Plant and Soil **170:** 99–113, 1995.

Fatty acid methyl ester (FAME) profiles as measures of soil microbial community structure

Michel A. Cavigelli[1], G. Philip Robertson[2] and Michael J. Klug[3]
[1]*W.K. Kellogg Biological Station and Center for Microbial Ecology, Michigan State University, Hickory Corners, MI 49060, USA;* [2]*W.K. Kellogg Biological Station and Department of Crop and Soil Sciences, Michigan State University, Hickory Corners, MI 49060, USA and* [3]*W.K. Kellogg Biological Station and Department of Microbiology and Public Health, Michigan State University, Hickory Corners, MI 49060, USA*

Key words: fatty acid methyl ester (FAME) analysis, geostatistics, principal components analysis, soil microbial community

Abstract

Analysis of fatty acid methyl ester (FAME) profiles extracted from soils is a rapid and inexpensive procedure that holds great promise in describing soil microbial community structure without traditional reliance on selective culturing, which seems to severely underestimate community diversity. Interpretation of FAME profiles from environmental samples can be difficult because many fatty acids are common to different microorganisms and many fatty acids are extracted from each soil sample. We used principal components (PCA) and cluster analyses to identify similarities and differences among soil microbial communities described using FAME profiles. We also used PCA to identify particular FAMEs that characterized soil sample clusters. Fatty acids that are found only or primarily in particular microbial taxa – marker fatty acids - were used in conjunction with these analyses. We found that the majority of 162 soil samples taken from a conventionally-tilled corn field had similar FAME profiles but that about 20% of samples seemed to have relatively low, and that about 10% had relatively high, bacterial:fungal ratios. Using semivariance analysis we identified 21:0 iso as a new marker fatty acid. Concurrent use of geostatistical and FAME analyses may be a powerful means of revealing other potential marker FAMEs. When microbial communities from the same samples were cultured on R2A agar and their FAME profiles analyzed, there were many differences between FAME profiles of soil and plated communities, indicating that profiles of FAMEs extracted from soil reveal portions of the microbial community not culturable on R2A. When subjected to PCA, however, a small number of plated communities were found to be distinct due to some of the same profile characteristics (high in 12:0 iso, 15:0 and 17:1 ante A) that identified soil community FAME profiles as distinct. Semivariance analysis indicated that spatial distributions of soil microbial populations are maintained in a portion of the microbial community that is selected on laboratory media. These similarities between whole soil and plated community FAME profiles suggest that plated communities are not solely the result of selection by the growth medium, but reflect the distribution, in situ, of the dominant, culturable soil microbial populations.

Introduction

Recent evidence suggests that culturable soil microorganisms may represent a tiny, possibly ecologically unimportant portion of the overall diversity present in most soils. DNA reassociation kinetics, for example, suggested the presence of at least 4,000 bacterial strains in 30g of a forest soil from southern Norway, with culturable strains representing less than 1% of the total present (Torsvik et al., 1990 a,b). Such studies are consistent with the typical finding that about 100 times more soil bacteria are seen with direct microscopy than are found using classical plating methods. Results such as these emphasize the need for new techniques that allow us to describe soil microbial community structure without the usual reliance on selective culturing, an approach that appears to severely underestimate the actual diversity of soil microbial communities. In recent years a number of useful approaches have been developed to address this challenge, including

16S rRNA probes, RFLP analyses of the products of PCR primers, %G + C profiles, and fatty acid methyl ester (FAME) profile analysis.

In this paper we present the results of an in situ assessment of the spatial distributions of soil microbial communities - part of an effort to better understand patterns, causes, and consequences of microbial diversity in soils of agronomic significance. We used FAME profile analysis for this task because of its unique ability to characterize whole communities rapidly and at relatively low cost.

The value of FAME analysis arises from the facts that there are a great number of different kinds of fatty acids in the lipids of microorganisms and that different organisms have different combinations of these fatty acids. Because fatty acids can be readily volatilized following methylation, they can be readily analyzed by gas chromatography (e.g. Moss et al., 1980; Moss 1981; Vestal and White, 1989). Many microbial isolates or taxa have unique FAME profiles (e.g. Mayberry et al., 1982; Viljoen et al., 1986), and the technique has been adapted to the examination of mixed communities from both sediments (Bobbie and White, 1980; Findlay and White, 1983; Hogg, 1984: Perry et al., 1979 and Volkman et al., 1980) and soils (pers. comm. H. Garchow and M. Klug, Michigan State University; Zelles et al., 1992; Zelles and Bai, 1993).

The interpretation of profiles from whole soil communities can be difficult because many fatty acids are common to different microorganisms and because there are hundreds of different fatty acids in environmental samples, especially in agricultural soils (Zelles et al., 1992). Thus far, FAME profile analysis has been largely limited to qualitative and univariate descriptions of the fatty acids present in environmental samples. We show in this paper that multivariate and geostatistical approaches to profile analysis can substantially aid the interpretation of FAME profile data. We also demonstrate the extent to which plated communities are representative of whole soil microbial communities by comparing results from whole soil FAME profiles with those of plated microbial communities extracted from the same soils.

Material and methods

Study site

This study was conducted at the W.K. Kellogg Biological Station's Long-Term Ecological Research Site in Row-Crop Agriculture, located in SW Michigan, in the northern portion of the U.S. corn belt. Soils of the site are Typic Hapludalfs (Kalamazoo series), sandy loams of moderate to high fertility. The Ap horizon has a mean CEC of 5.7 meq $100g^{-1}$, a mean pH of 7.0 (1:2 water), and a mean total C content of 1.0% (unpublished data).

Soil sampling

We collected 54 soil samples in July 1991 from an 80m transect in each of three replicate 1 ha agronomic plots (162 total samples) planted to conventionally-tilled corn (*Zea mays* L.). Transects were placed 10cm from and parallel to a chosen row of corn. Soils were sampled to 10cm depth at locations along each transect following a nested sampling scheme designed to minimize the number of samples required to capture statistically significant patterns of spatial autocorrelation; sample intervals were as small as 2 cm. Sampled soil was stored in plastic bags in a cooler immediately upon sampling and then later sieved (2 mm), mixed, and subsampled within 12 h for the analyses described below.

FAME analysis

Subsamples for FAME analyses (5 g) were stored at 4°C in ashed test tubes for 24 h and then processed according to the Microbial Identification System (MIS; Microbial ID Inc. 1992) standard protocol. First, lipids were saponified by adding 5.0 mL 3.25 *M* NaOH in methanol to each soil sample, then mixing, heating in a 100° water bath for 5 min, mixing again, and heating in the water bath for an additional 25 min. The samples were then methylated by adding 10 mL of 3.25 *N* HCl in methanol, mixed, placed in an 80°C water bath for 10 min, and then rapidly cooled in ice water. The FAMEs were extracted from this solution by adding 1.5 mL of one part methyl tert-butyl ether in one part (v:v) hexane and placing the closed tubes on a rotary mixer for 10 min. The top, organic phase was transferred by pipet to an ashed test tube and then washed using 3.0 mL of dilute NaOH. The organic phase was then transferred to a GC vial for subsequent analysis by gas-liquid chromatography using an HP 5890 (Hewlett Packard, Rolling Meadows, IL, USA) equipped with an HP Ultra 2 capillary column (crosslinked 5% Ph Me silicone, 25 m × 0.2 mm × 0.33 mm film thickness) and a flame ionization detector. One person can prepare 40 samples for analysis by gas chromatography in under 4 h using this protocol.

Table 1. Marker fatty acids. From Erwm (1973), White (1983), Harwood and Russell (1984), Jantzen and Bryn (1985), and Vestal and White (1989)

Eubacteria				
14:0 3 OH	15:0 br	17:0 br	17:0 cyc	18:1 cis 11
19:0 cyc	16:0 br 10	15:1 at 6	15:1 at 8	17:1 at 6
17:1 at 8	15:1 iso 3	15:1 iso 7		
trans monounsaturated branched and straight 16 and 17 C				
(eubacteria do not, in general, contain polyunsaturated fatty acids)				
Gram negative eubacteria				
OH fatty acids (usually 3 OH)				
Gram positive eubacteria (but also found in Gram negatives)				
branched fatty acids (iso, anteiso)				
Eukaryotes				
12:0	16:1 at 7	18:2 cis 9, cis 12	18:1 at 9	
α-18:3 cis 9, cis 12, cis 15	polyunsaturated fatty acids with > 20 C			
Protozoa				
20:3 at 6	20:4 at 6			
Microfauna				
γ-18:3				
All organisms				
14:0	16:0	18:0		

Since whole cell fatty acids have proven sufficient to distinguish microbial communities from soils which differ only in the type of agricultural management to which they have been subjected (Klug and Tiedje, 1993), we did not separate the phospholipids from the rest of the lipid fraction prior to saponifying the lipids.

We report our results (e.g. Table 1) using standard FAME nomenclature: the number of C atoms in the fatty acid is indicated by the number before the colon and the number after the colon indicates the number of double bonds. "Ante" means anteiso, and "br" means branched, either at the iso or anteiso positions. "Cyc" refers to the cyclo propane analogue, and numbers following a C indicate the location of the epoxy bond. Use of "at" indicates that cis or trans configuration is not known and the number(s) following the cis, trans, or at designation indicate(s) the position of the double bond(s) relative to the carboxyl end of the molecule. The number before an OH refers to the location of a hydroxy substitution relative to the carboxyl end of the molecule. A capital letter at the end of a monounsaturated acid indicates that the position of the double bond is not known but that the bond is at a different location than for other monounsaturated FAs of the same chain length and branching pattern that have a different letter designation. Sif means "sum in feature" and indicates that more than one FAME has a particular retention time. For example, sif 13 is a combination, in unknown proportions, of 19:1 trans 11, an unknown, and 19:0 cyc C9-10. Interpretation of profiles has been aided by the use of FA markers - those FAs found only or principally in particular groups of organisms (Vestal and White, 1989); marker FAs identified by various investigators are listed in Table 1.

Plate cultures

Subsamples used to culture microbial communities (1g) were refrigerated at 4°C for up to 24h prior to preparing a dilution series in phosphate buffer (pH 7.2). One ml, from each 10^{-2} dilution tube was transferred to a 150 × 10 mm plate containing R2A agar. Plates were incubated at room temperature and incubated for 36-40h; growth was rapid. Plates were then stored at 4°C for no more than 48h. The entire plated community was scraped into an ashed test tube and returned to the refrigerator for up to 48h. FAME analyses were conducted as for soil samples but using smaller quantities of reagents.

Microbial biomass

From the same dilution series as used for plating communities, acridine orange direct counts (AODC) were prepared for each sample from one of the transects. Acridine orange (0.5 mL of 1% solution) was added to 10 mL of the 10^{-3} dilution. Three mL of this solution were filtered through a 0.2μm Millipore membrane and the membrane was then washed with 6 mL of isopropanol. The filter was placed on a slide and 20 fields were counted at 63X. Counts were converted to bacterial biomass values based on an average size of 0.196 mm^3 $bacterium^{-1}$, calculated from measurements made previously on bacteria from the same site (pers. comm. D. Harris, Michigan State University).

Statistical analysis

We used principal components analysis (PCA; SAS Institute, 1991) to compare FAME profiles among soil samples and, separately, among plated communities. PCA, like other multivariate statistical methods, is

used to summarize data in which multiple variables have been measured for each sample. In PCA the original variables (FAMEs in this case) are orthogonally transformed into a new set of uncorrelated variables called principal components (PCs). Since each PC is a linear combination of the original variables, all original variables are represented in each PC. The degree to which each PC is influenced by each original variable is its eigenvector loading, the sign of which indicates the manner in which the original variable influences the PC. The variance associated with each PC decreases in order, so that much of the variability of the original variates may be accounted for in the first few PCs. A subset composed of those variables most important in discriminating among samples (i.e. with the highest loadings in the primary PCs) can be selected and used for further analysis and interpretation (Digby and Kempton, 1987; Joliffe, 1986 and Liu and Keister).

We used the correlation matrix rather than the covariance matrix in PCA since the standard deviations of some of the individual FAMEs were large compared to the means (Joliffe, 1986). Results of PCA may be presented as biplots - plots of eigenvector loadings of each variable, usually scaled and superimposed on a PC plot of the sample points. The farther a variable plots from the origin, the more influential it is in a particular PC. The relationship among variables can be determined by directional cosines of the angle formed between lines drawn from the origin to each variable. Variables at 90° to each other have no effect on each other (cos 90° = 0) and variables at 180° to each other have opposing effects (cos 180° = 1).

We also subjected the data to hierarchical cluster analysis using NTSYS (Applied Biostatistics Inc., 1992). We constructed dendrograms for most matrices for which we used PCA, using all combinations of Euclidean and average taxonomic distance metrics, and UPGMA (unweighted pair-groups method using arithmetic averages) and single linkage. We chose linkage methods that have been shown to be biased to the formation of different types of clusters (Everitt, 1980; Milligan and Cooper, 1987). Cluster analysis has the advantage over PCA in that all dimensions of a data matrix can be represented in a two dimensional dendrogram, but the disadvantage that variables contributing to the clustering cannot be identified (Everitt, 1980; Milligan and Cooper, 1987).

During processing and FAME analysis of samples, seven whole soil and 19 plated community subsamples were lost or did not provide sufficient material for analysis; multivariate statistical analyses were conducted on the remaining 136 samples shared by both soil and plated community data sets.

We used semivariance analysis, a geostatistical technique, to describe the degree of spatial autocorrelation of individual FAMEs, i.e. to quantify the degree to which soil samples taken close to one another are more similar with respect to quantities of individual FAMEs than samples taken farther apart. Such information can help to identify the scale at which controls on microbial populations containing particular FAMEs operate (Robertson and Gross, 1994; Trangmar et al., 1985) as well as to identify FAMEs unique to specific microbial communities.

Semivariance, γ(h), is defined mathematically as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where *z(xi)* is the value of a variable for a sample taken at point x_i and $z(x_i + h)$ is the value of the variable for another sample taken at a distance, or lag interval, h from point x_i. N(h) is the total number of sample pairs separated by a distance h (Webster, 1985).

Semivariance analysis results in a variogram (e.g. Fig. 4) of semivariance for all possible lag intervals within a given spatial domain. If a continuous variable is sampled at an appropriate scale, γ(h) decreases as h decreases to zero; a variate is perfectly autocorrelated with itself at the same location. The population variance is estimated by the sill, that portion of the curve where γ(h) is constant. The y-intercept is called the nugget variance. The portion of the population variance due to spatial structure is the difference between the sill and the nugget variance. The lag distance at which the sill is reached is the range and indicates the distance over which a variate is autocorrelated (Webster, 1985).

Results and discussion

We present FAME data in units of percent of total FAMEs within a sample. The MIS protocol we adapted to whole soil samples is highly reproducible; e.g. the standard deviations for five subsamples taken from a well-mixed sample from the same corn field were very low compared to the means (Fig. 1).

On average, 24 different FAMEs were detected in each whole soil sample. Fifty-six FAMEs were identified in total (Table 2) and more than 80% of these were detected in the first 12 samples, indicating that most of

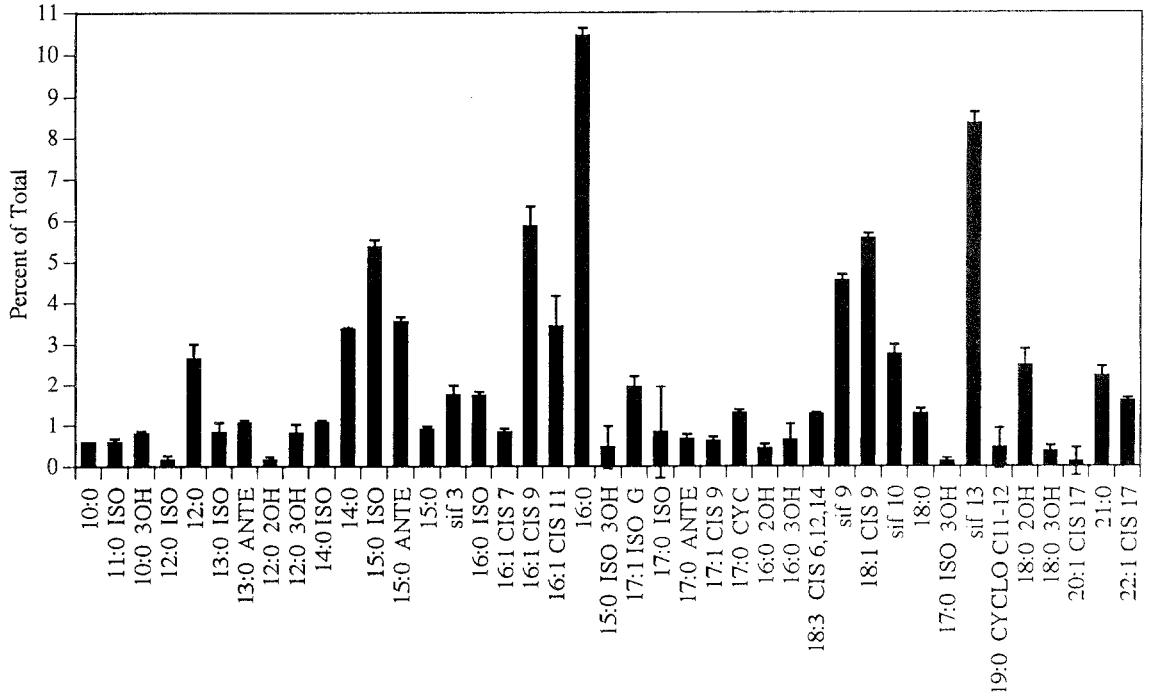

Fig. 1. Means and standard deviations (represented by error bars) of FAMEs for five subsamples from one homogenized sample of soil taken from a corn field.

the diversity detected in this study was present within a very small portion of the study area.

Zelles et al. (1992) found that phospholipid palmitic acid, 16:0, was highly correlated with a number of measures of microbial biomass. We did not find any relationship between AODC-derived biomass measurements and the proportion of 16:0 or any other FAME common to all organisms or any FAME specific to bacteria.

Principal components analysis

FAME profiles of whole soil communities included FAMEs with GC retention times of up to 23 min; those of plated communities included only FAMEs with retention times of less than 17 min. We conducted PCA separately on two whole soil community data matrices: a complete matrix including all 56 FAMEs and an abbreviated matrix that included only those 46 FAMEs with retention times less than 17 min. Because the results of the PCA were very similar for both the complete and abbreviated soil profiles, we present results only from the abbreviated profiles here. Since a number of FAMEs specific to plants (Vestal and White, 1989; White, 1983) have retention times on the GC greater than 23 min using this protocol, the lack of difference between complete and abbreviated profiles is evidence that FAMEs from roots and other plant materials were probably not an important component of the FAME profiles in these samples.

PCA of FAME profiles from whole soil communities showed that there were three distinct clusters of soil samples (Fig. 2A). The symbols used in Figure 2A and other PCA plots (Figs. 2B-C, 5A-B) serve only to identify soil samples that clustered together; since the same symbol is used for each sample throughout this paper comparisons can be made among plots. All but one of the 25 samples forming the small cluster of stars in Figure 2A were taken within a 58 m section of one of the three transects, indicating a patchy distribution of microbial communities at a large scale. The soil samples represented by diamonds were more evenly distributed across all three transects, indicating spatial variability on a different scale.

Since only 39% of the total variability was explained in the first three PC dimensions (Table 3), individual soil samples may be less similar than they appear in Figure 2A. Nonetheless, the separation of samples into three distinct clusters can be interpreted by identifying the individual FAMEs with the highest loadings in the eigenvectors of the PCs along which clusters are separated. This type of analysis is usually

Table 2. Fatty acids found in soil and in plated communities. Those fatty acids found in both communities are indicated in boldface. Fatty acids with retention times > 17 mins on the GC column and removed for the whole soil analyses are listed as a separate group

Soil communities			
Monounsaturated cis	*Saturated Even*	*Saturated Odd*	*3 and 12 OH*
16:1 cis 7	10:0	9:0	**10:0 3 OH**
16:1 cis 9	**12:0**	**15:0**	**12:0 3 OH**
16:1 cis 11	**14:0**	**17:0**	15:0 iso 3 OH
17:1 cis 10	**16:0**		16:0 3 OH
18:1 cis 9	**18:0**	*Branched: anteiso*	**17:0 iso 3 OH**
18:1 cis 15		**13:0 ante**	**18:0 3 OH**
	Branched: iso	**15:0 ante**	18:0 12 OH
Monounsaturated ND	**11:0 iso**	**17:0 ante**	
(*configuration*	**12:0 iso**	**19:0 ante**	*Retention Time*
not determined)	**13:0 iso**		*> 17 mins*
16:1 iso G	**14:o iso**	*Polyunsaturated*	20:1 cis 17
17:1 iso G	**15:0 iso**	18:3 cis 6, 12, 14	20:1 at 9
17:1 ante A	**16:0 iso**	20:4 cis 14	21:0 iso
16:1 2 OH	**17:0 iso**	20:5 cis 17	21:0
sif 10 (18:1 cis 11,	**19:0 iso**		21:1 cis 18
18:1 trans 9,		*2 OH*	21:1 cis 17
18:1 trans 6)	*Cyclo*	16:0 2 OH	22:0 3 OH
	17:0 cyc	18:0 2 OH	24:1 at 9
			25:0 2 OH
Plated communities			
Monounsaturated ND	*Monounsaturated cis*	*Branched: iso*	*Cyclo*
(*configuration*	**16:1 cis 9**	**11:0 iso**	**17:0 cyc**
not determined)	**18:1 cis 9**	**12:0 iso**	19:0 cyc C11-12
16:1 iso E		**13:0 iso**	
16:1 iso G	*Monounsaturated trans*	**14:0 iso**	*2 OH*
16:1 A	19:1 trans 7	**15:0 iso**	12:0 2 OH
17:1 iso E		**16:0 iso**	
17:1 iso H	*Saturated Even*	**17:0 iso**	*3 OH*
17:1 ante A	**12:0**	**19:0 iso**	**10:0 3 OH**
17:1 B	**14:0**		**12:0 3 OH**
sif 5 (17:1 iso I,	**16:0**	*Branched: anteiso*	**17:0 iso 3 OH**
17:1 ante B)	**18:0**	**13:0 ante**	**18:0 3 OH**
sif 7 (18:1 cis 11,	20:0	**15:0 ante**	
18:1 trans 9,		**17:0 ante**	
18:1 trans 6)	*Saturated Odd*	**19:0 ante**	
	15:0		
	17:0		

presented as a biplot, but the large number of variables in this data set precluded clear presentation of the biplot, so we tabulated those variables and their loadings that had the greatest influence (loadings of PC 1 > |0.20|) on the separation of samples into the two distinct clusters of circles and stars (Table 4). The principle differences between samples forming the small cluster of stars and the rest of the samples is that samples represented by stars had a greater proportion of 12:0 iso, 15:0, 16:1 iso G, 17:1 ante A, and 16:1 2 OH and a smaller proportion of 12:0, 14:0, 15:0 iso, 16:0, and 18:0 2 OH than did the rest of the samples. Two of the monounsaturated FAMEs more common in samples represented by stars (16:1 iso G and 17:1 ante A) are

Table 3. Proportion of variance accounted for by eigenvalues of the correlation matrices with values ≥ 1 for principal component analyses

	Eigenvalue	Proportion of Variance Explained by Eigenvalue (%)	Cumulative Proportion of Variance Explained by Eigenvalues (%)
Whole Soil:	9.14	19.9	19.9
full data	5.11	11.1	31.0
matrix	3.59	7.8	38.8
(Fig. 2A)	2.16	4.7	43.5
	2.03	4.4	47.9
	1.80	3.9	51.8
	1.62	3.5	55.3
	1.55	3.4	58.7
	1.47	3.2	61.9
	1.33	2.9	64.8
	1.24	2.7	67.4
	1.19	2.5	70.0
	1.05	2.3	72.3
	1.02	2.2	74.5
Whole Soil: 14	6.57	46.9	46.9
selected FAMEs	2.68	19.2	66.1
(Fig. 2B)	2.36	16.8	82.9
Whole Soil:	3.69	36.9	36.9
grouped FAMEs	1.72	17.2	54.0
(Fig. 2C)	1.27	12.7	66.7
Plated Communities:	12.5	30.5	30.5
full data matrix	5.14	12.5	43.0
(Fig. 5A)	3.07	7.5	50.5
	2.73	6.7	57.2
	2.55	6.2	63.4
	1.88	4.6	68.0
	1.64	4.0	72.0
	1.60	3.9	75.9
	1.37	3.3	79.2
	1.17	2.9	82.0
	1.09	2.7	84.7
Plated Communities:	4.37	39.8	39.8
11 selected FAMEs	3.31	30.1	69.8
(Fig. 5B)	2.03	18.4	88.3
Plated Communities:	4.57	45.7	45.7
grouped FAMEs	1.74	17.4	63.1
	1.13	11.3	74.4

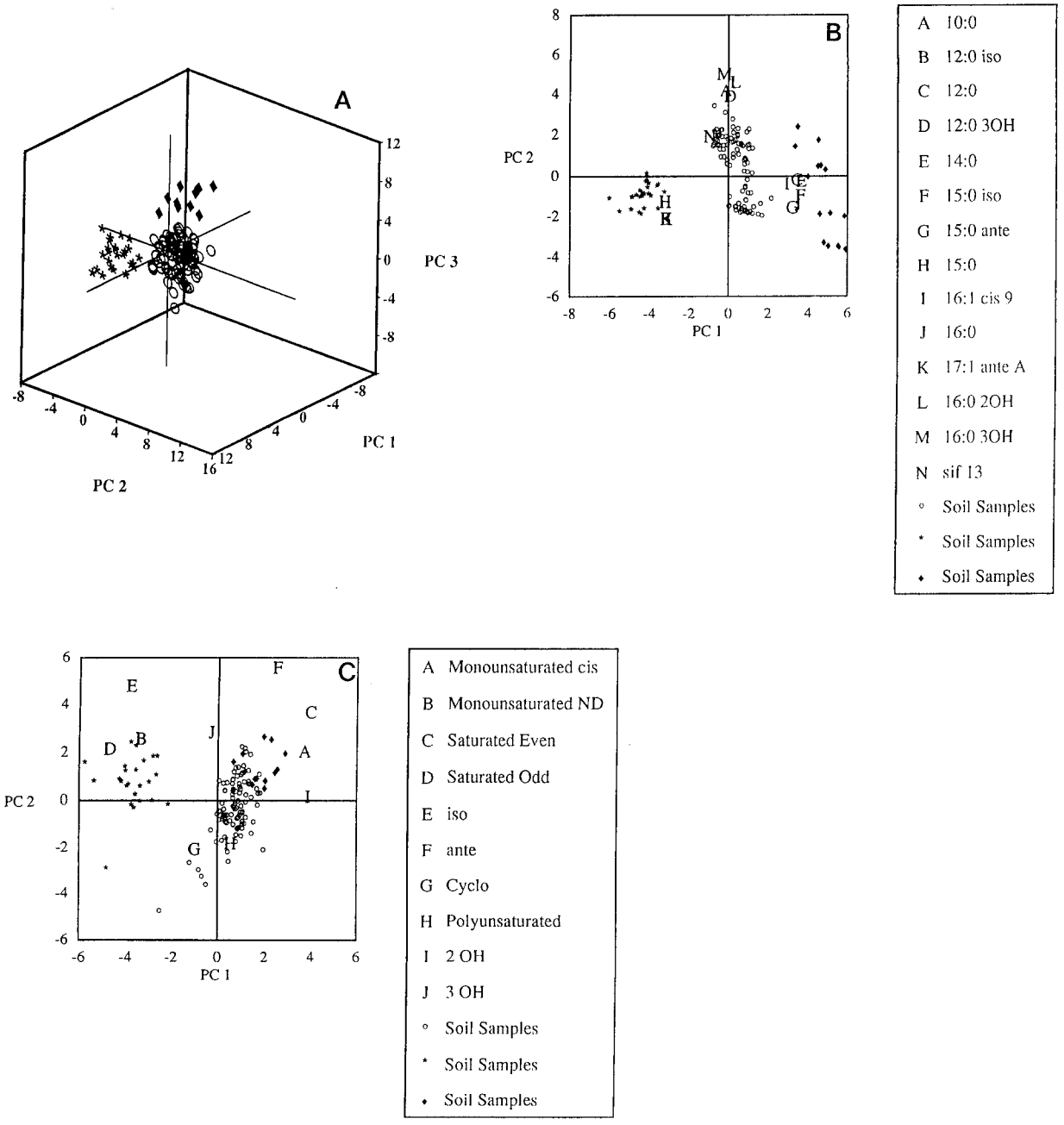

Fig. 2. PCA plot and biplots for soil community FAME profiles. **A.** PCA plot for whole soil community FAME profiles containing 46 FAMEs. **B.** Biplot for whole soil community FAME profiles containing 14 FAMEs selected using PCA. **C.** Biplot for whole community FAME profiles containing 10 groups of FAMEs. Parameters are listed in Table 3.

derived from FAs that are markers for bacteria (Perry et al., 1979), and one saturated FAME less common in these samples (12:0) is a marker for eukaryotes (White, 1983). Since fungi have much higher biomass than any other eukaryotes in agricultural soils (e.g. Brussaard et al., 1990), the soil samples found in the small cluster of stars probably have greater bacterial:fungal biomass ratios than do the majority of samples. Samples denoted by diamonds, which fall on the other end of the PC 1 axis, are, by similar reasoning, likely to have relatively low bacterial:fungal biomass ratios. In addition, our findings support Jantzen and Bryn's (1985) sug-

Table 4. FAME profile characteristics most influential in distinguishing the soil samples represented by stars from those represented by open circles in Figures 2A (whole soil communities) and 5A (plated communities). Those FAMEs that distinguished soil samples from both soil and plated communities are indicated in bold. All FAMEs which had loadings of more than |0.20| (an arbitrary threshold) in the relevant eigenvectors are included in this table. The value of the loadings are indicated in columns following each FAME

	Stars high in	*Loading*	*Open circles high in*	*Loading*
Whole	**12:0 iso** [Gram pos?]	0.29	12:0 [eukaryotes]	-0.20
Soil	**15:0**	0.30	14:0 [all organisms]	-0.21
	16:1 iso G [eubacteria]	0.27	15:0 iso	-0.23
	17:1 ante A [eubacteria]	0.29	16:0 [all organisms]	-0.29
	16:1 2 OH	0.24	18:0 2 OH	-0.21
	Outliers high in		*Majority high in*	
Plated	**12:0 iso** [Gram pos?]	0.40	10:0 3 OH [Gram neg]	-0.20
	15:0	0.40	16:1 cis 9	-0.24
	17:1 ante A [eubacteria]	0.40		
	19:0 iso [Gram pos?]	0.41		

gestion that branched fatty acids are not good markers for Gram positive bacteria since soil samples had both high and low proportions of some of these putative Gram positive markers (Table 4).

Since the amount of variance explained by the first three dimensions of the PCA was not very high, we ran PCA on matrices with fewer variables. We reduced the number of variables using two different approaches. First, we took advantage of the potential offered by PCA to eliminate variables that do not significantly explain total variance. We used Kaiser's rule (Joliffe, 1986) - that only as many variables as there are eigenvalues $\geqslant 1$ need to be kept for further analysis when the correlation matrix is used in PCA - to determine that 14 of the FAMEs should be retained. We chose to keep those 14 FAMEs with loadings > |0.20| in both PC 1 and PC 2 and those FAMEs with the largest loadings (absolute value) in either of the first two PC dimensions. Results and the FAMEs used in this analysis are presented in Figure 2B. Because of the greatly reduced dimensionality, presentation of results as biplots was possible, but only two dimensions are illustrated since three dimensional biplots did not show additional clustering of sample points, and two dimensional biplots were easier to read. In Figure 2B, samples identified as stars again formed a distinct cluster as did samples denoted by diamonds, and more than twice the matrix variability explained in Figure 2A was explained by reducing the dimensionality (Table 3). PC 1 is a contrast of 12:0 iso, 15:0, and 17:1 ante A vs. 12:0, 14:0, 15:0 iso, 15:0 ante, 16:1 cis 9, and 16:0. Since both 12:0 (White, 1983) and 16:1 cis 9 (Erwin, 1973) are markers for eukaryotes, samples represented by stars, again, seem to have relatively high bacterial:fungal biomass ratios and samples denoted by diamonds seem to have relatively low bacterial:fungal biomass ratios. Thus, differences in bacterial:fungal biomass ratios among samples were highlighted by eliminating less influential variables.

We also conducted PCA on a reduced dimension data matrix composed of FAMEs grouped according to similarities in their structure (Fig. 2C). These groupings, similar to those used by Zelles et al. (1992), reflect differences in the metabolic pathways required to produce FAs and, as such, may be taxonomically useful. We summed all FAMEs belonging to a particular group (Table 2) for each soil sample. The proportion of matrix variance explained was now 54 and 67% in two and three dimensions, respectively. Results were very similar to those for ungrouped data – samples represented by stars were distinctly different from the other samples, open circles formed a gradient along PC 2, and the diamonds were still grouped, albeit loosely - indicating that the groups into which we categorized FAMEs may be biologically meaningful.

Interpretation of the PCA of grouped FAMEs is consistent with that for ungrouped FAMEs. In the ungrouped data (Table 4), samples represented by stars harbored greater proportions than did the majority of samples of three monounsaturated FAMEs with

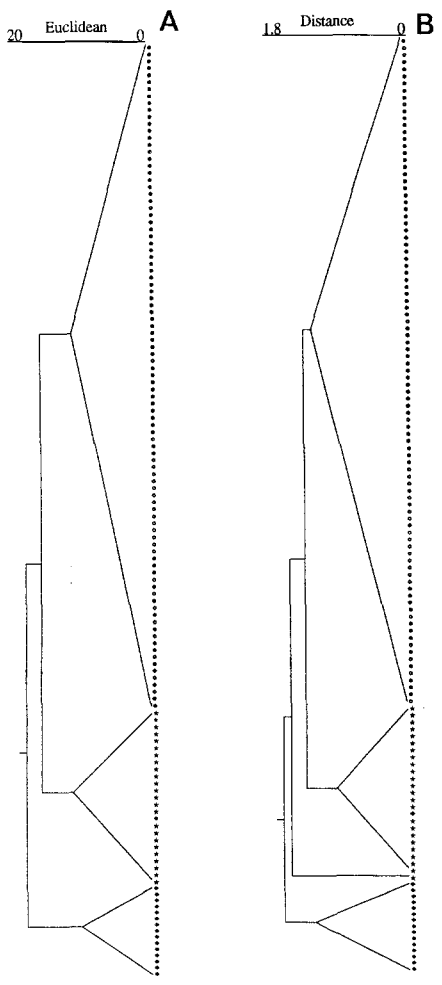

Fig. 3. Dendrograms derived from cluster analyses for whole soil communities. **A.** Euclidean clustering. UPGMA linkage. **B.** Average taxonomic distance clustering, single linkage. The same symbols used in Figures 2 and 5 are used for each soil sample.

unknown configuration (16:1 iso G, 17:1 ante A, 16:1 2 OH) and a saturated FAME with an odd number of C atoms (15:0). Samples represented by stars also had lower amounts of a FAME with a hydroxyl substitution at C number 2 (18:0 2 OH). These differences are reflected in Figure 2C in that monounsaturated ND and saturated odd FAMEs are contrasted with 2 OH FAMEs. Other differences among samples listed in Table 4 were not found when grouped FAMEs were subjected to PCA; this is because not all FAMEs belonging to a particular group were distributed equally among soil samples so that, for example, samples that were high in 12:0 iso were also low in another iso FAME, 15:0 iso.

Cluster analysis

Results from only two of eight cluster analyses conducted on soil community FAME profiles are presented (Fig. 3) since all dendrograms were essentially identical. Consistent results such as these, in light of the fact that UPGMA and single linkage methods are biased toward the creation of different types of clusters (Everitt, 1980; Milligan and Cooper, 1987), suggest that sample groupings identified by cluster analysis reflect real similarities and differences among FAME profiles of these soil samples. All eight cluster analyses also reflected cluster patterns found in PCA plots. Soil samples identified by stars and by open circles in PCA formed separate clusters in all dendrograms and samples identified as diamonds formed separate clusters only in the cluster analyses of the ungrouped data sets. Cluster analyses therefore corroborated the clustering identified in the first two and three PC dimensions for the PCA of ungrouped and grouped data sets. Unlike PCA, however, cluster analysis does not provide a direct means of identifying variables that contribute to the separation of samples into different clusters.

Geostatistical analyses

Our sampling scheme allowed us to capture previously unrecorded small-scale autocorrelation in some components of the numerically dominant portions of soil microbial communities (Fig. 4A). All eight FAMEs that exhibited spatial autocorrelation at the scale measured (minimum lag 0.07 m) exhibited a range < 1.00 m and most often < 0.20m (Table 5). This suggests that controls on the distributions of organisms with these FAMEs are also acting at these spatial scales (Trangmar et al., 1985). For example, the distribution of population(s) for which 15:0 ante is a marker appears to be controlled by some factor or set of factors that are acting at a scale of about 0.1m (Table 5, Fig. 4A). Two FAMEs that are known to be markers, 15:0 ante (eubacterial marker), and 16:0 2 OH (Gram negative marker) showed a significant pattern of spatial autocorrelation (Fig. 4A, Table 5). FAMEs that had a unique range of autocorrelation are probably also markers - at least among samples used to construct the variogram - since a unique range of autocorrelation identifies, by definition, a variable with a unique spatial distribution. This appears to be the case for 21:0 iso which was the only FAME with a range similar to 0.35m (Table 5). The population(s) for which a FAME identified as a marker in this manner, however, is not known.

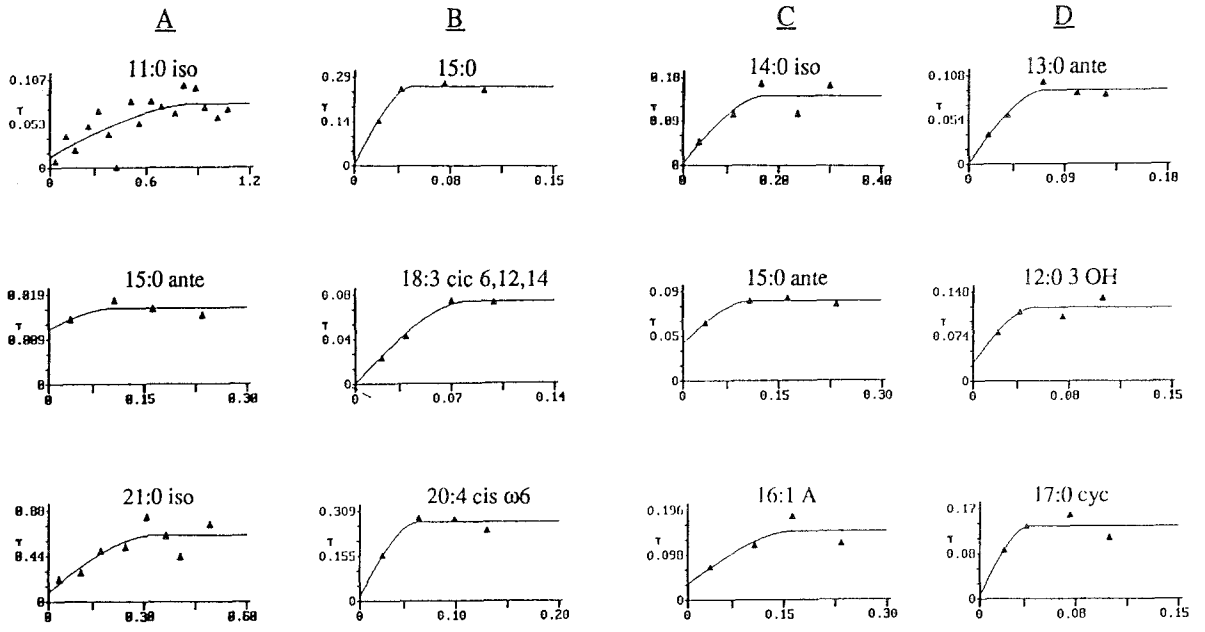

Fig. 4. Selected variograms for soil community FAMEs analyzed with a minimum step of 0.07 m (**A**) and 0.03 to 0.05 m (B), and plated community FAMEs analyzed with a minimum step of 0.07 m (**C**) and 0.03 to 0.05 m (**D**). The x axes are distance (m) and the y axes are semivariance (γ).

We found a high nugget:sill ratio for the remaining 48 FAMEs, indicating that most of the spatially structured variance in these FAMEs occurred at scales smaller than 0.07m (Webster, 1985). In fact for many FAMEs nugget variance approached zero [$C/(C_o + C) = 1$) when examined using a minimum lag interval of 0.03 to 0.05m (Table 5, Fig. 4B), although the small number of sample pairs for lag interval classes at this scale makes such patterns statistically tenuous. It is clear, however, that most spatially structured variance in these FAMEs - and presumably in the microbial communities identified by FAMEs - occurred at spatial scales analogous to individual soil peds and perhaps rhizospheres.

Whole soil vs. plated communities

On average, 20 different FAMEs were detected in each plated community. Forty-one FAMEs were identified in total from plated communities and only 28 of these were also found in the soil FAME profiles (Table 2).

The PCA plot of the plated organisms (Fig. 5A) was significantly different than that for the soil community (Fig. 2A), showing that FAME profiles captured large portions of soil microbial communities that were not culturable on R2A. The two plots, nonetheless, showed some surprisingly similar characteristics, although the low proportion of variance explained in the first three PC dimensions makes this conclusion somewhat tenuous. We therefore reduced the dimensionality of the data matrix for plated communities by the same means used for soil community FAME profiles. When the data were reduced by keeping only the most distinguishing variables (Fig. 5B), the clustering of soil samples in the first two PC dimensions was the same as for when the full data matrix was used, and 70 and 88% of the total variance was explained in two and three dimensions, respectively. Six of the seven outliers in both Figures 5A and B and the 25 soil samples forming the cluster of stars in Figures 2A and B were all characterized as being high in 12:0 iso, 15:0, and 17:1 ante A (Table 4). Four of the seven outliers in Figures 2A and B represent plated communities from the same soil samples that plotted in the cluster identified by stars in Figures 2A and B. These similarities are especially surprising given that the plated and abbreviated soil community data sets shared only 28 of 59 possible FAMEs (Table 2). That there were only seven outlier points in the PCA plot of plated communities (Figures 5A and B) but 25 sample points in the small cluster of stars in Figures 2A and B, and that the R2A used as a growth medium selects against fungi, supports our contention that the clustering in Figures 2A and B was due, at least in part, to a difference in the ratio of bacterial:fungal biomass

Table 5. Parameters for variograms of FAMEs that exhibited spatial autocorrelation, including variograms presented in Figure 4. $C/(C_o+C)$ is the proportion of population variance due to spatial structure and A_o is the range

FAME	Minimum lag (m)	$C/(C_o+C)$	A_o (m)	Number of sample pairs for first four points
Soils				
10:0	0.07	0.52	0.11	29, 32, 23, 19
11:0 iso	"	0.83	0.97	"
12:0 iso	"	0.68	0.10	"
15:0 ante	"	0.25	0.11	"
16:0 2 OH	"	0.46	0.12	"
sif 9	"	1.00	0.11	"
21:0 iso	"	0.86	0.35	"
21:1 cis 18	"	1.00	0.09	"
13:0 ante	0.03	1.00	0.08	10, 14, 13, 12
15:0	"	"	0.05	"
16:0 iso	"	"	0.05	"
16:1 cis 9	"	0.92	0.04	"
16:0 3 OH	"	1.00	0.04	"
17:0 ante	"	"	0.10	"
17:1 cis 10	"	"	0.06	"
18:0	"	"	0.05	"
18:3 cis 6, 12, 14	"	"	0.09	"
sif 3	"	"	0.06	"
sif 10	0.05	0.95	0.12	21, 19, 27, 12
20:4 cis ω6	0.04	0.95	0.06	18, 17, 14, 23
22:1 cis 17	0.03	1.00	0.05	10, 14, 13, 12
25:0 2 OH	"	"	0.09	"
Plates				
13:0 iso	0.07	0.45	0.17	26, 31, 23, 18
14:0 iso	"	1.00	0.19	"
15:0 ante	"	0.50	0.11	"
15:0 iso	"	0.50	0.28	"
16:1 A	"	0.77	0.20	"
17:1 iso E	"	0.63	0.20	"
13:0 ante	0.03	1.00	0.08	9, 12, 12, 12
12:0 3 OH	"	0.76	0.05	"
16:1 iso E	0.05	0.70	0.08	19, 17, 27, 12
sif 3	0.03	0.67	0.04	9, 12, 12, 12
sif 4	0.037	0.94	0.09	15, 12, 18, 16
17:0 cyc	0.03	1.00	0.04	9, 12, 12, 12
sif 7	0.04	1.00	0.07	16, 16, 13, 23

among soil samples. In addition, the seven outliers (Figures. 5A and B) seemed to have a greater Gram positive:Gram negative ratio than did the majority of plated communities based on the presence of more of the Gram negative marker 10:0 3 OH in the majority of plated communities (Table 4).

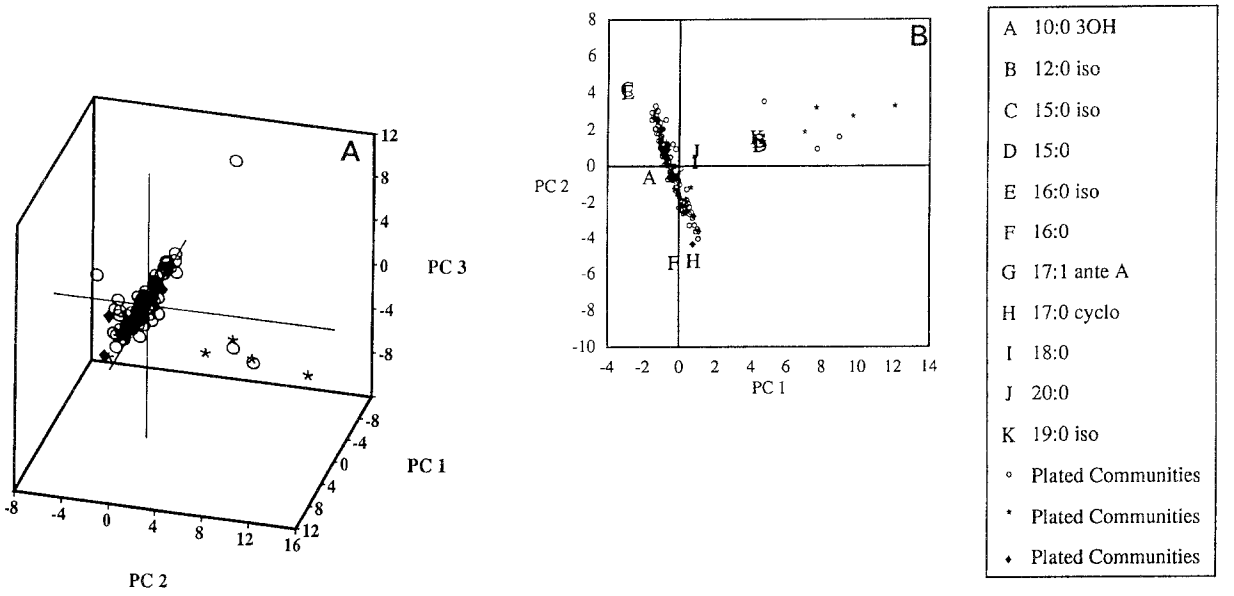

Fig. 5. PCA plot and biplots for plated community FAME profiles. **A.** PCA plot for plated community FAME profiles containing 41 FAMEs. **B.** Biplot for plated community FAME profiles containing 11 FAMEs selected using PCA. Parameters are listed in Table 3.

The effect of four other FAMEs on the distribution of sample points in the large cluster in Figure 5B is also evident. PC 2 is a contrast of 15:0 iso and 16:0 iso vs. 16:0 and 17:0 cyclo. That is, samples forming the linear cluster represent a gradient of these four FAMEs, with plated communities lying at the high PC 2 end of the cluster being high in 15:0 iso and 16:0 iso and low in 16:0 and 17:0 cyclo. Points at the other end of the cluster have opposite characteristics with respect to these four FAMEs. Since none of these FAs are currently known to be markers, further characterization of this gradient with respect to particular microbial taxa is not possible.

Grouping FAMEs from plated communities did not reveal any further information: the same samples which were outliers in Figures 5A and B were outliers in the third PC dimension. All cluster analyses, again regardless of distance metric or linkage method used, were consistent with clustering found in PCA.

Variograms for nine FAMEs from the plated communities exhibited spatial autocorrelation at a minimum step lag of 0.07m. Only one of these, 15:0 ante, also exhibited spatial autocorrelation when extracted from soil (Table 5, Fig. 4C). That spatial structure is revealed for some FAMEs from plated communities, but for different FAMEs than from soil communities, is evidence that, though laboratory media select for particular components of the soil microbial community, the resulting community is not solely a result of selection by the growth medium, but reflects the distribution - in soil - of those soil microbial populations that can be cultured. Further evidence that microbial distributions present in soil are reflected in plated communities is seen in that the variograms for 15:0 ante from both soil and plated communities had identical ranges (0.11m; Table 5). That one half of the population variance of 15:0 ante was explained by spatial structure among plated communities, but only one quarter was explained for the soil communities, is consistent with the fact that whole soil communities are more variable than plated communities. Again, some FAMEs that showed no spatial structure at a minimum step lag of 0.07m exhibited spatial autocorrelation when analyzed using smaller minimum lags (Table 5, Fig. 4D).

Conclusions

Whole cell FAME profiles of soil samples offer a rapid, inexpensive and reproducible means for characterizing numerically dominant portions of soil microbial communities, including those organisms not culturable. By taking advantage of current knowledge regarding marker fatty acids and using multivariate statistical procedures we have shown that the numerically dominant microbial communities of most soil samples taken from a corn field were similar. About 18% of samples, however, were distinct in having lower bacterial:fungal

biomass ratios and about 8% of samples seemed to have higher bacterial:fungal biomass ratios than the majority of samples. By using only those variables selected using PCA, we more than doubled the percent of variance explained by these differences. The paucity of current information regarding individual and grouped marker FAMEs reduced our ability to interpret similar patterns when PCA was conducted on a data matrix made smaller by grouping FAMEs.

Semivariance analysis may be a useful means of identifying new marker FAMEs within particular ecosystems. We found, for example, that 21:0 iso had a unique range of autocorrelation, indicating that it is a marker FAME, albeit for a currently unidentified microbial taxa. Concurrent use of geostatistical and FAME analyses may be a powerful means of revealing other potential marker FAMEs.

We were able to capture and describe small-scale patterns of distribution of microbial populations using semivariance analysis. Where autocorrelation occurred it was generally at scales < 0.2 m, a scale analogous to individual soil peds and rhizospheres. Sampling more intensely at very small scales should result in the discovery of spatial structure for more FAMEs.

FAME profiles of most plated communities were significantly different from those of whole soil communities, indicating that FAME profiles of soils reveal portions of the microbial community that do not grow on R2A. A small number of distinct plated communities, however, retained the same FAME profile characteristics (high in 12:0 iso, 15:0 and 17:1 ante A) that distinguished the whole soil samples from which plated communities were cultured. As indicated by the variograms for a number of individual FAMEs, the distribution of a portion of the soil community that is culturable seems to be retained when soil communities are cultured. Thus, microbial communities grown on laboratory media reflect not only specific laboratory growth conditions, but also the distribution of culturable microbial populations in soil.

Acknowledgements

We thank Lori Merrill, Deane Lehmann, and Chenn-Ching Chou for sample collection and laboratory assistance, and Helen Garchow and Peter Stahl for advice regarding FAME analyses. This project was supported by the National Science Foundation (NSF) via grants to Michigan State University for the Science and Technology Center for Microbial Ecology (BIR9120006), the Long Term Ecological Research Project in Agricultural Ecology (DEB9211771), and a Doctoral Dissertation Improvement Award (DEB9311380); support was also provided by the Michigan Agricultural Experiment Station.

References

Applied Biostatistics, Inc. 1992 NTSYS. Exeter Publishers, Setauket New York, USA.

Bobbie R J and White D C 1980 Characterization of benthic microbial community structure by high-resolution gas chromatography of fatty acid methyl esters. Appl. Environ. Microbiol. 39, 1212–1222.

Brussaard L, Bouwman L, Geurs M, Hassink J and Zwart K B 1990 Biomass, composition and temporal dynamics of soil organisms of a silt loam soil under conventional and integrated management. Neth. J. Agric. Sci. 38, 283–302.

Digby P G N and Kempton R A 1987 Multivariate Analysis of Ecological Communities. Chapman and Hall, New York, USA.

Erwin J A 1973 Fatty acids in eukaryotic microorganisms. *In* Lipids and Biomembranes of Eukaryotic Microorganisms. Ed. J A Erwin. pp 41–143. Academic Press, New York, USA.

Everitt B 1980 Cluster Analysis. Halsted Press, New York, USA.

Findlay R H and White D C 1983The effects of feeding by the sand dollar *Mellita quinquiesperforata* (Leske) on the benthic microbial community. J. Exp. Mar. Biol. Ecol. 72, 25–41.

Gillan F T and Hogg R W 1984 A method for the estimation of bacterial biomass and community structure in mangrove-associated sediments. J. Microbiol. Methods 2, 275–293.

Harwood J L and Russell N J 1984 Lipids in Plants and Microbes. Allen & Unwin, London.

Jantzen E and Bryn K 1985 Whole-cell and lipopolysaccharide fatty acids and sugars of Gram-negative bacteria. *In* Chemical Methods in Bacterial Systematics. Eds. M Goodfellow and D E Minnikin. pp 145–171. Academic Press, London.

Jolliffe I T 1986 Principal Component Analysis. Springer-Verlag,.

Klug M J and Tiedje J M 1993 Response of microbial communities to changing environmental conditions: chemical and physiological approaches. *In* Trends in Microbial Ecology. Eds. R Guerrero and C Pedros-Alio. pp 371–374. Spanish Society for Microbiology, Barcelona, Spain.

Liu C J and Keister T D 1978 Southern pine stem form defined through principal component analysis. Can. J. For. Res. 8, 188–197.

Mayberry W R, Lambe D W and Ferguson K P 1982 Identification of *Bacteroides* species by cellular fatty acid profiles. Int. J. Syst. Bacteriol. 32, 21–27.

Microbial ID Inc. 1992 Microbial Identification System Operating Manual, Version 4. Newark, DE, USA.

Milligan G W and Cooper M C 1987 Methodological reviews: clustering methods. Appl. Psychol. Measurement 11, 329–354.

Moss C W 1981 Gas-liquid chromatography as an analytical tool in microbiology. J. Chrom. 203, 337–347.

Moss C W, Dees S B and Guerrant G O 1980 Gas-liquid chromatography of bacterial fatty acids with a fused-silica capillary column. J. Clin. Microbiol. 12, 127–130.

Perry G J, Volkman J K, Johns R B and Bavor H J 1979 Fatty acids of bacterial origin in contemporary marine sediments. Geochim. Cosmochim. Acta 43, 1715–1725.

Robertson G P and Gross K 1994 Assessing the heterogeneity of belowground resources, quantifying scale. *In* Exploitation of Environmental Heterogeneity by Plants. Eds. M M Caldwell and R W Pearcy. pp 237–253. Academic Press, New York, USA.

SAS Institute 1991 SAS User's Guide Version 6.03. Ed.SAS Inst., Cary, NC, USA.

Torsvik V, Goksoyr J and Daae F L 1990a High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56, 782–787.

Torsvik V, Salte K, Sorheim R and Goksoyr J 1990b Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl. Environ. Microbiol. 56, 776–781.

Trangmar B B, Yost R S and Uehara G 1985 Application of geostatistics to spatial studies of soil properties. Adv. Agron. 38, 45–94.

Vestal J R and White D C 1989 Lipid analysis in microbial ecology. Bioscience 39, 535–541.

Viljoen BC, J L F Kock and P M Lategan 1986 Fatty acid composition as a guide to the classification of selected genera of yeasts belonging to the endomycetales. J. Gen. Microbiol. 132, 2397–2400.

Volkman J K, Johns R B, Gillan F T, Perry G J and Bavor H J 1980 Microbial lipids of an intertidal sediment - I. Fatty acids and hydrocarbons. Geochim. Cosmochim. Acta 44, 1133–1143.

Webster R 1985 Quantitative spatial analysis of soil in the field. Adv. Soil Sci. 3, 1–70.

White D C 1983 Analysis of microorganisms in terms of quantity and activity in natural environments. *In* Microbes in Their Natural Environments. Eds. J H Slater, R Whittenbury and J W T Wimpenny. pp 37–66. Cambridge University Press, Cambridge, UK.

Zelles L and Bai Q Y 1993 Fractionation of fatty acids derived from soil lipids by solid phase extraction and their quantitative analysis by GC-MS. Soil Biol. Biochem. 25, 495–507.

Zelles L, Bai Q Y and Beese F 1992 Signature fatty acids in phospholipids and lipopolysaccharides as indicators of microbial biomass and community structure in agricultural soils. Soil Biol. Biochem. 24, 317–323.