PERGAMON

Soil Biology & Biochemistry 34 (2002) 1853–1860

Soil Biology & Biochemistry

www.elsevier.com/locate/soilbio

High resolution characterization of soil biological communities by nucleic acid and fatty acid analyses

Karen P. Dierksen[a], Gerald W. Whittaker[a], Gary M. Banowetz[a,*], Mark D. Azevedo[a], Ann C. Kennedy[b], Jeffrey J. Steiner[a], Stephen M. Griffith[a]

[a]*USDA/ARS, 3450 S.W. Campus Way, Corvallis, OR 97331, USA*
[b]*USDA/ARS, Washington State University, 215 Johnson Hall, Pullman, WA 99164-6421, USA*

Received 10 October 2001; received in revised form 2 September 2002; accepted 5 September 2002

Abstract

Fatty acid methyl ester (FAME) and length heterogeneity-polymerase chain reaction (LH-PCR) analyses were used to generate 'fingerprints' of FAMEs and eubacterial 16S rDNA sequences characteristic of agricultural soil communities. We hypothesized that pooling data from two methods that characterized different components of soil biological communities would improve the resolution of fingerprints characterizing the effects of contrasting tillage and ground cover practices. By using supervised classifications of FAME and LH-PCR, a discriminant analysis procedure distinguished soils from contrasting tillage and ground cover management and predicted the origin of soil samples. Used independently, FAME provided higher resolution of tillage, ground cover, and field location than LH-PCR, but LH-PCR was effective at identifying field location. Pooling data from both methods did not enhance the predictive power. A comparison of linear discriminant analysis, quadratic discriminant analysis, and nonparametric density estimation demonstrated that minimizing assumptions about data distribution improved the capacity of FAME analysis to resolve differences in soil types. Use of a purely statistical Bayesian method to select a subset of fatty acids (FA's) as variables in discriminant analyses identified FA's that represented signature FA's for specific groups of organisms.
Published by Elsevier Science Ltd.

Keywords: Fatty acid methyl ester; MIDI-FAME; Length heterogeneity PCR; Tillage; Microbial ecology; Statistics; Diversity; Nucleic acid analysis

1. Introduction

Changes in soil biological community composition might provide a measure of the effects of land-use practices on soil health. Although some of these changes can be characterized using microbial isolation procedures, it is estimated that less than 10% of soil microorganisms are culturable using existing techniques (Kennedy, 1999). Alternative approaches based upon structural component analysis have been developed to characterize changes in soil community diversity (Kennedy and Gewin, 1997). One approach, fatty acid methyl ester (FAME) analysis, utilizes gas chromatography to quantify FA's extracted from soil samples (Ratledge and Wilkinson, 1988; Ritchie et al., 2000). The chromatogram generates unique FAME profiles that represent complex populations. These 'fingerprints' are representative of specific sites and can be used as a reference to monitor the effects of land-use on changes that affect soil properties and biological communities (Cavigelli et al., 1995; Buyer and Drinkwater, 1997; Griffiths et al., 1999). This approach has been used as an indicator of the effects of agricultural practices on soil quality (Wander et al., 1995; Bossio et al., 1998; Ibekwe and Kennedy, 1998) and to characterize microbial communities from groundwater (Glucksman et al., 2000).

A second approach uses polymerase chain reaction (PCR) to characterize differences in the length and sequence of nucleic acids (Torsvik et al., 1990 a,b; Avaniss-Aghajani et al., 1994; Bruce, 1997; Dunbar et al., 1999; McCaig et al., 1999). Methods include length heterogeneity PCR (LH-PCR), which discerns natural variations in the length of the target gene, and terminal restriction fragment polymorphism (T-RFLP), which discriminates between gene sequence differences based on restriction enzyme digestion patterns. The PCR-based procedures generate fingerprints representing

* Corresponding author. Tel.: +1-541-750-8722; fax: +1-541-750-8750.
E-mail address: banowetg@onid.orst.edu (G.M. Banowetz).

0038-0717/02/$ - see front matter Published by Elsevier Science Ltd.
PII: S0038-0717(02)00198-0

microbial nucleic acid diversity (Suzuki et al., 1998). Both FAME and LH-PCR can distinguish profiles of soil community members from different sites (Ritchie et al., 2000). No assessment has combined FAME and LH-PCR profiles to attempt to resolve soil community differences in adjacent areas of a field under different cultivation practices.

Most FAME or PCR fingerprints have been analyzed using principle component analysis (McCune and Mefford, 1997; Ritchie et al., 2000). These types of analyses are referred to in pattern recognition literature as unsupervised classification because prior knowledge about the actual classification of observations is not used. In contrast, supervised classification methods use prior knowledge about data categories and the existence of components that differentiate these categories. One example of supervised classification applied canonical discriminant analysis to ester-linked and phospholipid-linked FAMEs to classify soils subjected to contrasting tillage practices (Drijber et al., 2000). The results of their study showed that soil disturbance provided the greatest dissimilarity in FAME profiles that differentiated wheat and fallow systems. A single FAME, C16:1 (*cis*11), a biomarker for soil arbuscular mycorrhizal fungi, was particularly significant in their capability to differentiate these soils.

The objectives of this research were: (1) assess the effects of combining PCR and FAME data; (2) compare different scales of analysis; and (3) determine whether use of a Bayesian variable selection method and nonparametric density discriminant analysis would improve fingerprint resolution. We evaluated the capability of each approach, independently and combined, to provide soil community fingerprints that classified soils at higher spatial resolution than previously attempted, and classified soils subjected to contrasting tillage practices.

2. Materials and methods

2.1. Biological community analysis

2.1.1. Soil sample collection

Soil samples (Woodburn silt loam, a fine-silty, superactive, mixed, mesic Aquultic Argixeroll) were collected in February 2000 from within and between the planted rows of tall fescue (*Festuca arundinacea* Schred.) field plots at the Oregon State University Hyslop Research Farm that were established with contrasting conventional tillage vs. no-tillage practices. A 2.5 cm corer was used to collect samples to a depth of 10 cm. The corer was cleaned between samples with water followed by methanol (800 ml l^{-1}). The samples were placed in zip-lock plastic bags and stored at −20 °C. Samples were also collected from an experimental wheat plot in a tilled Walla Walla silt loam (coarse-silty, mixed, mesic Typic Haploxeroll) soil north of Pendleton, OR.

2.1.2. FAME procedure

Total community FAMEs were extracted from 1 g dry weight equivalent samples using the method of Kennedy and Busacca (1995). Briefly, each sample was saponified for 30 min at 100 °C in 3.75N NaOH in methanol (500 ml l^{-1}), methylated for 10 min at 80 °C by addition of 6N HCl in methanol, extracted into two 1 ml aliquots of 1:1 hexane/methyl *tert*-butyl ether (MTBE), fractions combined and washed in 0.3N NaOH prior to concentration by overnight volatilization of the hexane/MTBE. Hexane/MTBE (150 μl) was added to the volatilized vials and transferred to 200 μl inserts in 2 ml crimp top vials. The extracted samples were analyzed on a HP 5890 series 2 gas chromatograph (Hewlett Packard, Palo Alto, CA) using an Ultra 2 crosslinked 5% PH ME siloxane capillary column 25 m long × 0.20 mm (Agilent Technologies, Palo Alto, CA). The instrument was programmed and controlled with MIDI Sherlock software package, version 3.0 (MIDI Inc, Newark, DE). Fatty acids identified by the MIDI software (MIDI-FAMEs) using MIDI microbial calibration standards were denoted using MIDI nomenclature and designated by the number of carbon atoms followed by a colon, the number of double bonds, and the position of the first double bond from the ω end of the molecule. Cis- and trans-isomers were indicated by c or t, respectively. An alpha character following the designations indicated a unique fatty acid in which the location of the double bond was not determined by the MIDI software. 'Summed features' indicated overlapping peaks for more than one identified fatty acid at a particular retention time (e.g. the summed feature 16:1 ω7/15 ISO 2OH peak cannot distinguish 16:1 ω7 from 15 ISO 2OH). Unknown peaks were noted and listed by retention time only.

2.1.3. PCR procedure

Microbial DNA was extracted from soil samples using a Fast DNA SPIN kit (BIO 101, La Jolla, CA). Purified DNA was stored at −20 °C. A fluorescent-labeled forward primer (5′-(6FAM) AGAGTTTGATCTGGCTCAG-3′) and an unlabeled reverse primer (5′-GCTGCTCGTAGGAGT-3′) were used to amplify conserved eubacterial 16S rDNA sequences present in the purified soil extracts (Ritchie et al., 2000). Amplification was conducted on 50 μl reaction volumes that contained 0.06% bovine serum albumin, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μM of each primer, and two units of *Taq* polymerase. A 3 min denaturation at 94 °C was followed by 25 cycles of 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 2 min. A final extension was conducted at 72 °C for 7 min. Fluorescent-labeled products were detected using an automated DNA sequencer (Applied Biosystems, Foster City, CA) equipped with GeneScan software that converted fluorescence into electropherograms representing distinct fragment lengths. Fragment lengths were calculated using an internal standard in each lane (GeneScan 400ROX, Applied Biosystems). GeneScan software was used to calculate LH-PCR product base pair

lengths of each 6-FAM-labeled product based on known sizes of a ROX-labeled internal standard run in each lane of the gel. We used nonparametric density estimation to identify the integer value associated with each fragment length. Specifically, we used the average shifted histogram estimator (Scott, 1992) with a very small bandwidth to identify the clusters.

2.2. Statistical analysis

2.2.1. Data preparation

Output from the MIDI (FAME) and Genescan (LH-PCR) software were stored in ASCII files and processed with an AWK program available for free from GNU Software, (http://www.gnu.org/directory/gawk.html) to create files for import into the statistical package R. R is a freely available "computer language like the S language developed at AT&T Bell Laboratories by Rick Becker, John Chambers and Allan Wilks " (http://www.R-project.org).

The imported data sets were created using the area under each peak as a percentage of the total peak areas. The data were normalized using the mean and variance of the whole data set, and data converted to a data frame within R for analysis.

2.2.2. Variable selection

A simulated annealing routine (Brown et al., 1999) was used to identify a subset of MIDI-FAMEs and LH-PCR fragments suitable for explaining variation between experimental groups or categories. Originally developed to model the physical process of annealing metals (Metropolis et al., 1953), this iterative routine has since been applied to wide variety of optimization problems. Although the variables in the routine are still labeled to identify parameters specific to the metal annealing process (i.e. 'temperature', 'cost'), as used for our application, these parameters only serve as place-holding variables that were arbitrarily altered to increase the stringency of the variable selection process. We used Matlab software (The Math Works, Natick, MA, USA) and code for this procedure available at http://stat.tamu.edu/~mvannucci/webpages/codes.html. The simulated annealing used $\gamma^0 = 1$ and the starting temperature was chosen by reversing the annealing. The temperature value was increased in 10° increments until the acceptance ratio equaled 1. The starting temperature used in the final selection was 500°. The adding, deleting, and swapping steps were chosen with probabilities P_A and $P_D = 1/3$. The acceptance ratio was calculated every 500 iterations ($m = 2000$), and the search stopped when the acceptance ratio reached 0. The cost parameter (C) was the major determinant for the number of variables selected with $C = 0.275$ and 0.25 for the LH-PCR and MIDI-FAME data, respectively. The determinism parameter (w) was 0.01 for the LH-PCR and 0.125 for MIDI-FAME. These parameters were used to guide the search algorithm in selection or rejection of updated solutions.

2.2.3. Discriminant analysis

We used discriminant analysis, also referred to as supervised pattern recognition to investigate FAME and LH-PCR sample profiles. A set of replicated MIDI-FAME and LH-PCR observations from known soil sources, called the training set, was used to develop rules to assign future observations (MIDI-FAME and LH-PCR data from unknown soil sources) to group membership. We used three variations of discriminant analysis including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and nonparametric density estimation. LDA was used to apply linear classification rules under the assumption of normal variance for all categories. QDA was used to apply nonlinear classification which permitted different variances for each category (McLachlan, 1992). Kernel density estimation was used to apply a model that made no assumptions about distributions. Kernel density estimation was conducted by calculating a probability density function from MIDI-FAME and LH-PCR observations (Scott, 1992).

We used a 10-fold cross-validation procedure to estimate LDA predictions. By random selection, the data were roughly split into ten pieces and in each fold, one of the ten pieces was predicted using the remaining 90% of the data (Efron and Tibshirani, 1993).

2.2.4. Software

The operating system used was RedHat 6.2 Linux, kernel version 2.2.12-20. Statistical software included MatLab version 5.3 (Mathworks, Inc. Natick, MA) and R version 1.1.1. The discriminant analyses were conducted using the *lda* and *qda* functions. The *lda* and *multinom* functions are part of the MASS library available at http://www.stats.ox.ac.uk/pub/R/, and are described in Venables and Ripley (1999). We used a variation of the code available in the online compliments at that site for the cross-validation and estimation procedures.

3. Results

3.1. Identification of MIDI-FAME and LH-PCR peaks

The 73 MIDI-FAMEs detected in soil samples represented fatty acid chain lengths ranging from 9 to 20 carbons and included signature fatty acids for Gram positive and Gram-negative bacteria (Zelles, 1995, 1997; Table 1). Thirty-four LH-PCR products ranging in size from 316 to 369 bp were amplified from the soil samples (Fig. 1). Due to differences in electrophoretic mobility of the two fluorophores, many of the LH-PCR product lengths were reported in the output file as fractional values. Because we knew these lengths were integers, we treated the differences as variation about a true value and applied cluster analysis to assign integer values to each observation. Many of the clusters were distinct, but the region from 340 to 360 bp

Table 1
Fatty acids detected in soil samples

9:0	15:1 ISO G[a]	18:1 2OH
9:0 3OH	16:0	18:1 ISO H
10:0	16:0 10 methyl[b]	18:1 ω5c
10:0 2OH	16:0 2OH	18:1 ω7c
10:0 3OH	16:0 3OH	18:1 ω9c
11:0	16:0 ANTEISO	18:3 ω6c (6,9,12)
11:0 ISO	16:0 ISO[b]	unknown 18.814
11:0 ISO 3OH	16:0 N alcohol	19:0
12:0	16:1 2OH	19:0 ANTEISO
12:0 2OH	16:1 ISO G	19:0 ISO
12:0 3OH	16:1 ISO H	19:1 ISO I
12:0 ANTEISO	16:1 ω5c	20:0
12:0 ISO	17:0	20:0 ISO
12:1 3OH	17:0 10 methyl	20:1 ω7c
13:0 ANTEISO	17:0 ANTEISO	20:1 ω9c
13:0 ISO	17:0 CYCLO	20:2 ω6c (6,9)
14:0	17:0 ISO	20:4 ω6c (6,9,12,15)
14:0 ISO	17:0 ISO 3OH	Summed features/peak overlaps
Unknown 14.263	17:1 ω7c[c]	14:0 3OH/16:1 ISO I
15:0	17:1 ω8c	16:1 ω7c/15 ISO 2OH[c]
15:0 2OH	ISO 17:1 ω9c	17:1 ANTEISO B/i I
15:0 ANTEISO	ANTEISO 17:1 ω9c	18:2 ω6,9c/18:0 ANTEISO
15:0 ISO	18:0	19:1 ω9c/19:1 ω11c
15:0 ISO 3OH	18:0 3OH	19:0 CYCLO ω10c/19ω6
TBSA 10Methyl 18:0		18:1 ω8c/18:3 ω3c

FAMEs were identified and designated using MIDI Sherlock software.
[a] A unique 15:1 fatty acid in which the location of the double bond was not determined by MIDI software.
[b] Gram-positive bacterial signature fatty acids.
[c] Gram-negative bacterial signature fatty acid.

required some method of classification. As an alternative to cluster analysis, we simply rounded the fragment length values to the nearest integer. This produced slightly different classifications, but the discriminant analysis was not affected. The key subsets of MIDI-FAME and LH-PCR variables identified by the simulated annealing technique and used for the analyses are listed in Table 2.

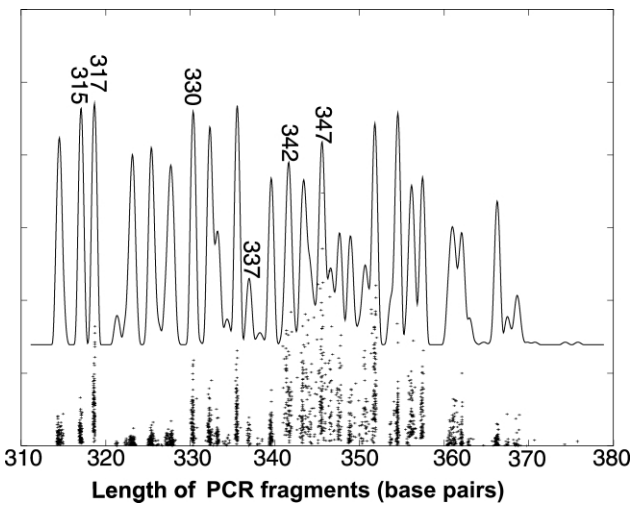

Fig. 1. Lengths of eubacterial 16S rDNA fragments amplified by LH-PCR from soil samples. A Bayesian approach was used to select a subset of these (315, 317, 330, 337, 342, and 347 bp) for comparison of soil samples by discriminate analyses. Peaks represent scanned electrophoregrams containing PCR products of discrete lengths indicated. ++++ indicate band intensities from which peaks were calculated.

Table 2
Key FAME and LH-PCR products selected as variables by a Bayesian approach

Fatty acid peaks		
15:1 ISO G[a]	16:0 10 methyl	
16:0 ISO	17:1 ω7c	
18:0	16:1 ω7c/15 ISO 2OH	
LH-PCR lengths in base pairs		
315	317	330
337	342	347

A simulated annealing routine was used to select a subset of MIDI-FAMEs and LH-PCR fragments for use as variables in pattern recognition statistical procedures.
[a] A unique 15:1 fatty acid in which the location of the double bond was not determined by MIDI software.

3.2. Comparison of MIDI-FAME and LH-PCR

After trying a large number of combinations including a stepwise discriminant analysis procedure, we found that two MIDI-FAMEs [16:0 10 methyl (Gram-positive) and 17:1 ω7c (Gram-negative)] explained most of the classification. LDA had a cross-validated error rate of 13% using these two variables. QDA had an identical cross-validated error rate of 13%. Because the fatty acid 18:0 can be used as a marker for total community biomass (Cavigelli et al., 1995), it was selected as a third variable. The use of these three MIDI-FAMES provided accurate classifications of soils subjected to contrasting tillage and ground cover (Table 3). The cross-validated error rate with these three MIDI-FAMEs was 10.9% using LDA and 8.7% using QDA (classification results not shown). Relative to FAME analysis, the predictive power of LH-PCR for land-cover and tillage practices

Table 3
Accuracy of FAME in classifying samples from soils subjected to contrasting tillage

	Predicted number of samples by			
	Conventional till		No-till	
	Between row	On row	Between row	On row
Conventional till, between row	*11*	0	1	0
Conventional till, on row	1	*10*	0	0
No-till, between row	1	0	*10*	1
No-till, on row	0	1	0	*10*

FAME profiles from soil samples were compared to profiles developed for soils subjected to known tillage practices using LDA. Rows display total numbers of soil samples classified for each soil category. Italicized numbers represent total correct predictions, Error rate = 10.9%.

Table 4
Accuracy of LH-PCR in classifying samples from soils subjected to contrasting tillage

	Predicted number of samples by			
	Conventional till		No-till	
	Between row	On row	Between row	On row
Conventional till, between row	*6*	2	3	1
Conventional till, on row	3	*6*	0	2
No-till, between row	3	0	*6*	3
No-till, on row	2	1	6	*2*

LH-PCR profiles from soil samples were compared to profiles developed for soils subjected to known tillage practices using LDA. Rows display total numbers of soil samples classified for each soil category. Italicized numbers represent total correct predictions, Error rate = 56.5%.

was poor (classification error rates of 56.5% using LDA) when samples were extracted from adjacent areas in the same field, even when using all six variables (Table 4). There was no significant correlation between the MIDI-FAME and LH-PCR variables, an expected result given that the two data sets provided information about different components of the soil communities. Combined MIDI-FAME and LH-PCR data sets did not improve the prediction of land-use and tillage categories. The LH-PCR had greater accuracy in the classification of soils from distant locations, with error rates of approximately 26% for LDA and QDA (Table 5). The two locations were classified with no error, however, tillage practices were identified with a 30% error.

Table 5
Accuracy of FAME and LH-PCR in classifying soils collected from distant sites

	Predicted number of samples by		
	Corvallis		
	Conventional till	No-till	Pendleton
A. FAME classification results			
Conventional till	*19*	4	0
No-till	1	*22*	0
Pendleton	0	0	*8*
Error rate = 9.3%			
B. LH-PCR classification results			
Conventional till	*15*	8	0
No-till	6	*17*	0
Pendleton	0	0	*8*
Error rate = 25.9%			

Samples of soils subjected to contrasting tillage practices and to soils collected from distant sites were analyzed by FAME analysis and LH-PCR and classified using LDA. Rows display total numbers of soil samples classified for each soil category and italicized numbers represent total correct classifications.

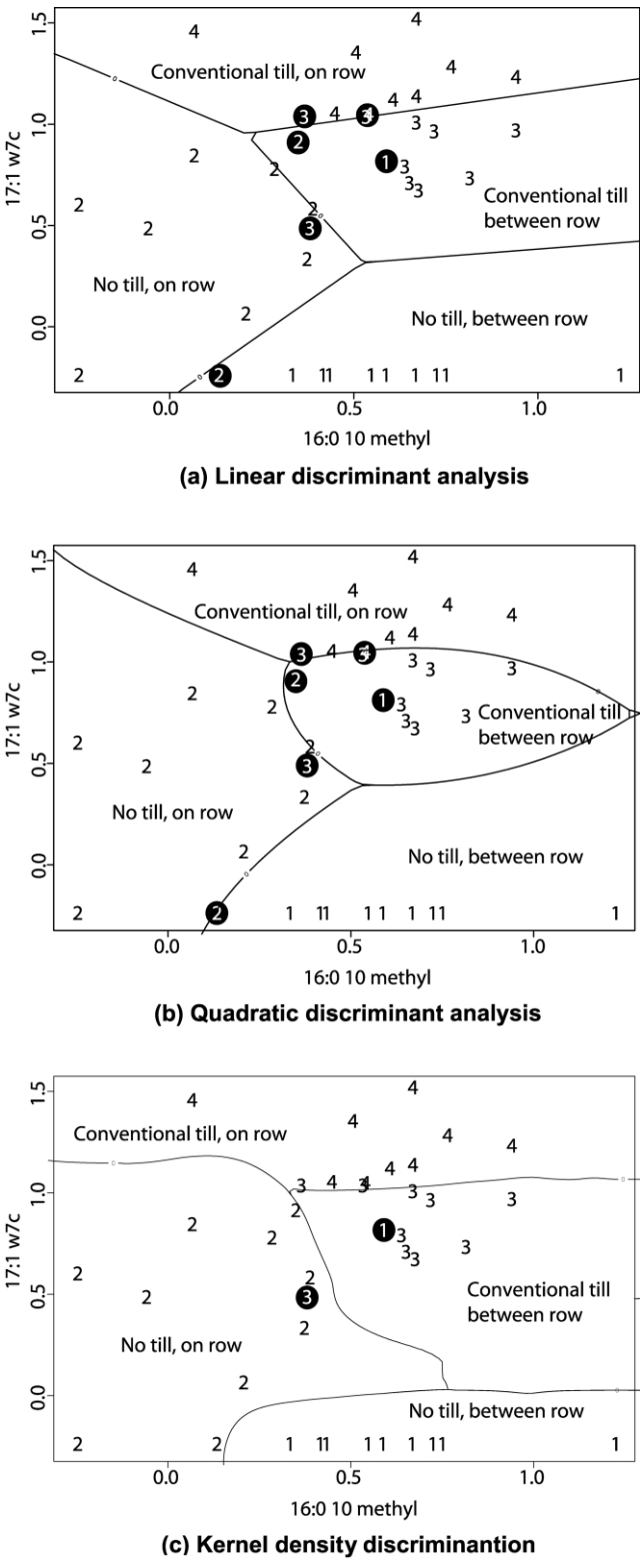

Fig. 2. Classification of soil samples collected from contrasting tillage plots by utilization of contour plots of relative amounts of two FAMEs (17:1 Ω7c vs 16:0 10 methyl) generated with contrasting statistical analyses. (a) LDA; (b) QDA; (c) nonparametric density estimation. (1) no-till, between row, (2) no-till, on row, (3) conventional till, between row, (4) conventional till, on row. Numbers in black circles represent observations that were classified incorrectly.

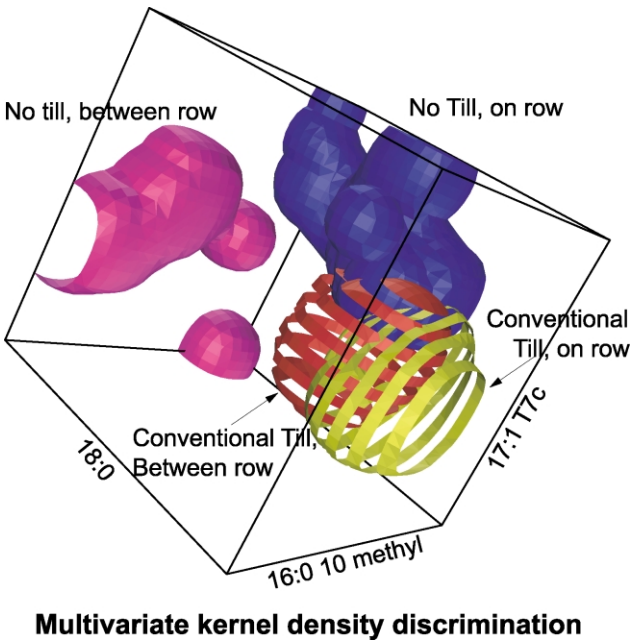

Fig. 3. Kernel density estimation (90% probability surfaces) using three FAMEs to differentiate soils samples collected within field plots subject to contrasting tillage.

3.3. Comparison of statistical approaches

In order to compare discriminant analyses that used contrasting assumptions about the data distribution, we developed contour plots based on the abundance of two FAMEs selected for analysis. Fig. 2(a) shows the classification decisions plotted within the straight lines characteristic of the LDA method with six incorrect predictions. A comparison of Fig. 2(b) with Fig. 2(a) shows that relaxation of the requirement for identical variances did not improve the classification decision for observations close to boundaries and also provided six incorrect predictions. Kernel density estimation, where assumptions about the underlying distributions are dropped, provided the best results with only two incorrect predictions (Fig. 2(c)). The kernel density estimate of the probability density functions of the four categories when the 18:0 fatty acid was added to the analysis is shown in Fig. 3. The shell of each category enclosed an area where there was a 90% probability of observing a particular combination of FAMEs associated with that category. Estimates representing soil from no-till and between rows sampling were widely separated from the other three categories.

4. Discussion

4.1. Selection of variables from MIDI-FAME and LH-PCR data sets

Our hypothesis was that utilization of statistical procedures combining FA and DNA component analyses would increase the capability for spatial resolution among soil biological communities. The first task in the study was to develop an algorithm for analysis that used all prior information. The problem had two parts, variable selection and pattern recognition. Selection of a reduced number of variables was necessary for application of traditional statistical methods to FAME and LH-PCR data because the data sets contained more variables (FA's and PCR products) than observations. In our study we had 46 soil samples that represented four categories along with 72 FAMES and 34 LH-PCR fragments as potential variables to explain differences between the categories. We used a recently developed Bayesian variable selection procedure based on simulated annealing, to select a subset of these variables for potential regressors. The variable selection procedure, simulated annealing, found a useful subset of variables for discriminant analysis. Four of the six FA's identified in our subset (Table 2) have been suggested as signature FA's for specific groups of microorganisms (Zelles, 1995, 1997). The FA's 16:0 ISO, 16:1 ω7c/15 ISO 2 OH (Summed feature), and 16:0 10 methyl are associated with Gram-positive bacteria and 17:1 ω7c with Gram-negative bacteria. Use of the latter two provided analyses with a classification error rate of only 13%. The addition of a marker for total microbial biomass (18:0) to the variable subset decreased the classification error rate to 8.7%. The fact that four of six FAMEs in our subset are 'signature' FA's for bacteria and the other two FAMEs are not specific for fungi suggests that differences in bacterial communities are primarily responsible for community differences in this sample set. In general, both no-till and conventional tillage favor the development of fungal-dominated communities, however, specific community members tend to differ (Holland and Coleman, 1987). Spatial stratification of organic matter exerts a dominant selective pressure in no-till soils while the predominant influence in conventional-tilled soils is the relatively low soil moisture (Paul and Clark, 1989). In contrast, two reports suggest that bacteria tend to dominate in organic soils relative to that of conventionally tilled soils (Scow et al., 1994; Gunapala and Scow, 1998). The soil samples we analyzed were of similar composition and differed primarily in tillage practice and plant cover. Woodburn silt loam is relatively high in clay content and is saturated most of the winter at the Hyslop site so soil conditions prevail that do not favor fungal dominance of the soil community.

4.2. Comparison of statistical analyses of MIDI-FAME, LH-PCR and combined data sets

We chose discriminant analysis for supervised classification of observations. Three variants were applied, each with a greater relaxation of underlying assumptions: LDA, based on assumption of a continuous distribution (i.e. means and variances are identical and normal for all variables),

QDA, which assumes a discrete distribution for the classes (variances may be different for each variable), and nonparametric density estimation, where means, variances, and shape of distribution are all estimated (rather than assumed) for each variable. Our primary application used data that represented multiple samples from a single experimental plot with differences in tillage and ground cover. Consequently, the resolution required of the discriminant analysis was much higher than attempted in previous studies that sampled geographically separated plots. We compared the results of both methods of discriminant analysis with MIDI-FAME, LH-PCR, and combined data sets. Although FAME and LH-PCR measure different components of soil communities, expression of data from both procedures as % of total peak areas permitted the use of combined data sets in the discriminant analyses.

Comparisons of methods for analyzing changes in soil communities have been made (Franzluebbers et al., 1995), but reports of combining data sets generated from two different methods for increased resolution are rare (Engelen et al., 1998). In order to test our hypothesis that combined analyses would permit greater spatial resolution between contrasting soil communities, we chose two methods (MIDI-FAME and LH-PCR) that demonstrated similar resolving capability. A previous comparison of these two methods found no difference in their respective resolutions when measuring populations from discrete soil samples (Ritchie et al., 2000). In contrast, we found LH-PCR unable to discriminate between land use practices in adjacent soils whereas MIDI-FAME provided high resolution. When a geographically distant soil was analyzed, the predictive resolution of LH-PCR improved (Table 5) although not to the level of accuracy using MIDI-FAME alone. One possible explanation is that variable selection for the two methods was conducted from a small number of LH-PCR product sizes (34 lengths) relative to the 74 different FA's identified by MIDI-FAME. It should be noted that the NaOH procedure used to extract FA's from the soil also extracted FA's representing nonmicrobial soil organic matter. Consequently, there is no assurance that all 74 MIDI-FAME peaks represented soil microbial communities. Although LH-PCR bands are not directly interpreted in phylogenetic terms, their fingerprints represent the frequency distribution of sequences in the sample. Changes in fingerprints indicate variations in the community composition. The assignment of one band to one species is not possible since nonidentical sequences may be found at the same position by chance and bias in community members may be introduced by different numbers of 16S RNA genes per genome (Farelly et al., 1995). Changes in bacterial community members might result in the substitution of one LH-PCR band for another of the same length. It may be necessary to isolate and sequence products identified by the Baysian approach to identify responsive community members. Perhaps the resolving power of LH-PCR in discriminating subtle differences in bacterial populations in similar soils subjected to differing land cover and tillage practices could be enhanced if only the metabolically active fraction of the population was assayed. This might be accomplished by bromodeoxyuridine (BrdU) incorporation and immunocapture prior to LH-PCR as reported by Urbach et al. (1999) with the caveat that the range of bacteria that fail to incorporate BrdU is not known.

The nonparametric method provided an error free separation of soil collected from a no-till/between row location from the other treatments (Figs. 2 and 3). The 18:0 MIDI-FAME clearly added to the discrimination power to separate no-till/between row from the other soil samples. No-till/on row overlapped the other two categories to some degree, although less with conventional-till/on row than conventional-till/between row (Fig. 3). The two categories of conventional-till (on row and between row), representing the greatest amount of soil disturbance, had the most overlap. In agreement with Drijber et al. (2000), these analyses suggest that the greater the difference between soil disturbance among categories, the more readily the categories were discriminated by MIDI-FAME.

Linear methods worked well for these data, but we concluded that minimizing assumptions made about the probability structure (i.e. the distribution of the data) improved the accuracy of the predictions. It is noteworthy that purely statistical methods of variable selection resulted in the choice of biologically meaningful (i.e. signature) FA's to explain the observed categories. The MIDI-FAME analyses provided high resolution of tillage type, ground cover, and location. Although use of LH-PCR did not discriminate as well in ground cover, it was effective at identifying location. The combination of these two analytical approaches that provide information about different aspects of the soil biological community did not enhance our capacity to identify the source of soil.

Acknowledgements

We wish to thank Nancy J. Ritchie, Oregon State University, for assistance with the LH-PCR analyses and Gene Scan. We also wish to acknowledge the Oregon State University Center for Gene Research for the use of their DNA sequencer and Gene Scan software.

References

Avaniss-Aghajani, E., Jones, K., Chapman, D., Brunk, C., 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. Biotechniques 17, 144–149.

Bossio, D.A., Scow, K.W., Gunapala, N., Graham, K.J., 1998. Determinants of soil microbial communities: effects of agricultural management, season, and soil type on phospholipid fatty acid profiles. Microbial Ecology 36, 1–12.

Brown, P.J., Fearn, T., Vannucci, M., 1999. The choice of variables in

multivariate regression: a non-conjugate Bayesian decision theory approach. Biometrika 86, 635–648.

Bruce, K.D., 1997. Analysis of *mer* gene subclasses within bacterial communities in soils and sediments resolved by fluorescent-PCR-restriction fragment length polymorphism profiling. Applied and Environmental Microbiology 63, 4914–4919.

Buyer, J.S., Drinkwater, L.E., 1997. Comparison of substrate utilization and fatty acid analysis of soil microbial communities. Journal of Microbiological Methods 30, 3–11.

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

Drijber, R.A., Doran, J.W., Parkhurst, A.M., Lyon, D.J., 2000. Changes in soil microbial community structure with tillage under long-term wheat-fallow management. Soil Biology & Biochemistry 32, 1419–1430.

Dunbar, J., Takala, S., Barns, S.M., Davis, J.A., Kuske, C.R., 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Applied and Environmental Microbiology 65, 1662–1669.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap, Chapman and Hall, New York.

Engelen, B., Meinken, K., von Winzingerode, F., Heuer, H., Malkomes, H.-P., Backlaus, H., 1998. Monitoring impact of a pesticide treatment on bacterial soil communities by metabolic and genetic fingerprinting in addition to conventional testing procedures. Applied and Environmental Microbiology 64, 2814–2821.

Farelly, V.F., Rainey, F., Stackerbrandt, E., 1995. Effect of genome size and gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Applied and Environmental Microbiology 61, 2798–2801.

Franzluebbers, A.J., Zuberer, D.A., Hons, F.M., 1995. Comparison of microbiological methods for evaluating quality and fertility of soil. Biology and Fertility of Soils 19, 135–140.

Glucksman, A.M., Skipper, H.D., Brigmon, R.L., Santo Domingo, J.W., 2000. Use of the MIDI-FAME technique to characterize groundwater communities. Journal of Applied Microbiology 88, 711–719.

Griffiths, B.S., Ritz, K., Ebblewhite, N., Dobson, G., 1999. Soil microbial community structure: effects of substrate loading rates. Soil Biology & Biochemistry 31, 145–153.

Gunapala, N., Scow, K.W., 1998. Dynamics of soil microbial biomass and activity in conventional and organic farming systems. Soil Biology & Biochemistry 30, 805–816.

Holland, E.A., Coleman, D.C., 1987. Litter placement effect on microbial and organic matter dynamics in an agroecosystem. Ecology 68, 425–433.

Ibekwe, A.M., Kennedy, A.C., 1998. Phospholipid fatty acid profiles and carbon utilization patterns for analysis of microbial community structure under field and greenhouse conditions. FEMS Microbial Ecology 26, 151–163.

Kennedy, A.C., 1999. Bacterial diversity in agroecosystems. Agricultural Ecosystems and the Environment 74, 65–76.

Kennedy, A.C., Busacca, A.J., 1995. Microbial analysis to identify source of PM-10 material. In: Weiskircher, R.J., Zwikl, J.R. (Eds.), Particulate Matter: Health and Regulatory Issues, Air and Waste Management Association, Pittsburgh, pp. 670–675.

Kennedy, A.C., Gewin, V.I., 1997. Soil microbial diversity: present and future considerations. Soil Science 162, 607–617.

McCaig, A.E., Glover, L.A., Prosser, J.I., 1999. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Applied and Environmental Microbiology 65, 1721–1730.

McCune, B., Mefford, M.J., 1997. PC-ORD, Multivariate analysis of ecological data. Version 3.0. MJM Software, Gleneden Beach, OR.

McLachlan, G.J., 1992. Discriminant Analysis and Statistical Pattern Recognition, Wiley, New York.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. Journal of Chemistry and Physics 21, 1087–1092.

Paul, E.A., Clark, F.E., 1989. Soil Microbiology and Biochemistry, Academic Press, San Diego.

Ratledge, C., Wilkinson, S.G., 1988. Microbial Lipids, vol. 1. Academic Press, London.

Ritchie, N.J., Schutter, M.E., Dick, R.P., Myrold, D.D., 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. Applied and Environmental Microbiology 66, 1668–1675.

Scott, D.W., 1992. Multivariate Density Estimation, Wiley, New York.

Scow, K.M., Somasco, O., Gunapala, N., Lau, S., Venette, R., Ferris, H., Miller, R., Shennan, C., 1994. Changes in soil fertility and biology during the transition from conventional to low-input and organic cropping systems. California Agriculture 48, 20–26.

Suzuki, M., Rappe, M.S., Giovannoni, S.J., 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Applied and Environmental Microbiology 64, 4522–4529.

Torsvik, V., Goksoye, J., Daae, F.L., 1990a. High diversity in DNA of soil bacteria. Applied and Environmental Microbiology 56, 782–787.

Torsvik, V., Salte, K., Surheim, R., Goksoyr, J., 1990b. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Applied and Environmental Microbiology 56, 776–781.

Urbach, E., Vergin, K.L., Giovannoni, S.J., 1999. Immunochemical detection and isolation of DNA from metabolically active bacteria. Applied and Environmental Microbiology 65, 1207–1213.

Venables, W.N., Ripley, B.D., 1999. Modern Applied Statistics with S-Plus, Springer Verlag, New York.

Wander, M.M., Hendrick, D.S., Kaufman, D., Traina, S.J., Stinner, B.R., Kehrmeyer, S.R., White, D.C., 1995. The functional significance of the microbial biomass in organic and conventionally managed soils. Plant and Soil 170, 87–97.

Zelles, L., Bai, O.Y., Rackwitz, R., Chadwick, D., Beese, R., 1995. Determination of phospholipid- and lipopolysaccharide-derived fatty acids as an estimate of microbial biomass and community structures in soils. Biology and Fertility of Soils 19, 115–123.

Zelles, L., 1997. Phospholipid fatty acid profiles in selected members of soil microbial communities. Chemosphere 35, 275–294.