ELSEVIER

FEMS Microbiology Ecology 44 (2003) 335–346

FEMS MICROBIOLOGY Ecology

www.fems-microbiology.org

Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field

Rima B. Franklin *, Aaron L. Mills

Laboratory of Microbial Ecology, Department of Environmental Sciences, University of Virginia, Clark Hall, 291 McCormick Road, P.O. Box 400123, Charlottesville, VA 22904-4123, USA

Received 18 December 2002; received in revised form 25 February 2003; accepted 28 February 2003

First published online 21 March 2003

Abstract

To better understand the distribution of soil microbial communities at multiple spatial scales, a survey was conducted to examine the spatial organization of community structure in a wheat field in eastern Virginia (USA). Nearly 200 soil samples were collected at a variety of separation distances ranging from 2.5 cm to 11 m. Whole-community DNA was extracted from each sample, and community structure was compared using amplified fragment length polymorphism (AFLP) DNA fingerprinting. Relative similarity was calculated between each pair of samples and compared using geostatistical variogram analysis to study autocorrelation as a function of separation distance. Spatial autocorrelation was found at scales ranging from 30 cm to more than 6 m, depending on the sampling extent considered. In some locations, up to four different correlation length scales were detected. The presence of nested scales of variability suggests that the environmental factors regulating the development of the communities in this soil may operate at different scales. Kriging was used to generate maps of the spatial organization of communities across the plot, and the results demonstrated that bacterial distributions can be highly structured, even within a habitat that appears relatively homogeneous at the plot and field scale. Different subsets of the microbial community were distributed differently across the plot, and this is thought to be due to the variable response of individual populations to spatial heterogeneity associated with soil properties.

Keywords: Spatial variability; Microbial community; Soil; Community structure; Geostatistics; Multi-scale; Nested variability

1. Introduction

Microorganisms are not distributed uniformly in the environment, rather their abundance and activity change along environmental gradients. Even within a homogeneous system, biological processes (e.g., growth or colony formation) may produce aggregations of organisms at various spatial scales. Soil systems are particularly heterogeneous, and this heterogeneity arises as a result of the interaction of a hierarchical series of interrelated variables that fluctuate at many different spatial and temporal scales. The factors that affect microbial survival and community structure in soils are known to be both biotic (e.g., predation and competition) and abiotic (e.g., temperature, pH, or substrate availability). Some of these processes are primarily important at microscopic scales (e.g., particle size and pore space structure), whereas others act over larger distances (e.g., vegetation cover and precipitation). These soil properties do not vary independently; rather, the general perception is that any such variable measured at a certain point in space and time is the outcome of several physical, chemical, and biological processes, all of which are spatially variable.

Given that environmental factors do not necessarily operate independently, or at distinct spatial scales, studying microbial systems using a single analytical scale cannot provide a complete understanding of community dynamics. Multi-scale comparisons, in which patterns are analyzed at several different spatial scales, may be more useful when trying to identify the factors that control community development. Conclusions about the organization of microbial communities, the effect of disturbance, or the roles of various limiting factors are likely to differ at different spatial scales [1]. Moreover, the characterization of micro-

* Corresponding author. Tel.: +1 (434) 924-0537; Fax: +1 (434) 982-2137.
E-mail address: rbfranklin@virginia.edu (R.B. Franklin).

0168-6496/03/$22.00
doi:10.1016/S0168-6496(03)00074-6

bial communities at several different scales may help explain paradoxes that arise when different investigators, studying similar communities but at different scales, arrive at different conclusions about the factors that structure those communities. These disagreements may reflect viewpoints of different scales, and not differences in the way communities are organized [2].

Previous work studying spatial organization in soil microbial systems has primarily focused on the distribution of individual cells [3–6], specific types of organisms [7–11], or collective parameters such as bacterial abundance or total biomass [12–17]. There are fewer studies that have considered variations in community structure [15,18–21] or function/activity [13,22–24]. In general, these studies have concentrated on understanding spatial variability at a single analytical scale, and have found significant spatial autocorrelation at a variety of separation distances, ranging from μm to km depending on the spatial extent studied. Recently, scientists have begun to focus on multi-scale comparisons, and have found evidence for nested scales of spatial structure [18,25–28]. For example, Nunan et al. [3] studied the spatial distribution of soil bacteria at three different scales, ranging from μm to meters, and found that the distribution of individual bacterial cells was organized at two scales in the subsoil, and at a single scale in the topsoil. Studies conducted in agricultural and shrub–steppe ecosystems suggest that microbial biomass and activity may be spatially dependent at scales less than 1 m, nested within a larger scale related to variations at the landscape level [14,16,29]. The presence of nested scales of variation suggests that the various factors regulating the development of microbial communities in the soil ecosystems may operate at different scales [27], and a simultaneous analysis of the multi-scale spatial variability of microbial community structure and soil microenvironment could help identify these factors and determine their relative influence.

The present study was designed to address the general need for increased research into multi-scale patterns of spatial organization in soil systems. In particular, the research focused on quantifying the spatial patterns associated with microbial community structure at the cm to meter scale using geostatistical techniques. Nested levels of spatial autocorrelation were observed (ranging from 30 cm to more than 6 m), and, in some locations, up to four distinct ranges of spatial influence were quantified.

2. Materials and methods

2.1. Site description and sample collection

Soil samples were collected from an agricultural field on the eastern shore of Virginia (USA) in May 2000 (37°17.62′N, 75°55.53′W). The field was planted with durum wheat (*Triticum turgidum*), and the crop was approximately 75 days old on the day of sampling. Samples were collected with separation distances ranging from 2.5 cm to 11 m, using the sampling scheme detailed below. At each sampling location, the loose layer of surface material was removed, and a small hole (1.5 cm diameter) was dug to collect 5–10 g of soil. The samples were placed on ice for transport to the lab, where they were sifted (approximately 750-μm mesh size) to remove gravel, plant, and root material, and stored at −80°C.

2.2. Sampling scheme

The basic sampling design was a square with 7.1-m edges and 10-m diagonals (Fig. 1). Samples were collected at regular intervals around the perimeter of the block (1.8-m separation distance), and at 1-m intervals along the diagonals. At each node (A, B, C, D, and X), more concentrated sampling efforts were employed.

Nested within the original sampling grid, a second set of samples were collected at 10-cm increments in a cross shape surrounding each node. Five samples were collected in each direction – north, south, east, and west – from the center node. Nested within this area, a third set of samples was collected at 2.5-cm increments around each node, following the same pattern (2.5, 5.0, 7.5, and 10 cm in each direction). A total of 193 soil samples were collected, 33 at each node and 28 at larger separation distances.

2.3. DNA extraction and quantification

Whole-community DNA was extracted from 0.25-g subsamples of soil with the MoBio Soil DNA isolation kit (Solana Beach, CA, USA) using the alternative heat shock lysis procedure described in the kit documentation. Purified DNA was resuspended in 10 mM Tris buffer and stored at −20°C. The concentration of DNA in each sample was determined using the PicoGreen reagent (Molecular Probes, Eugene, OR, USA).

2.4. AFLP

Amplified fragment length polymorphism (AFLP) analysis was performed using the Perkin Elmer Microbial Fingerprinting Kit (PE Applied Biosystems, Foster City, CA, USA). For community analysis, the manufacturer's instructions for analysis of individual bacterial strains were modified as described below. For details regarding the primer and adapter sequences, and an explanation of primer selection criteria, readers should consult the kit documentation.

With AFLP, a restriction digest is performed on a DNA sample (similar to restriction fragment length polymorphism (RFLP)), and then a set of primer recognition sequences (adapters) is used to amplify the restriction fragments using polymerase chain reaction (PCR) [30]. The primers and restriction enzymes used are not specific for

a gene or group of genes, but can, theoretically, interact in numerous random places throughout a genome. AFLP is very similar in premise and application to randomly amplification of polymorphic (RAPD) DNA fingerprinting, which has been used a number of times to compare microbial community structure [31–33]; the specific use of AFLP for community analysis is discussed in Franklin et al. [34]

2.4.1. Restriction/ligation procedure and preselective amplification

The restriction and ligation steps of the AFLP reaction were performed simultaneously by adding 10 ng of DNA, 2 U of *Mse*I, 4 U of *Eco*RI, and 10 U of T4 DNA ligase (enzymes purchased from New England Biolabs, Beverly, MA, USA) to a reaction mixture containing: 1×T4 DNA ligase buffer (50 mM Tris–HCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, and 25 μg ml^{-1} bovine serum albumin (BSA, New England Biolabs)), 0.05 M NaCl, 0.5 μg BSA, 0.2 μM *Eco*RI adapter, and 2 μM *Mse*I adapter (PE Applied Biosystems); the total reaction volume was 11 μl. The reactions were incubated for 6 h at 37°C, and then diluted by adding 189 μl of $TE_{0.1}$ buffer (20 mM Tris–HCl, 0.1 mM EDTA, pH 8.0).

Preselective amplification was performed following the manufacturer's protocol, though the PCR mixture was supplemented with 400 μg ml^{-1} BSA. Successful amplification was verified by agarose gel electrophoresis of 10 μl of PCR product in a 1.5% agarose gel. The remaining product from the preselective amplification (10 μl) was then diluted with 190 μl $TE_{0.1}$ buffer.

2.4.2. Selective amplification

For the selective amplification, several different combinations of primers were tested; in each case, one *Eco*RI primer, labeled with a fluorescent dye, was paired with one *Mse*I primer. However, the AFLP patterns obtained using the bacterial primer pairs were too complex, and primers from the AFLP Plant Mapping Kit (PE Applied Biosystems) were also tested. These primers were identical to those designed for the bacterial samples, but contained an additional selective nucleotide at the 3′ end of the primer. After screening several pairs of primers, two sets were selected for use in this study; the selection was based on the number and intensity of the peaks in the final AFLP fingerprint, as well as the reproducibility of these fingerprints. The primer pairs used were: *Eco*RI-ACA (FAM-labeled) with *Mse*I-CAA, and *Eco*RI-AAC (NED-labeled) with *Mse*I-CTC.

Selective amplification was performed as directed in the kit documentation, with two modifications. Firstly, the reaction volume was doubled (20 μl total), and, secondly, the PCR reaction mixture was supplemented with 800 μg ml^{-1} BSA. Successful amplification was confirmed by agarose gel electrophoresis (8 μl of PCR product, 1.5% agarose gel). The remaining PCR product was then purified using the QIAquick PCR Purification kit (Qiagen, Valencia, CA, USA). To elute DNA from the QIAquick column, 20 μl of elution buffer was added to the center of the membrane, allowed to stand for 1 min, and then centrifuged for 1 min at 13 000 rpm in a tabletop microcentrifuge.

2.4.3. Electrophoresis and data collection

After purification, the selective amplification products were resolved using an ABI Prism 310 Genetic Analyzer. For the FAM-labeled products, 10 μl of PCR product was mixed with 1 μl of size standard (GeneScan 500 ROX, PE Applied Biosystems) and 14 μl of deionized formamide. For the NED-labeled products, 1.5 μl of PCR product was mixed with 1 μl of the size standard and 22.5 μl of deionized formamide. These mixtures were denatured by heating to 95°C for 5 min, and then quick-chilled on ice. The samples were analyzed with the following electrophoresis parameters: 10-s injection time, 15-kV injection voltage, 13-kV run voltage, and 30-min run time.

The electropherograms of the AFLP products were analyzed using the Genotyper software (PE Applied Biosystems), and the presence or absence of each peak in each sample was coded as 1 or 0. The data from the two primer pairs were pooled into a single large dataset for all further analyses. Collectively, these primers produced a total of 331 bands, and an individual sample contained between 20 and 210 bands. The average number of bands observed for an individual sample was 88.

The Jaccard coefficient was used to calculate the relative similarity between each set of samples, based on the proportion of positive bands shared by a sample pair [35]. The similarity matrix was then converted into a dissimilarity matrix by subtracting each value from 1. The dissimilarity matrix represents the relative difference in microbial community genetic structure between each pair of soil samples.

2.5. Geostatistical analyses

In most geostatistical analyses, a variance term (usually semi-variance) is calculated between each pair of samples and graphed versus spatial separation to produce a variogram. When the overall spatial structure of a multivariate dataset is of interest, researchers may generate plots using a 'resemblance coefficient' for the *y*-axis (e.g., a similarity or dissimilarity matrix [15,36,37]) or information derived from a principal components analysis (PCA) [13,18,38,39], rather than a conventional variance estimate. Since the AFLP analyses generated multivariate binary data, it was not possible to calculate semi-variance between sample pairs; instead, pseudo-variograms were created using the 'relative dissimilarity' values calculated from the Jaccard similarity matrix. These pseudo-variograms were constructed and analyzed using the same techniques as traditional variograms.

2.5.1. Analytical approach

An analytical approach was developed to explore two distinct aspects of spatial variability in these soil microbial communities. First, the overall spatial autocorrelation structure was analyzed in order to quantify the relationship between community variability and spatial separation (lag distance). Data from all sampling locations were included to provide an average portrait of the spatial relationships in the plot. This system was analyzed multiple times, changing the size of the observational window, to study this relationship at different spatial scales. The second portion of the analysis was directed toward trying to understand any changes in spatial pattern and community organization associated with different locations in the field.

For the first set of analyses, data from all of the sampling locations were analyzed to obtain an average portrait of the spatial relationships in this plot. Subsets of these data, varying in maximum separation distance, were then analyzed to quantify autocorrelation at different spatial scales. These scales were named based on relative size, and the following designations were used: plot scale (all sampling locations), large scale (separation distances up to 5 m), small scale (up to 1 m), and fine scale (up to 0.4 m). For each of these different sample groupings, geostatistical analyses of the overall difference in community structure were performed.

For the second set of analyses, local spatial autocorrelation was quantified by analyzing samples from different sections of the plot. These results were used to help understand whether the spatial autocorrelation structure was different in different areas of the field. An analysis of each scale was performed at each of the five different nodes (A–X, Fig. 1): large scale (all samples located within a 2.5-m radius surrounding each node (maximum separation distance of 5 m)), small scale (all samples located within a 0.5-m radius), and fine scale (all samples located within a 0.2-m radius).

Lastly, in order to determine whether the pattern of spatial variability changed with direction in the field (anisotropy), the data were also grouped into two additional categories. The 'north–south' (N/S) analysis included all of the samples collected along the axis between nodes B and D, and the 'east–west' (E/W) analysis included all of those points along the line between nodes A and C.

2.5.2. Guidelines used in variogram construction

Prior to constructing each variogram, it was first necessary to segregate the data into distance classes by calculating the appropriate number of bins and the appropriate bin width (lag distance). The purpose of 'binning' the data was to obtain the maximum resolution (the most detail) at small distances, without being misled by structural artifacts resulting from whatever particular size class was chosen. This approach allowed us to quantify the dominant spatial pattern at each scale, but obscured the autocorrelation structure at smaller distances. For each analysis, lag distance was calculated by considering the maximum separation distance between sample pairs, as discussed in Franklin et al. [15]. Several variograms were then produced and modeled, for a range of different lag distances surrounding this initial estimate, and the results with the best fit (highest R^2), the most reasonable parameter estimates, and the maximum detail are presented here.

One problem with using equal distance classes to construct a variogram is that the number of points in some bins may be quite small (especially bins at the far right of the variogram), so the average associated with one of these classes may not be a valid estimate of the mean at that distance. To help avoid this problem, all bins that contained fewer than 1% of the total number of points (pairwise comparisons) in each analysis were removed from the experimental variograms prior to statistical modeling.

2.5.3. Modeling the experimental variograms

In general, variograms may take one of three different forms: nugget (sometimes called 'nugget effect'), linear, and linear-sill. A variogram that is categorized with a *nugget* model is flat, indicating a lack of spatial structure in the data at the scales measured. A variogram that displays a *linear* pattern represents a system where samples are autocorrelated at all of the separation distances measured, and may be modeled with a linear equation: $y = C_0 + bx$, where y is the variance term (in this case, dissimilarity in genetic community structure), x is the spatial separation distance, C_0 is the y-intercept, and b is the slope of the variogram model. Most of the time, these variograms do not pass though the origin of the graph, but display some variability even at separation distances of zero; this value, C_0, is also referred to as 'nugget' because it represents the variability in the data that cannot be modeled using the spatial autocorrelation function. This phenomenon may occur because of random sampling variance, experimental error, or variability at other spatial scales.

A *linear-sill* model is a general term used to describe variograms that display increasing variance with increasing separation (conceptually similar to the linear model), and then level off at a *sill*. In our study, all variograms displaying this pattern were modeling using the exponential equation:

$$y = C_0 + C_1\left[1 - \exp\left(-3\frac{x}{a}\right)\right]$$

where y is the variance term, and x is the spatial separation distance. C_0 is a parameter quantifying the nugget effect (the amount of variability at distance = 0), C_1 is a spatially structured component of the model, and a is the range (the distance beyond which variance is no longer a function of spatial separation). The range is sometimes referred to as the correlation length scale (CLS). In the

exponential model, the (semi-)variogram approaches its maximum asymptotically, and the range is therefore defined as the distance where the (semi-)variance equals 95% of the sill. The sill (C) is the y value at which the variogram levels off, and can be calculated as: $C = C_0 + C_1$. The ratio of the spatially structured component of the model to the total variability captured by the model (C_1/C) represents the proportion of variability in the dataset that was modeled by the autocorrelation structure function, and is commonly referred to as *spatial dependence*. This value approaches 1 in a strongly spatially structured system, and 0 when no spatial structure is detected with the sampling scale used.

Each of our experimental variograms was modeled using either a linear or an exponential equation. All regressions were performed in SigmaPlot (Version 5.0) and R^2 was used to measure the fit of the model to the data; P values less than 0.05 were considered statistically significant. From the model, C_0, C_1, and a were estimated, and the sill (C) and spatial dependence (C_1/C) were calculated. Situations in which a linear or exponential model could not be successfully applied ($P > 0.05$) were categorized as 'nugget'.

2.5.4. Kriging

Kriging is a family of generalized least-squares regression algorithms that may be used to estimate the values of a given parameter at unsampled locations, by considering the spatial autocorrelation structure of the variable as determined for the sampled locations. In the variogram analyses discussed above, we used the similarity matrix to describe the overall relationship between samples, and determined the autocorrelation associated with variation in the composition of the entire community. However, it was not possible for us to generate maps using the similarity matrix (we cannot plot paired values across all locations) or to use the original data matrix (1s and 0s) in the kriging. Instead, a PCA was performed on the original data in order to generate numerical values describing community structure at each sampling location. The first three principal components (PCs) were used in the kriging, and each describes a portion of the variability in microbial community structure.

A separate geostatistical semi-variogram analysis was performed on each of the first three PCs to quantitatively describe the spatial autocorrelation structure for each PC. This information was then used in the kriging to generate maps of the PC scores with the SADA statistical package (Spatial Analysis and Decision Assistance, Version 3.0.80, University of Tennessee). This approach was also used to generate maps of microbial community structure along each of the main axes of the sampling grid (N/S and E/W, as analyzed in the directional analysis). In this case, maps were only produced for the first PC.

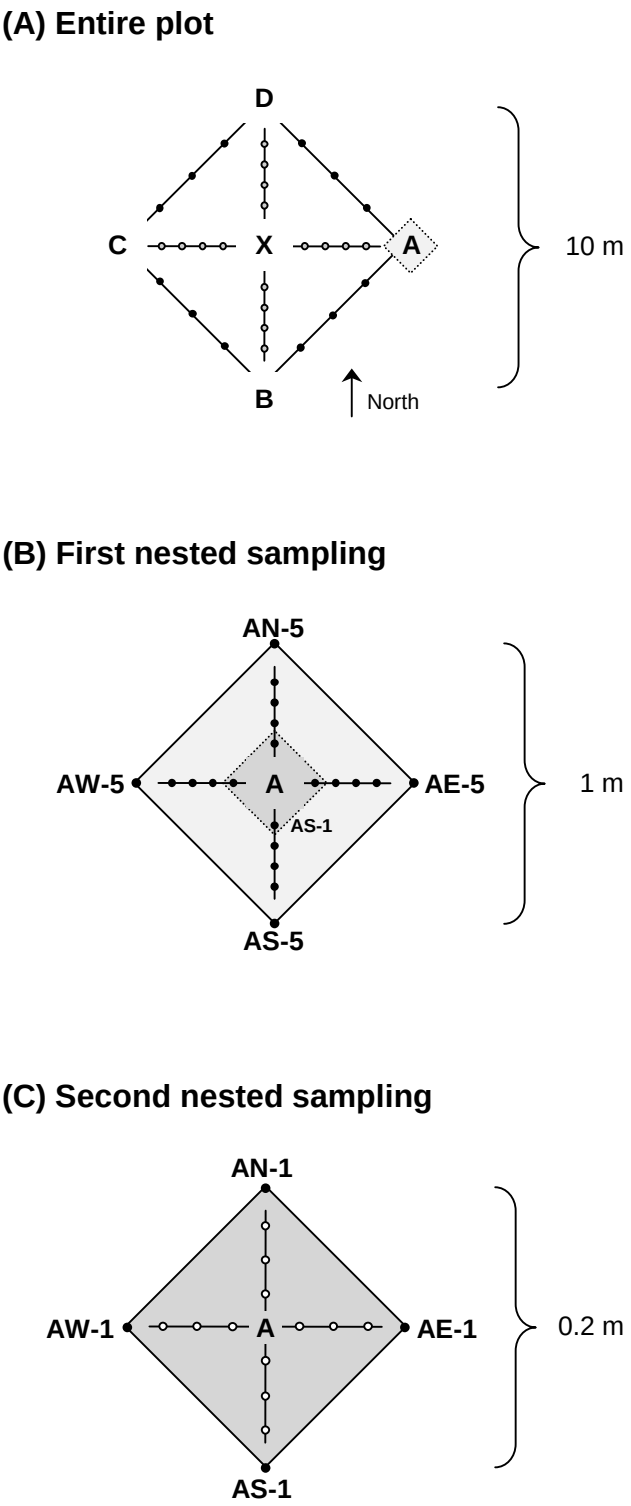

Fig. 1. Map of the sampling scheme. A: The sampling area was a 50-m^2 square (diamond) with 10-m diagonals. Around the perimeter of the square, samples were collected at 1.8-m increments, and at 1-m increments along the diagonals. At each node (A–X), more concentrated sampling efforts were employed. A nested sampling pattern was applied at each location, and node A is presented in the figure as an example. Additional samples were collected at 10-cm increments (B) and 2.5-cm increments (C) in a cross shape surrounding each node

3. Results

3.1. Quantifying multiple scales of spatial autocorrelation within the plot

In the first set of analyses, data from all of the sampling locations were considered in order to obtain an average portrait of the spatial relationships in the plot (Fig. 2). For the plot-scale analysis, a bin size of 0.5 m was used, and the number of points included in each bin varied from a minimum of approximately 100, for very large separation distances (>10.5 m), to more than 3000 points for intermediate separation distances. On average, each point on the variogram is the mean of approximately 800 pairwise comparisons. The points on the variogram that were averages of a small number of comparisons (<1%) were excluded from the graph and the geostatistical modeling. A similar approach was used for the large- (Fig. 2B), small- (Fig. 2C), and fine-scale analyses (Fig. 2D), where 8288, 2860, and 1670 pairwise comparisons were used, respectively.

Significant spatial autocorrelation was detected at each analytical scale (Table 1), and could be modeled using either the exponential (plot and large scale) or linear equations (small and fine scale). The plot-scale analysis showed that the overall spatial pattern was organized with a range of 6.3 m at this site, and the large-scale analysis showed another level of organization with a range of 2.0 m. The small- and fine-scale analyses displayed spatial autocorrelation, but range estimates could not be made because a sill was not reached within either analytical extent.

Because of the techniques and conventions used to construct these variograms, the smaller-scale autocorrelation structure of these communities was usually not visible in the variograms constructed for larger spatial extents. For example, in the plot-scale analysis, the data were binned with a lag distance of 0.5 m; this action made it impossible to detect the autocorrelation structure at the fine and small scales (less than 1 m). Similarly, the resolution associated with this bin size was not sufficient to allow us to accurately model spatial autocorrelation at the large scale. In order to study the autocorrelation structure at these other spatial scales, only the relevant sections of the variogram were considered. It is generally acceptable to analyze subsets of a variogram in this way, so long as there are enough data. A geostatistical 'rule of thumb' suggests that each distance class should contain at least 30 pairs of points; however, the greater the number of points, the greater the statistical reliability [40]. This guideline was far exceeded in all analyses, except the fine-scale analysis for the individual nodes, which only included 21 sampling locations.

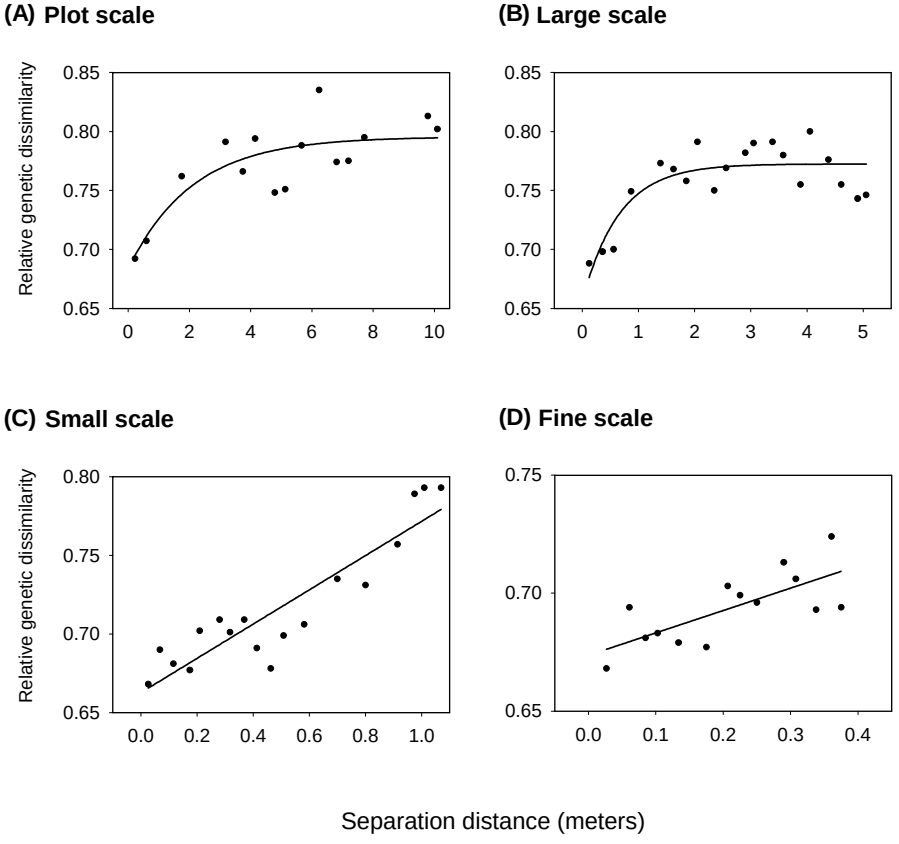

Fig. 2. Variograms used to model the overall spatial autocorrelation structure at each analytical scale. Data from all of the sampling locations were included to obtain an average portrait of the spatial relationships in the plot. A: Plot scale, all sampling locations. B: Large scale, separation distances less than 5 m. C: Small scale, separation distances less than 1 m. D: Fine scale, separation distances less than 0.4 m.

Table 1
Summary of results from geostatistical analyses of community structural similarity (AFLP profiles)

Scale and extent	Node	Model type	R^2	P	Nugget	Sill	Spatial dependence	Range (m)
Plot (0.025–11 m)	Entire plot	Exponential	0.67	0.001	0.68	0.80	0.14	6.3
Large (0.025–5 m)	Entire plot	Exponential	0.70	<0.0001	0.66	0.78	0.15	2.0
	A	Nugget			0.72			
	B	Exponential	0.74	0.002	0.67	0.80	0.16	1.3
	C	Exponential	0.93	<0.0001	0.63	0.76	0.17	3.3
	D	Exponential	0.93	<0.0001	0.65	0.83	0.21	1.9
	X	Exponential	0.96	0.002	0.56	0.82	0.32	1.8
Small (0.025–1 m)	Entire plot	Linear	0.83	<0.0001	0.66			>1.0
	A	Nugget			0.73			
	B	Exponential	0.92	0.006	0.66	0.76	0.12	0.6
	C	Exponential	0.63	0.03	0.60	0.70	0.15	0.6
	D	Nugget			0.73			
	X	Nugget			0.65			
Fine (0.025–0.4 m)	Entire plot	Linear	0.52	0.0034	0.67			>0.4
	A	Nugget			0.73			
	B	Exponential	0.50	0.03	0.66	0.72	0.09	0.3
	C	Linear	0.76	0.01	0.61			>0.4
	D	Linear	0.65	0.05	0.72			>0.4
	X	Linear	0.45	0.009	0.60			>0.4

3.2. Comparing patterns of spatial autocorrelation in different regions of the field

When the data from the different regions of the plot were analyzed separately, using each node (A–X) as a center point, other CLSs were detected (Table 1). For each scale and each sampling location, variograms were constructed using either 741, 528, or 210 pairwise comparisons (per node) for the large-, small-, and fine-scale analyses, respectively. In general, each dataset displayed an obvious linear or linear-sill pattern, and the appropriate model was applied. However, in a few cases, when visual interpretation of the data was unclear, it was necessary to fit the variogram with both a linear and an exponential equation, and use statistical criteria (R^2 and P value) to determine the most appropriate model.

For any given scale of analysis, the results for the different nodes were usually similar (Table 1). At the large scale, spatial autocorrelation was modeled at four of the five nodes, using an exponential equation; range estimates varied between 1.3 and 3.3 m, and the average of the different estimates was 2.0 m. For the small scale, significant models of spatial autocorrelation were only determined at two of the nodes, which produced identical range estimates of 0.6 m. At the fine scale, the exponential model was applied to node B, and the range estimate was 0.3 m. At nodes C, D, and X, a linear model was appropriate, and indicated that the communities are spatially autocorrelated with a range greater that 0.4 m, which was the maximum separation distance used at that level of analysis.

In general, the results for the different sampling locations (nodes) were similar, though each node displayed a unique multi-scale pattern of organization (Table 1). The same patterns of spatial organization were found at nodes D and X (identical CLSs for each scale), and the patterns observed at nodes B and C were very similar. At node A, spatial autocorrelation was not detected for any of the analytical scales.

3.3. Directional variograms

In order to determine whether the pattern of spatial variability changed with direction in the field, the data were also analyzed along each axis of the sampling grid. For the E/W transect, the variogram could be modeled using a linear equation (Table 2, Fig. 3B). For the N/S

Table 2
Summary of results of the geostatistical analyses of the directional variograms

Direction	Portion of variogram modeled	Model type	R^2	P	Nugget	Sill	Spatial dependence	Range (m)
North–south[a]								
Section 1	0–4 m	Exponential	0.92	<0.0001	0.65	0.84	0.23	1.8
Section 2	5–11 m	Exponential	0.64	0.01	0.68	0.84	0.18	1.8
East–west[b]								
Entire length	0–11 m	Linear	0.72	<0.0001	0.69	–	–	>11.0

[a]North–south refers to all points along the line from node B to node D.
[b]East–west refers to all points along the line from node A to node C.

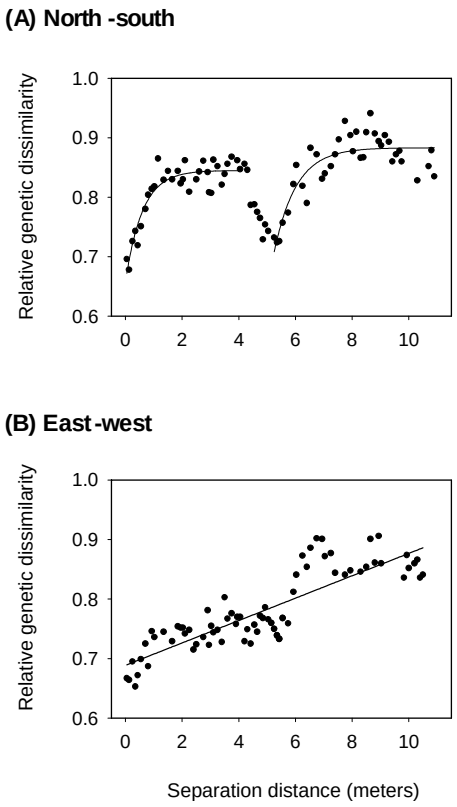

Fig. 3. 'Directional variograms' used to model the overall spatial autocorrelation structure along each axis of the sampling grid.

transect, multiple scales of spatial autocorrelation were observed within a single variogram (Fig. 3A); the variogram was divided into two regions, which were analyzed separately to estimate the range and spatial dependence using an exponential model (Table 2). A repeating pattern was evident, and each of the patches had a CLS of 1.8 m beyond the minimum (nugget) value. The nugget (section 1: 0.65, section 2: 0.68) and sill values (both are 0.84) for the two patches were identical, suggesting that the same amount of total variability was associated with each patch.

3.4. Kriging

Kriging was used to generate maps of the spatial distribution of microbial community structure for the entire plot (Fig. 4), and for each axis in the directional analyses (Fig. 5). First, the AFLP profiles were analyzed using PCA, and the sample scores from the first three PCs were used as derived variables in the geostatistical modeling (results not shown). Three separate maps were generated at the plot scale, one based on each PC. Together, these components explained 26.4% of the variance in microbial community structure (PC1: 12.6%, PC2: 8.3%, PC3: 5.5%). For each directional analysis, an additional PCA was performed using data from only the respective sampling locations. The first PC from each analysis was then used to generate a map of microbial community structure along each axis using ordinary kriging. The first PC explained 18% of the variance among the samples located along the N/S axis, and 15% of the variance among the samples located along the E/W axis.

4. Discussion

In order to more fully characterize the spatial variability of microbial systems, studies that make use of several different scales of measurement are necessary. In this research, multi-scale analysis of the spatial distribution of a soil microbial community revealed several different scales of organization, ranging from 30 cm to more than 6 m. In some locations, it was possible to identify and quantify up to four different CLSs. The patch size estimates varied some at the different sampling locations across the plot (different nodes) indicating that the patterns of spatial organization at a particular level (spatial scale) are not necessarily fixed across this system.

When the multi-scale approach was applied to analyze the entire plot, two distinct scales of organization were detected (large scale: 2.0 m, plot scale: 6.3 m). Multiple scales of spatial organization were also visible on the kriged maps, and each map showed a different spatial pattern (Fig. 4). The PCA used in the construction of these maps reduced the complex AFLP fingerprints into a set of derived variables, each of which explains a portion of the pattern present in the AFLP data. In this way, each PC describes a different aspect of the variability among the microbial communities, and the kriging results indicate that those distinct aspects have different patterns of spatial organization. These distinct patterns may develop as separate populations or groups of organisms respond to the spatial distribution of different environmental variables.

The map generated for the first PC shows a patchy structure organized around the center of the plot (node X). The patch in the center of the plot has a diameter of approximately 1.5–2 m, the CLS detected in the large-scale variogram analysis, and the next surrounding ring has a diameter of 5–6 m, which corresponds well to the CLS detected at the plot scale. One possible explanation for this bull's-eye pattern is that some aspect of the environment at the center of the plot is unique (e.g., a different vegetation patch, a hill or mound, or a large application of fertilizer), and the map shows the variation in community structure along a gradient away from this aberration. In contrast, the maps generated from the second and third PCs reveal a very different spatial pattern. The portion of the communities represented on these maps may be responding to a suite of variables that are spatially structured as a gradient extending from the NE corner to the SW corner of the plot.

In addition to quantifying the overall pattern of spatial organization in this system, our study was designed to evaluate how variable the autocorrelation structure was

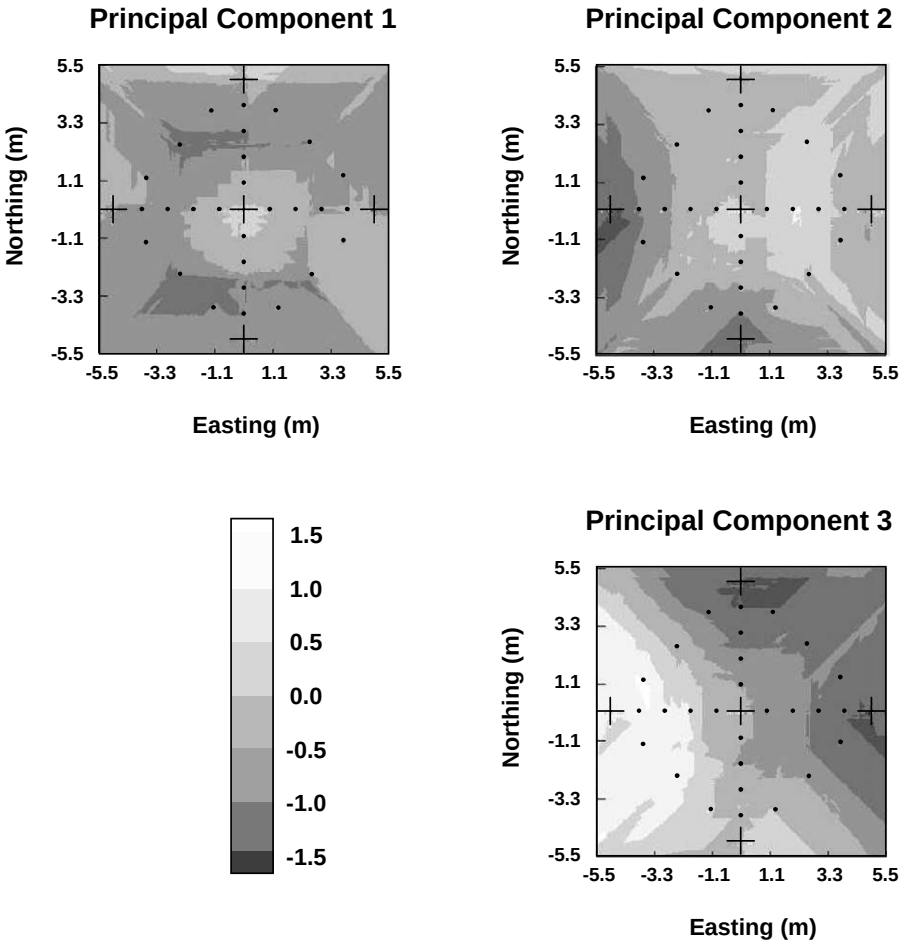

Fig. 4. Kriged maps showing the distribution of microbial community structure across the plot, as described by the first three PCs. Each graph is oriented to match Fig. 1A; sampling nodes are indicated with '+', and perimeter and diagonal sampling locations are marked with '●'.

in different locations. In general, the results for the different sampling locations (nodes) were similar and multiple CLSs were detected, though spatial autocorrelation was not detected at node A. The samples collected at this location contained an unusually large amount of plant and root material, which should have been removed by sieving, but could have contaminated the DNA extracts. However, given that the AFLP profiles for node A are not significantly different from those obtained at other locations, we feel this is unlikely. Instead, these results may be correlated with some environmental heterogeneity that altered the spatial organization, but not the overall composition, of the communities in this region of the field.

Spatial dependence is the percent of total model variance that is explained by the spatial autocorrelation function. When this value is low (all variance in nugget), it indicates that most of the spatial dependence occurs at distances greater or smaller than the scale of study, or that the measurement error associated with the analysis is high [41]. In the present study, the spatial dependence for any single location (node) or particular analytical scale ranged from 0.09 to 0.32. These values are consistent with other studies considering community structure [15,19]. For example, when Saetre and Bååth [18] performed a geostatistical analysis of the overall microbial community structure in a forest soil (using phospholipid fatty acid profiling (PLFA), the spatial dependence varied between 0.12 and 0.25 for an analytical scale of 0.2–20 m. The values discussed above refer to the spatial dependence for the analysis of a single spatial scale. However, if the spatial dependence in the present study is summed across each of the four analytical scales (either for each node, or for the entire plot), the total estimate increases to between 0.35 and 0.45.

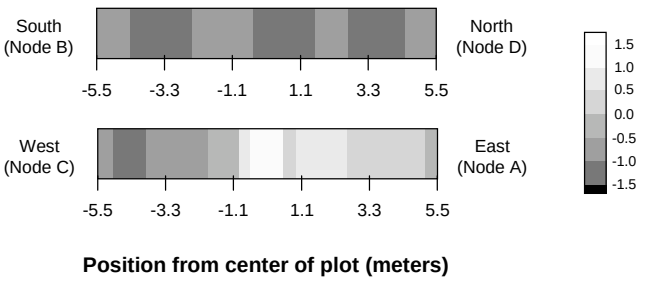

Fig. 5. Kriged maps showing the distribution of microbial community structure, as described by the first PC, for each directional analysis.

Communities at this site are expected to display additional spatial structure at scales larger than the maximum separation distance used (11 m) and at distances smaller than the minimum sampling interval (2.5 cm). In this study, it would have been useful to analyze multiple subsamples from each field sample to determine the variability within a sampling unit. This would have provided additional information regarding community variability at small spatial scales, and would have functionally represented a separation distance of zero. Unfortunately, increasing the number of analyses was not feasible for this study. Recently, Ellingsøe and Johnsen [42] considered the influence of soil sample size on the analysis of bacterial community structure (using denaturing gradient gel electrophoresis (DGGE)), and found that sample size did influence their assessment of community structure for smaller sample sizes (0.01 and 0.1 g of soil), but less so for larger sample sizes (1.0 and 10.0 g). In our study, each DNA extraction was performed on a single 0.25-g subsample, which may not have been sufficient to completely categorize the community variability of each field sample (5–10 g).

In the directional analyses, the variograms (Fig. 3) matched up very well with the kriged maps generated from the PC scores (Fig. 5). In the E/W variogram, a linear pattern was observed; communities that were nearby along this axis were more similar to one another than they were to communities at greater separation distances, though all samples along the transect were spatially autocorrelated. On the map, patches/communities that are nearby have similar PC scores (more similar grayscale values), but there is very little repetition of an individual community 'type'. There is a general gradient along the axis that corresponds to the gradient observed in the maps for the entire plot (Fig. 4), and it is likely that the same environmental factors are correlated/responsible for this pattern in both situations.

Along the N/S axis, two types of communities dominate on the map of community structure (Fig. 5). The patch size for these communities is about 2 m, which is the CLS calculated from the nested variogram (Fig. 3). It is important to note that, though the kriged map shows only two dominate communities, with PC scores ranging from −1.5 to −0.5, a frequency histogram of the entire set of PC scores showed a normal distribution with values ranging from −1.5 to 2.5 (results not presented). The presence of a regular and repeating spatial pattern along this axis suggests that the microbial communities may be partially structured in response to some agricultural or land management activity that occurs at fixed intervals in the field. For example, it has been suggested that spatial structure may exist in agricultural soils in association with crop rows and aisles [16,25,43], and compaction due to wheel traffic may impact microbial activity [44]; however, we feel that the CLS of 2 m is too large to correspond with these particular features.

Often, when scientists research and discuss the existence of multiple scales of spatial organization in microbial systems, they are referring to the presence of patterns over a very wide range. For example, Parkin [44] discussed four main scales of interest: microscale, plot scale, field or landscape scale, and regional scale, and Ettema and Wardle [25] primarily focused their recent review on the distribution of soil properties and biota in distance classes of tens to hundreds of meters, cm to meters, and at microscopic scales. The research presented here demonstrates that a single variable can manifest an incredible amount of spatial structure, at multiple scales, within these broad classifications. For example, at node B, four different CLSs were detected: 30 cm, 60 cm, 1.3 m, and 6.3 m. This is a remarkable degree of spatial variability for a pedagogically homogeneous site that has been plowed and cropped as a single field for several years. Variability such as this is likely to exist in most ecosystems, and should be considered when making inferences about ecological relationships and when developing sampling strategies [16].

While many ecological theories and models acknowledge that elements that are close to one another in space or time are more likely to be influenced by the same generating processes, the classical statistical procedures employed to analyze these phenomena assume an independence of observations. Violations of the assumption of independence and inappropriate application of these statistical procedures to spatially autocorrelated data may lead to incorrect conclusions [45–47]. For example, Franklin et al. [15] found that estimates of microbial abundance obtained using spatially autocorrelated data were significantly different from those obtained using independent samples. The varying degrees of autocorrelation shown in the present study emphasize that sampling approaches and experimental designs may need to consider the impact of spatial autocorrelation, depending on the ecological question of interest. Because it is not always feasible to first do an extensive reconnaissance survey, and because the results presented here suggest that one is not likely to avoid the impact of spatial autocorrelation, a possible solution is to include spatial separation as a part of routine data collection. An initial analysis of this information can then be used to determine the influence of spatial autocorrelation on the dataset. If significant spatial structure is found, this information must be considered as a variable and incorporated into the subsequent data analysis; if not, traditional parametric statistical techniques may be appropriate.

Most of the previous work examining spatial variability in agricultural systems has been performed by soil scientists who were interested in understanding the distribution of physical and chemical properties in order to assess soil quality or determine the impact of land management practices [16,26,48–50]. These studies have shown that spatial autocorrelation is a common feature of these systems. The

spatial variability associated with microbial communities is less frequently studied [16,23,49], and, in general, efforts to link agricultural soil properties and microbiological properties have been unsuccessful. To our knowledge, the work presented here is one of the first studies of spatial organization of community structure in this type of environment, and is unique in its consideration of multiple scales of autocorrelation. The results indicate that microbial communities may have several nested levels of organization, even within the cm- to meter-scale analysis. Different subsets of the community were distributed differently across the plot, and this is thought to be due to the variable response of individual populations to the spatial heterogeneity associated with different soil properties (or groups of properties). Future studies that focus on comparing the spatial structure of microbial communities with that of environmental properties may yield new insights into how communities develop in soil systems, and what factors may be important in maintaining and regulating soil ecosystem function.

Acknowledgements

This work was funded by a NASA GSRP (Grant NGT10-52620) and a University of Virginia, Department of Environmental Sciences, Moore Research Award. Access to the study site was provided by the Nature Conservancy and the Virginia Coast Reserve Long Term Ecological Research (VCR-LTER) Program.

References

[1] Wiens, J.A., Addicott, J.F., Case, T.J. and Diamond, J. (1986) The importance of spatial and temporal scale in ecological investigations. In: Community Ecology (Diamond, J. and Case, T.J., Eds.), pp. 145–172. Harper and Row, New York.
[2] Rahel, F.J. (1990) The hierarchical nature of community persistence – a problem of scale. Am. Nat. 136, 328–344.
[3] Nunan, N., Wu, K., Young, I.M., Crawford, J.W. and Ritz, K. (2002) In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. Microb. Ecol. 44, 296–305.
[4] Nunan, N., Ritz, K., Crabb, D., Harris, K., Wu, K.J., Crawford, J.W. and Young, I.M. (2001) Quantification of the in situ distribution of soil bacteria by large-scale imaging of thin sections of undisturbed soil. FEMS Microbiol. Ecol. 37, 67–77.
[5] Dandurand, L.M., Knudsen, G.R. and Schotzko, D.J. (1995) Quantification of *Pythium-Ultimum Var Sporangiiferum* zoospore encystment patterns using geostatistics. Phytopathology 85, 186–190.
[6] Fendorf, S.E., Li, G.C., Morra, M.J. and Dandurand, L.M. (1997) Imaging a pseudomonad in mineral suspensions with scanning force and electron microscopy. Soil Sci. Soc. Am. J. 61, 109–115.
[7] Grundmann, G.L. and Debouzie, D. (2000) Geostatistical analysis of the distribution of NH_4^+ and NO_2^--oxidizing bacteria and serotypes at the millimeter scale along a soil transect. FEMS Microbiol. Ecol. 34, 57–62.
[8] Dandurand, L.M., Schotzko, D.J. and Knudsen, G.R. (1997) Spatial patterns of rhizoplane populations of Pseudomonas fluorescens. Appl. Environ. Microbiol. 63, 3211–3217.
[9] Bending, G.D., Shaw, E. and Walker, A. (2001) Spatial heterogeneity in the metabolism and dynamics of isoproturon degrading microbial communities in soil. Biol. Fertil. Soils 33, 484–489.
[10] Both, G.J., Gerards, S. and Laanbroek, H.J. (1992) Temporal and spatial variation in the nitrite-oxidizing bacterial community of a grassland soil. FEMS Microbiol. Ecol. 101, 99–112.
[11] Felske, A. and Akkermans, A.D.L. (1998) Spatial homogeneity of abundant bacterial 16S rRNA molecules in grassland soils. Microb. Ecol. 36, 31–36.
[12] Morris, S.J. (1999) Spatial distribution of fungal and bacterial biomass in southern Ohio hardwood forest soils: fine scale variability and microscale patterns. Soil Biol. Biochem. 31, 1375–1386.
[13] Saetre, P. (1999) Spatial patterns of ground vegetation, soil microbial biomass and activity in a mixed spruce-birch stand. Ecography 22, 183–192.
[14] Smith, J.L., Halvorson, J.J. and Bolton, H. (1994) Spatial relationships of soil microbial biomass and C and N mineralization in a semiarid shrub-steppe ecosystem. Soil Biol. Biochem. 26, 1151–1159.
[15] Franklin, R.B., Blum, L.K., McComb, A.C. and Mills, A.L. (2002) A geostatistical analysis of small-scale spatial variability in bacterial abundance and community structure in salt marsh creek bank sediments. FEMS Microbiol. Ecol. 42, 71–80.
[16] Robertson, G.P., Klingensmith, K.M., Klug, M.J., Paul, E.A., Crum, J.R. and Ellis, B.G. (1997) Soil resources, microbial activity, and primary production across an agricultural ecosystem. Ecol. Appl. 7, 158–170.
[17] Kuperman, R.G., Williams, G.P. and Parmelee, R.W. (1998) Spatial variability in the soil foodwebs in a contaminated grassland ecosystem. Appl. Soil Ecol. 9, 509–514.
[18] Saetre, P. and Bååth, E. (2000) Spatial variation and patterns of soil microbial community structure in a mixed spruce-birch stand. Soil Biol. Biochem. 32, 909–917.
[19] Cavigelli, M.A., Robertson, G.P. and Klug, M.J. (1995) Fatty-acid methyl-ester (FAME) profiles as measures of soil microbial community structure. Plant Soil 170, 99–113.
[20] Ranjard, L., Poly, F., Combrisson, J., Richaume, A., Gourbiere, F., Thioulouse, J. and Nazaret, S. (2000) Heterogeneous cell density and genetic structure of bacterial pools associated with various soil microenvironments as determined by enumeration and DNA fingerprinting approach (RISA). Microb. Ecol. 39, 263–272.
[21] Acosta, D. and Lynn, D.H. (2002) A preliminary assessment of spatial patterns of soil ciliate diversity in two subtropical forests in Puerto Rico and its implications for designing an appropriate sampling approach. Soil Biol. Biochem. 34, 1517–1520.
[22] Parkin, T.B., Starr, J.L. and Meisinger, J.J. (1987) Influence of sample size on measurement of soil denitrification. Soil Sci. Soc. Am. J. 51, 1492–1501.
[23] Robertson, G.P., Huston, M.A., Evans, F.C. and Tiedje, J.M. (1988) Spatial variability in a successional plant community: patterns of nitrogen availability. Ecology 69, 1517–1524.
[24] Gorres, J.H., Dichiaro, M.J., Lyons, J.B. and Amador, J.A. (1998) Spatial and temporal patterns of soil biological activity in a forest and an old field. Soil Biol. Biochem. 30, 219–230.
[25] Ettema, C.H. and Wardle, D.A. (2002) Spatial soil ecology. Trends Ecol. Evol. 17, 177–183.
[26] Stenger, R., Priesack, E. and Beese, F. (2002) Spatial variation of nitrate-N and related soil properties at the plot-scale. Geoderma 105, 259–275.
[27] Robertson, G.P. and Gross, K.L. (1994) Assessing the heterogeneity of belowground resources: quantifying pattern and scale. In: Exploitation of Environmental Heterogeneity by Plants (Caldwell, M.M. and Pearcy, R.W., Eds.), pp. 237–253. Academic Press, San Diego, CA.
[28] Bruckner, A., Kandeler, E. and Kampichler, C. (1999) Plot-scale spatial patterns of soil water content, pH, substrate-induced respiration and N mineralization in a temperate coniferous forest. Geoderma 93, 207–223.

[29] Smith, J.L., Halvorson, J.J. and Bolton Jr., H. (2002) Soil properties and microbial activity across a 500 m elevation gradient in a semi-arid environment. Soil Biol. Biochem. 34, 1749–1757.
[30] Zabeau, M. and Vos, P. (1993) Selective Restriction Fragment Amplification: A General Method for DNA Fingerprinting. European Patent Office, EP 0534858 A1.
[31] Franklin, R.B., Taylor, D.R. and Mills, A.L. (1999) Characterization of microbial communities using randomly amplified polymorphic DNA (RAPD). J. Microbiol. Methods 35, 225–235.
[32] Wikström, P., Andersson, A.-C. and Forsman, M. (1999) Biomonitoring complex microbial communities using random amplified polymorphic DNA and principal component analysis. FEMS Microbiol. Ecol. 28, 131–139.
[33] Xia, X., Bollinger, J. and Ogram, A. (1995) Molecular genetic analysis of the response of three soil microbial communities to the application of 2,4-D. Mol. Ecol. 4, 17–28.
[34] Franklin, R.B., Garland, J.L., Bolster, C.H. and Mills, A.L. (2001) Impact of dilution on microbial community structure and functional potential: Comparison of numerical simulations and batch culture experiments. Appl. Environ. Microbiol. 67, 702–712.
[35] Sneath, P.H. and Sokal, R.R. (1973) Numerical Taxonomy. W.H. Freeman and Company, San Francisco, CA.
[36] Mackas, D.L. (1984) Spatial autocorrelation of plankton community composition in a continental shelf ecosystem. Limnol. Oceanogr. 29, 451–471.
[37] Underwood, A.J. and Chapman, M.G. (1998) A method for analyzing spatial scales of variation in composition of assemblages. Oecologia 117, 570–578.
[38] Sokal, R.R., Bird, J. and Riska, B. (1980) Geographic variation in *Pemphigus populicaulis* (Insecta: Aphididae) in eastern North America. Biol. J. Linn. Soc. 14, 163–200.
[39] Kuzyakova, I.F., Romanenkov, V.A. and Kuzyakov, Y.V. (2001) Application of geostatistics in processing the results of soil and agrochemical studies. Eurasian Soil Sci. 34, 1219–1228.
[40] Rossi, R.E., Mulla, D.J., Journel, A.G. and Franz, E.H. (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecol. Monogr. 62, 277–314.
[41] Isaaks, E.H. and Srivastava, R.M. (1989) An Introduction to Applied Geostatistics. Oxford University Press, New York.
[42] Ellingsoe, P. and Johnsen, K. (2002) Influence of soil sample sizes on the assessment of bacterial community structure. Soil Biol. Biochem. 34, 1701–1707.
[43] Stoyan, H., De-Polli, H., Bohm, S., Robertson, G.P. and Paul, E.A. (2000) Spatial heterogeneity of soil respiration and related properties at the plant scale. Plant Soil 222, 203–214.
[44] Parkin, T.B. (1993) Spatial variability of microbial processes in soil – a review. J. Environ. Qual. 22, 409–417.
[45] Dutilleul, P. (1993) Spatial heterogeneity and the design of ecological field experiments. Ecology 74, 1646–1658.
[46] Legendre, P. and Legendre, L. (1998) Numerical Ecology. Elsevier Science, Amsterdam.
[47] Murray, C.J. (2001) Sampling and data analysis for environmental microbiology. In: Manual of Environmental Microbiology (Hurst, C.J., Crawford, R.L., Knudsen, G.R., McInerney, M.J. and Stetzenbach, L.D., Eds.), pp. 166–177. American Society for Microbiology Press, Washington, DC.
[48] Dobermann, A., Goovaerts, P. and George, T. (1995) Sources of soil variation in an acid ultisol of the Philippines. Geoderma 68, 173–191.
[49] Savin, M.C., Gorres, J.H., Neher, D.A. and Amador, J.A. (2001) Biogeophysical factors influencing soil respiration and mineral nitrogen content in an old field soil. Soil Biol. Biochem. 33, 429–438.
[50] Castrignanò, A., Giugliarini, L., Risaliti, R. and Martinelli, N. (2000) Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97, 39–60.