Evaluation of the Benchmark Soil Concept in Relation to Soil Microbial Community Structure and Function

David Crowley*¹, Rosario Pineda¹, Jean Paul Baquiran¹

Project Objectives

- Evaluate the effects of soil chemical and physical variables on microbial community composition under standard conditions using the California benchmark reference soil series.
- Evaluate the stability (resistance to change) of soil microbial communities in relation to organic matter, clay content, and other physical and chemical variables.

Approach and Procedures

To develop bioindicators of soil quality, it is necessary to understand the physical and chemical factors that shape microbial communities in different soils as well as the influence of plant cover. The objectives of this project were to examine the relationship between soil chemical, physical, and biological properties using PLFA biochemical fingerprinting methods and to evaluate the use of neural network modeling procedures for study of these relationships. The research conducted here specifically addressed the following hypotheses:

I. Chemical and physical properties used to classify benchmark groups are reliable determinants of microbial community composition in different soils when all other dynamic variables are brought to standard conditions.

II. Microbial community structures and activities for soil series within benchmark groups are more similar to each other than to series from different soil benchmark groups when compared under standardized environmental conditions.

III. Stability (resistance to change) of the soil microbial community composition (as determined by phospholipid fatty acid profiling) and function is directly proportional to the organic matter and clay content of a soil series.

Field samples of benchmark soils were collected and brought to the laboratory for growth chamber experiments to examine the effects of different plant species and environmental variables on microbial community structures. After 45 days, the soils were destructively analyzed to determine microbial community structures by PLFA (phospholipid fatty acid analysis). A total of 13 soils from benchmark locations across southern California were contrasted. Subsamples from each location were analyzed to determine texture, pH, salinity and other chemical and physical properties. The resulting data were analyzed using neural network modeling procedures using the program Synapse (Peltarion Inc.).

Two types of artificial neural network models were used to analyze the data. One used an unsupervised model called a Kohonen Self Organizing Map (KSOM). The second was a function model in which the model is trained to predict the relationship between selected input and output variables. The general concept for generation of KSOM models is shown below (Figure 1),

¹Department of Environmental Sciences, University of California, Riverside, CA 92521
*Principal Investigator
For more information contact Dr. David Crowley (Crowley@ucr.edu)
which illustrates how neural networks generate self-organizing maps using computational methods. The model consists of two layers comprising an input layer that has nodes representing each input variable, and an output layer (Kohonen map) that is a 2-dimensional array of nodes. The input and output layers are linked by mathematical functions between every node in each layer. The computational model uses randomly selected input data (sample units) and calculates the distance between the input data and every node in the output layer. The program uses an ordering phase and a tuning phase to group all the input variables that vary similarly. The resulting map contains color coded or shaded hexagons that represent individual abiotic and biotic variables, site locations, and other input information. The summary diagram is displayed as a “U-matrix” or unified matrix. Each of the component planes can be displayed separately and compared visually to examine their clustering relationships with all of the other variables.

The second type of ANN model that was employed uses a supervised training, in which selected input variables are used to predict one or more output variables. The model is trained using a randomly selected portion of the data set for a specific number of cycles in which the program analyzes the error between the true and predicted values and adjusts the weights of the mathematical functions linking the hidden layer to the input and output layers. In most cases, approximately 5000 cycles were run to generate a model that fit the data, but which was not overtrained to the extent that it could no longer fit the validation data set. This is empirically determined by examining the error and confidence intervals for the training and validation data sets over a range of training cycles. In the post-processing phase, the output page generated by Peltarion synapses provides a set of slide bars for all of the input variables that can be adjusted individually to examine the effects of individual variables or combinations of variables on the selected output variable. In this manner, response curves can be generated and plotted to illustrate the response of an output variable to each factor while the remaining input variables are held constant. To describe the community structure, it is useful to reduce the multidimensional data set generated by PLFA to principal components (Stenberg et al., 1998). Here, we also examined variation for microbial biomass and fungal bacterial ratios in relation to clay, organic matter, and soil pH. The PLFA data were also analyzed by cluster analysis and discriminant analysis to determine the extent to which more traditional statistical analyses could associate differences between the microbial community structures and the different benchmark soil series.
Results

Analysis of the microbial communities yielded data that enabled comparison of microbial community structures in relation to specific chemical and physical variables and the type of plant that was grown. The results showed that fescue, corn, and bean each produced different microbial community structures, but that the soil chemical and physical properties also interacted with plant species to shape the communities. The stabilities of the community (resistance to change) were examined with respect to changes in plant species, and revealed that each soil-plant combination responded differently. Figure 2 contains the plots that were generated using a KSOM model, which allows visual comparison of the correlations between all of the variables.

As shown in Figure 2, each box represents an individual input or output variable. Elevated values are indicated by red, and low values are indicated by blue. Areas within the boxes that are similar in color indicate positive correlations between variables; whereas opposite colors in the same region reflect inverse relationships. In this manner, all of the data can be visualized in a single pass that enables the study of relationships. Following this procedure, observed patterns can be further evaluated using other statistical procedures, or by testing in further experiments. As an example in the figure, color patterns in the boxes for soil clay and sand are inversely related. Similarly Figure 2 indicates that the plant species and soil series each generated strong, distinct effects that segregated out from the data set.

Many other relationships can be explored using the KSOM output display. As shown in Figure 2 in the section for soil series, each soil presented between 1 to 3 different pattern groupings. This variation in the number of groupings likely corresponds to the influence exerted by the three different plant species, but also corresponds to the influence of all other variables that are included in the model. In soils showing only 1 pattern cluster (Crouch, Milham, San Emigdio), there was no discernable influence of plant species (or other variables) that caused the communities to shift in structure. These soils can thus be considered as highly stable (resistant to change). In those showing two pattern clusters, this generally corresponded to one community structure that formed under the bean and corn, and a second that formed under fescue, and is likely related to
Figure 2. Kohonen self organizing map of physical, chemical, and biological variables for 13 benchmark soils from S. California following growth of fescue, corn, or bean plants. Color patterns in each box indicate ranges for each variable (red=high, blue=low). Comparison of patterns between boxes revealed simultaneous correlations between all variables as determined using an artificial neural network model.
Evaluation of the Benchmark Soil Concept in Relation to Soil Microbial Community Structure and Function—Crowley

differences in the fungal:bacterial ratios (discussed below). In soils with three pattern groups (Los Osos, Pico, and Sorrento), each plant species apparently exerted a distinct effect on microbial community structure. There was no correspondence between microbial community stability and soil clay or organic matter content. These data thus do not support hypothesis three, which proposed that clay and organic matter increase microbial community stability (resistance to change), at least in response to the type of plant cover.

Using the PLFA data alone, it was not possible to distinguish the different benchmark soils using a discriminant analysis. Only 20% of the soils were correctly classified to their corresponding soil series using the PLFA data alone (data not shown). Based on these results, the data do not support hypothesis two, which proposed that soils within a benchmark series were more similar than across series; as inability to distinguish between broad level benchmark groups negates the ability to compare fine level resolution within a benchmark series. On the other hand, when combined with the physical and chemical variables, it is possible to distinguish each soil series individually. The microbial community structures and soil similarity groupings are further determined by which plant species was cultivated in the soil. Multiple cluster analyses were run using different combinations of variables for the soils when cultivated with different plant species. The resulting cluster analyses showed that cluster groupings were highly dependent on which plants were grown. Bean plants in particular exerted a strong influence on microbial community structures, reducing overall differences in similarities among the communities. This may be due to differences in the root mass and amounts of exudate that were secreted in soils after six weeks as compared to the smaller plant mass in soils planted with fescue.

Analysis of the interactions among the physical and chemical variables showed that there were complex interactions among the variables that could not be separated by traditional statistical analyses, but which could be separated using the pattern recognition capability of artificial neural network models. After generating a model, the influence of individual variables such as clay, pH, and organic matter content could then be inspected by setting all variables at fixed values (average values for the entire data set), and then examining the effects of selected individual variables individually or in pairs. These results were used to generate three-dimensional surface plots for selected dependent variables including microbial biomass and fungal bacterial ratio.

Using the combined data set across all three plant experiments, an artificial neural network function model was run to examine the effect of organic matter and clay content on the microbial biomass (Figure 3). These data showed that microbial biomass increased with increasing clay and organic matter, and that there was an additive effect of the two independent variables. At low organic matter content, biomass increased with increasing clay content up to 40% clay, after which it leveled off. In contrast, in soils with higher organic matter content, the microbial biomass continued to increase with clay content up to 60% clay, the upper end of the range represented by these soils.

After training of the ANN model, sensitivity analyses were conducted to examine the effects of individual variables on broad level properties including a marker for arbuscular mycorrhizae, total bacterial and fungal biomass, fungal biomass, and the ratio of fungal to bacterial biomass. The results showed that AM fungi increased with clay content, organic
Evaluation of the Benchmark Soil Concept in Relation to Soil Microbial Community Structure and Function—Crowley

**Figure 3.** Microbial biomass (total PLFA) as controlled by soil clay content and organic carbon. Values predicted using an ANN model for 13 benchmark soils. Predicted values for biomass obtained by fixing all other variables at mean values measured for the entire data set.

Clay and organic matter content also affected the broad community structure as reflected by the fungal:bacterial ratio (F:B ratio) generated using PLFA markers (Figure 4). In this case, the F:B ratio was further influenced by plant species as shown in the three separate surface response plots generated for soils under each plant type. Under fescue, the F:B ratio increased with increasing clay content up to 30% clay, after which bacteria became increasingly predominant in soils with high clay. Organic matter exerted little influence on the F:B ratio in soils with fescue, with only a slight increase predicted to occur with increasing organic matter content.

In contrast to the data for fescue, soils planted with bean or corn both showed synergistic relationships between clay and organic matter content with respect to the F:B ratio. At low clay or organic matter content, soils were dominated by bacteria and there was no effect of increasing organic matter in soils with low clay content, or of increasing clay content in soils with low organic matter. However, as clay and organic matter content increased together, the F:B ratio values increased rapidly.

The ANN model was also used to examine several other factors and their influence on community structure and further confirmed the complex interactions that occurred between soil physical and chemical variables on microbial biomass and community structure. Figure 5 shows...
Evaluation of the Benchmark Soil Concept in Relation to Soil Microbial Community Structure and Function—Crowley

a surface response plot for microbial biomass as affected by organic matter content and pH in soils under fescue and bean. In soils under fescue with low organic matter content, the soil pH appeared to have a strong effect, such that microbial biomass increased nearly five fold as soil pH increased from 6.5 to 8.5. At higher organic matter levels, the effect of increasing pH was slightly diminished. A similar pattern was observed with soils with bean plants, although the responses were greatly muted as compared to those obtained in soils with fescue.

Figure 4. Predicted relationship between soil organic matter content and soil clay content on the fungal:bacterial ratios in S. California benchmark soils under fescue (top left), bean (top right), and maize (bottom left).

Figure 5. Predicted microbial biomass response to differences in soil organic matter and pH for fescue (left) and bean (right).
Discussion

Results of this research strongly support the hypothesis that soil chemical and physical variables are strong determinants of soil microbial community structure. However, this and prior studies show that the community structure is also strongly influenced by the type of plant species cover and there are complex and largely unpredictable interactions that occur in response to different combinations of variables (Bossio et al., 1998; Balser et al., 2005). In some cases, there were strong effects of soil-plant combinations on community structure, whereas in other soils, the community structure was relatively stable and non-responsive to changes in plant species. Soil pH, clay, and organic matter content were shown to be strong determinants of community structure, and interacted in complex ways. These patterns were readily separated using the artificial neural network modeling approach which allowed the analysis of responses to individual variables and combinations of variables across the entire data set.

PLFA analysis inherently provides a relatively low resolution method for studying changes in the composition of microbial communities. Here some twenty PLFAs were used to describe changes in microbial communities that consist of 10s of thousands of bacterial and fungal species, and there are many assumptions that such PLFA markers reflect differences in the taxonomic composition of microbial communities. At this time, there is still little information in the literature on the correspondence between shifts in PLFA composition and taxonomic composition as determined by higher resolution markers such as 16S rRNA genes, or the evolutionary distance of the predominant types of microorganisms. In this respect, the best use of PLFA is for examining broad level shifts associated with F:B ratios in soil, but even here PLFA is not precise as bacteria and fungi may share common marker fatty acids. Certainly changes in PLFA indicate that communities have shifted in composition, and communities with similar PLFA profiles can be hypothesized as likely similar in species composition. PLFA thus provides a useful tool for examining overall impacts of variables on stability of the microbial community.

In the original proposal, it was hypothesized that PLFA signatures could be used to distinguish different benchmark soils, and soil series within benchmark groups. This hypothesis (Hypothesis II) was not supported by the results of this research, which instead showed that physical and chemical variables distinguish these soils, but that there are no consistent patterns in the marker PLFAs, or principal components generated from PLFAs, that are associated with these physical and chemical variables. On the other hand, this research showed that broad community parameters such as biomass and F:B ratio could be separated by pattern analysis.
using ANN models and thereafter predicted from specific combinations of the key drivers: pH, clay, and organic matter content. When these models are generated, a portion (20%) of the data was randomly reserved for use as a validation data set, and the models provide a confidence interval for each of the output variables that were analyzed. This modeling approach thus provides a very powerful tool for separating out the effects of individual input variables on selected parameters.

The underlying approach using artificial neural network models that was used for this research is especially applicable for extracting patterns among data sets in which there may be complex nonlinear (chaotic) interactions that cannot be detected with traditional methods for multivariate analyses (Kampichler et al., 2000; Noble et al., 2000; Lentzsh et al., 2005). The results from the analysis provide a 2- or 3-dimensional map of clusters that simultaneously display possible co-relationships between all of the variables that are introduced into the model. This information can then be used to generate hypotheses as to cause and effect relationships, and for forecasting changes in response variables over time. Although widely used in areas of investigation ranging from medical diagnostic analysis, economic forecasting, and an array of military applications, there are few studies in which ANN have been applied for environmental analyses. Among the range of applications for environmental studies, ANNs have been shown to provide a powerful approach to classify soil structure and physical properties (Levine et al., 1996; Chang et al., 2000), for soil and vegetation mapping (Tan et al., 1996; Deadman et al., 1997), prediction of non-point source pollution (Muleta et al., 2005), prediction of soil salinization (Patel et al., 2002), and for assessment of soil organic carbon accumulation under different land-use schemes (Somaratne et al., 2005). For studies on soil biological data, neural network analyses have most recently been used to relate microbial biomass and 16S rRNA gene banding patterns to soil texture and other chemical properties (Lentzsh et al., 2005; Ramadan et al., 2005).

In using ANN models, it is tempting to conclude that the input variables used to describe a particular parameter such as biomass, F:B ratio or soil protein content, have a cause and effect relationship. Nonetheless, any input variable such as soil pH, organic matter content, or clay content exerts its influence on the microbial community via other factors that come into play such as soil aeration, aggregate stability, nutrient concentrations, and other intermediate factors that directly influence microbial physiology and selection for the most fit species under those conditions. As we delve deeper into studies on the factors that shape microbial community structure and function, pattern recognition modeling using artificial neural networks will enable deeper insight into possible relationships that exist between all of the physical and chemical variables that are currently used to describe soils. The results of this research suggest that there are very complex (nonlinear) interactions that may be impossible to predict using traditional statistical analysis and multivariate models that cannot separate out chaotic patterns. If so, prediction of model community structures using PLFA and sets of input variables such as plant cover, soil texture, organic matter, etc. may be impossible, and will require empirical analysis of individual plant-soil combinations. Future research will determine the extent to which specific soil functions such as disease suppression, phosphorus solubilization, and N mineralization are associated with different chemical, physical, and biological markers. This will likely be a huge challenge, but one in which ANN models will facilitate better predictions for use of this information in analyzing the relationship between soil quality, productivity, and management practices.
Evaluation of the Benchmark Soil Concept in Relation to Soil Microbial Community Structure and Function—Crowley

References


This research was funded by the Kearney Foundation of Soil Science: Understanding and Managing Soil-Ecosystem Functions Across Spatial and Temporal Scales, 2006-2011 Mission (http://kearney.ucdavis.edu). The Kearney Foundation is an endowed research program created to encourage and support research in the fields of soil, plant nutrition, and water science within the Division of Agriculture and Natural Resources of the University of California.