The samples were collected during a cultivation operation from organic, low-input, and conventional management plots, each with corn seedlings. Six respirable dust cyclones with HFS-513A Air Sampling System pumps were attached to the cultivator bar, located approximately 20 cm above the soil surface. Dust samples were collected during cultivation for approximately 60 min from a total of three replicate plots of each management regime (designated Org, Low, or Conv Filter). Twenty soil samples were collected from the 0-5 cm layer in a diagonal transect of each plot, combined, and sieved (2 mm) to make a single composite soil sample (designated Org, Low, or Conv Soil). The single composite sample represented the total area cultivated during respirable dust collection for each of the three treatment types. Two laboratory replicates are included in the analyses (e. g. Org Soil 1 and 2)

For lipid analysis, three soil samples were collected (SAFS1, SAFS2, SAFS3). The samples were collected from field surfaces (top 10 to 15 cm), sieved, air dried and frozen at -20° C until extracted. These soils and three soils collected spring, 2000 from the SJV in the vicinity of Fresno were used in the lipid detection limit study, and were chosen to vary by crop type, texture and geographic location. Amounts of these soils (dry weight) extracted using the PLFA method were 8 g, 5 g, 3 g, 2 g, 1 g, 0.5 g, 0.3 g and 0.1 g. Amounts of soil (dry weight) extracted using the SFAME method were 500 mg, 300 mg, 200 mg, 100 mg, 50 mg, 25 mg and 10 mg. These were extracted in duplicate for both methods and the replicate data were combined prior to final statistical analysis.

2.1.3 Central Valley soils

Dr. Randall J. Southard at U. C. Davis made a collection of soils sampled in the fall of 1994 for the USDA PM10 project available to us. These samples were collected during harvest operations (top 10-15 cm) from soils cropped with cotton, almond, figs, or walnut. The soils had been air-dried, sieved to 2 mm, and stored in cardboard cartons at room temperature. Particle size analysis data and information about the sample sites (e.g., crop, geographic location) were used to select samples for analysis.

2.1.4 Small-scale dust generation and collection

To compare TSS-12 soil microbial DNA fingerprints to those of dust, a small-scale dust generation and collection system was devised (Figure 2). The chamber was manufactured, with modifications, after the design of Carvacho and coworkers (1996) with reference to Chow, et al. (1994). The chamber was designed for qualitative studies to confirm that microbial DNA could be extracted and analyzed from laboratory-generated respirable dust collected from whole soil (sieved to 2 mm). Thus, modifications were required to: i) integrate a rotating sample chamber for producing dust from air-dried whole-soil samples, more closely approximating conditions in the field where fugitive dust is produced from energy applied to exposed soil surfaces as in tillage or by the action of tires on unpaved roads, ii) create a completely closed, cleanable system to avoid the introduction of potential air-borne contaminants or the retention of biological residues that would confound results of the sensitive, PCR-based analyses, and iii) simultaneously collect six respirable dust samples on sterile quartz filters in cyclone samplers to provide laboratory replicates for analysis. A detailed description of the chamber is provided in Appendix 10.2.

Respirable dust samples were captured on sterilized (autoclaved), pre-weighed, 37 mm QM-A quartz filters (Whatman International Ltd., Maidstone, England) housed in 37 mm 3-piece conducting cassettes (Omega Specialty Instrument Co., Chelmsford, MA) fitted to the cyclones. Five-gram aliquots of soil were placed in the rotating sample holder for 20-60 min and changed 6-12 times until net filter deposits of 2-4 mg of respirable dust were collected. Samples of different types and textures produced varied amounts of suspended material. Between runs, the entire assembly was dismantled and cleaned with a mild soap solution containing bleach, rinsed with sterile water, and bathed in 95% ethanol.

The six filter replicates for laboratory-generated respirable dust were extracted individually. Three concentrations of DNA were used in the PCR, including the amount that could be extracted from a single filter, and from each of either two or three filter exacts, which were combined and concentrated. The PCR was conducted on each DNA concentration as well as on extracts of filter blanks, and on an extraction blank. The PCR was also run on concentrated filter blank extracts. Although the quantity of DNA was not measurable for the filter extracts, an estimate of the amount of dust required to produce a positive PCR signal was back-calculated by multiplying the volume of extract used in the PCR by the concentration factor, if any, and the original amount of laboratory dust collected on the filter. This calculation provides a measure of the amount of dust required to give a positive PCR signal with the extraction and PCR methods used, assuming an extraction efficiency of 50%. While the absolute minimum detection limit was not tested, the three concentrations of filter extracts provided a basis for relative comparisons.

Figure 2. Small-scale respirable dust generation and collection chamber. 1 = air pressure regulator, 2 = filter assembly for house air, 3 = electric motor (50 rpm), 4 = rotating sample holder, 5 = ground wire, 6 = dust collection chamber, 7 = Magnehelic differential pressure gauge, 8 = BGI-4 respirable dust cyclones with fitted 3 piece cassettes with 37 mm QM-A quartz filters, 9 = Gilian HFS-513A Air Sampling pumps.

2.1.5 Large-scale dust generation and collection

To generate larger amounts of dust, a large-scale dust generator was designed and constructed at U. C. Davis (Figure 3). In principle, this chamber is similar to the small-scale chamber. The dust generation portion of the chamber consists of a large rotating drum with fins, into which source soil was added. A fan-driven air stream deposits entrained dust in the collection portion of the chamber into which cyclone samplers, and eventually, the top portion of a high volume (Hi-Vol) PM_{10} sampler was fitted. PM_{10} samples were collected on autoclaved, pre-weighed 8" X 10" QM-A quartz filters. One-kilogram amounts of soil were placed in the rotating drum and samples were collected until enough PM_{10} was collected for lipid analysis. Between runs, the entire assembly was dismantled and cleaned using a vacuum.

Figure 3. Large-scale dust generation and collection chamber. 1 = Dust generation chamber, 2 = Rotating sample drum, 3 = Dust collection chamber, 4 = Vacuum pump, 5 = cyclone respirable dust samplers.

2.2 DNA-Based Methods

DNA from soil microorganisms is obtained by subjecting soil and dust samples to chemical and physical treatments, which lyse (break open) microbial cells and allow microbial DNA to go into solution. After DNA purification steps, the DNA is quantified (only for soil extracts, which have higher DNA content) with a spectrophotometer to standardize concentrations in preparation for the polymerase chain reaction (PCR). PCR primers direct the creation of multiple copies of (amplify) targeted portions of microbial DNA (DNA template). This work uses PCR primers designed to amplify the last portion of the small subunit ribosomal RNA (rRNA) gene, the first portion of the large-subunit rRNA gene, and the internal transcribed spacer (ITS) region between these two genes. Ribosomal RNA genes are present in every living organism. Two types of PCR primers are used to amplify ITS regions from either soil bacteria or eucaryotes (fungi, protozoa, and others). Recent research has shown that the eucaryotic DNA fingerprints may also resolve DNA from plants (Altschul et al., 1997).

The ITS region contains transfer RNA (tRNA) genes in the case of bacteria and the 5.8S rDNA in the case of eucaryotes, along with intervening non-coding, highly variable (variable sequence) regions. Thus, ITS regions vary both in sequence and length among microorganisms. The products of the PCR, multiple copies of ITS fragments (bands) from whole-community DNA, range in size from 300 to 1400 base pairs. These fragments are separated in a polyacrylamide gel matrix on the basis of fragment length and visualized with DNA-binding stain. The varied position and intensity of the bands constitute the DNA fingerprint. The gels are photographed to capture a digital image of the DNA fingerprint pattern. The pattern, which looks similar to a bar code, is the genetic profile that constitutes the data used for distinguishing sources and for soil-dust comparisons. All profiles are analyzed using cluster analysis and multivariate statistics to allow inferences to be made about sample similarities and relationships. These data may also be analyzed in conjunction with various soil property data (e. g. clay or carbon contents, pH) and to environmental data (e. g. geographic location, crop type) to assess the influences of soil and environment on microbial community composition.

2.2.1 Extraction and purification of DNA from sources and dust

DNA from soil microorganisms was extracted and purified from 500 mg soil, from filter blanks, and from filters containing the respirable dust samples with the FastDNA™ Spin Kit for soil and the FastPrep Instrument (Bio 101, Inc., Vista, CA) according to the manufacturer's guidelines. An empty extraction tube was included in each extraction as a blank. The quantity of DNA extracted from each soil sample was estimated by absorbance at 260 nm on a Lambda 10 UV/Vis Spectrophotometer (Perkin Elmer Applied Biosystems, Foster City, CA). The quantity of DNA extracted from filter-collected dust was below the detection level of the instrument and could not be estimated. The six filter replicates for laboratory-generated respirable dust were extracted individually. The lowest quantity of DNA for PCR was obtained from a single filter extract. Then, two filter extracts were combined and concentrated for the second-highest quantity, then, the remaining three extracts were combined and concentrated to provide DNA template theoretically three times more concentrated than from a single filter extract. The PCR was conducted on each of the resulting filter extracts, on extracts of filter blanks, and on an extraction blank. The PCR was also run on concentrated filter blank extracts.

2.2.2 PCR amplification of DNA

The polymerase chain reaction (PCR) targeting the Intergenic Transcribed Spacer (ITS) region was used to obtain DNA fingerprints of soil microbial communities. As mentioned previously, PCR primers were designed to complement either bacterial or eucaryotic DNA. For bacterial DNA fingerprints, PCR amplification was directed by primers 1406f, 5'-TGYACACACCGCCCGT-3' (Universal, 16S rRNA gene) and 155r, 5'-GGGTTBCCCCATTCRG-3' (bacterial-specific, 23S rRNA gene) (Borneman and Triplett, 1997). For eucaryotic DNA fingerprints, primers ITS1 5'-TCCGTAGGTGAACCTGCGG-3' (18S rRNA gene) and ITS4 5'-TCCTCCGCTTATTGATATGC-3' (23S rRNA gene) (White et al., 1990) were used. The ITS1-ITS4 primers, originally thought to amplify ITS sequences only for Basidiomycete and Ascomycete fungi, also amplify ITS regions from microeucaryotes and a variety of flowering plants (Altschul et al., 1997).

Replicate PCRs were completed for both primer sets for all samples with serial dilutions of DNA extracts that yielded 1-4 ng of DNA template. PCR optimization was completed following the guidelines offered by Palumbi (1996) using DNA extracted from *Bacillus subtilis* (ATCC# 6051), *Escherichia coli* (ATCC# 10798), and *Saccharomyces cervisiae* (ATCC# 204680). DNA extracts of these species were also used as positive and negative control DNA for PCR. As their compete genomic sequences are available (The Institute for Genomic Research (TIGR), 2001), the numbers and sizes of their ITS regions are known. The resolution of appropriately sized bands from these control strains confirmed optimal PCR (and electrophoresis) conditions both initially and throughout the study.

The 50 µl reaction mixture, consisting of 25 pmol of each primer, 200 µM each dNTPs, 1X PCR buffer, 2.0 mM $MgCl_2$, and 1.5 U AmpliTaq Gold™ DNA polymerase (Perkin Elmer Applied Biosystems, Foster City, CA) was combined with either 2 µl (bacterial PCR primers) or 4 µl (eucaryotic PCR primers) of the 2-fold dilutions of template DNA. In addition, positive and negative control DNA (see above), solutions from DNA extraction blanks, and sterilized nanopure water as a PCR blank were included in each PCR. After a pre-incubation step to activate the AmpliTaq Gold (95° C for 10 min), thermocycling consisted of 30 cycles of denaturation at 94° C (30 s), annealing at either 60° C (30 s) for bacterial primers or at 55° C for eucaryotic primers (30 s), extension at 72° C (1 min), and a final extension at 72° C (10 min). All PCR products were examined by agarose gel electrophoresis with ethidium bromide stain to estimate the volumes of PCR products to load in polyacrylamide gels.

PCR product solutions (4-10 µl) were loaded in 4% polyacrylamide/1X TBE gels and electrophoresed in the DCode™ System (Bio-Rad Laboratories, Inc., Hercules, CA) at 150V (6.8 V cm^{-1}) for 3.5 hours at 25° C. After staining with 0.01% SYBR Green (BioWhittaker Molecular Applications, Rockland ME) for 30 min, gels were illuminated with UV light for image capture with a charge-coupled-device (CCD) camera equipped with a 520 nm bandpass filter (Corion Corp., Franklin, MA). Two DNA fingerprints, representing two dilutions of template DNA for a single sample, were included in the image and statistical analyses.

2.2.3 DNA fingerprinting and analysis

Digital images of DNA fingerprints were imported to Gelcompar II, a state-of-the-art software package designed for DNA fingerprint analysis (Applied Maths, Kortrijk, Belgium). Images were processed as recommended by the Gelcompar II designers with reference to image processing points from Rademaker and De Bruijn (1997). Details are provided in Appendix 10.3. DNA fingerprint data for each of the soil and dust samples consist of the number, location, and intensity of DNA fragments produced from PCR amplification of microbial community DNA. These data were used in profile comparisons among sources, or between soil and dust pairs. The specific methods of applying these data in statistical analyses are described in Section 2.4.

2.3 Lipid-Based Methods

Lipids were directly extracted from sources using two different methods: phopholipid fatty acid (PLFA) extraction and soil fatty acid methyl ester (SFAME) extraction. Phospholipids, a subset of fatty acids, are essential components of all living cells and rapidly degrade once disassociated from the cell membrane. Only viable microbes have intact membranes and thus PLFAs represent living organisms (White et al. 1979). In contrast, the SFAME method extracts total fatty acids, recovering not only membrane fatty acids, but also fatty acids derived from extracellular lipids and cellular storage compounds. Although these total fatty acids were originally derived from living organisms, they may be preserved over long periods in the form of soil organic matter. Fatty acids derived from samples by both methods are analyzed using a gas chromatograph to yield a series of peaks, which are compared to a bacterial database of known fatty acids for identification. Once identified, profiles are standardized using internal standards of known concentration, peak areas are converted to nanomole concentrations, and profiles are compared using correspondence analysis (CA). Please refer to Section 2.4 for details.

For lipid-based analyses, six soils were initially selected for differences in crop, soil treatment, soil type, and location. They were extracted over a range of dry weight amounts to compare extraction efficiency and profile reliability, and to determine the detection limits of each method. Large amounts of PM_{10} that were produced from the six soils using the large-scale dust generator were compared to each other and to their source soils. The masses required to meet detection limits, as determined initial study of six soils, were applied to a larger set of 55 soils collected from the Central Valley. These soils encompassed four crop types and several soil texture combinations. This sample set was also used for PLFA to SFAME comparisons to see if conclusions arrived at in the smaller study held true, and to assess the importance of crop type and soil texture in lipid profile relationships. PLFA data from 353 Central Valley agricultural sources were used to classify sources with artificial neural net techniques.

2.3.1 PLFA extraction and analysis

Lipids were extracted from soil samples using a mixture of chloroform, methanol, and a phosphate buffer (White et al. 1979). Extracted lipids were reconstituted and separated into

three lipid classes (neutral lipids, glycolipids, and phospholipids) using silicic acid columns. The phospholipids were retained, dried with nitrogen, trans-esterified to form fatty acid methyl esters. We used the automated Microbial Identification Diagnostic System (MIDI, Inc., Newark, NJ), which consists of a Hewlett Packard gas chromatograph and software for the identification of fatty acids. A gas chromatograph-mass spectrometer in another U. C. Davis laboratory was used for confirmation of fatty acid identity. The final protocol is presented in Appendix 10.4.

2.3.2 SFAME extraction and analysis

The SFAME (Soil Fatty Acid Methyl Ester) method was developed as an alternative to the PLFA method to produce lipid fingerprints from smaller masses of soil. This protocol requires approximately 1 day to process 16 samples in contrast to the 3 days required for the same number of samples by PLFA analysis, and requires smaller samples sizes. The final SFAME protocol is described in Appendix 10.5.

We discuss issues with methodology and quality assurance in the Results section.

2.4 Statistical Analyses

Data from both DNA- and lipid-based methods were rigorously analyzed by a number of statistical methods. The complete range of methods applied to DNA fingerprint data include the derivation of similarity matrices and their application to cluster analysis, Principal Components Analysis (PCA) and Redundancy Analysis (RDA), discriminant partial least squares (D-PLS), and back propagation-artificial neural net (BP-ANN) analysis. The statistical methods applied to lipid profile data include correspondence analysis (CA) and canonical correspondence analysis (CCA), discriminant partial least squares (D-PLS), and regularized discriminant analysis (RDisA). These statistical methods were evaluated to determine the most effective way to analyze data gained from DNA- and lipid-based methods, and to select which variables are the most useful in sample identification. Finally, these statistical tools aid in the assessment of detection limits.

2.4.1 Similarity Matrices and Cluster Analysis

Initially, individual lanes in the polyacrylamide gels containing bacterial or eucaryotic PCR products were catalogued by sample in the GelCompar II database, and analyzed collectively as "bacterial DNA fingerprints" or "eucaryotic DNA fingerprints." For each fingerprint type a whole-pattern analysis (of the 320 pixel densitometric curve that comprises a fingerprint pattern) was completed with the Pearson product moment correlation (Pearson, 1926), which directly compares samples based on densitometric curve data. The Pearson product moment correlation was previously demonstrated as the appropriate method for analysis of complex DNA fingerprints (Haene et al., 1993; Rademaker and J., 1997). Cluster diagrams were then generated by the un-weighted pair group method using arithmetic averages (UPGMA) algorithm (Sneath and Sokal, 1973), a standard method for finding groups in data. The consistency of the clusters in the diagram was evaluated by cophenetic correlation, in which

limited credence is attributed to the hierarchical structure below a value of 70% (Sackin and Jones, 1993).

2.4.2 Principal Components, Redundancy, and Correspondence Analyses

To conduct principle components and redundancy analyses, densitometric curves of bacterial DNA fingerprints were reduced to their salient features, namely to the visually observable array of bands of varied position and intensity that were identified by the software in the gel image processing step. To retrieve the data, the GelCompar II bandmatching function was used with 1.0% optimization and 2.0% position tolerance to assign bands to "band classes." Band classes provide discrete position designations for bands that migrate to the same position within a gel. Band classes are identified by comparing the position of a band to the migration of bands of known sizes in the 20 bp standard ladder (e.g., 909 bp or 399 bp). Band classes that contained small numbers of light bands were filtered using the band class filter with a setting of 20% minimum area. A total of 32 band class variables were exported to a spreadsheet to standardize the data for PCA and RDA.

A preliminary observation of the band class variables revealed that not all band classes were represented in each sample, and some bands were below the detection threshold of the software, resulting in a data matrix with many zero values. This feature is incompatible with PCA and RDA. Thus for each sample, values from the raw densitometric curve (also exported from GelCompar II) were included where bands had not been selected. The PCR replicates were then averaged and the data set was evaluated by observing the histogram distribution of band class values. As the distribution appeared to conform to a log-normal shape, a $\log_{10}$ transformation was applied to the entire data set. Finally, to ensure that each band class variable had equal weight in the analysis, the band class data matrix was standardized to a mean of zero and a standard deviation of one (Song et al., 1999). Soil property values for nitrogen, carbon, DNA content, and electrical conductivity were also log-transformed, and all soil property variables were standardized to a mean of zero and a standard deviation of one before performing multivariate analyses.

PCA is a multivariate statistical analysis technique used to project the maximum variance of the bacterial DNA fingerprint band class data optimally in multiple dimensions, (e.g., axis 1 and axis 2) in an unconstrained ordination. This method presents a very large data matrix as points in a single diagram. In a PCA ordination diagram, samples with similar DNA fingerprints are located close to one another, and those dissimilar are located far apart. Band class variables with the highest recorded peak areas are located adjacent to the samples that contribute those values to the data matrix. These band class variables represent possible biomarkers for adjacent samples. To explore the relationships of the environmental variables (soil properties) to the bacterial DNA fingerprints, RDA directly regresses the soil property measurements with the multivariate band class data resulting in an ordination diagram with axes constrained to be linear combinations of the soil properties. In an RDA ordination diagram, the DNA fingerprint data matrix and the environmental variable data matrix are presented as points on a single diagram. With the Monte Carlo permutation test, the significance of the soil properties in accounting for the observed variance of the band class multivariate data can be assessed with p-values. Thus, RDA can be used to test hypotheses, such as which soil properties are significant

in explaining grouping observed in an ordination diagram. In the RDA diagram, positively correlated soil property variables are shown as arrows pointing in the same direction, negatively correlated variables point in opposite directions, and perpendicular vectors are uncorrelated. In addition, the length of the arrow is a measure of the relative importance of the soil property in explaining the band class data. Both PCA and RDA were performed in Canoco 4.02 software (Center for Biometry, Wageningen, Netherlands), and are discussed by Jongman, et al. and ter Braak (1995; 1994).

For lipid-based analyses, Correspondence Analysis (CA), also projects the maximum variance of data optimally in multiple dimensions, (e.g., axis 1 and axis 2) in an unconstrained ordination. In this case, the data are nanomole concentrations of fatty acids. CA, also known as dual scaling and reciprocal averaging, is an alternative to principal components analysis (PCA). This method does not require data to be transformed to a proportional or percentage basis prior to analysis. Canonical Correspondence Analysis (CCA) is a constrained ordination, or direct gradient analysis that uses supplemental data in the form of environmental variables. It is an extension of CA where the sample scores derived from the fatty acid scores are used as the dependent variables in a multiple linear least-squares regression against the environmental variables of the samples. The sample scores from the least squares regression are then used to assign new fatty acid scores by weighted averaging. The algorithm continues to iterate until the sample scores from the least squares regression are stable. The axes created by the CCA analysis are linear combinations that are constrained to maximize the correlations between samples and environmental variables. These correlations can used to test the importance of various environmental variables on the fatty acid profiles of samples. CA was performed in SAS (Version 8.0, SAS Institute, Cary NC) and CCA was performed in Canoco 4.02 software (Center for Biometry, Wageningen, Netherlands), both methods are fully discussed by Jongman, et al., and ter Braak (1995; 1994).

2.4.3 Classification Models

Dr. Philip K. Hopke, our collaborator at Clarkson University in New York, is an expert in the use of artificial neural networks and other data classification methods. He used these classification methods to determine the groupings of soil sources based on DNA and lipid data. The initial examination of these data using PCA suggested that classes could not always be cleanly separated using methods like PCA that maximally reproduce variance. Thus, partial least-squares (PLS) methods were examined since they maximize covariance. In addition, for some of the classification problems, the initial examination of the data suggested that the classes could not be separated with linear methods. PCA and PLS find linear structures (hyperplanes in the reduced dimensional space). However, artificial neural networks (ANN) can model non-linear structures and provide good classification of the samples. Data analysis methods typically assume that each of the input variables contains information that is useful in solving the problem at hand, and is linearly independent of the other measured variables. However, typically there are redundancies in the data as well as measurements that are not directly related to the problem of separating the classes. Thus, variable selection or data compression prior to the application of the classification techniques can often provide better identification and separation of the groups in the data set.

Dr. Hopke's group first determined which input variables (e.g., bands in a DNA fingerprint) are most important for prediction of soil sources using discriminant partial least square (D-PLS) and back propagation-artificial neural network (BP-ANN) models. The variable selection methods investigated were the stepwise variable selection method and genetic algorithms (GAs). Two hundred and twenty-three DNA fingerprints were used in the analysis. Based on the brightness of the bands, densitometric curves of the selected DNA band pattern were extracted from the gel images. The curves were smoothed using Savitsky-Golay method and scaled to the DNA standard markers. The prediction results based on the two variable selection methods for PLS and Neural Network models were compared. These approaches and data were also used in a second study to classify sources, and to determine relationships between DNA fingerprints and soil properties, such as texture, nitrogen, and organic carbon.

Microbial lipids in agricultural soils from 352 sites in California's Central Valley were extracted and analyzed. The variables most important for classifying samples were derived from the complete set of PLFA data based on partial least squares regression coefficients. With the selected set of fatty acid variables, both discriminant partial least squares (D-PLS) and regularized discriminant analysis (RDisA) were run and compared for their relative performance in classifying samples.

3.0 RESULTS AND DISCUSSION

3.1 Construction of Dust Generation Chambers

An important and unanticipated objective of this contract was the design and construction of chambers for the generation and collection of dust from source samples. The original contract indicated that we would obtain samples of PM_{10} collected in the Crocker Nuclear Laboratory dust resuspension chamber. However, it became evident that we would be unable to obtain adequately sized samples of PM_{10} for our analyses. Therefore, we designed and used chambers customized for our requirements.

Two chambers were constructed, one to generate and collect respirable dust from relatively small source sample sizes (small-scale chamber) and the other to collect respirable dust or PM_{10} from large source sample sizes (large-scale chamber). The small-scale chamber (Figure 2) was fitted with OSHA-approved personal air samplers to collect respirable dust (50% cut of 4 μm aerodynamic diameter). Six respirable dust samples were collected simultaneously on sterile quartz filters in cyclone samplers to provide laboratory replicates for analysis. The system included a rotating sample chamber for producing dust from air-dried whole-soil samples in a completely closed, cleanable system to avoid contamination and permit disinfections between runs. The large-scale chamber was originally fitted with personal air samplers (Figure 3). However, when a Hi-Vol sampler became available on loan (through the efforts of Tony VanCuren), an additional port was constructed adjacent to the original chamber to accommodate this EPA-approved collection system. The Material and Methods section provides a detailed description of the small-scale chamber.

3.2 DNA-Based Approaches

A major emphasis of this contract was to improve DNA techniques for fingerprinting dust. After an evaluation of several DNA fingerprinting methods including Random Amplified Polymorphic DNA (RAPD) and thermal gradient gel electrophoresis (TGGE) analyses in the previous contract (No. 94-321), we decided to use Intergenic Transcribed Spacer (ITS) DNA fingerprinting to characterize samples because it is a relatively rapid, highly repeatable, and an inexpensive method. In addition, ITS DNA fingerprinting was used to evaluate different fractions of the microbial community. Specifically, we focused on DNA fingerprinting of bacterial (prokaryotic) and eucaryotic (fungi, protozoa, and plants) organisms. This technique was designed to amplify signature patterns of microbial DNA from small sample sizes (microgram and milligram quantities) to ascertain if the approach held potential for future application in field-based PM_{10} monitoring studies.

3.2.1 Reproducibility of PCR method for source and dust samples

In terms of sources, we confirmed that DNA fingerprints were reproducible among replicate extractions, for varied PCR template concentrations, and across multiple gels, in an initial reproducibility study with three soil samples from the San Joaquin Valley (COT1B, TOM1A, and ALM1A). Each sample was extracted 3 times, and 3 independent PCR reactions were completed for several dilutions of template DNA from each extraction, and a total of 5 gels were run for each sample. As errors may change with migration distance in the gel matrix, consistency was evaluated by calculating the size and relative area of three bands in positions approximately one quarter, one-half and three-quarters of the total migration distance of the gel. A “consensus” fingerprint, consisting of bands present across all replicates was used to estimate the variability in area estimates of the bands. The relative areas were estimated by dividing the intensity of the selected peak by the total intensity of all peaks in the consensus pattern, multiplied by 100.

Two measures of reproducibility were evaluated, fragment size (proxy for migration distance) and peak area (Table 1). The coefficient of variation (CV) for estimates of fragment size was low, less than 1% for all peaks. There was no increase in the CVs associated with migration of the fragments at the bottom portion of the gels. The relative area estimates of the peaks showed higher CVs. The lowest CV was recorded for Peak 2 of TOM1A (12.3), and the highest was recorded for Peak 3 of COT2B (38.9).

In terms of dust samples, reproducibility was evaluated by individually extracting and DNA fingerprinting five filter samples collected simultaneously in the small-scale chamber from a single SJV source (SPF10, an agricultural soil recently planted to grape). The DNA fingerprints from the filter samples were nearly 90% similar for both bacterial and eucaryotic types (Figure 4).

Table 1. The variability of estimates of fragment sizes and their relative abundances in DNA fingerprints. Fragment size is derived from a standard curve generated from the 20bp ladder using a cubic spline fit (logarithmic dependence). Estimates of fragment relative abundance (calculated by dividing individual peak areas by the total integrated area) are average values for a total of 3 replicate extractions and 4 PCRs on 5 gels for each sample (ALM1A, COT2B, TOM1A).

Sample and Peak	Fragment Size (bp)		Fragment Relative Abundance	
	Mean	SD/CV	Mean	SD/CV
ALM1A				
Peak 1	785.29	6.12/0.78	15.39	3.67/23.82
Peak 2	634.56	3.43/0.54	8.48	1.31/15.48
Peak 3	518.36	3.10/0.60	7.33	1.82/24.82
COT2B				
Peak 1	796.16	2.37/0.30	10.29	1.53/14.89
Peak 2	694.99	2.27/0.33	2.39	0.73/30.63
Peak 3	527.89	2.17/0.41	12.85	4.94/38.94
TOM1A				
Peak 1	796.56	2.62/0.33	6.01	1.40/34.49
Peak 2	592.24	2.16/0.37	11.34	1.40/12.34
Peak 3	491.41	1.48/0.30	7.71	1.81/23.44

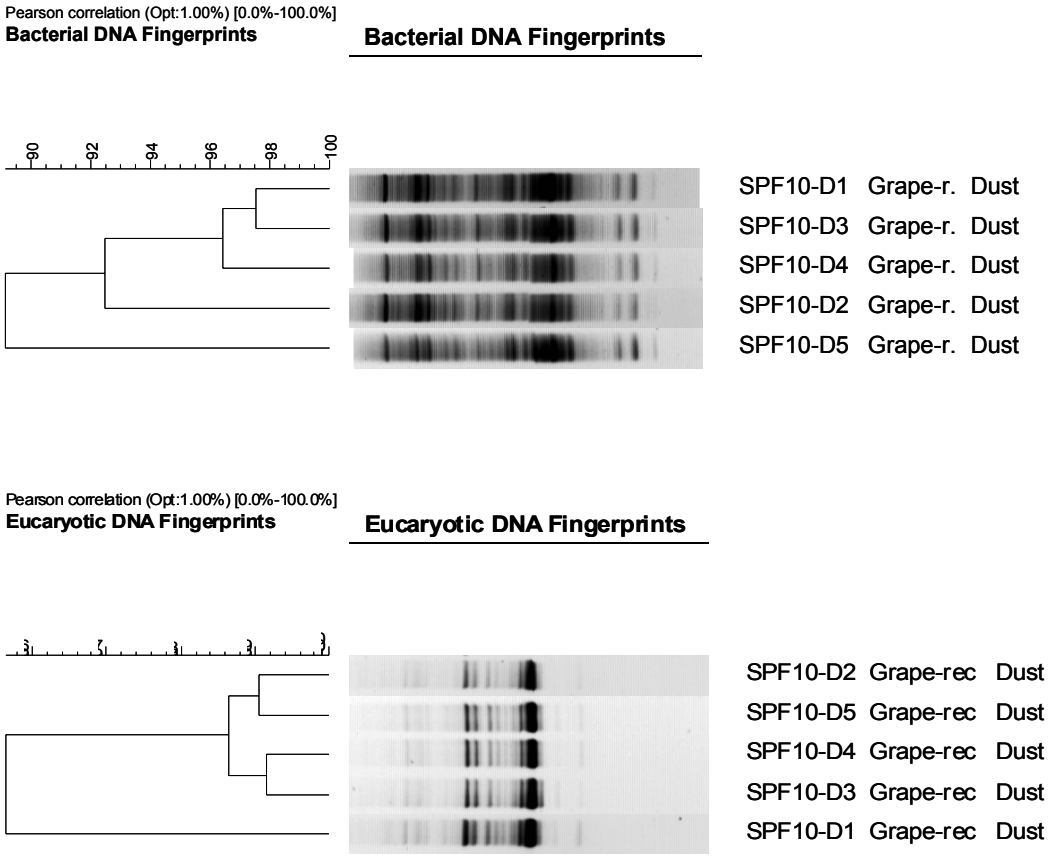


Figure 4. Reproducibility of laboratory replicates of filter-collected dust from small-scale chamber. SPF10 = Grape, recently planted. D1-D5 = Filter replicate designations. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

3.2.2 Analysis of source material from the San Joaquin Valley and comparison to laboratory-generated dust

As proposed, samples from Technical Support Study 12 were obtained for DNA fingerprint analysis. In addition, several other samples were collected from the SJV in spring, 2000 by the U. C. Davis research team (SPF samples). Microbial DNA was extracted from fugitive dust source samples, and from dust collected in the small-scale chamber. DNA fingerprinting was performed on all extracts, and DNA fingerprints among sources and versus laboratory dusts were compared. These results are presented in the following subsections: i) reproducibility on the field scale, ii) capacity of DNA fingerprinting to differentiate individual sources, iii) degree of similarity between source and dust DNA fingerprints, and iv) potential of DNA fingerprinting to differentiate source categories.

3.2.2.1 Reproducibility on the field scale: source samples

In the TSS-12 study, three field replicates were taken from a total of five agricultural fields (one almond, three cotton, and one tomato), to determine within field variability of analytical methods potentially useful in characterizing fugitive dust sources. The similarities of these samples based on DNA fingerprint data were computed with the Pearson product moment correlation and cluster analysis. As described in the Statistical Analyses section, the percent similarity indicated on the scale for branches encompassing the field replicated samples and their laboratory replicates (indicated by –S1 and –S2) basically provides a measure of how many DNA fingerprint bands are shared among samples contained in the same group.

For bacterial DNA fingerprints, field replicates range in similarity from 78% (COT3A-C) to 92% (ALM1A-C), as shown in Figure 5. This finding suggests that despite the high degree of spatial heterogeneity revealed by soil property analysis (Table 2), the samples are relatively homogeneous by DNA fingerprint analysis. This finding does not extend to the eucaryotic DNA fingerprints (Figure 6), for TOM1A-C (for which a lower similarity of about 67% was recorded) and COT1A-C (for which field replicates are not contained within the same group). Overall, differences between bacterial DNA fingerprints of multiple samples collected within a single field were less than differences between fields, an indication that individual fields have relatively unique profiles.

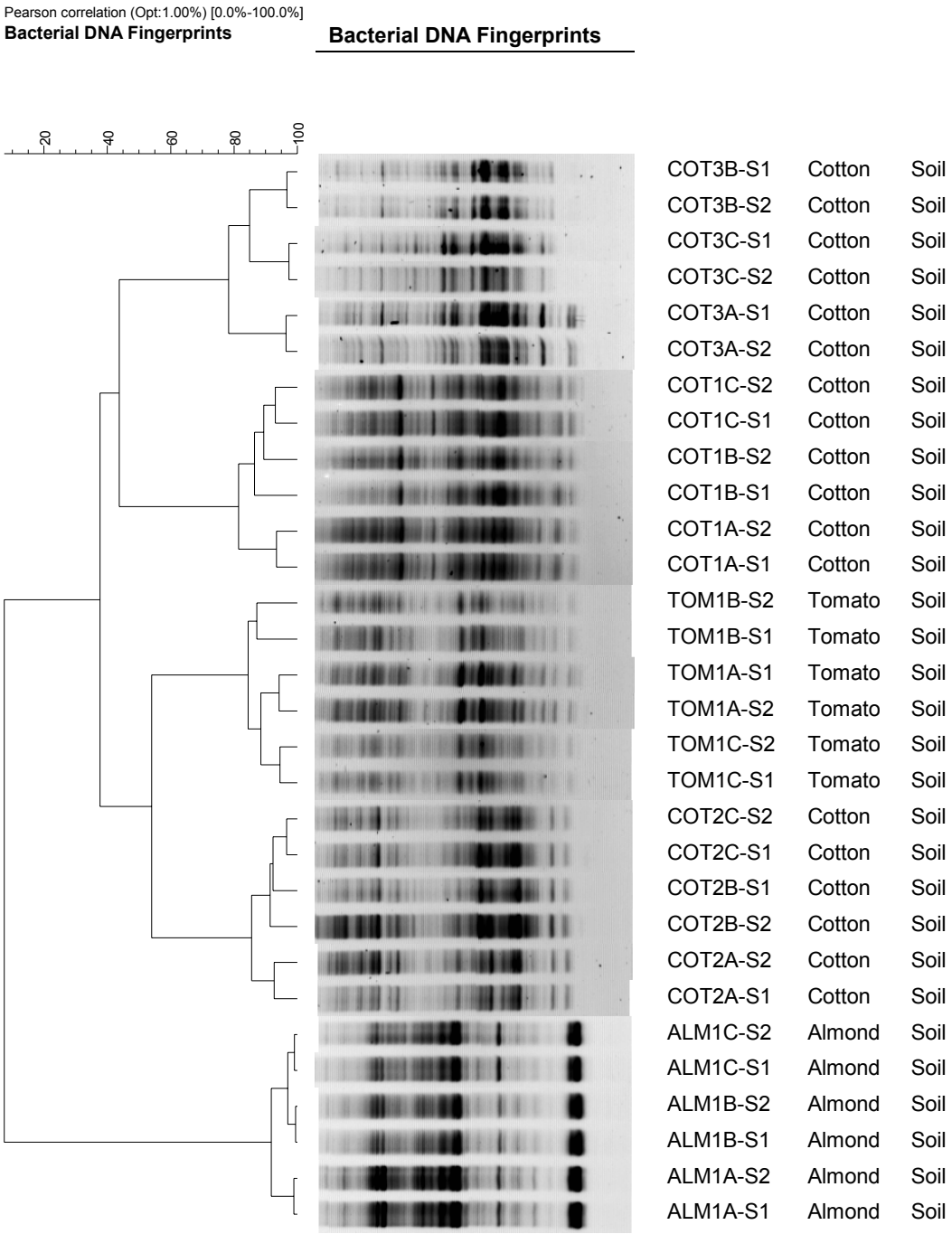


Figure 5. Reproducibility on the field scale: sources. Bacterial DNA fingerprints of field replicated samples from TSS-12. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

Table 2. Selected properties[a] of TSS-12 fugitive dust source samples.

Sample ID	Source Category	Sand (%)	Silt (%)	Clay (%)	N (%)	InC (%)	OC (%)	DNA µg gsoil^{-1}	pH[b]	EC[b] dS m^{-1}
Crop Production samples										
ALM1A	Almond	80	7	13	0.597	0.032	8.81	10.0	7.1	2.81
ALM1B	Almond	78	10	12	0.852	0.036	12.57	5.5	7.3	3.30
ALM1C	Almond	83	6	10	0.612	0.027	8.56	8.5	7.5	4.00
ALM2	Almond	82	11	7	0.254	0.025	2.18	2.3	6.8	2.76
ALM3	Almond	70	12	19	0.740	0.061	9.27	18.6	7.6	0.60
ALM4	Almond	89	6	5	0.201	0.006	2.30	5.5	6.7	2.77
COT1A	Cotton	81	12	7	0.045	0.009	0.37	2.4	7.4	0.43
COT1B	Cotton	71	17	11	0.062	0.020	0.55	2.7	7.3	0.57
COT1C	Cotton	64	23	13	0.065	0.058	0.61	2.2	7.5	0.73
COT2A	Cotton	28	31	42	0.073	0.081	0.66	0.8	7.3	0.44
COT2B	Cotton	41	24	35	0.063	0.063	0.51	0.9	7.3	0.43
COT2C	Cotton	28	33	39	0.095	0.036	0.73	0.9	7.3	0.61
COT3A	Cotton	57	19	24	0.115	0.020	1.13	2.8	6.9	1.08
COT3B	Cotton	10	39	52	0.272	0.002	2.86	2.5	6.7	1.38
COT3C	Cotton	19	31	50	0.221	0.000	2.32	1.9	6.8	0.86
COT4	Cotton	12	37	51	0.190	0.007	2.01	2.0	7.5	1.60
COT5	Cotton	42	26	33	0.107	0.011	0.87	2.0	7.5	0.91
COT6	Cotton	6	38	55	0.170	1.289	3.32	2.2	7.8	2.14
GRA1	Grape	73	21	6	0.060	0.008	0.61	2.2	6.9	0.89
GRA2	Grape	66	19	15	0.074	0.001	0.89	3.6	7.0	0.45
GRA3	Grape	84	12	4	0.043	0.000	0.46	3.5	6.1	0.38
SAF1	Safflower	7	42	51	0.126	1.146	2.40	2.5	7.7	1.68
SAF3	Safflower	39	25	36	0.186	0.002	1.90	3.3	7.1	2.92
STA1	Staging Area	51	22	27	0.094	0.069	1.35	NA	NA	NA
TOM1A	Tomato	22	31	47	0.133	0.169	1.15	1.8	7.6	1.79
TOM1B	Tomato	24	32	44	0.132	0.098	1.15	1.9	7.6	1.90
TOM1C	Tomato	13	32	55	0.137	0.155	1.21	1.7	7.6	1.22
TOM2	Tomato	29	29	43	0.092	0.022	0.51	1.6	7.3	1.07
TOM3	Tomato	49	26	24	0.075	0.053	0.43	1.9	7.5	2.27
UPR1	Ag Unpaved Road	75	15	10	0.107	0.085	0.35	0.9	8.2	15.60
UPR2	Ag Unpaved Road	9	42	49	0.180	1.062	2.65	1.2	7.7	4.59
UPR3	Ag Unpaved Road	82	7	11	0.043	0.002	0.33	0.5*	6.5	3.90
Non-crop Samples										
CON1	Construction/Earthmoving	78	19	4	0.018	0.055	0.30	0.5*	8.5	0.74
CON2	Construction/Earthmoving	73	18	9	0.026	0.004	0.29	0.5*	7.3	0.49
CTD1	Dairy	72	12	16	1.304	0.294	17.81	2.0	8.4	14.92
CTD2	Dairy	NA	NA	NA	1.902	0.412	20.74	11.8	NA	NA
CTF1	Feedlot	71	15	15	0.506	0.244	4.98	10.0*	8.2	25.50
CTF2	Feedlot	na	NA	NA	1.542	0.364	17.80	9.0	8.1	22.20
DIS1	Dist Land Salt Buildup	31	33	37	0.039	0.424	0.46	0.5*	10.1	78.00
DIS2	Dist Land Salt Buildup	53	28	20	0.112	0.205	0.82	2.8	8.5	29.70
PVR1	Urban Paved Road	88	7	5	0.161	0.131	3.71	0.5*	NA	NA
PVR2	Urban Paved Road	NA	NA	NA	NA	NA	NA	0.1*	NA	NA
PVR3	Rural Paved Road	57	29	14	0.258	0.022	4.31	0.1*	5.4	3.37
PVR4	Rural Paved Road	56	22	22	0.304	0.199	3.35	3.1	NA	NA
UPR4	Pub/Res Unpaved Road	92	4	4	0.043	0.002	0.41	1.4	7.4	0.24
UPR5	Pub/Res Unpaved Road	89	6	5	0.024	0.480	0.21	0.1*	8.1	1.12
UPR6	Pub/Res Unpaved Road	86	7	6	0.037	0.052	0.39	1.4	7.5	0.57

[a] N = nitrogen, InC = inorganic carbon (carbonate) OC = organic carbon, DNA = deoxyribonucleic acid, EC = electrical conductivity.

[b] pH and EC of a 1:1 soil:water mixture. [c] NA = not available (insufficient sample, or not measured).

*No spectrophotometric reading recorded, based on visual examination of DNA in agarose gels.

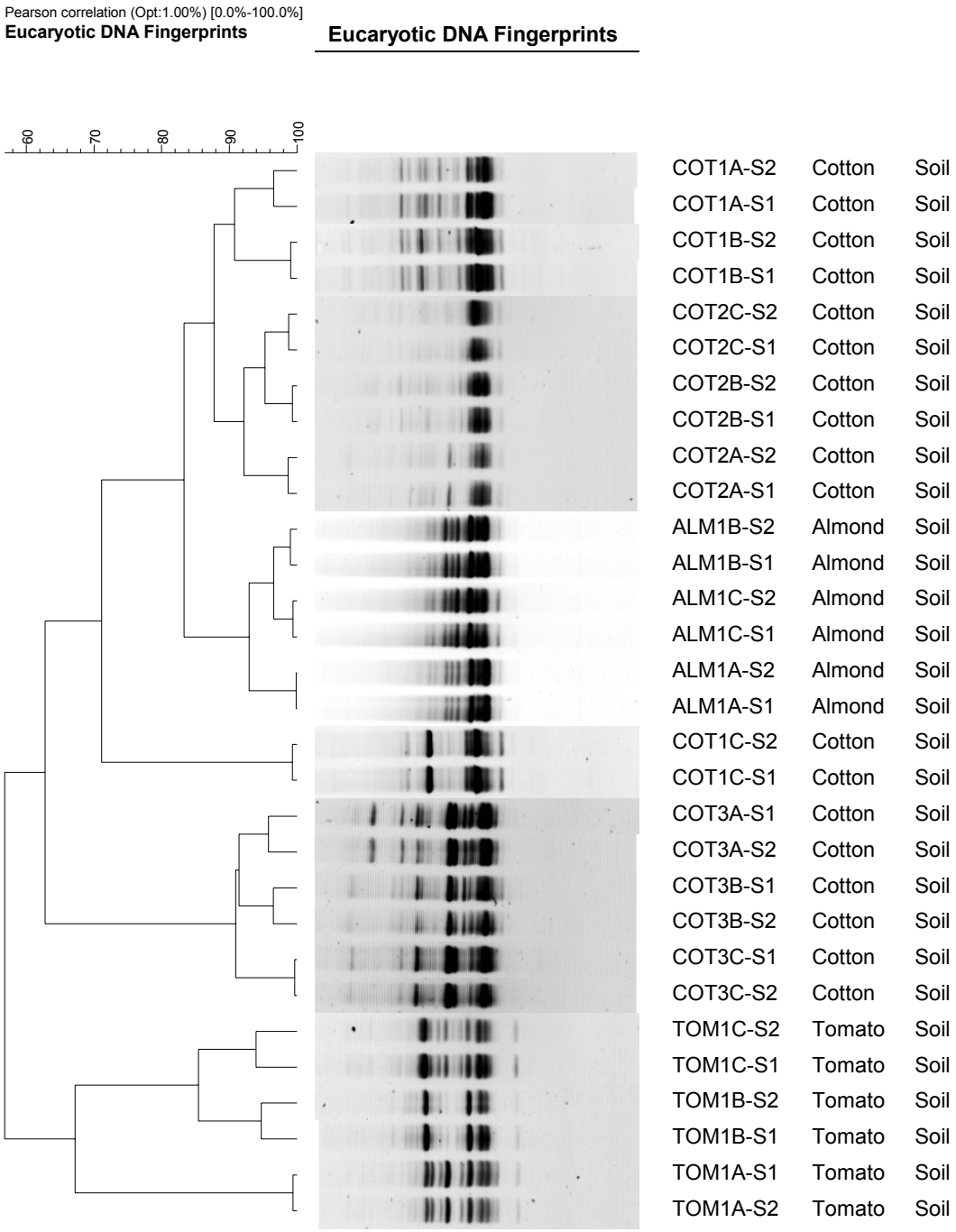


Figure 6. Reproducibility on the field scale: sources. Eucaryotic DNA fingerprints of field replicated samples from TSS-12. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

3.2.2.2 Reproducibility on the field scale: laboratory-generated dust

The trends observed for bacterial and eucaryotic DNA fingerprints of laboratory-generated dust from field replicates were similar to those observed for source material (Figures 7 and 8). For bacterial DNA fingerprints, field replicated samples were 60% similar or greater, and sources could be differentiated (with the exception of COT2A). For eucaryotic DNA fingerprints, field replicates were generally 90% similar or greater, but sources were not as well differentiated (e.g. COT1 and COT3) as with bacterial fingerprinting. This finding may have been due to the presence of fewer bands in the eucaryotic DNA fingerprint providing fewer data points for statistical analysis. In addition, it is likely that eucaryotic organisms, being larger than bacteria, are more variable in density and diversity across small sample sizes and thus less consistently fingerprinted.

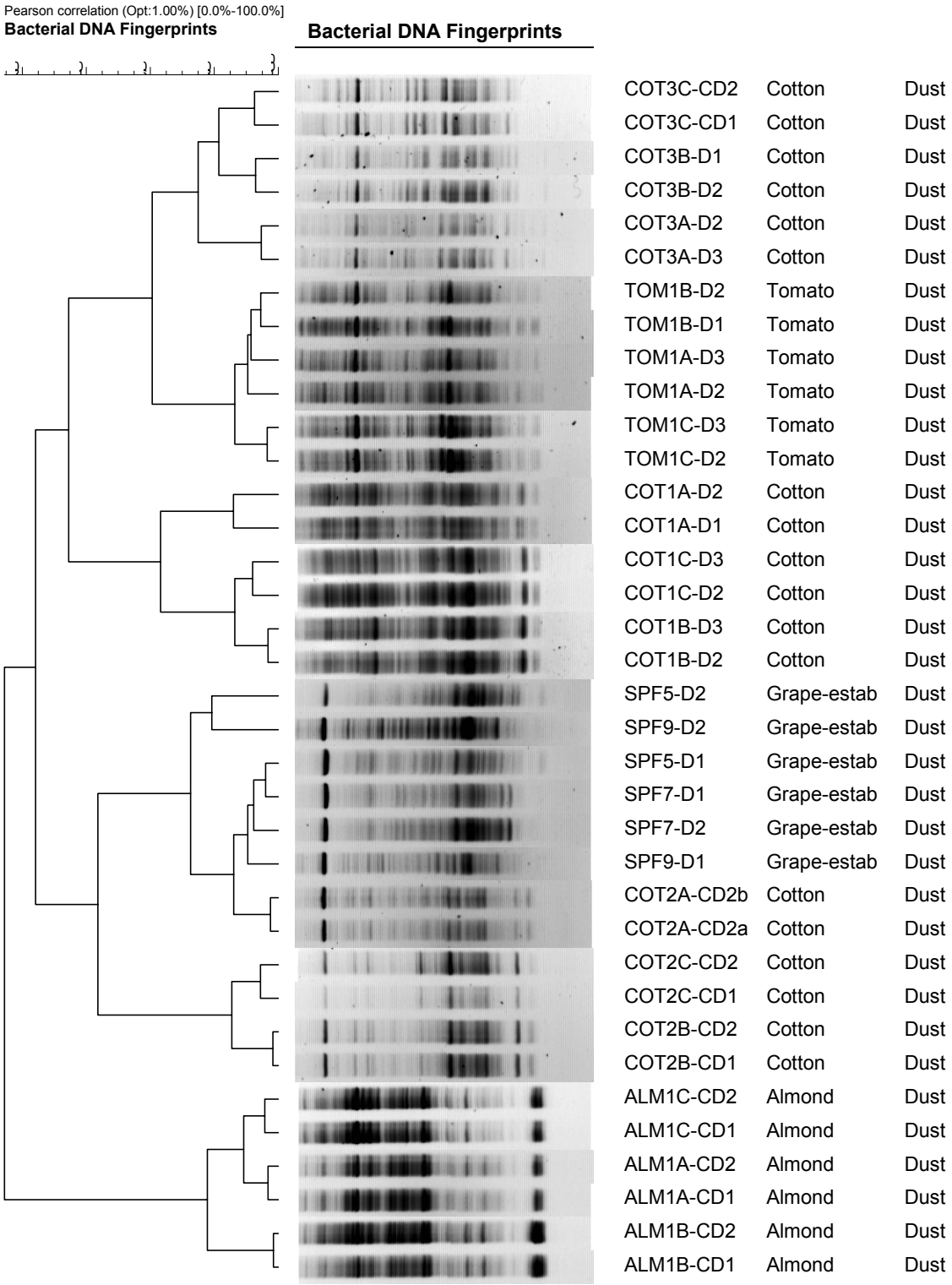


Figure 7. Reproducibility on the field scale: dust. Bacterial DNA fingerprints of field replicated laboratory-generated dust from TSS-12 samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

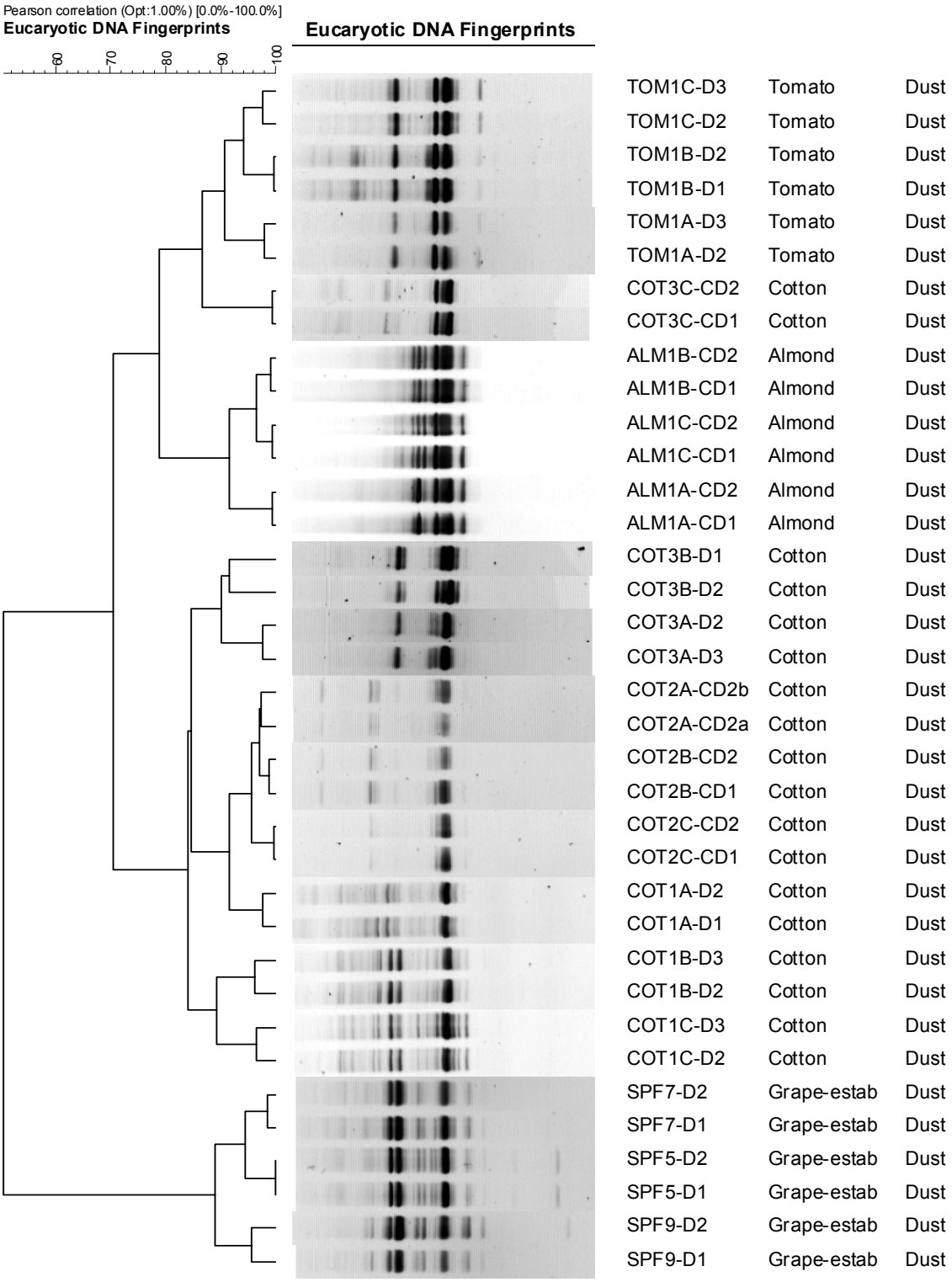


Figure 8. Reproducibility on the field scale: dust. Eucaryotic DNA fingerprints of field replicated laboratory-generated dust from TSS-12 samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

3.2.2.3 Differentiation of sources

The purpose of this analysis was to compare TSS-12 SJV source samples in two major categories of source samples, crop and non-crop (as defined in Materials and Methods).

TSS-12 crop samples

A preliminary inspection of bacterial versus eucaryotic DNA fingerprints for crop samples (Figures 9 and 10) reveals that the bacterial DNA fingerprints are more complex (contain 25-30 bands) than eucaryotic fingerprints, which contain 8-15 bands. The highest degree of similarity was observed for PCR replicates, (90-99% except for bacterial COT1B and TOM1B). Samples that comprise field replicates, (ALM1A-C, COT1A-C, COT2A-C, COT3A-C, TOM1A-C), grouped at the next level with similarities ranging from 78% (COT3A-C) to 92% (ALM1A-C) for bacterial DNA fingerprints (Figure 9). In contrast, eucaryotic fingerprints for field replicates COT1A-C and TOM1A-C bore little resemblance by cluster analysis (Figure 10). In general, eucaryotic DNA fingerprints were less consistent across a single field than the bacterial patterns. The next level of groupings, at 50-80% similarity, produced four major groups in the bacterial DNA fingerprints. The groups are inclusive for samples from the same crop (ALM1, 3 and GRA1, 3, and all TOM samples) or from adjacent locations within several km (e.g., ALM2-UPR1 and SAF1-COT6).

Almond and tomato bacterial DNA fingerprints were most similar among the crop sources. Field replicates of almonds (ALM1A-C) were 92% similar. Two almond soil samples from other areas (ALM3 and ALM4) had slightly different fingerprints, but contained enough bands in common with ALM1 to form a group with ALM1 (50% similarity). This feature was noted in the eucaryotic fingerprints as well. In contrast, ALM2 showed a very different pattern from the other almond samples, and grouped most closely with UPR1 (51%), an unpaved agricultural road sample collected in the vicinity (within several kilometers) of ALM2. This feature was not noted in the eucaryotic fingerprints. Bacterial DNA fingerprints of replicated field samples of tomato (TOM1A-C) bacterial fingerprints grouped together, and TOM2 and TOM3 shared 70% similarity with TOM1. Conversely, tomato samples (even field replicates) did not group based on eucaryotic DNA fingerprinting.

Twelve of the 32 crop samples were collected from six fields of different textures that supported cotton. While all field replicated cotton samples grouped by bacterial DNA fingerprinting, COT1 field replicates were split in the eucaryotic diagram. Cotton samples as a whole did not group separately from other samples by either fingerprint type. The remaining samples were grape, tomato, and safflower soils. Of the three grape samples, each of which was collected from different locations, GRA1 and GRA3 grouped together, distinct from GRA2 in the bacterial fingerprints. Grape samples did not group by eucaryotic DNA fingerprints. Two samples from safflower fields did not group with either fingerprinting method. Bacterial patterns of SAF1 grouped with COT6 (72% similarity), which was collected from an immediately adjacent field. However, this result was not noted for SAF3 and COT4, also collected from adjacent fields.

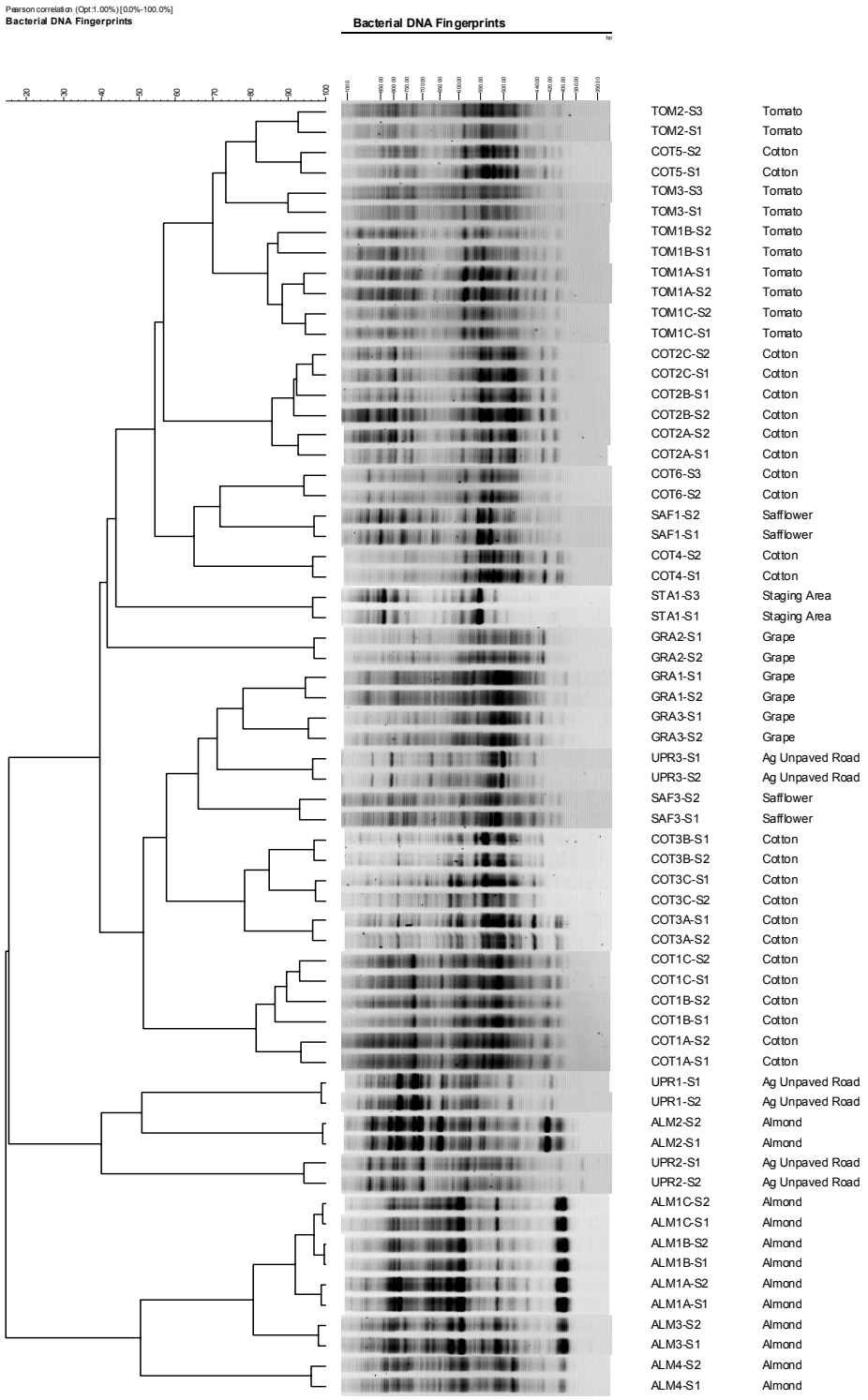


Figure 9. Cluster diagram of bacterial DNA fingerprints for crop samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

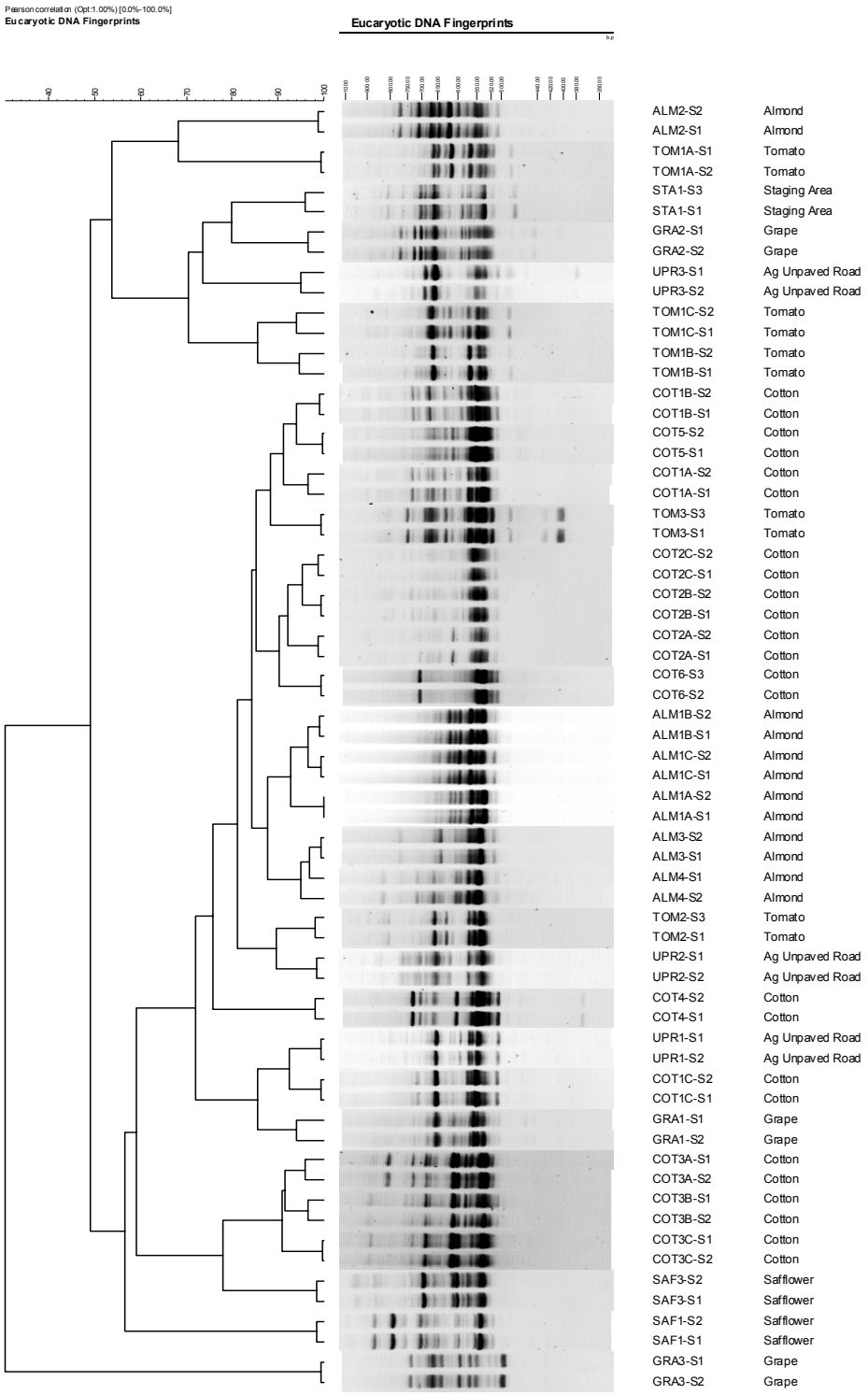


Figure 10. Cluster diagram of eucaryotic DNA fingerprints for crop samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

TSS-12 non-crop samples

The bacterial and eucaryotic DNA fingerprints for the non-crop samples, presented in Figures 11 and 12, are notable for their higher degree of individuality. There were no apparent groupings of the sources beyond those for PCR replicates, perhaps because these samples were not collected in close proximity. One notable feature for both fingerprint types is that samples with the lowest DNA contents (e.g., paved road and construction), displayed fingerprints on the same order of complexity as crop sample fingerprints. Although cattle feedlot sample CTF2 had a high DNA content (9 μg DNA g soil^{-1}), its bacterial fingerprint was lowest in complexity (fewest bands), and the corresponding eucaryotic fingerprint could not be generated from this sample with the standardized PCR conditions used in this study. Possible explanations for this result are that the sample held low numbers of microorganisms, or the DNA extract contained high salts or humic residues due to the high content of organic matter (manure), which in turn reduced the PCR efficiency.

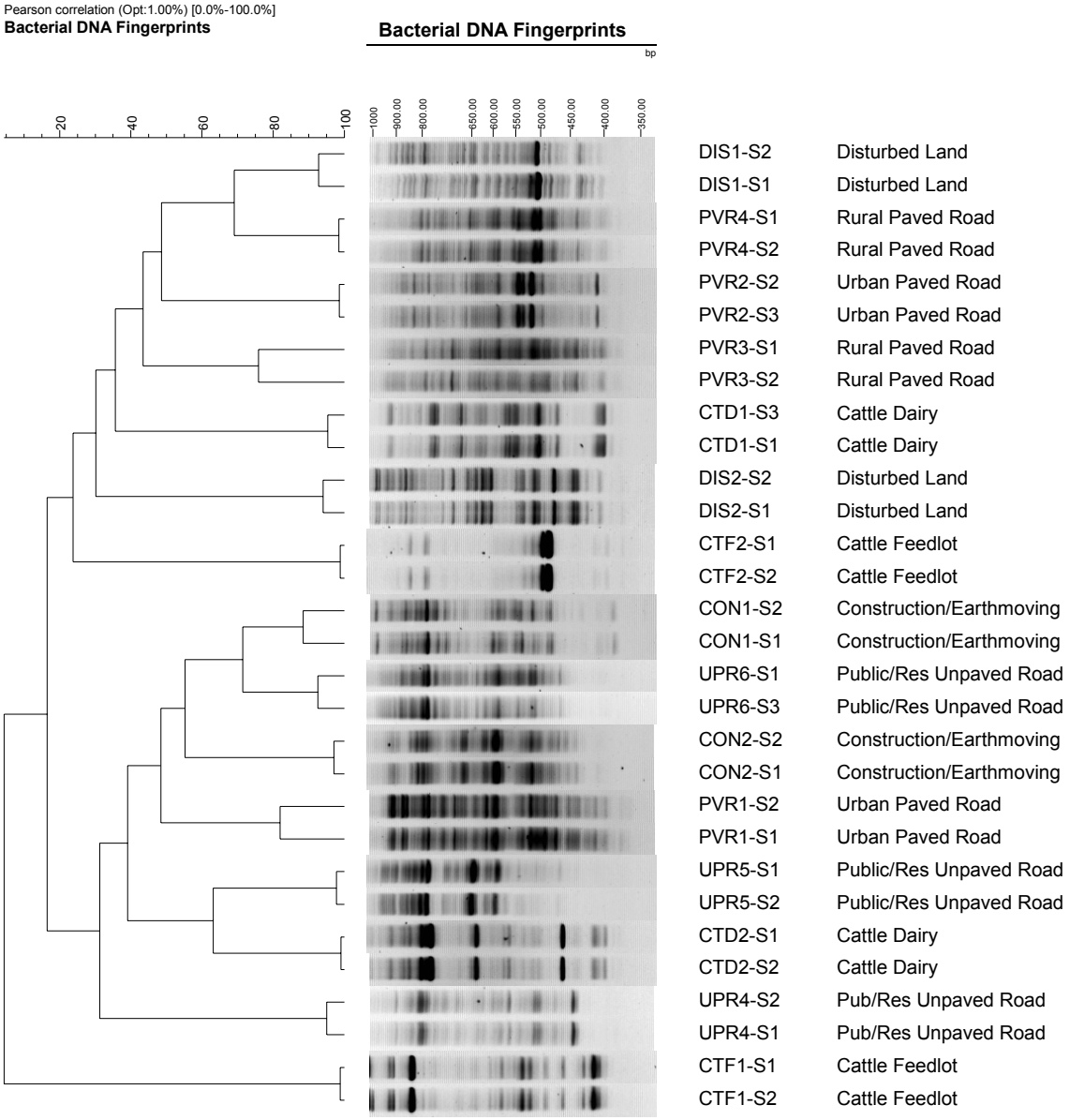


Figure 11. Cluster diagram of bacterial DNA fingerprints for non-crop samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

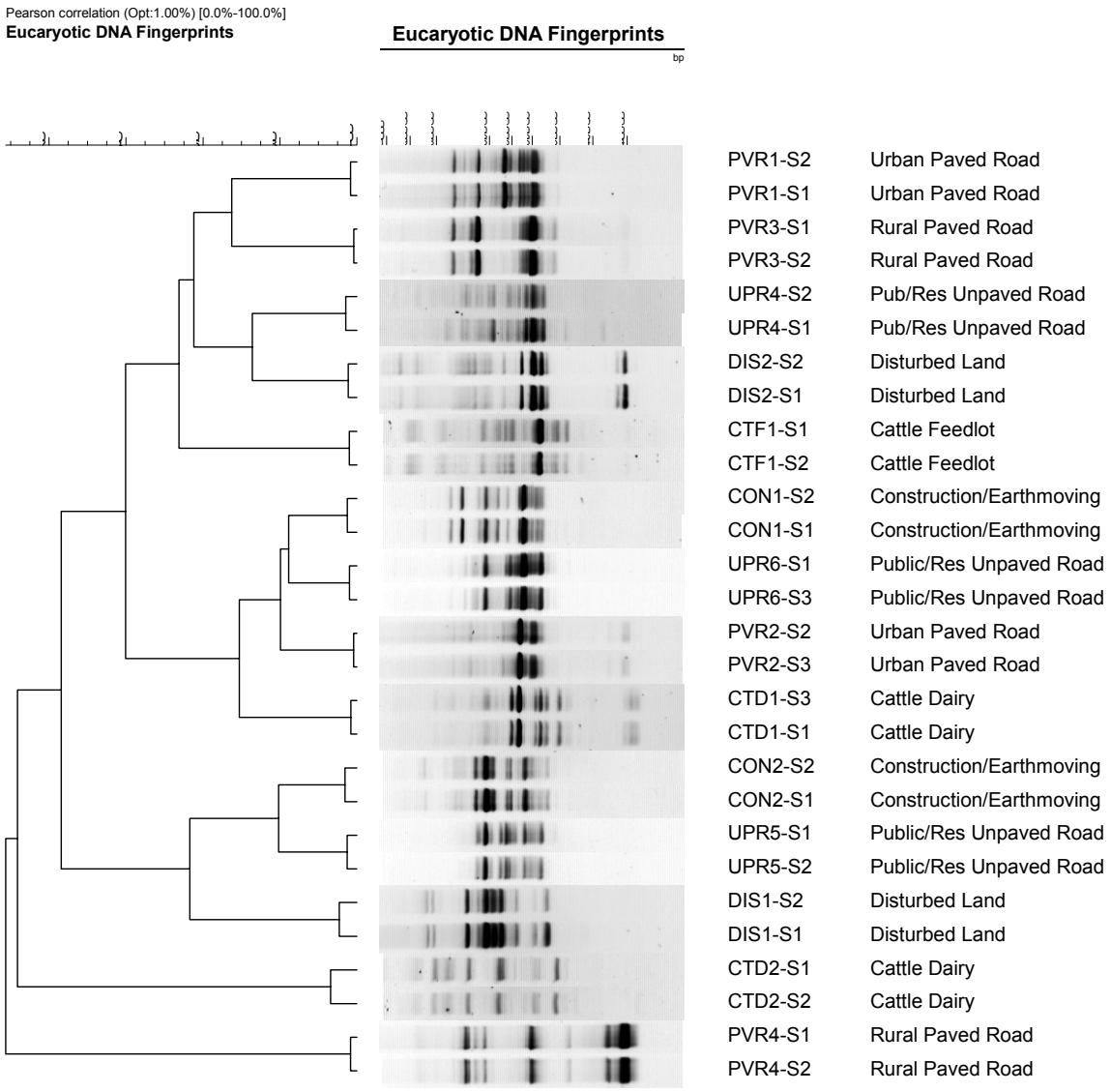


Figure 12. Cluster diagram of eucaryotic DNA fingerprints for non-crop samples. Profiles were analyzed by the Pearson product moment correlation and the un-weighted pair group method using arithmetic averages (UPGMA) algorithm.

Almond and tomato bacterial DNA fingerprints appeared to show the highest consistency among crop sources. There are two reasons why almond fingerprints tended to be highly similar. First, there was a visibly high leaf litter content in the almond samples; perhaps microbial populations from the almond leaf surface contributed to a distinctive signal for almond samples. A recent DNA-based study has shown that the leaf surfaces of citrus and annual crops including cotton and beans support unique and complex bacterial communities (Yang et al., 2001). Second, almond orchards are cultivated over a number of years in contrast to annual rotations of other crops. Perhaps the soil microbial communities associated with almond roots and decomposing leaf litter are more stable in almond orchards compared with those associated with annual crops rotated yearly. Two of three grape samples, the other perennial crop, also grouped. These samples were also collected close to one another, so the influence of local soil properties could not be ruled out.

Another example where different land uses from the same location grouped together was in samples collected from the northeast part of the SJV. In this case, construction site samples were closely associated with grape samples by DNA fingerprinting.

Tomato sample DNA fingerprints were more similar to one another than was the case for other annual crops sampled. However, all tomato samples were taken from the same general area (West Fresno County), so the influence of location-specific factors may be important, and could not be confirmed in this study. As a source category, cotton samples did not group, except by field. It appears that in this case, the influence of crop is subordinate to other undetermined factors. Perhaps the defoliants applied to cotton vegetation prior to harvest differentially disrupt soil microbial communities typically associated with cotton (presumably in different ways in different soils), masking potential similarities in the cotton DNA fingerprints. Alternately, there may be location-specific influences on microbial community composition, which could not be evaluated for this study because cotton samples were dispersed throughout the valley.

Associations between non-crop sources were not revealed, probably because they were collected from single points across very broad areas. The DNA fingerprints obtained for these samples were highly individual, except for the previously mentioned construction samples. In some respects, this result is not surprising given that the microbes detected on a paved road, for example, may not specifically or universally inhabit a paved road. Rather, they could be tracked on to the paved road surface from a nearby exposed soil, or originate from aerial deposition. Thus, while DNA fingerprints of paved road samples in this study share few common elements, they may be representative of a particular location, which could be investigated by sampling a number of sites along a road, including adjoining unpaved shoulders, which have also been demonstrated as a significant emission source for PM_{10} (Moosmuller et al., 1998). It may possible to specifically assay roadway samples for bacteria that degrade petroleum products.

While the DNA fingerprints of unpaved road sources were easily detected, those of the corresponding dust were weak or undetectable. Chemicals linked to dust suppression, detected in several of the unpaved road samples, could hamper the extraction or detection of microbial DNA in laboratory-generated dust. This observation should be evaluated further as it has

implications for the successful detection of microbial DNA from roadways treated with dust suppression chemicals.

3.2.2.4 Comparison of laboratory-generated dust to source samples: San Joaquin Valley soils

We compared source samples to dust samples generated in the small-scale laboratory chamber. The approach was advantageous because we could ensure that the dust was derived from a particular source sample. This step is a necessary prelude to the analysis of field dust, which consists of mixtures of sources. We tested the hypothesis that fingerprints of DNA extracted from dusts suspended under laboratory conditions were similar to those of the source samples. An assumption of this aspect of the work is that a representative fraction of microorganisms present in soil become entrained during dust generation, are captured on the quartz filters, and are detected in the DNA fingerprints of dust.

Dust was generated and collected for each of the TSS-12 samples, and paired source-dust DNA fingerprints were compared based on the number of common bands they shared. This procedure was completed for both bacterial and eucaryotic DNA fingerprints to assess their relative degrees of utility for characterizing the biological material in both source samples and dust. Almond soil and dust fingerprints are presented in Figure 13. Almond DNA fingerprints were very similar in most soil-dust pair-wise comparisons (greater than 80%) for both fingerprint types. ALM4 samples were less similar, 73% and 67% for bacterial and eucaryotic DNA fingerprints, respectively. Results for the remainder of the source-dust comparisons (cotton, grape-safflower, tomato, cattle feedlot-dairy, construction-disturbed land, and roadway samples) are depicted in Figures A1-A6 within Appendix 10.6.

Generally, bacterial DNA fingerprints appear to hold greater potential for characterizing fugitive dust and its sources for several reasons. First, there are more DNA bands in bacterial DNA fingerprints, hence a greater number of data points available for analysis. Second, across the range of TSS-12 samples, bacterial DNA was more consistently detected than eucaryotic DNA, for which several samples could not be fingerprinted. Third, the overall degree of similarity in source-dust pair-wise comparisons was higher for the bacterial versus eucaryotic DNA fingerprints indicating that a more representative fraction of the bacterial community is entrained and detected in laboratory-generated dust.

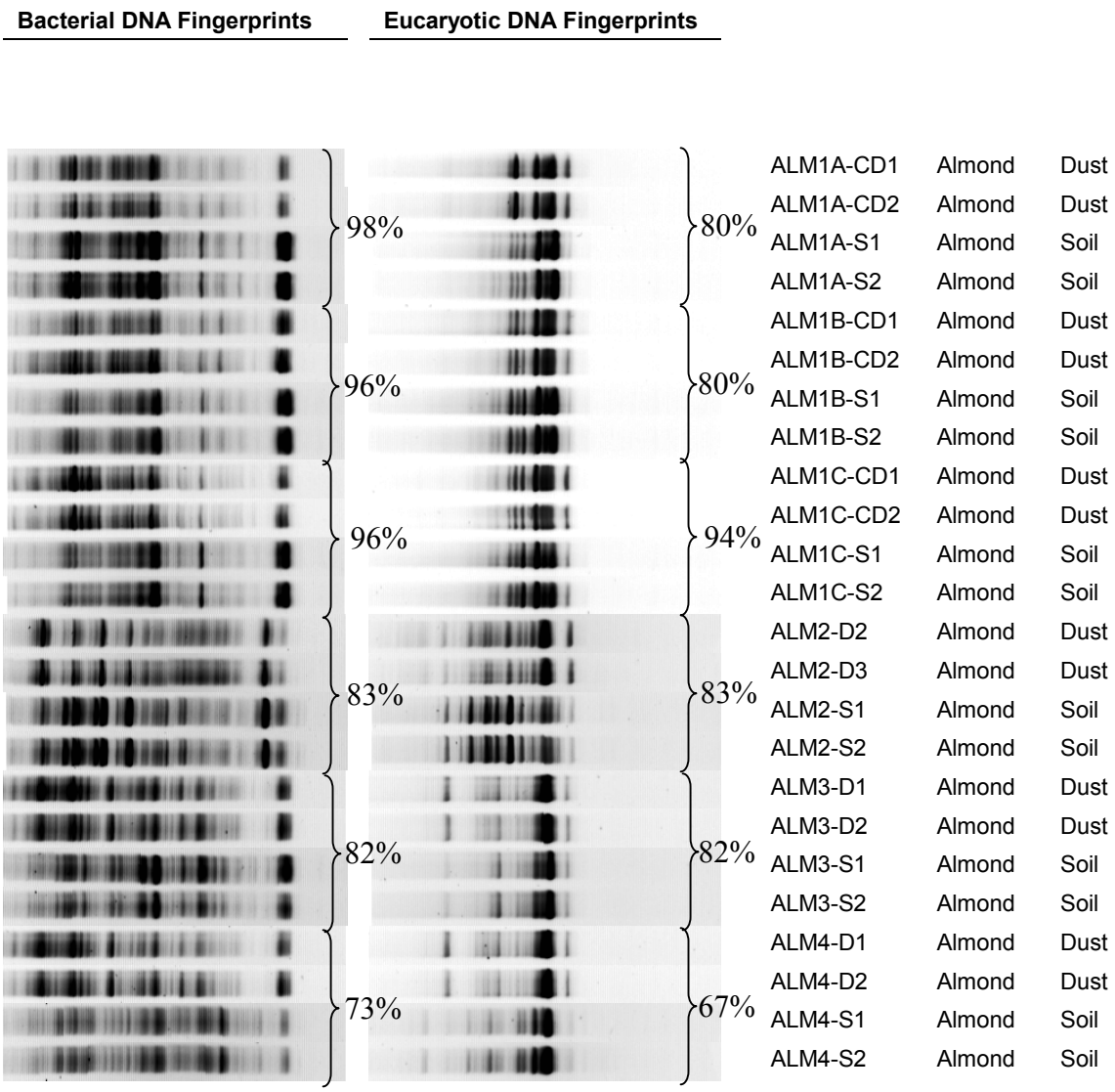


Figure 13. Bacterial and eucaryotic DNA fingerprints of soil and respirable dust from almond samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample.

3.2.2.5 Comparison of field-generated dust to source samples: SAFS soils

Source and dust samples were collected from organically and conventionally managed agricultural plots at the SAFS field site at U. C. Davis. In this case we wanted to compare DNA fingerprints of source samples to those of dust collected in the field. Focusing on a particular field event, a cultivation operation, reduced the number of potential sources and made it possible to directly compare field-generated dust to source material.

Similarities for SAFS corn cultivation soil and field-collected respirable dust pairwise comparisons ranged from 71% to 83% for bacterial DNA fingerprints and ranged from 59-71% for eucaryotic DNA fingerprints (Figure 14). In addition, bacterial DNA fingerprints across samples for all management regimes shared a number of prominently featured bands. Thus, the influence of soil type on microbial community structure was greater than the influence of management regime. This result was confirmed a result reported in our previous contract in which PLFA analysis of source material indicated that influences of management on microbial communities were smaller than seasonal or location differences.

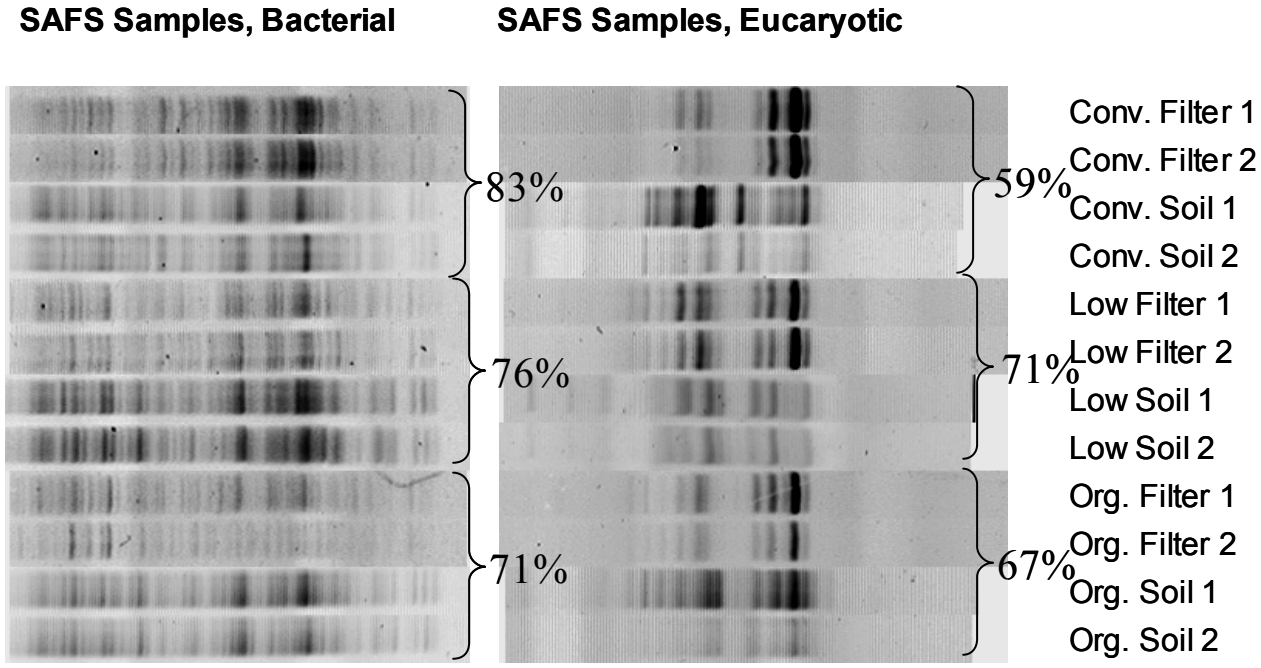


Figure 14. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from samples collected during a corn cultivation operation at the Sustainable Agriculture Farming Systems Project. Conv. = Conventional, Low = Low-input, and Org = Organic management regimes.

3.2.3 Differentiation of dust samples and potential biomarkers

In this analysis, DNA fingerprints of the entire set TSS-12 dust samples were compared by whole-pattern analysis using the Pearson product moment correlation-UPGMA (shown for bacterial DNA fingerprints in Figure 15, and eucaryotic DNA fingerprints in Figure 16). Again, the dendogram forms groups of samples with similar DNA fingerprints, with the highest degrees of similarity depicted by tight groupings on the right. Specific portions of interest within the figure are highlighted and discussed below.

In the bacterial dendogram (Figure 15), tomato dust samples (TOM1-3) were highly similar (74%), and grouped separately from the other samples (highlighted area 1). Two grape dust fingerprints (GRA1 and GRA3) grouped together in highlighted area 2 whereas the other grape sample (GRA2) did not. GRA1 and GRA3 sampling locations were located fairly close together (Figure 1). One safflower (SAF3) and one cotton sample (COT4), collected from adjacent fields, grouped together (highlighted area 3). Highlighted areas 4a, b, and c designate the field-replicated cotton samples, which group, except for COT2A. In contrast to the tomato fingerprints, the cotton DNA fingerprints did not form distinct groupings beyond field replicates. The fifth highlighted area includes the majority of almond samples, which form a distinct group (65% similarity). One almond sample (ALM2) was grouped with one of the unpaved road samples (UPR1); the samples were collected from within several km. Sample groupings of DNA fingerprints from laboratory-generated dust were more likely to be associated with crop type, field replicate, or location, rather than source category.

For eucaryotic DNA fingerprints (Figure 16), some of the same groupings occur for field-replicated samples, including COT2A, which did not group by bacterial DNA fingerprints. Otherwise there are few other groupings of note, perhaps because of the limited amount of band data generated in eucaryotic DNA fingerprints.

Another way of reviewing the dendogram is by looking at broader groupings, denoted by the highlighted areas surrounding portions DNA fingerprint image (Figure 15). Using this strategy, the entire set of fingerprints may also be divided into 3 major groups, which prominently display either band 1 (B1), band 2 (B2), or band 3 (B3). A fourth group could consist of assorted fingerprints that lack any of the three bands. The group of samples that prominently displays B1 includes annual crop (cotton, tomato, safflower), and grape DNA fingerprints. The inclusion of both construction samples appears anomalous, but these were collected in the same geographic region as the grape samples. The fingerprints with B2 consist of almond samples except ALM2. The group sharing B3 generally consists of roadway DNA fingerprints, with the exception of ALM2 and COT2. The undesignated samples do not prominently display any of these bands. The bands that separate TSS-12 samples into four groups represent potential biomarkers.

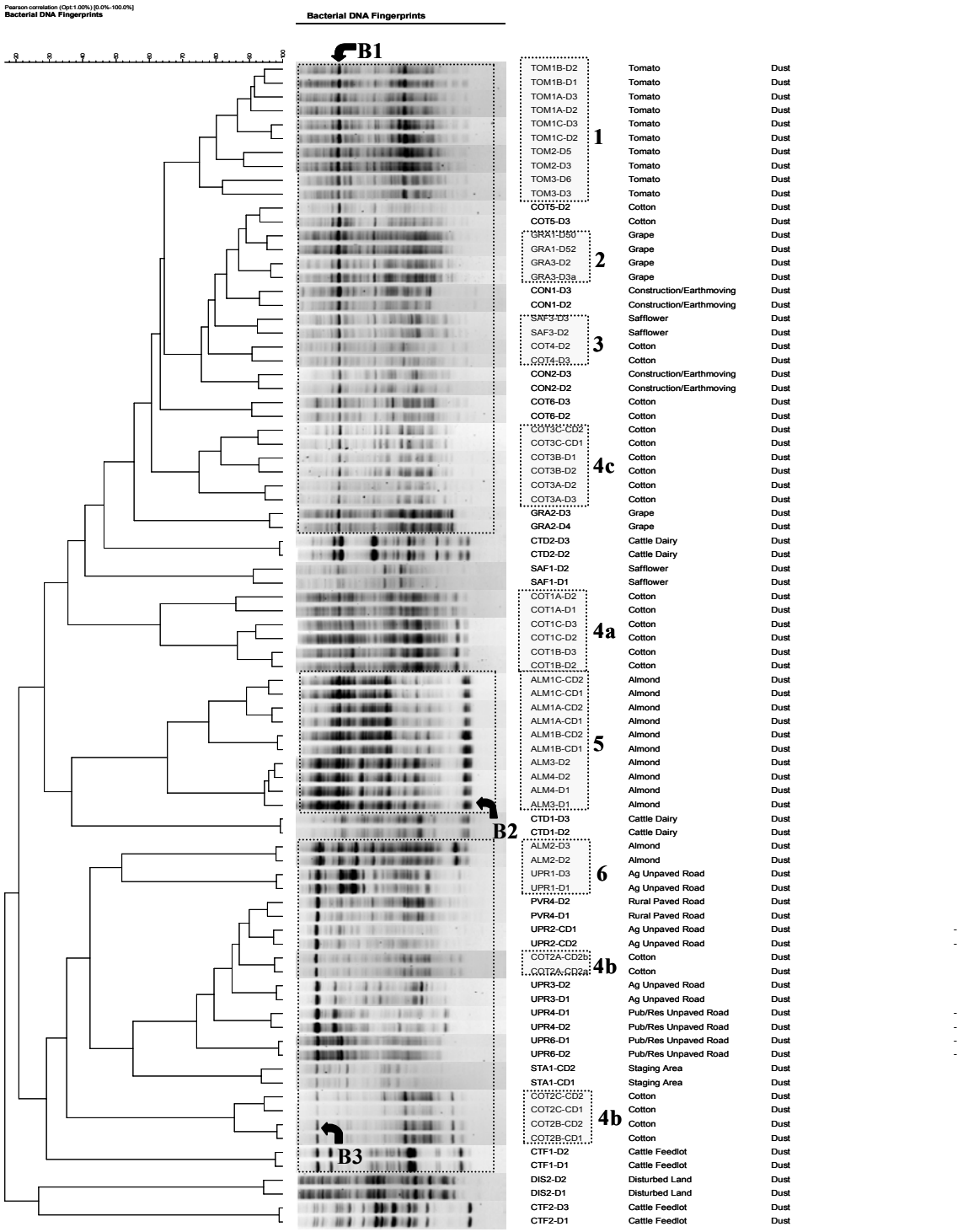


Figure 15. Bacterial DNA fingerprints of respirable dust from all source samples. Please refer to text for a detailed explanation of numerically-designated shaded boxes of interest, and the symbols B1, B2, and B3.

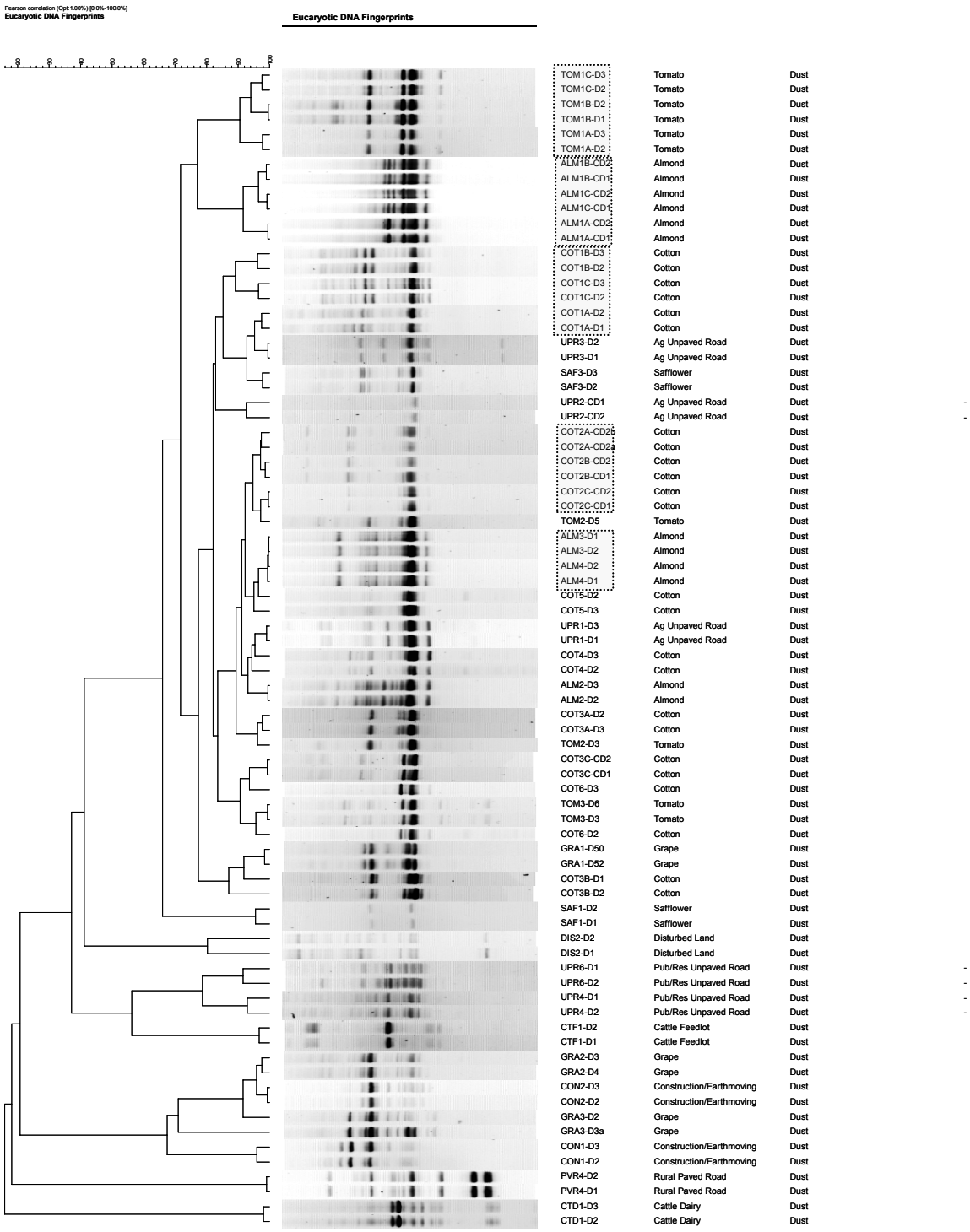


Figure 16. Eucaryotic DNA fingerprints of respirable dust from all source samples. Shaded boxes denote groupings of interest, please refer to text for details.

Bacterial DNA fingerprints of grape soil and laboratory-generated dust samples, collected from four locations in the SJV in two sampling events, also hold potential biomarkers for grape sources. They are designated as G1 and G2 in Figure 17. Potential biomarker G1, a fragment size of 902 bp, although not detected in several of the soil samples, is detected in every dust sample. Potential biomarker G2, a fragment of size 489 bp, is present in all soil and dust samples.

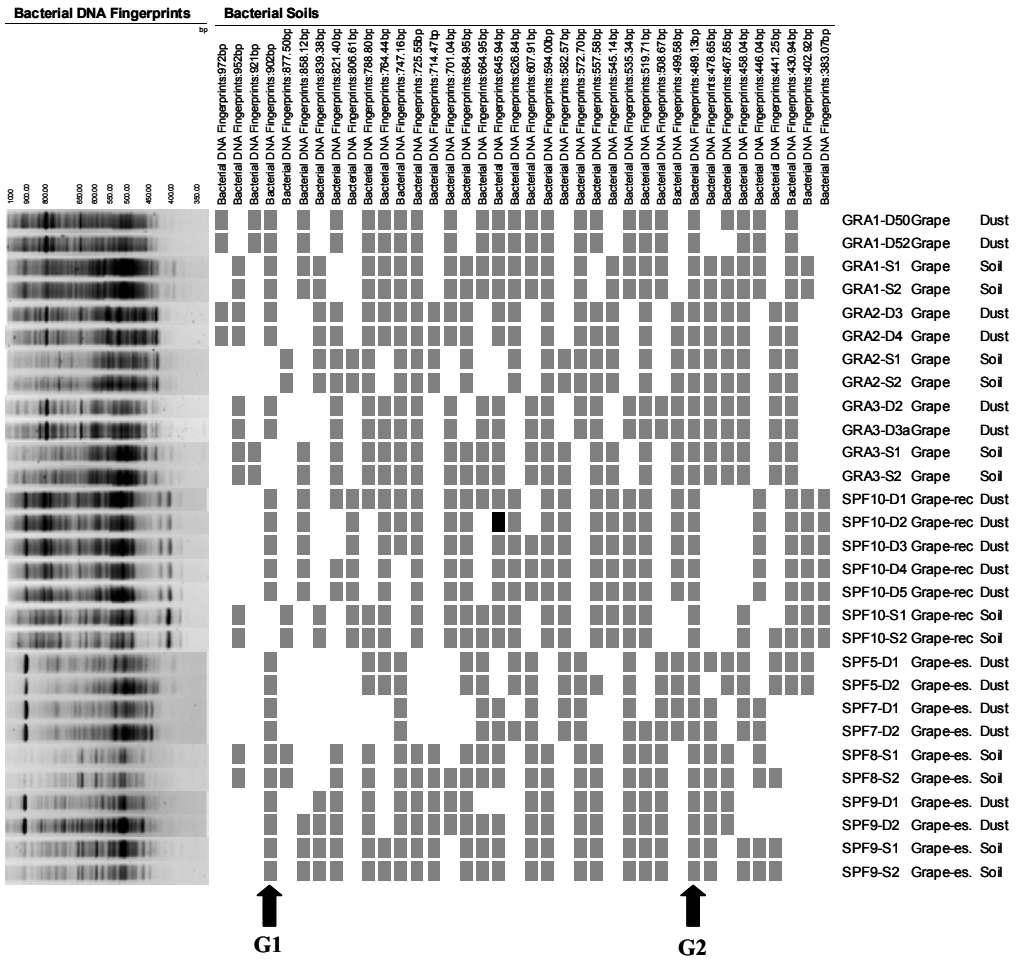


Figure 17. Bacterial DNA fingerprints of sources and respirable dust from all grape samples, and potential biomarkers G1 and G2.

Future research that involves the genetic sequencing of bands within DNA fingerprints should focus first on potential biomarkers. Thus, it will be possible to determine that in addition to being the same size (migrating to similar positions in the gel matrix), they also are of the same DNA sequence, the most specific and definitive biological characteristic of a source possible. The final requirement for the use of biomarkers is to confirm that this specific DNA sequence is absent in all other sources, a well-founded assumption, based on the breadth of microbial genetic diversity.

3.2.4 Detection limits of DNA fingerprinting

Using DNA fingerprinting it was possible to analyze a relatively small sample size, while maintaining a sufficient amount of data for classification analyses. As samples sizes are reduced, fewer microorganisms present in the sample. The biological fingerprint of the sample becomes less complex, contains less data, and usually becomes less distinguishable from that of another sample. Thus the lower limit on sample size is that which maintains a biological fingerprint complex enough to distinguish it from those of other samples. Although the absolute lower limits of detection were not tested, a general understanding of the relative amounts of filter dust required to produce a DNA fingerprint from myriad sources in SJV was gained. We did not determine the detection limit for source samples because larger quantities of samples are easily attainable and we would expect results from dust to also apply to sources.

DNA fingerprints were obtained for a majority of agricultural samples with an estimated 50-100 μg of laboratory-generated dust (Table 3). A majority of non-crop samples required an estimated of 100-200 μg of dust. The lowest amounts of dust (20 μg or less) were required for all almond and GRA3 samples. Higher amounts of dust (mg levels) were required for detection of STA1, UPR2, and UPR3. DNA fingerprints were not detected, despite concentration of extracts, for CTD2, DIS1, PVR3, and UPR5.

The source samples with higher extractable DNA contents tended to require the lowest masses of filter dust for the production of a DNA fingerprint (Table 2). However, sufficient data are not yet available for developing predictions about detection limits founded on DNA contents or organic matter contents of source material. The final report of our last contract (# 94-321) suggested that minimum sample sizes could be related to the amount of organic carbon in a source. In a broad sense this may be true. In particular, for crop samples, the amount of DNA extracted (Table 3) was positively ($r = 0.74$) and significantly ($p < 0.001$) correlated with soil organic carbon content, as found by Zhou, et al. (1996). Almond samples tended to have the highest DNA and organic carbon contents, and required the smallest amounts of laboratory-generated dust to produce a DNA fingerprint. However, that milligram quantities of laboratory-generated dust were required to produce DNA fingerprints from agricultural unpaved roads and the staging area cannot be solely explained by low DNA contents (construction site DNA contents were just as low, but microgram quantities of filter dust were sufficient for DNA fingerprinting). In summary, it may be difficult to develop generalizations that cover the gamut of source materials. Intensive sampling in additional research efforts may resolve these issues.

Table 3. The estimated amounts of laboratory-generated dust required to produce a DNA fingerprint. (Euc) = Eucaryotic DNA fingerprint not obtained (bacterial was).

20 μg or less	50-100 μg	100-200 μg	Milligrams (mg)
Almond 1	Cotton 1	Cotton 2	Staging Area
Almond 2	Cotton 3	Safflower 1	(6.63)
Almond 3	Cotton 4	Construction 1	Ag. Unpaved Rd. 2
Almond 4	Cotton 5	Construction 2	(1.67)
Grape 3	Cotton 6	Disturbed Site 2	Ag. Unpaved Rd. 3
	Grape 1	Paved Road 4	(2.60)
	Grape 2	Ag. Unpaved Rd. 1	
	Safflower 3	Pub/Res Unpaved Rd. 4	
	Tomato 1	Pub/Res Unpaved Rd. 6	
	Tomato 2		
	Tomato 3		
	Cattle Feedlot 1		
	Cattle Feedlot 2		
	Cattle Dairy 1		

DNA fingerprint was not obtained	Sample was not available for dust generation
Cattle Dairy 2 (Euc)	Safflower 2
Cattle Feedlot 2 (Euc)	Paved Road 1
Disturbed Site 1	Paved Road 2
Paved Road 3	
Pub/Res Unpaved Rd. 5	

3.3 Lipid-Based Approaches

We continued the work initiated in our previous contract on lipid-based approaches for characterizing microbial communities. Objectives of the current contract were to continue quantitative assessments of the detection limits for lipid-based methods and attempt to improve the sensitivity of the method. Our initial strategy was to improve the sensitivity by use of a gas chromatograph-mass spectrometer (GC-MS). However this method proved too labor intensive to be a cost-effective option for this application and was not considered further.

For quantitative assessments of detection limits, it was important to consider how well a lipid profile represents the breadth of fatty acids present in the sample. Specifically, as decreasing masses of source material are extracted, fatty acids of low original amounts are not detected in, nor quantified from, the gas chromatograph profile. Thus, in detection limit studies, the lipid profile from masses of source material used in standard practice (8 g for PLFA and 0.5 g for SFAME) defined a "reliable" profile. A reliable profile accurately represents the breadth of fatty acids present in a source sample. As lower and lower masses of sample are extracted, the profile becomes less reliable in terms of accurately representing the source sample. This factor eventually influences the outcome of multivariate statistical analyses. Before these detection limit studies, however, the lower limits of source material required to produce a reliable profile were not known. The following sections evaluate profile reliability for both lipid-based methods.

3.3.1 Analysis of source samples: Comparison of PLFA and SFAME methods

We compared PLFA and SFAME with respect to the types and amounts of data they provided and their ability to differentiate soil samples. As mentioned previously, the PLFA method extracts only fatty acids from the cell membranes of living organisms. In contrast, the SFAME method extracts fatty acids from whole cells and non-living biological material (and thus includes PLFAs as one subset of fatty acids). Two groups of soils were tested in the lipid-based analyses of source material. The first group (soil set 1) was composed of six soils, three from the Sustainable Agriculture Farming Systems (SAFS) project at the University of California, Davis (SAFS1, SAFS2, SAFS3), and three from cotton and grape agricultural fields surrounding Fresno (Table 4a). These soils were used in the detection limit study and differed with respect to crop type, soil texture, and geographic region. The second group (soil set 2) was composed of 55 soils obtained from the U. C. Davis Air Quality Group's USDA PM_{10} project. Soil set 2 consisted of a collection of air-dried soils from privately owned fields in the San Joaquin Valley (Table 4b**)**. These samples were used to validate the results from the detection limit study and to evaluate the potential influence of crop type and soil texture on lipid profile analyses. These samples were chosen to represent major crops and soil texture types within the valley (cotton, almond, walnut and fig).

Table 4. Descriptions of samples in soil set 1 and soil set 2: lipid analysis.

4a.
Summary of the details of the six soil sample types of soil set 1.

Sample name	Crop	Texture	Sampling location	Notes
Cotton	Cotton	Clay	17 miles E of Fresno, CA	Fallow field
Vin3	Grapes	Sandy loam	5 miles S of Fresno, CA	Established grapes
Vin10	Grapes	Sandy loam	14 miles SW of Fresno, CA	Recently planted grapes, field ripped (5').
SAFS1	Safflower	Silty loam	SAFS, U. C. Davis, CA	Conventional treatment
SAFS2	Fallow	Silty loam	SAFS, U. C. Davis, CA	Organic treatment
SAFS3	Safflower	Silty loam	SAFS, U. C. Davis, CA	Organic treatment

4b.
Summary of the details of the 55 soil samples of soil set 2.

Sample name	Number of samples	Crop	Texture	Site code(s)	Number of sites
ACL	7	Almond	Clay loam	PF6	1
AL	4	Almond	Loam	PF7/PF8	2
ASL	5	Almond	Sandy loam	PF2/PF3	2
CC	8	Cotton	Clay	BR3/BV1/BV2/CR1/NB1	5
CCL	2	Cotton	Clay loam	BR3	1
CL	3	Cotton	Loam	K61/K62	2
CSL	2	Cotton	Sandy loam	K61	1
CSiL	3	Cotton	Silty loam	KL3	1
WL	7	Walnut	Loam	WH1/FU1	2
WLS	1	Walnut	Loamy sand	FU1	1
WSL	2	Walnut	Sandy loam	WH1/FU1	2
FLS	2	Fig	Loamy sand	DE1/DE2	2
FSL	9	Fig	Sandy loam	DE2	1
Total	55	4 types	6 types	17 different sites	

3.3.1.1 Fatty acid yields and the mass of source material required for analysis: soil set 1

Different masses of each of the six soils of soil set 1 were extracted and analyzed by both methods. The amounts extracted ranged from masses near the GC detection limit to masses far in excess of those needed for reliable analysis (8 g to 100 mg for PLFA, 500 mg to 10 mg for SFAME). Two measures fatty acids were considered, the fatty acid yields in nanomoles (nM), and the numbers of peaks detected, both on a per gram dry weight basis. The mass yields of fatty acids extracted and identified per gram soil were at least seven times greater with the SFAME than PLFA method (Table 5) for each of the six samples tested. PLFA yield was lowest in the cotton sample, which also held the highest content by SFAME.

Table 5. Summary of the total amounts of fatty acids detected in each sample of soil set 1, and the lower limits of the mass required to extract reliable profiles. Values shown are means from three laboratory replicates.

Soil type	Type of lipid analysis	Fatty acid content[a] in 1g dry weight soil (nM)	Lower weight limit[b] for analysis
Cotton	PLFA	8	>5 g
	SFAME	217	300 mg
Vin3	PLFA	17	1 g
	SFAME	169	200 mg
Vin10	PLFA	28	1 g
	SFAME	196	200 mg
SAFS1	PLFA	28	1 g
	SFAME	206	100 mg
SAFS2	PLFA	53	2 g
	SFAME	307	100 mg
SAFS3	PLFA	29	1 g
	SFAME	215	100 mg

[a]Total nM from all peaks detected and identified using Midi-GC system.
[b]Determined by projecting outliers on ordination diagrams after correspondence analysis (CA).

The lower weight limits are the amounts of source material required to produce a reliable lipid profile, as determined by collectively analysing fatty acid data from the extraction of different masses of soil. As previously described, lipid profiles from masses of soil extracted under standard procedures (8 g for PLFA and 0.5 g for SFAME) constitute an accurate biological characterization of the source. With correspondence analysis, the samples that group with the standard mass samples are also considered to accurately characterize a source. In general, samples with higher fatty acid contents required less sample mass to generate reliable profiles by either method (Table 5). Figure 18 illustrates how these estimates were determined for Vin3 soil. The amount of Vin3 soil required for a reliable profile by PLFA is 1 g (Pe), and by SFAME is 200 mg (Fd), as indicated by their position in groups with high mass samples on the right side of the vertical axis.

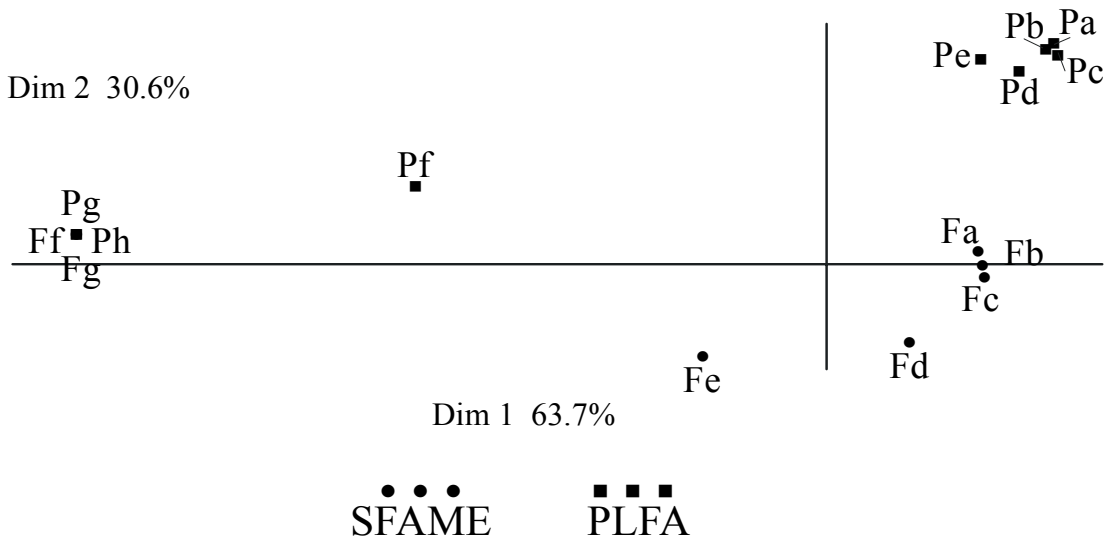


Figure 18. Correspondence analysis of Vin3 soil using either PLFA or SFAME lipid extraction on varying amounts of soil. PLFA extraction soil amounts (dry weight): Pa=8g; Pb=5g; Pc=3g; Pd=2g; Pe=1g; Pf=0.5g; Pg=0.3g; Ph=0.1g. SFAME extraction soil amounts (dry weight): Fa=0.5g; Fb=0.3g; Fc=0.2g; Fd=0.1g; Fe=0.05g; Ff=0.025g; Fg=0.01g. Lipids used in analysis were present in at least 50% of all samples.

After evaluating the extraction yields of soil set 1 individually, the extraction yields were averaged across the sample set, as were the number of peaks. The mean number of peaks identified by the MIDI software from the gas chromatographs (corresponding to individual fatty acids) was approximately the same for PLFA and SFAME methods although smaller masses of soil are extracted for SFAME. The mean number of peaks identified in the two highest masses extracted by the PLFA method (8 g and 5 g) was 34.8 (+/- 4.9). The mean number of peaks identified after SFAME extraction (0.5 g and 0.3 g soil) was 35.3 (+/- 8.7). Both methods generate large data sets suitable for statistical and classification analyses, provided the appropriate amounts of sample are extracted. The mean fatty acid yield for all six samples in soil set 1 differed greatly by method (Table 6). The mean lower weight limits (the amount of soil required to generate a reliable profile) averaged 1.32 g for PLFA and 0.18 g for SFAME among soil set 1 samples.

Table 6. Summary of the mean total amounts of fatty acids detected across all samples in soil set 1, and the mean lower weight limits for analysis. Standard amounts of source material used for each method are included for reference.

Lipid analysis	Mean fatty acid content[a] in 1g dry weight soil (CV[b])	Mean lower weight limits for analysis		Standard amounts of source material for analysis	
		Dry wt.	nM (CV)	Dry wt.	nM[a]
PLFA[c]	31 nM (38.1%)	1.32 g	41.0 nM (89.5%)	5.0 g	155 nM
SFAME	219 nM (19.5%)	0.18 g	40.0 nM (40.3%)	0.5 g	110 nM

[a] Total nM from all peaks detected and identified using Midi-GC system.
[b] Mean coefficient of variation from all samples and replicates in soil set 1.
[c] 'Cotton' soil type values not used in PLFA calculations.

3.3.1.2 Variability

The SFAME method yields more fatty acids, and more variable data than does the PLFA method. An illustration is provided by the PLFA- versus SFAME-extracted replicate sample points for soil set 1 (Figure 19). For example, SFAME-extracted SAFS2 samples (in the upper left quadrant of the diagram) are widely spread compared with the PLFA-extracted SAFS2 samples (depicted in the upper right quadrant). This result is consistent across all samples with the exception of the Vin3 sample, in which both extraction methods showed low variability. Overall, PLFA extraction produced slightly more consistent fatty acid profiles, indicating that fatty acids are more consistently extracted from the living fraction of soil organic matter (cell membranes from viable microbes) than with SFAME (fatty acids from whole cells and non-living biological material). Extractions using the PLFA technique can however take four or five times longer to perform than SFAME extractions, and require up to ten times more sample on a dry weight basis.

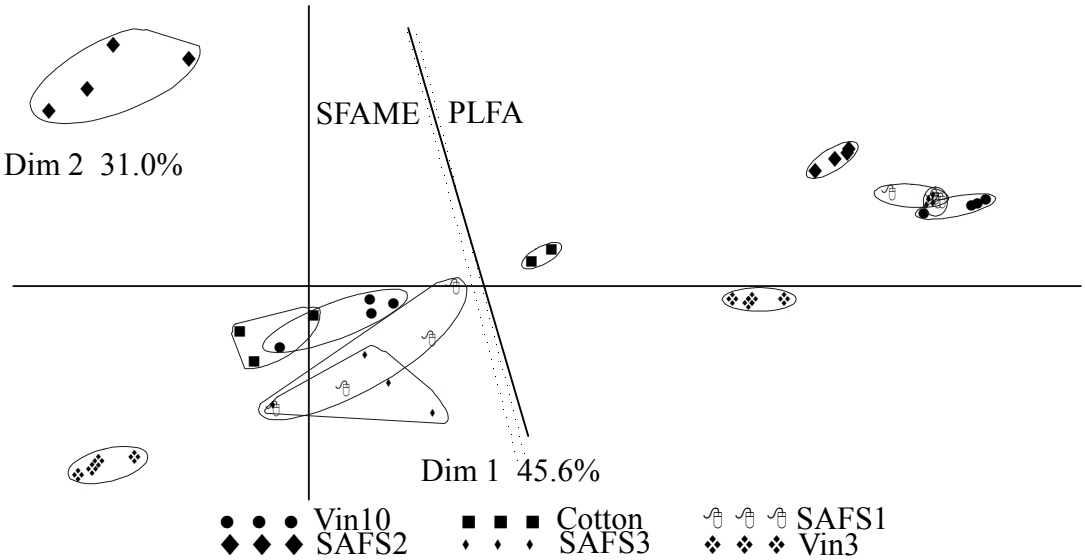


Figure 19. Correspondence analysis of six samples in soil set 1, each with two to four replicates, using either PLFA or SFAME lipid extraction. Lipids used in analysis were present in all samples.

3.3.2 Comparison of laboratory-generated dust to source samples: Comparison of PLFA and SFAME methods

Dust was generated from the six agricultural soils described above. In the large-scale chamber, PM_{10} was collected on quartz filters in a Hi-Vol sampler. Filter samples were PLFA- or SFAME-extracted, and lipid profiles were evaluated in terms of reproducibility and detection limits. Lipid profiles of PM_{10} samples were also compared to those of their parent soils (soil set 1). The cotton sample could not included in PLFA analysis, due to the very limited amount of dust generated. The fatty acid content in PM_{10} samples was generally higher than the fatty acid content in source samples on a dry weight basis (Table 7). This result is attributed to the presence of sand particles (biologically inert) in the source samples compared with entrained material, which includes silt, clay, and associated microorganisms (biologically rich). In addition, the amounts of sample required for the reliable generation of profiles were lower for PM_{10} versus source samples. The lower weight limits are the amounts of source material required to produce a reliable lipid profile, as determined by collectively analysing fatty acid data from the extraction of different masses of soil or PM_{10} (refer to section 3.3.1 for details), are also shown in Table 7.

Table 7. Comparison of the total fatty acid content and the lower weight limit for analysis of source material or PM_{10} from soil set 1.

Soil type	Type of lipid analysis	Fatty acid content[a] in 1 g dry weight soil (nM)	Fatty acid content[b] in 1 g dry weight PM_{10} (nM)	Lower weight limit for analysis: Source	Lower weight limit for analysis: PM_{10}
Cotton	PLFA	8	NA[b]	>5 g	NA
	SFAME	217	740	300 mg	50 mg
Vin3	PLFA	17	94	1 g	NA
	SFAME	169	563	200 mg	100 mg
Vin10	PLFA	28	75	1 g	NA
	SFAME	196	483	200 mg	50 mg
SAFS1	PLFA	28	29	1 g	750 mg
	SFAME	206	312	100 mg	115 mg
SAFS2	PLFA	53	68	2 g	750 mg
	SFAME	307	899	100 mg	50 mg
SAFS3	PLFA	29	53	1 g	750 mg
	SFAME	215	1486	100 mg	100 mg

[a] Total nM from all peaks detected and identified using Midi-GC system.
[b] Not Available, insufficient sample for analysis.

3.3.3 Detection limits of PLFA and SFAME

After evaluating the extraction yields of PM_{10} from soil set 1 individually, the extraction yields were averaged across the sample set. The mean fatty acid yield for all six samples in soil set 1 was an order of magnitude greater by SFAME versus PLFA methods (Table 8).

Table 8. Summary of the mean total amounts of fatty acids detected across all PM_{10} samples generated from in soil set 1, and the mean lower weight limits for reliable analysis.

Lipid analysis	Mean fatty acid content[a] in 1g dry weight PM_{10} (CV[b])	Mean lower weight limits for analysis		Recommended mean mass of PM_{10} required for analysis	
		Dry wt.	nM (CV)	Dry wt.	nM[a]
PLFA[c]	64 nM (38.3%)	0.69 g	44 nM (39.7%)	0.75g	48 nM
SFAME	747 nM (55.6%)	0.04g	28 nM (48.1%)	0.10g	75 nM

[a] Total nanomoles from all peaks detected and identified using Midi-GC system.
[b] Mean coefficient of variation from all samples and replicates in soil set 1.
[c] 'Cotton' soil type values not used in PLFA calculations.

Mean lipid amounts required for reliable analysis for PLFA and SFAME were comparable with those of parent soils, indicating no loss in profile quality in dust analysis. Coefficients of variation for dust replicate extractions indicated slightly lowered reproducibility of dust profile than soil profile, but comparable levels between dust methods.

PM_{10} generated from the six soils used for detection limit analysis showed over a two-fold increase in lipid amounts yielded from dusts as compared to soils using PLFA extraction, and over a three-fold increase with respect to SFAME extraction. Profiles from PM_{10} also showed a less pronounced decline in reliability as amounts extracted were reduced than did soils. Consequently, recommended amounts for analysis were much lower than for soils, with PLFA extractions requiring 0.75 g of PM_{10}, and SFAME extractions requiring 0.1 g. Coefficients of variation for replicates of dust extraction for both methods were higher than those for soil, although differences between methods were reduced, so that the consistency of lipid extraction using SFAME approached that of PLFA. Relationships between profiles derived from each method showed differences, similar to those found in the six source soils. However, relationships between dusts and source soils were only apparent in two of five soils for PLFA, and on none for SFAME analysis. DNA analysis indicated relationships between samples for soil and dust to be similar to those found for PLFA soil and PLFA dust. This was not the case for relationships found by SFAME analysis of soil or dust.

3.3.4 Summary

In conclusion, the extraction limits for the 6 soils in sample set 1 ranged from 1g to 5g for PLFA analysis, and 0.1 g to 0.3 g for SFAME analysis. PLFA was slightly less variable than SFAME in its extraction of fatty acids across replicates. The low mean variation present in the PLFA-extracted samples is similar to that found in previous studies (Macalady et al., 2000; Saetre and Baath, 2000). An advantage of the SFAME method is that it requires smaller sample sizes than does PLFA analysis. The amount of fatty acid (nM) extractable per gram soil was approximately seven times higher in the SFAME than PLFA method. Regardless of extraction method however, the same concentrations and peak numbers of fatty acids were required by both methods to reliably differentiate soil samples.

3.4 Statistical Approaches for Classifying Source and Dust Samples

3.4.1 Principal Components and RedundancyAnalyses

The multivariate analytical techniques PCA and RDA were applied only to bacterial DNA fingerprint data for agricultural samples. Bacterial fingerprints for other land uses and eucaryotic fingerprints were not analyzed because of the inconsistencies among replicates shown by the Pearson correlation and cluster analysis. Bacterial DNA fingerprints of almond soils, by PCA, were distinctly different from those of other crops (Figure 20), and formed two distinct subgroups (ALM1A-C, ALM3, ALM4 versus ALM2 and UPR1). Groupings of samples were similar to the four major groups indicated by cluster analysis (Figure 9), these four groups are delineated in the ordination diagram by ellipses. The band class variables located in close proximity to particular samples or sample groups represent portions of the fingerprint pattern that provide distinguishing features for those samples. For example, band class variable 399 bp (the bottom band of ALM1A-C, ALM3 and ALM4) is present as a unique band of higher intensity in those samples. Other band class variables that provide distinguishing features include band class variables 648 bp (ALM2, UPR1), 423 bp (ALM2, UPR1, COT1), 884 bp (ALM2, UPR1, COT1, COT3), and 547 bp (present in nearly every sample but most prominently featured in every sample located in the lower left quadrant of the PCA diagram). The ordination first two axes of this diagram display a total of 34% of the variance in the band class data for all crop samples.

Redundancy analysis of bacterial DNA fingerprints in conjunction with specific soil property variables (Figure 21) depicts and ranks associations that may be intuitive, but are difficult to glean from a simple visual survey of values in the multivariate data set. The DNA fingerprints from almond samples, UPR1, and UPR2 are positively correlated with higher electrical conductivity (except ALM3) and sand content (except UPR2). Higher nitrogen, organic carbon, and DNA contents are positively correlated with ALM1, ALM3, and ALM4 fingerprints. Conversely, clay content is positively correlated with all of the tomato, safflower, and cotton fingerprints (except COT1 and COT3), and negatively correlated with almond fingerprints. Inorganic carbon, highest in COT6, SAF1, and UPR2, does not correlate with those samples in the RDA diagram, perhaps because the association is depicted on other than the first two axes. Specifically, the soil properties significantly associated with the fingerprint data, as tested by Monte Carlo permutation include electrical conductivity ($p<0.005$), DNA extracted ($p<0.005$), sand and clay content ($p<0.05$), inorganic carbon and nitrogen ($p<0.05$), but neither organic carbon nor pH. These results suggest that groupings based on bacterial DNA fingerprints reflect not only crop type, but also physical properties of the soils.

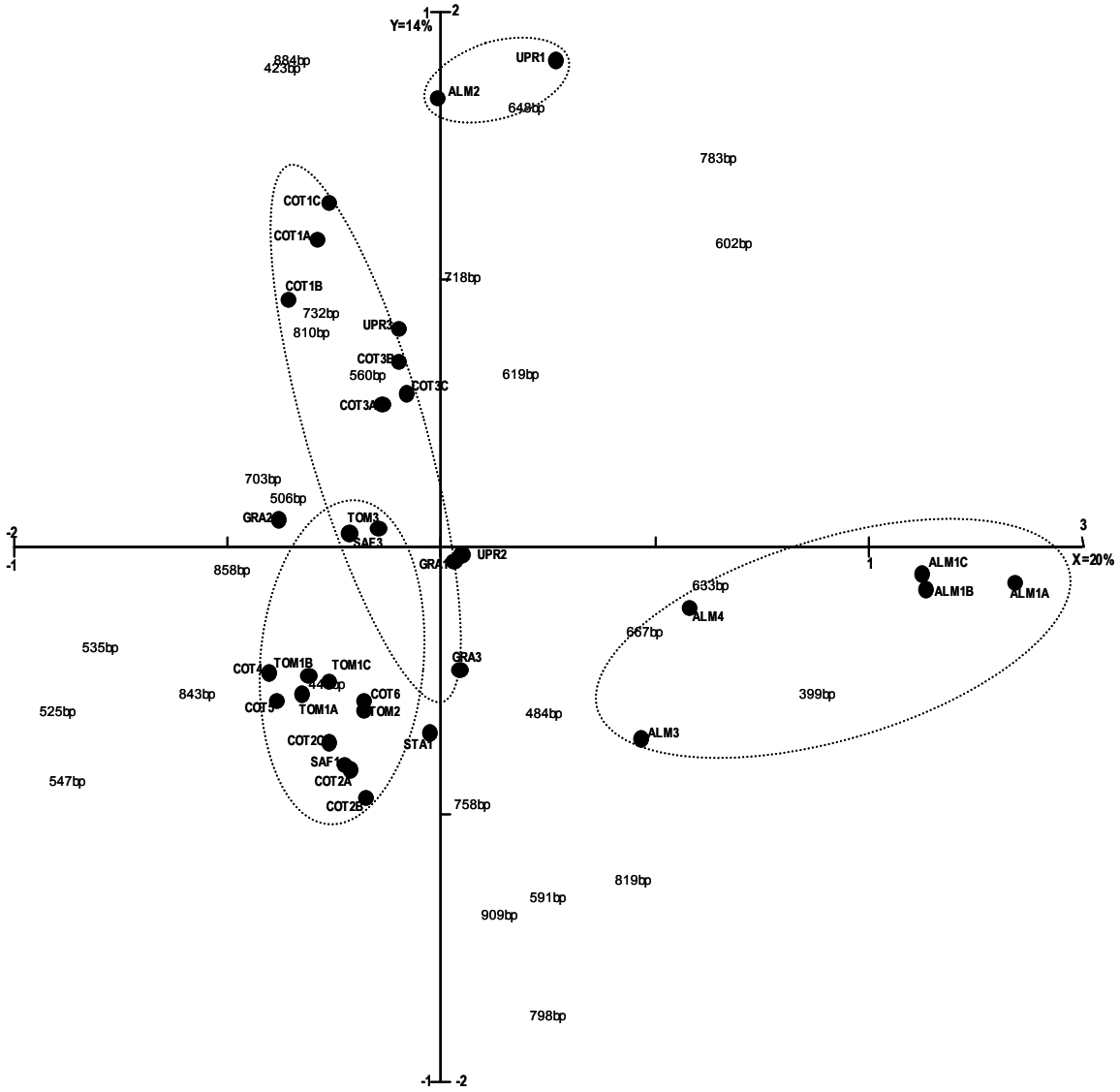


Figure 20. Principal components analysis of TSS-12 crop samples. Upper and right scales on the ordination axes relate to sample scores, lower and left scales relate to base pair variable scores. A total of 34% of the total variance in the data set is explained by the first two axes. Dotted ellipses outline the four major groups observed for bacterial DNA fingerprints by cluster analysis.

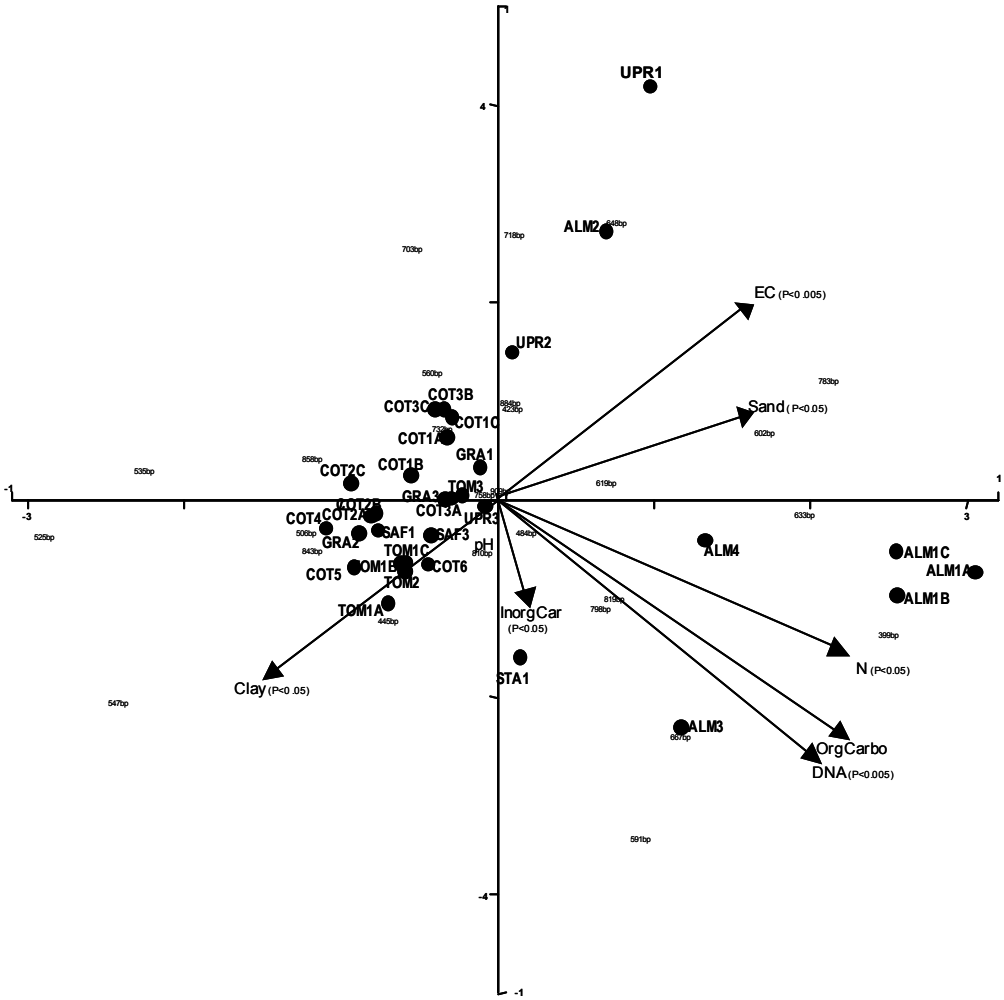


Figure 21. Redundancy analysis of TSS-12 crop samples. Upper and right scales on the ordination axes relate to sample scores, lower and left scales relate to base pair variable scores. A total of 30% of the total variance in the species data set is explained by the first two axes, the sum of all canonical eigenvalues is .39. Environmental variables are labeled with their significant p-values, if present.

3.4.2 Classification Models

DNA fingerprints from TSS-12 sources (Table 9) were classified with all 320 variables (pixel data from the whole DNA fingerprint), and after two variable selection methods (stepwise variable selection with partial least square (PLS) and genetic algorithm (GA) variable selection). The entire set of samples was split into a training set and a test set (Table 10). Two classification models, discriminant partial least square (D-PLS) and back propagation-artificial neural network (BP-ANN), were run and compared for their accuracy in classifying sources before variable section, and after variable selection by PLS or by GA. Both D-PLS and BP-ANN gave reasonably accurate classification results for the test set before any variable selection methods, with the BP-ANN (91%) being better than D-PLS (84%) (Table 11). The prediction performance of the test set improved for both methods after variable selection. Classification results after PLS variable selection increased to 89% (D-PLS) and 96% (BP-ANN) (Table 12). Classification results after GA variable selection were the highest, 95% (D-PLS) and 99% (BP-ANN) (Table 13). In summary, the classification results were most accurate with BP-ANN after the GA variable selection technique. With GA, a subset of 108 input variables from the original 320 provided a significant improvement of the predictive ability of the models.

Table 9. Number of samples of different source types used in DNA-based classification models.

Source Type	Abbreviation	Number of Samples
Cotton	COT	54
Almond	ALM	25
Tomato	TOM	36
Ag Unpaved Road	AUR	19
Construction/Earthmoving	CEM	14
Dairy	DIR	5
Dist Land Salt Buildup	DLS	6
Feedlot	FED	9
Grape	GRP	13
Pub/Res Unpaved Road	PRU	14
Rural Paved Road	RPR	9
Safflower	SAF	13
Urban Paved Road	UPR	6
Total		223

Table 10. Source types used in training and test sets of classification models.

Source Type	Number of Samples (Training Set)	Number of Samples (Test Set)
COT	36	18
ALM	17	8
TOM	24	12
AUR	13	6
CEM	9	5
DIR	4	1
DLS	4	2
FED	6	3
GRP	8	5
PRU	9	5
RPR	6	3
SAF	9	4
UPR	4	2
Total	149	74

Table 11. PLS and Neural Net classification results for full-variable data.

	Training		Test	
Source Type	PLS	BP-ANN	PLS	BP-ANN
COT	31	36	15	16
ALM	17	17	8	8
TOM	21	24	11	12
AUR	13	13	6	6
CEM	9	9	4	5
DIR	4	4	1	1
DLS	2	4	1	1
FED	6	6	3	3
GRP	7	8	5	5
PRU	8	9	5	5
RPR	2	6	0	1
SAF	5	9	2	4
UPR	2	4	1	0
Correctly Classified (%)	85	100	84	91

Table 12. Classification of PLS and BP-ANN model results after variable selection using D-PLS.

	Training		Test	
Source Type	D-PLS	BP-ANN	D-PLS	BP-ANN
COT	33	36	16	18
ALM	17	17	8	7
TOM	22	24	10	10
AUR	13	13	6	6
CEM	9	9	5	5
DIR	4	4	1	1
DLS	4	4	2	2
FED	6	6	3	3
GRP	8	8	5	5
PRU	9	9	5	5
RPR	2	6	0	3
SAF	9	9	3	4
UPR	4	4	2	2
Correctly Classified (%)	94	100	89	96

Table 13. Classification of PLS and BP-NN model results after variable selection using Genetic Algorithms.

	Training		Test	
Source Type	D-PLS	BP-ANN	D-PLS	BP-ANN
COT	33	36	16	18
ALM	17	17	8	7
TOM	23	24	12	12
AUR	13	13	6	6
CEM	9	9	5	5
DIR	4	4	1	1
DLS	4	4	2	2
FED	6	6	3	3
GRP	8	8	5	5
PRU	9	9	5	5
RPR	6	6	1	3
SAF	9	9	4	4
UPR	4	4	2	2
Correctly Classified (%)	97	100	95	99

In the second analysis of DNA fingerprints of TSS-12 sources, partial least square (PLS), and back propagation artificial neural networks (BP-ANN) used to predict properties of source samples from the fingerprint data. The entire set of samples was split into a training set (149 samples) and a test set (74 samples), as in the previous study (Table 109). In this case, however, the models developed in the training set were used to predict soil properties in the test set samples. The data for each sample included 320 variables (pixel data from the whole DNA fingerprint) and 6 scaled dependent variables (%Sand, %Silt, %Clay, and Nitrogen, Organic Carbon, and DNA contents). The predicted soil properties were compared with the actual soil properties by calculating the root mean square error of prediction (RMSEP). As an example, measured and predicted best fit lines for organic carbon are presented for PLS (Figure 22) and BP-ANN (Figure 23), and BP-ANN after principle components analysis (Figure 24). For measured versus predicted organic carbon content, BP-ANN following principal components analysis provided the greatest best line fit (R=0.944), and the lowest root mean square error of prediction (RMSEP=1.664). Hopke and coworkers determined that microbial DNA fingerprints can provide a quantitative indicator of the soil property, as well as which crop had been grown on it. The application of PCA improved the prediction capability of the BP-ANN model and gave the lowest values for the root-mean-square error for the independent test set.

Figure 22. Measured versus predicted organic carbon content of test samples using Partial Least Squares. RMSEP = 2.5884

Figure 5: Measured versus Predicted Organic Carbon percentage of 1 Samples using PLS, RMSE

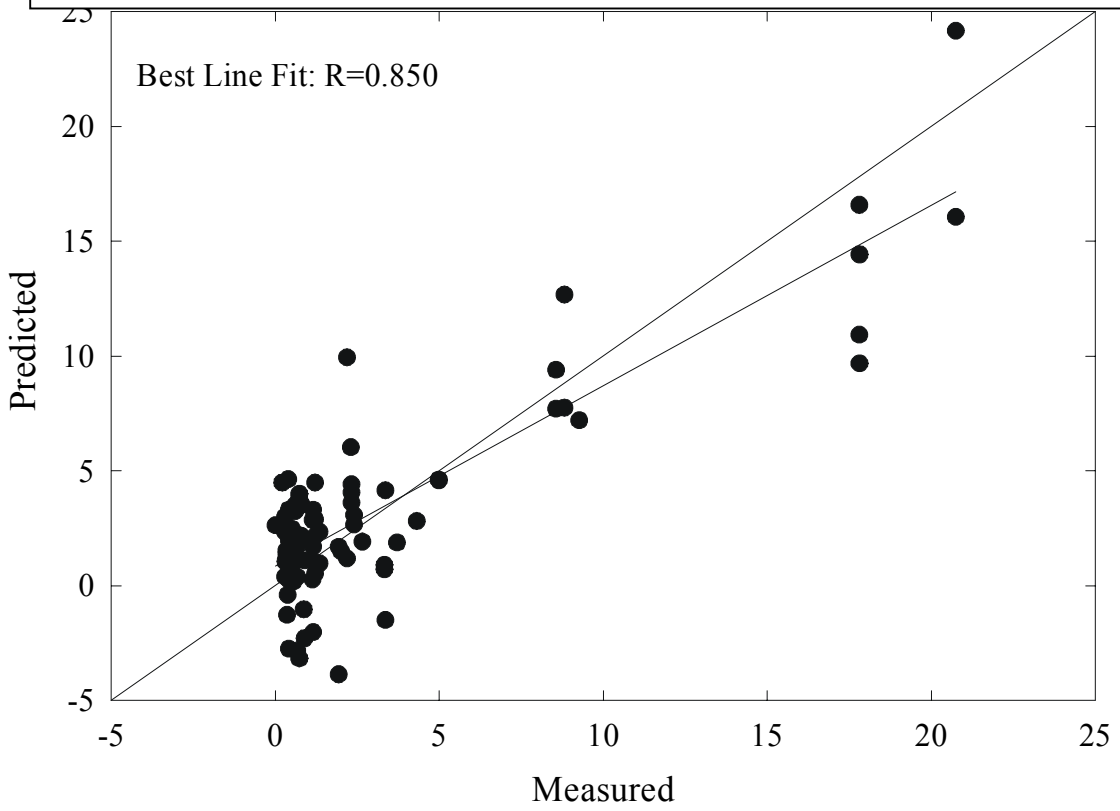


Figure 23. Measured versus predicted organic carbon content of test samples using Artificial Neural Net (ANN). RMSEP = 2.5978.

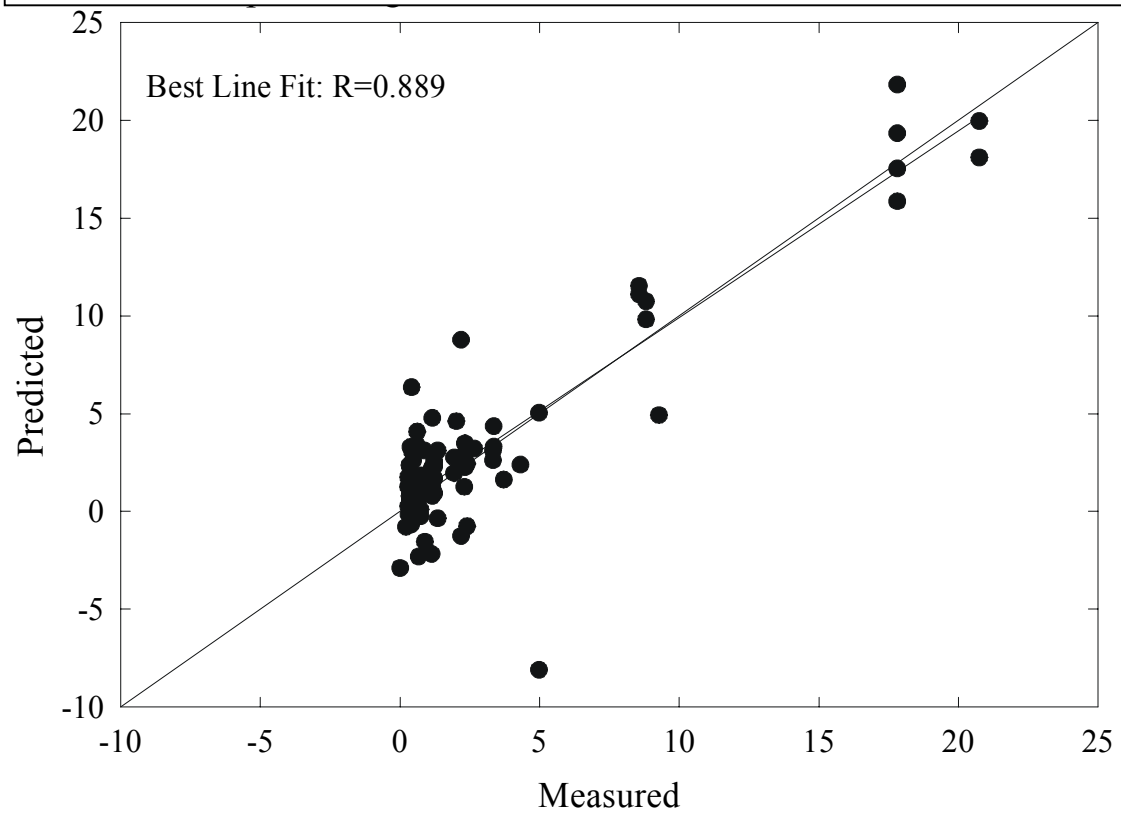
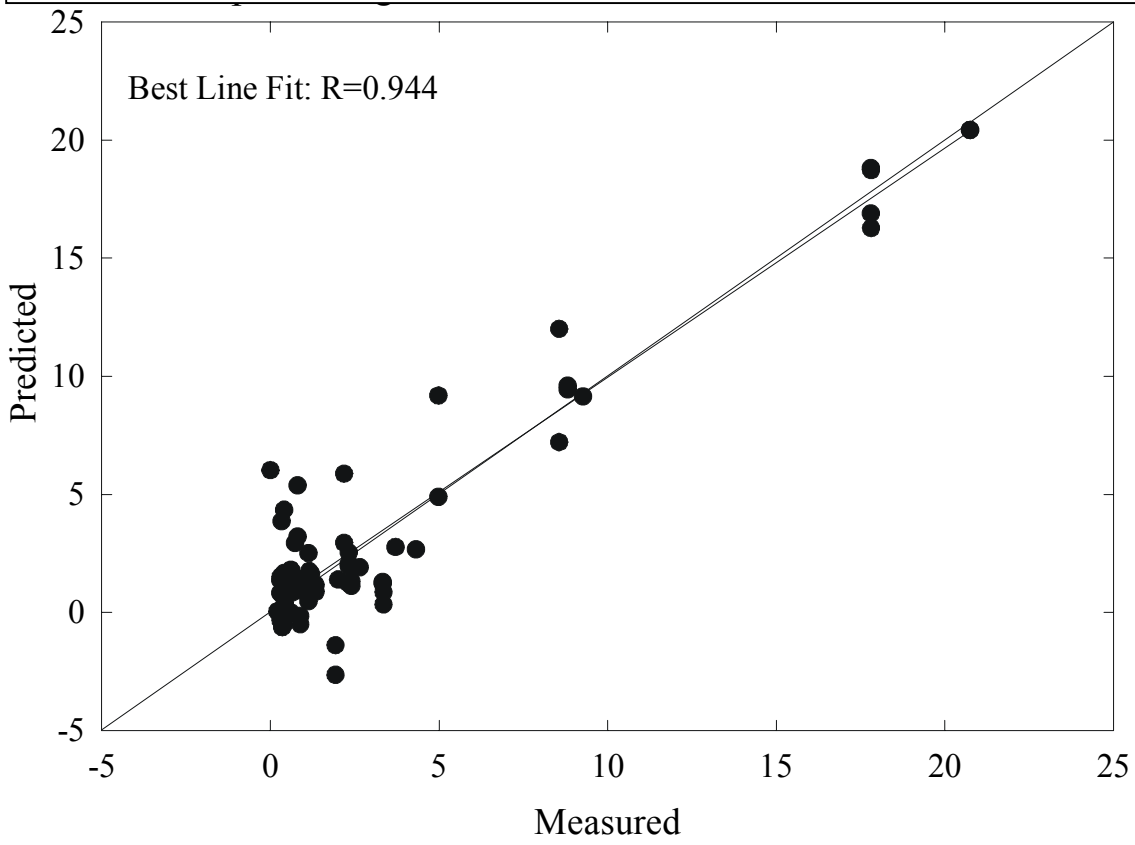


Figure 24. Measured versus predicted organic carbon content to test samples using Principal Components Analysis and Artificial Neural Net (ANN). RMSEP = 1.664



Phospholipid fatty acid (PLFA) data from 353 Central Valley agricultural sources were used to distinguish soils planted with different crops. Sample class designations are found in Table 14. The training set (177 samples) and test set (176 samples) were first used in discriminant partial least squares (D-PLS) and regularized discriminant analysis (RDisA) on the original 72 PLFA variables (Table 15). Then a variable selection approach based on the PLS regression coefficients identified the most important PLFA variables for classifying samples. Twenty-eight of the original 72 PLFA variables were retained, and the misclassification rate was reduced (Table 16). Both the D-PLS and RDisA methods provided satisfactory performance in classifying soil samples, with RDisA being slightly better. The majority of samples misclassified by both methods were within two specific sample classes (fig and walnut), the two classes with the least number of samples, and thus the least training information.

Table 14. Source sample classes used for training and test sets in PLFA-based classification models.

Class	Source Type	Number of Samples
1	pasture	25
2	vegetable	28
3	tomato	54
4	grass	43
5	rice	112
6	almond	14
7	cotton	16
8	fig	11
9	walnut	10

Table 15. D-PLS and RDisA classification results for full-variable data.

		Training (misclassifications)		Test (misclassifications)	
Class	Source Type	D-PLS	RDisA	D-PLS	RDisA
1	pasture	0	0	0	0
2	vegetable	0	0	0	0
3	tomato	2	1	1	3
4	grass	0	0	3	0
5	rice	0	0	0	1
6	almond	0	0	0	0
7	cotton	0	0	0	0
8	fig	6	3	5	3
9	walnut	5	1	5	3
	Total Misclassifications	13	5	14	10

Table 16. D-PLS and RDisA classification results for selected-variable data.

		Training (misclassifications)		Test (misclassifications)	
Class	Source Type	D-PLS	RDisA	D-PLS	RDisA
1	pasture	0	0	0	0
2	vegetable	0	0	1	1
3	tomato	0	1	4	1
4	grass	0	0	2	1
5	rice	0	0	0	0
6	almond	0	0	0	0
7	cotton	0	0	0	0
8	fig	1	3	0	0
9	walnut	0	1	3	3
	Total Misclassifications	1	4	10	8

4.0 COMPARISON OF DIFFERENT BIOLOGICAL APPROACHES FOR CHARACTERIZING SOURCES AND DUST

DNA fingerprinting of bacterial communities provides a reproducible and highly sensitive method for characterizing source and dust samples. DNA fingerprints are relatively similar at different locations within an individual agricultural field (i.e., field variability is low). Agricultural source samples grouped to some degree according to crop type and/or geographical location; however, each agricultural field had a relatively unique fingerprint and could be differentiated from the other fields. Bacterial DNA fingerprints of sources tend to closely match those of laboratory-generated dust. This finding supports the contention that biochemical information in source soils is preserved at least for the short term, and is detectable in dust suspended from source material. Research beyond the scope of this contract will be needed to determine whether such a relationship holds true in the field. Under field conditions, there will be a longer period of time between when dust is generated and when it is sampled. Also, microorganisms in field dust will be exposed to adverse conditions (e.g., desiccation, UV radiation). Depending upon the extent of possible DNA damage, biological fingerprints of field dust may deviate more strongly from the source signal.

Microbial lipid profile analysis is also a reproducible approach, under some conditions, for characterizing source and dust samples, although less promising than DNA-based techniques. Although the PLFA method is more reproducible than SFAME, it requires larger sample masses. Even if these greater masses of dust could be collected in the field, the relationships between dust and source material are not as strong as those with DNA-based techniques. Generally, however, there was good agreement between the PLFA- and DNA-based techniques. Consequently, PLFA is valuable as an independent, reproducible method, which can be used to support or confirm results of more sensitive DNA-based methods.

We are currently involved in an ARB-supported project (Fall 2000 CRPAQS field campaign, Corcoran, CA), in which we are using the methods developed in this contract. Comparisons of agricultural source and field PM_{10} samples collected during cultivation operations, additional source characterization in the vicinity of Corcoran, as well as results from DNA-based studies of PM_{10} collected at the Corcoran anchor site should be available in October 2001.

5.0 SUMMARY AND CONCLUSIONS

1. DNA fingerprinting can discriminate both sources of dust (agricultural and other categories) as well as dust samples themselves. A major advantage of DNA fingerprinting is that the small masses typical of dust samples can be analyzed using this method. Identifying and sequencing DNA bands unique to a specific source (e.g., biomarkers) may increase the sensitivity of this method and greatly simplify the statistical analysis.

2. We have identified the minimum sample sizes of source material or dust needed for biological characterization by DNA or fatty acid fingerprinting. For DNA fingerprinting, the minimum sample size for laboratory-generated dust from agricultural sources is 20 μg, but more commonly in the range of 50-100 μg. The majority of other source types (e.g., roads) required 100-200 μg of dust. Lower detection limits may be possible if biomarkers sequences are present in, and characteristic of, specific sources. For lipid profiling of sources, PLFA required 1-5 g and SFAME required 100-300 mg. For lipid profiling of laboratory-generated PM_{10}, PLFA required 700 mg and SFAME required 40 mg of material.

3. Dust samples generated in laboratory chambers can be matched to their respective source soils, using either DNA or fatty acid methods of analysis. These results were supported in a field trial in which dust samples collected during a cultivation operation at U. C. Davis could be related back to their source soils by DNA fingerprinting.

4. Comparing the relative strengths and weaknesses of two fatty acid methods of analysis, we found PLFA analysis provides a more reliable characterization of samples (based on comparisons of sources versus laboratory-generated PM_{10}), but requires greater sample mass for analysis. In contrast, SFAME analysis is more sensitive and can be used for small sample sizes but less successful at matching sources and laboratory-generated PM_{10}.

5. Artificial neural network and other classification models can accurately classify source and dust samples using DNA and PLFA data. Because the data obtained by either method are extensive, variable selection strategies are required for optimal model performance. For instance DNA fingerprint variables were reduced from 320 to 108, and PLFA variables were reduced from 72 to 28.

6. We have collected dust samples on high volume filters under field conditions (as part of Corcoran Fall 2000 study), concurrently with source samples. This set of samples will provide a unique opportunity to test whether our laboratory conclusions can be supported in field conditions. We anticipate that these data will be analyzed by October 2001.

6.0 FUTURE WORK AND RECOMMENDATIONS

The following lists new research directions that expand on the results of this contract and could extend progress in the use of biological methods for characterizing fugitive dust.

1. **Identify of biomarkers from DNA fingerprints of fugitive dust sources.** Microbial and plant biomarkers could be identified from existing DNA fingerprint data and genetically sequenced. Genetic sequence information is required for constructing source-specific genetic probes and PCR primers with enhanced specificity (e.g., can target specific types of bacteria, rather than all bacteria in a source). Genetic sequence information is also required for the development of DNA microarray technology, which could be used in the future to rapidly and inexpensively analyze samples. Microarray technology would also be valuable for implementing validation studies on a large, such as statewide scale.

2. **Determine the quantitative relationships among DNA fingerprint data, biomarkers, and the original biological material in source samples and fugitive dust.** The actual quantities and distributions of biological markers in the soil environment are not yet known. These data will be required before biological techniques can be appropriately applied to studies of source apportionment. In addition, the absolute or relative quantities of biological markers may be subject to seasonal or yearly fluctuations, which should be considered in future research.

3. **Assess the utility of DNA fingerprints versus DNA markers for their application in source apportionment.** Known sources could be mixed in different proportions, in controlled laboratory based studies, to determine how well original sources can be identified, quantified, and/or predicted from composite dust samples. Dust generation and collection chambers, such as those constructed in our laboratory could be useful for these efforts. This work would require continued collaboration with statisticians well-versed in classification methods.

4. **Investigate the stability of biological markers when dust samples are exposed to UV, desiccation, and open air factors (OAFs), as would be encountered in the field.** In preliminary studies, biological material contained in dust collected on quartz filters could be exposed under precisely controlled conditions in the laboratory. The next phase of research could be conducted in an outdoor chamber designed to resuspend, expose, and collect PM_{10} previously fractionated from source samples.

5. **Continue the use of classification methods (D-PLS, BP-ANN) to compare different biological methods of characterizing dust and soil samples.** The primary requirement for the appropriate application of these statistical methods is a large sample set (100s or 1000s of samples). Extensive sample collection and analysis should continue with rigorously standardized procedures, preferably using automated technologies.

6. **Extend field studies of microbiological fingerprinting methods.** Comparative analyses of biological fingerprints or biomarkers from fugitive dust generated from known sources in the field and collected using current air quality monitoring procedures will determine whether the methods successful in the laboratory can be implemented on the field scale.

7.0 REFERENCES

Altschul, S. F., T. L. Madden, A. A. Schaeffer, 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.

Baath, E., A. Frostegard, T. Pennanen and H. Fritze (1995). Microbial community structure and pH response in relation to soil organic matter quality in wood-ash fertilized, clear-cut or burned coniferous forest soils. Soil Biology and Biochemistry **27**: 229-240.

Bartley, D. L., C.-C. Chen, R. Song and T. J. Fischbach (1994). Respirable aerosol sampler performance testing. American Industrial Hygiene Association Journal **55**: 1036-1046.

Borneman, J. and E. W. Triplett (1997). Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. Applied and Environmental Microbiology **63**: 2647-2653.

Bossio, D. A., K. M. Scow, N. Gunapala and K. J. Graham (1998). Determinants of soil microbial communities: Effects of agricultural management, season, and soil type on phospholipid fatty acid profiles. Microbial Ecology **36**: 1-12.

Buyer, J. S. and L. E. Drinkwater (1997). Comparison of substrate utilization assay and fatty acid analysis of soil microbial communities. Journal of Microbiological Methods **30**: 3-11.

Carvacho, O. F., L. L. Ashbaugh and R. G. Flocchini (1996). Characteristics of the UC Davis dust resuspension chamber. Presented at the Annual Meeting of the American Association of Aerosol Research, Orlando, FL.

Cavigelli, M. A., G. P. Robertson and M. J. Klug (1995). Fatty acid methyl ester (FAME) profiles as measures of soil microbial community structure. Plant and Soil **170**: 99-113.

Chow, J. C., J. G. Watson, J. E. Houck, L. C. Pritchett, C. F. Rogers, C. A. Frazier, R. T. Egami and B. M. Ball (1994). A laboratory resuspension chamber to measure fugitive dust size distributions and chemical compositions. Atmospheric Environment **28**: 3463-3481.

Chow, J. C., J. G. Watson, D. H. Lowenthal, P. A. Solomon, K. L. Magliano, S. D. Ziman and L. W. Richards (1993). PM(10) and PM(2.5) Compositions in California's San Joaquin Valley. Aerosol Science and Technology **18**: 105-128.

Garcia-Martinez, J., S. G. Acinas, A. I. Anton and F. Rodriguez-Valera (1999). Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. Journal of Microbiological Methods **36**: 55-64.

Gunapala, N. and K. M. Scow (1998). Dynamics of soil microbial biomass and activity in conventional and organic farming systems. Soil Biology & Biochemistry **30**: 805-816.

Haene, B. G., K. Jaeger and H. G. Drexler (1993). The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. Electrophoresis **14**: 967-972.

Ibekwe, A. M. and A. C. Kennedy (1999). Fatty acid methyl ester (FAME) profiles as a tool to investigate community structure of two agricultural soils. Plant and Soil **206**: 151-161.

Jensen, M. A., J. A. Webster and N. Straus (1993). Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Applied and Environmental Microbiology **59**: 945-952.

Jongman, R. H., C. J. F. ter Braak and O. F. R. Van Tongeren (1995). Data Analysis in Community and Landscape Ecology. Cambridge, Cambridge University Press.

Macalady, J. L., E. E. Mack, D. C. Nelson and K. M. Scow (2000). Sediment microbial community structure and mercury methylation in mercury-polluted Clear Lake, California. Applied and Environmental Microbiology **66**: 1479-1488.

Moosmuller, H., J. A. Gillies, C. F. Rogers, D. W. DuBois, J. C. Chow, J. G. Watson and R. Langston (1998). Particulate emission rates for unpaved shoulders along a paved road. Journal of the Air & Waste Management Association **48**: 398-407.

Palumbi, S. R. (1996). Nucleic acids II: The polymerase chain reaction. Molecular Systematics. D. M. Hillis, C. Moritz and B. K. Mable. Sunderland, MA, Sinauer Associates**:** 205-247.

Pearson, K. (1926). On the coefficient of racial likeness. Biometrika **18**: 105-117.

Rademaker, J. L. W. and D. B. F. J. (1997). Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis. DNA markers: Protocols, Applications, and Overviews. G. Caetano-Anolles and G. P. M. New York, Wiley-Liss, Inc.**:** 151-171.

Sackin, M. J. and D. Jones (1993). Computer-assisted classification. Handbook of New Bacterial Systematics. M. Goodfellow and A. G. O'Donnell. New York, Academic Press**:** 281-313.

Saetre, P. and E. Baath (2000). Spatial variation and patterns of soil microbial community structure in a mixed spruce-birch stand. Soil Biology & Biochemistry **32**: 909-917.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical Taxonomy; The Principles and Practice of Numerical Classification. San Francisco, W. H. Freeman.

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**: 3524-3530.

ter Braak, C. J. F. (1994). Canonical community ordination. Part I: Basic theory and linear methods. Ecoscience **1**: 127-140.

The Institute for Genomic Research (TIGR) (2001). Comprehensive Microbial Resource, Microbial Genome Program, U. S. Department of Energy.

U. S. EPA (1995). Compilation of Air Pollutant Emission Factors, AP-42. Research Triangle Park, N.C., U.S. Environmental Protection Agency.

United States Department of Agriculture (1996). Soil Survey Laboratory Methods Manual. National Soil Survey Center, United States Department of Agriculture.

White, D. C., W. M. Davis, J. S. Nickels, J. D. King and R. J. Bobbie (1979). Determination of the sedimentary microbial biomass by extractable lipid phosphate. Oceologia **40**: 51-61.

White, T. J., T. Bruns, S. Lee and J. Taylor (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols : A Guide to Methods and Applications. M. A. Innis. San Diego, Academic Press.

Wintzingerode, F. V., U. B. Goebel and E. Stackebrandt (1997). Determination of microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis. FEMS Microbiology Reviews **21**: 213-229.

Yang, C.-H., D. E. Crowley, J. Borneman and N. T. Keen (2001). Microbial phyllosphere populations are more complex than previously realized. Proceedings of the National Academy of Sciences of the United States of America **98**: 3889-3894.

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

8.0 PUBLICATIONS PRODUCED

Elliott, G. N., Till, A. M., and K. M. Scow. 2001 Comparison of Phospholipid Fatty Acid (PLFA) and direct fatty acid methylation (SFAME) in characterization of soil microbial communities. Submitted to Soil Biology and Biochemistry.

Elliott, G. N., Till, A. M., and K. M. Scow. 2001 Comparison of PM_{10} and source soil lipid profiles using PLFA, SFAME and ITS-DNA profiling techniques. Submitted to Atmospheric Environment.

M. J. Johnson. 2001. DNA fingerprinting microbial communities in respirable dust and soils from California's San Joaquin Valley. Dissertation. University of California at Davis.

Johnson, M. J., Lee, K. Y., and K. M. Scow. 2001. DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. Submitted to Geoderma.

Johnson, M. J., Lee, K. Y., and K. M. Scow. 2001. DNA fingerprinting respirable dust generated from fugitive dust sources of California's San Joaquin Valley. Submitted to Atmospheric Environment.

Ramadan, Z., Song, X-H., Hopke, P. K., Johnson, M. J., and K. M. Scow. 2001. Variable selection in classification of environmental soil samples for partial least square and neural network models. Analytica Chimica Acta 21225:1-12.

Ramadan, Z., Song, X-H., Hopke, P. K., Johnson, M. J., and K. M. Scow. 2001. Application of PLS and Back Propagation Neural Networks for the estimation of properties of soil samples. Submitted to Journal of Chemometrics.

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

9.0 GLOSSARY OF TERMS, ABBREVIATIONS, AND SYMBOLS

ANN	Artificial neural net
BP-NN	Back-propagation neural net
CA	Correspondence analysis
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
D-PLS	Discriminant partial least squares
GA	Genetic algorithm
ITS	Intergenic transcribed spacer
PCA	Principal components analysis
PCR	Polymerase chain reaction
PLFA	Phospholipid fatty acid
PLS	Partial least squares
PM_{10}	Particulate matter of 10μM or less in aerodynamic diameter
RDA	Redundancy Analysis
RDisA	Regularized discriminant analysis
rDNA	Ribosomal deoxyribonucleic acid
RNA	Ribonucleic acid
SFAME	Soil fatty acid methyl ester
UV	Ultra-violet

Appendix 10.1

Technical Support Study 12 Source Sampling Protocol

In the context of Technical Support Study 12, the agricultural sources are termed crop samples, and all other sources are termed non-crop samples. Crop samples were collected accordance with U.S. Environmental Protection Agency guidelines (U. S. EPA, 1995). One sub-sample collected in the center of an agricultural field, and four others collected 100 meters away in orthogonal directions (North, South, East and West), were mixed to obtain a single composite sample. Single composite field samples have no letter designation (e.g., ALM2).

Field replicate composite samples were collected by the same method from three different regions of the same field (e.g., the NW, NE, and SW corners) and are designated by the suffix A, B, and C (e.g., ALM1A-C). Typically, field replicate samples are located within one kilometer of each other. The crop samples were collected from the top 2 cm of cultivated or fallow soils that support(ed) almond (ALM1-4), cotton (COT1-6), grape (GRA1-3), safflower (SAF1 and SAF3), and tomato (TOM1-3) crops during the current (or previous) growing season. Safflower sample SAF2 was not available for DNA analysis. Three samples were also collected from agricultural unpaved roads (UPR1-3), and one sample (STA1) was collected from an agricultural staging area (an unpaved thoroughfare for farm machinery), each adjacent to agricultural field collection points (within several km). The agricultural unpaved road and staging area samples, collected from a single rectangular area by brush and dustpan, are included in the analyses of crop samples.

Samples from non-crop sources were collected from the top 2 cm of cattle feedlots (CTF1-2), cattle dairies (CTD1-2), salt-affected sites, including one dry irrigation drainage basin (DIS1) and one irrigation runoff channel (DIS2), and construction (CON1-2) sites (exposed subsurface soil). Samples from three Public/Residential unpaved roads (UPR4-6) were collected by brush and dustpan. Finally, two each urban and rural paved road samples (PVR1-4), were collected from high vehicular traffic areas with a vacuum.

Appendix 10.2

Design and Construction of the Small-scale Dust Generation and Collection Chamber

Overview

The small-scale dust generation and collection chamber was specifically designed to accommodate samples of limited mass, as in the TSS-12 samples. It provided the means to analyze respirable dust samples and make direct comparisons to source samples in the laboratory.

Materials and Construction

In the dust generation portion of the chamber (labels 1-4 in Figure 2), particles were suspended from the whole-soil sample via airflow and mechanical energy. Pressure regulated, filtered air (Product # EW-98252, Cole-Parmer Instrument Company, Vernon Hills, IL) was set at a seven liter per minute ($l\ m^{-1}$) flow by a 150 mm high-flow correlated meter with a valve and stainless steel float (U-03217-26, Cole-Parmer Instrument Company, Vernon Hills, IL). The air flows through an encased aluminum sample holder attached via a shaft to a small electric motor rotating at 50 rpm. Four small fins welded to the interior of the cylindrical sample holder agitate the sample by lifting and dropping the soil sample across the path of the air stream. Entrained particles enter a vertically oriented expansion section created from aluminum or stainless steel QF hardware (Kurt J. Lesker Company, Clariton, PA). The assembly includes a 25 mm hose adapter (QF25-100-SH) welded to the sample holder encasement, a 25 mm 90° elbow (QF25-100-E90), a conical reducer (QF40xQF25CA), a 150 mm nipple (QF40-150-N), another conical reducer (QF40xQF16CA), and a 16 mm hose adapter (QF16-075-SH). All fittings were sealed with appropriately sized aluminum QF centering rings and hinged QF clamps with thumbscrews. The end of the assembly dispenses aerosolized particles into the collection chamber.

The dust collection portion of the chamber (labels 5-9 in Figure 2) consists of an electrical circuit box (total volume approximately 30.6 liters) adapted to accommodate a glass viewing port, the dust generation assembly outlet, and sealed 1/4" I.D. Tygon tubing which fits between the respirable dust cyclones (BGI-4, BGI Incorporated, Waltham, MA) and the HFS-513A Air Sampling System pumps (Gilian Instrument Corp, West Caldwell, NJ). The Gilian pumps were calibrated at a flow of 2.2 $l\ m^{-1}$ by a 150 mm correlated flow meter with an aluminum float (U-0327-20, Cole-Parmer Instrument Company, Vernon Hills, IL). At a flow of 2.2 $l\ m^{-1}$, the 50% cut is for particles with 4 μm aerodynamic diameter (Bartley et al., 1994). Particles greater than 4 μm aerodynamic diameter are removed by inertial separation and collected in the "grit cup." Six Gilian pumps operating at 2.2 $l\ m^{-1}$ produce a total suction of 13.2 $l\ m^{-1}$. As 7 $l\ m^{-1}$ flowed through the rotating sample holder, makeup air was required to balance the flow in and out of the closed system. To accomplish this, approximately 6.2 $l\ m^{-1}$ of filtered air regulated by a brass needle valve (U-688831, Cole-Parmer Instrument Company, Vernon Hills, IL) was delivered through 1/4" I.D. Tygon tubing perforated at 10 cm intervals. This system was designed as an alternative to a single

make-up air inlet, which created very high turbulence in the chamber. A positive pressure of less than 1 psi was maintained with use of a Magnehelic Differential Pressure Gage (part # 2302, Dwyer Instruments, Inc., Michigan City, IN). This slight positive pressure ensured that ambient laboratory air containing potential microbial contaminants was not introduced into the chamber.

Appendix 10.3

Image Processing Points for DNA Fingerprint Analysis

Overview

To compare DNA fingerprints of more than 16 samples (the number of lanes within a single gel), a standardized system of producing and processing multiple gel images is required. This task was completed with the acquisition and use of a state-of-the-art software package for processing DNA fingerprints called Gelcompar II (Applied Maths, Kortrijk, Belgium). The specific settings used in image analysis of DNA fingerprint patterns are detailed below.

DNA fingerprint gel image processing

The settings used to process inverted gel images included 35 point, 3 node image strip extraction for the raw data, an averaging thickness of 13 points with 3 nodes for curve extraction, a rolling ball background subtraction of 10% (based on spectral curve analysis of all gels with the Fourier method), and arithmetic average and least square filtering with a cutoff below 1% and a power of 2.0. The vertical dimension of the gel consisted of 312 pixels. All gels lanes were normalized to a common reference pattern, a 20 bp DNA ladder (BioWhittaker Molecular Applications, Rockland ME), which was loaded in at least 4 positions on every gel. To designate bands in the profiles, the band search filter was set to a minimum profiling of 5% relative to maximum value, a minimum area of 0.5%, and a shoulder sensitivity of two. The positions of bands within the fingerprints were located by fitting the peaks of intensity by regression (cubic spline fit with logarithmic dependence) against the migration of bands in the 20 bp DNA ladder. Band positions were only derived for a specific region on the gel, namely within the bounds of the 20 bp ladder (1000-300 bp range). Quantitative values for the bands in a profile were obtained by integrating one-dimensional band areas (Gaussian fit). The array of bands selected by the software in each lane was inspected to remove false information (e.g., small illuminated spots from dust that met band choosing criteria of the software), or to include appropriate information (e.g., visually detected bands that did not meet the band choosing criteria of the software), and to confirm that one-dimensional band areas were integrated accurately.

Appendix 10.4

PLFA Extraction and Purification Protocol

Overview

Analysis of PLFAs provides insight into the structure and status of the soil microbial community. Phospholipids are quickly degraded upon the death of a microorganism. Fatty acids derived from phospholipids represent the potentially viable members of a community. Soil lipids are extracted directly from the sample. Phospholipids are separated from the other lipid classes for analysis by gas chromatography.

Sample precautions

Sample contamination can be reduced by proper laboratory practices. Reagents are of the highest grade possible e.g. optima grade solvents. All procedures are carried out in either Teflon or glass. Rinse labware with hexane prior to use. Fatty acids from oils on the skin may appear in the analysis. Never directly handle samples or anything that will come in contact with the sample. Use nitrile gloves at all times. They are more resistant to the solvents used than latex, or most other common glove materials. Keep samples frozen (-20° C or lower) until ready for analysis. Avoid thawing and refreezing of samples.

Personal safety

Read the material safety data sheet for all reagents used. Work in an approved fume hood and wear gloves.

Reagents and Supplies

Glacial acetic acid (for 1.0 M Acetic Acid)
Acetone
Chloroform
Hexane
Methanol
Methyl Nonadecanoate (recommend internal standard for GC)
Nano-Pure® water, or equivalent
Potassium phosphate, dibasic (for 0.05 M phosphate buffer)
Sodium hydroxide pellets (for 0.2 M KOH in MeOH)
Toluene
Silica gel solid phase extraction cartridges, 500 mg, 3 ml
Disposable Pasteur pipette, 5¾ inch and 9 inch
11 mm GC vial, Teflon lined crimp top cap, 250 ml insert
Disposable Micro-Pipettor glass tubes
¼" disposable vials with Teflon lined caps
99.9995% Pure hydrogen, helium and air for GC
99.99% Pure nitrogen for sample evaporation

Equipment and Instrumentation

35 ml Teflon centrifuge tubes
Shaker
Centrifuge
125 ml separatory funnels with Teflon stopcocks

13x100 mm test tubes with Teflon lined phenolic caps
10x50 mm test tubes with Teflon lined phenolic caps
Sample drying apparatus
Solid phase extraction cartridge rack
Water bath
Freezer
50-250 µl Digital Micro-Pipettor with glass capillary tubes
10 ml x 1 ml pipette
1 ml x 0.1 ml pipette
GC vial crimper
Gas chromatograph with:
flame ionization detector
25 M x 0.20 mm I.D x 0.33 µm
autosampler
chromatography and peak identification software

Procedure

Use 35 ml Teflon centrifuge tubes that have been washed, dried, and rinsed with hexane.
For 8 grams, dry weight, of soil for an analysis:

$8.0 \times (1 + \theta)$ = Mass of moist soil to be weighed out.

Bring total H_2O in the initial extraction to 5 ml using PO_4 buffer, while accounting for the soil's water content. The volume of water in the soil is equal to the mass of moist soil minus the 8 grams of dry soil, based on 1 gram / 1 ml. Subtract amount of water in soil from 5 ml of P buffer to determine the amount used in the centrifuge tube.
Example of Calculation: Soil "X" is at 16% soil moisture content,

8 x 1.16 = 9.28 grams of moist soil and 3.72 ml of P buffer required.

<u>Extraction</u>

1) Add total of 5 ml of P buffer (see above, be sure to account for soil moisture content), plus 6 ml of $CHCl_3$ and 12 ml of MeOH.
2) Shake for 2 hours.
3) Centrifuge at 2500 rpm for 10 min. at 25° C.
4) Decant to separatory funnel.
5) Add 23 ml of Extractant ($CHCl_3$:MeOH:Buffer in a 1:2:0.8 ratio) to soil remaining in tube, vortex.
6) Shake for ½ hour.
7) Add 12 ml of $CHCl_3$ and 12 ml of P buffer to sep funnel (Add this while waiting for step # 6).
8) Centrifuge at 2500 rpm for 10 min. at 25° C.
9) Decant this to the same sep funnel.
10) Shake sep funnel for 2 minutes.
11) Let stand overnight for separation.
12) Clean centrifuge tubes: Fill half full with water, cap, vortex, dispose of soil in waste. Wash tubes with soap and hot water, rinse w/ hot tap 5x, DI 5x, nanopure 3x.

<u>Next Day</u>

1) Drain bottom layer from sep funnel into large diameter long glass test tubes.

2) Evaporate with N_2 at 30° - 32° C in water bath.

Conditioning Solid Phase Extraction cartridges (SPE), Transfer of lipids, and Fractionation

1) Use 10 x 50 mm test tubes for SPE fraction collection.
2) Add 3 ml of $CHCl_3$ to condition column.
3) Transfer lipids with four (4X) 250µl transfers of $CHCl_3$, using digital micro-pipetor.
4) Add 5 ml of $CHCl_3$.
5) Add 10 ml of Acetone.
6) Change collection tubes.
7) Add 5 ml of MeOH, Be sure to save this fraction.
8) Evaporate with N_2 at 32° C in water bath.

Transesterification

1) Add 1 ml of 1:1 MeOH:Toluene, and 1 ml of 0.2 M KOH, to the dried sample. Vortex.
2) Heat at 37° C for 15 min. in the water bath.
3) After heating, add 0.3ml of 1.0 M acetic acid, then 2 ml of hexane, then 2 ml of nanopure water, then cap and shake for 10 minutes on low setting.
4) Remove the upper layer to small disposable screw top vials.
5) Repeat the 2 ml of Hexane, shake for another 10 minutes.
6) Remove this upper layer and add it to the first hexane fraction.
7) Dry with N_2. (No water bath required).

Preparation for GC

1) Use small crimp seal G.C. vials with inserts.
2) Transfer (Use and save glass pipettes) with two 75 µl additions of 19:0 internal standard. The concentration of the internal standard depends on the expected concentration of fatty acids in your sample. Recommend 25 ng/µl.
3) Purge with N_2 and seal.
4) Store sealed G.C. vials in the freezer until analysis.

APPENDIX 10.5

SFAME Extraction and Purification Protocol

Overview

Soil Fatty Acid Methyl Ester (SFAME) analysis is a means to survey a wider cross-section of the carbon pool soil. This method resolves fatty acids from both microorganisms and plant residues, which form the active portion of a soil's carbon pool. The direct extraction of this pool, and the subsequent analysis of chemically modified compounds, can be performed on samples of limited mass. Fatty acids derived from samples are analyzed using a gas chromatograph to yield a series of peaks (as in PLFA), which are compared to a bacterial database of known fatty acids for identification.

Sample precautions

Sample contamination can be reduced by proper laboratory practices. Reagents are of the highest grade possible e.g. optima grade solvents. All procedures are carried out in either Teflon or glass. Rinse labware with hexane prior to use. Fatty acids from oils on the skin may appear in the analysis. Never directly handle samples or anything that will come in contact with the sample. Use nitrile gloves at all times. They are more resistant to the solvents used than latex, or most other common glove materials. Keep samples frozen (-20° C or lower) until ready for analysis. Avoid thawing and refreezing of samples.

Personal safety

Read the material safety data sheet for all reagents used. Work in an approved fume hood and wear gloves.

Reagents and Supplies

Hexane
Methanol
Methyl-tert-buytl ether (MTBE)
Methyl Nonadecanoate (recommend internal standard for GC)
Nano-Pure® water, or equivalent
Potassium hydroxide pellets
6.0 N Hydrochloric acid
Sodium chloride
Disposable Pasteur pipette, 5¾ inch and 9 inch
11 mm GC vial, Teflon lined crimp top cap, 250 µl insert
Disposable Micro-Pipettor glass tubes
¼" oz. disposable vials with Teflon lined caps
99.9995% Pure hydrogen, helium and air for GC
99.99% Pure nitrogen for sample evaporation

Equipment and Instrumentation

50 ml Teflon centrifuge tubes
Shaker
Centrifuge
Sample drying apparatus
Water bath

Hot Plate
Freezer
50-250 µl Digital Micro-Pipettor
10 ml x 1 ml pipette
1 ml x 0.1 ml pipette
GC vial crimper
Gas chromatograph with:
flame ionization detector
25 M x 0.20 mm I.D. x 0.33 µm
autosampler
chromatography and peak identification software

Procedure (based on 500 mg samples)

Saponification

1) Add 1.0 ml of 3.75N Alkaline methanol, tighten caps.
2) Vortex 5-10 seconds.
3) Place in 100°C water bath for 5 minutes. Ensure methanol is not boiling.
4) Remove and vortex for 5-10 seconds. Check tightness of caps.
5) Place in 100°C water bath for additional 25 minutes.
6) Remove and place in room temperature water bath.

Methylation

Add 2.0 ml of 3.25N Acidic methanol.

1) Cap and vortex 5-10 seconds.
2) Place in 80°C water bath for 10 minutes.
3) Remove and place in room temp water bath.

Extraction

1) Add 1.25 ml of hexane MTBE mix (1:1).
2) Gently shake for 10 minutes.
3) Centrifuge for 10 minutes at 2000 R.P.M.
4) Remove upper layer to disposable glass vial.
5) Repeat steps 1 through 4.

Transfer to GC vial

1) Evaporate sample with nitrogen.
2) Transfer with two 75 µl transfers using the internal standard.

Purge with nitrogen and cap.

APPENDIX 10.6

Comparison of Laboratory-Generated Dust to Source Samples: San Joaquin Valley Soils

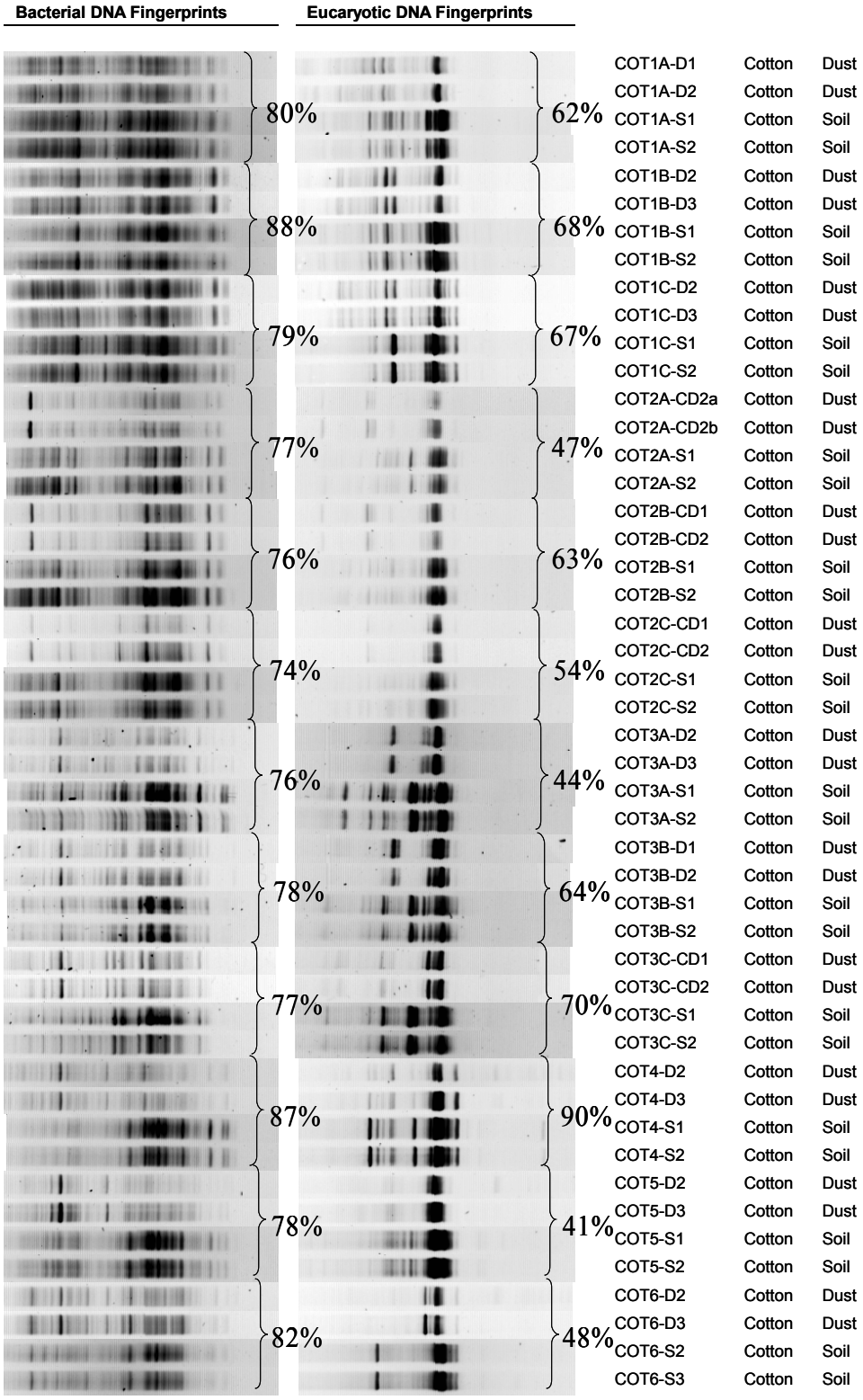


Figure A1. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from cotton samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample.

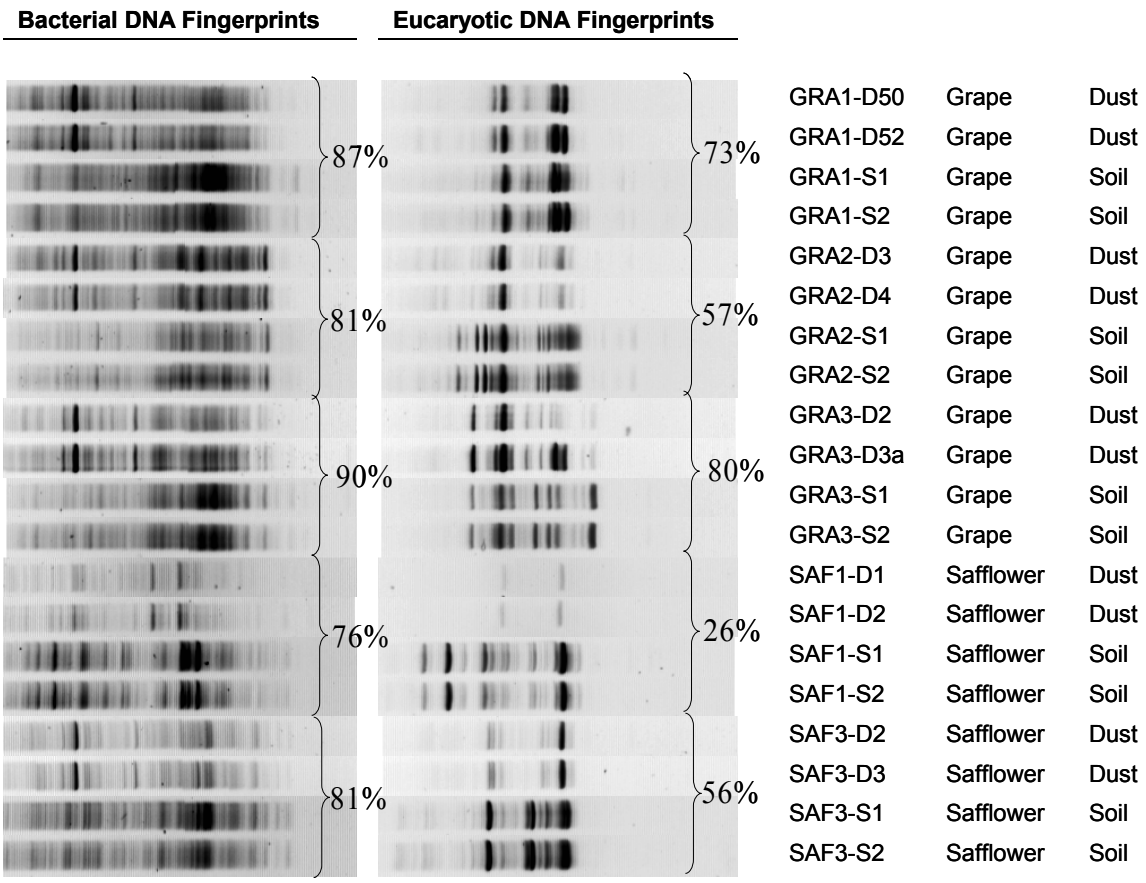
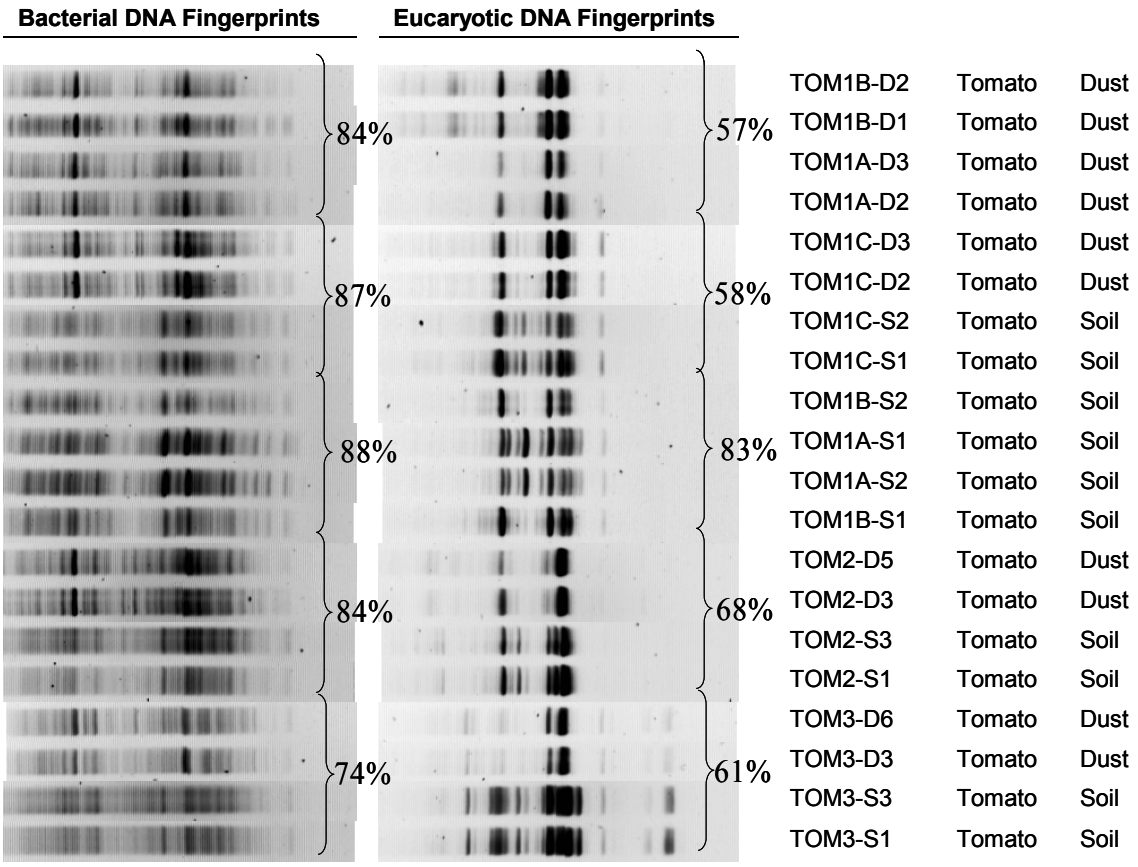


Figure A2. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from grape and safflower samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample.



FigureA3. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from tomato samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample.

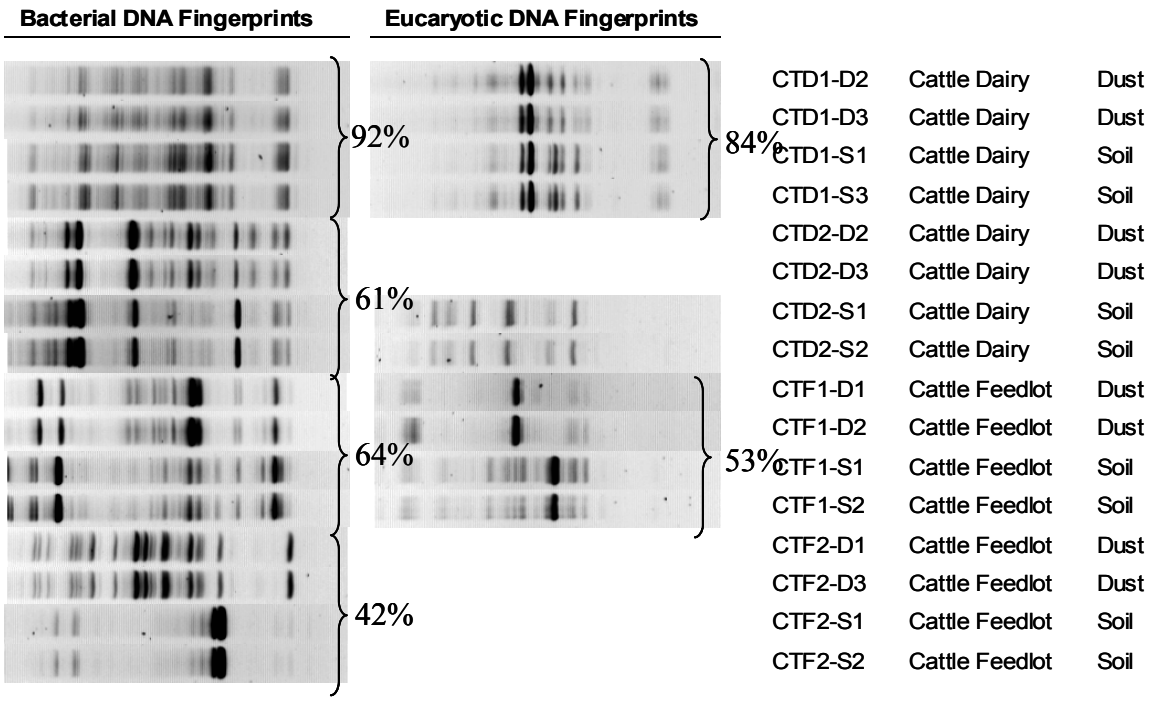


Figure A4. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from cattle dairy and cattle feedlot samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample. Eucaryotic DNA fingerprints were not produced for cattle dairy 2 dust or for either of the cattle feedlot 2 samples, as indicated by the blank spaces in the figure.

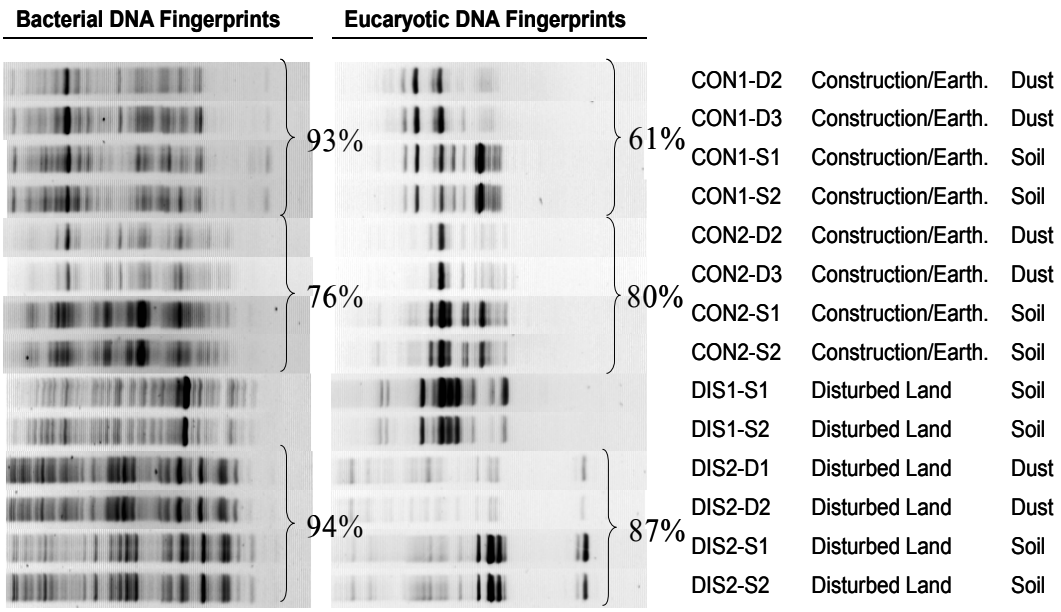


Figure A5. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from construction and disturbed land (irrigation ditch and dry irrigation drainage pond) samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample. Patterns of respirable dust from disturbed site 1 were not resolved for either DNA fingerprint type.

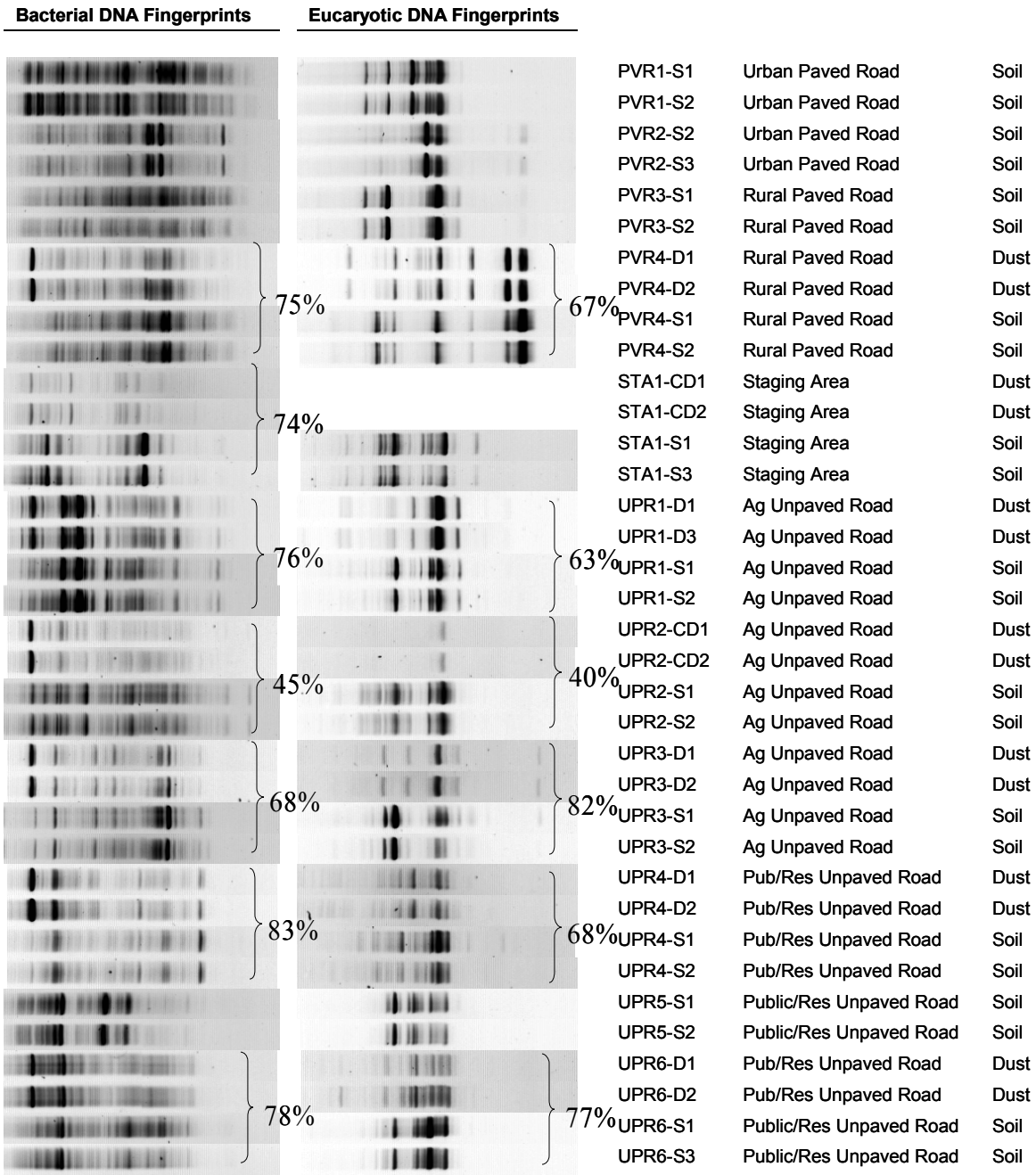


Figure A6. Bacterial and Eucaryotic DNA fingerprints of soil and respirable dust from roadway samples. The average percent similarity recorded for 4 pair-wise comparisons using the Dice coefficient is recorded to the right of soil and dust DNA fingerprints for each sample. PVR = paved road, STA = staging area, UPR = unpaved road. A eucaryotic DNA fingerprint was not produced for the staging area or unpaved road 5 respirable dust samples. Paved road samples 1-3 were not of sufficient quantity to suspend respirable dust.