Applied and Environmental Microbiology, Feb. 2000, p. 694–699 Vol. 66, No. 2
0099-2240/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Application of Neural Computing Methods for Interpreting Phospholipid Fatty Acid Profiles of Natural Microbial Communities†

PETER A. NOBLE,[1]* JONAS S. ALMEIDA,[2] and CHARLES R. LOVELL[3]

Belle W. Baruch Institute for Marine Biology and Coastal Research[1] and Department of Biological Sciences,[3] University of South Carolina, Columbia, South Carolina 29208, and Chemistry Department, FCT/UNL, and ITQB/UNL, 2780-901 Oeiras, Portugal[2]

Received 27 August 1999/Accepted 28 October 1999

The microbial community compositions of surface and subsurface marine sediments and sediments lining burrows of marine polychaetes and hemichordates from the North Inlet estuary (near Georgetown, S.C.) were analyzed by comparing ester-linked phospholipid fatty acid (PLFA) profiles with a back-propagating neural network (NN). The NNs were trained to relate PLFA inputs to sediment type outputs (e.g., surface, subsurface, and burrow lining) and worm species (e.g., *Notomastus lobatus*, *Balanoglossus aurantiacus*, and *Branchyoasychus americana*). Sensitivity analysis was used to determine which of the 60 PLFAs significantly contributed to training the NN. The NN architecture was optimized by changing the number of hidden neurons and calculating the cross-validation error between predicted and actual outputs of training and test data. The optimal NN architecture was found to be four hidden neurons with 60-input neurons representing the 60 PLFAs, and four output neurons coding for both sediment types and worm species. Comparison of cross-validation results using NNs and linear discriminant analysis (LDA) revealed that NNs had significantly fewer incorrect classifications (2.7%) than LDA (8.4%). For the NN cross-validation, both sediment type and worm species had 3 incorrect classifications out of 112. For the LDA cross-validation, sediment type and worm species had 7 and 12 incorrect classifications out of 112, respectively. Sensitivity analysis of the trained NNs revealed that 17 fatty acids explained 50% of variability in the data set. These PLFAs were highly different among sediments and burrow types, indicating significant differences in the microbiota.

Natural microbial communities are complex assemblages of organisms composed of a variety of different physiological groups of bacteria, archaea, and microeucaryotes, including fungi, algae, and protozoa. Recent molecular biological studies have shown that the assemblage of organisms present in an environmental sample can contain thousands of distinct biotypes of bacteria (39, 40) and that many of these organisms represent undescribed microbial lineages, some of which lack even a single representative species available in pure culture (4). This complexity is further compounded by the high spatial and temporal variability of species represented in many natural communities (1, 7) and by the still-open question of what proportion of the microscopically observable assemblage of organisms is viable or even metabolically functional (10, 25, 27, 42). The characterization of natural microbial communities is clearly a difficult problem. Many molecular biological methods for community analysis are either very time consuming, such as the construction and screening of clonal libraries (4, 13, 19), or yield an index of community composition that does not reveal organism taxonomic affiliations without additional analysis, such as denaturing gradient gel electrophoresis (14, 30, 31, 36). There is a clear need for methods that can be used to rapidly and accurately examine microbial community composition. Ideally, such methods would include all (or at least most) microbial domains and produce a standardizable community profile, allowing comparisons across different sample types and facilitating the development of broad hypotheses concerning microbial community dynamics.

One approach to microbial community characterization that has been successful in a wide variety of applications (e.g., those described in references 5, 6, 11, 15–17, 20, 32, 34, 38) is phospholipid fatty acid (PLFA) analysis (15). Phospholipids are structural components of all biological membranes. These compounds have no storage function and thus represent a consistent fraction of cell mass. They also degrade quickly upon an organism's death, and current extraction and derivatization methods permit recovery of PLFAs exclusively from living organisms (41). Extraction and subsequent analysis by gas chromatography and mass spectroscopy provide precise resolution, sensitive detection, and accurate quantification of a broad array of PLFAs. Analyses can be performed rapidly and efficiently, and each yields a profile composed of numerous PLFAs defined on the basis of compound structure and the quantity of each compound present in the sample. Different PLFA classes and specific molecules are produced by bacteria and eucarya, allowing both domains to be examined through the same analysis. The PLFA profile obtained thus constitutes a "fingerprint" of the living bacterial and microeucaryotic microbial community and reflects its species composition. Comparisons across sample types are simplified by normalizing the quantity of each PLFA to the total recovered, expressing the result in terms of moles percent (mol%) PLFA. One drawback of the PLFA analysis is that archaea are not represented in the analysis since their membrane lipids employ ether rather than ester bonds. Nonetheless, PLFA profiles can be used to characterize microbial community compositions and provide useful

* Corresponding author. Mailing address: Belle W. Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC 29208. Phone: (803) 777-3928. Fax: (803) 777-3935. E-mail: noble@biol.sc.edu.

† This is contribution number 1192 of the Belle W. Baruch Institute for Marine Biology and Coastal Research.

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries

information on the dynamics of viable bacteria and microeucaryotes.

Comparing PLFA profiles of microbial communities is often complicated by variability in the types and quantities of compounds present in replicate samples from a given sample site. This, together with the complexity of the PLFA profiles, constitute substantial challenges to our ability to quantitatively compare microbial community compositions by this approach. Previous studies have employed cluster analysis and principal components analysis (PCA) to group-like samples (37) and to determine which PLFAs are most diagnostic for the groups defined (11, 20, 23, 36, 37). There are two problems with these approaches: (i) they do not account for the nonlinear nature of biological data which is due to inherent variability, and (ii) they do not permit details of the profiles to be conveniently examined. An approach that provides the advantages of cluster analysis and PCA while accounting for the nonlinear nature of the biological data and facilitating a more detailed profile examination would be very advantageous.

Artificial neural networks (NNs) provide a useful tool for recognizing patterns in complex, nonlinear data sets such as those associated with PLFA profiles. NNs are particularly advantageous over conventional statistical methods because they can deal with the inherent variability associated with biological data. NNs are constructed by using computer software and consist of layers of neurons that make independent computations and pass on their outputs to other neurons (33). Each neuron in a layer is connected to neurons in the next layer so that the output of each neuron affects the activation of all neurons to which it is connected. Neurons are adaptable, and through the process of learning from examples, store knowledge and make it available for use (2). In a training technique called back propagation (35), a pattern is presented to an input layer of a network and the network produces output based on the sum of the weighted inputs. When the pattern of the output layer is compared to target values, the errors between them are computed. An error function is used to readjust the weights of each neuron. The adjusted weights of a trained network can be used to recognize patterns such as those in PLFA profiles of microbial communities and to provide information on functional relations between components of a profile and target values. For example, sensitivity analysis of the adjusted weights can be used to determine the relative contribution of individual input neurons (i.e., PLFA molecules) to target values (9).

The focus of this study was to investigate the utility of NN for interpreting complex data, such as PLFA profiles of microbial communities in surface and subsurface nearshore sediments and in the linings of burrows of three species of marine polychaetes and hemichordates. We compared results obtained by this approach with those obtained by conventional PCA and linear discriminant analysis (LDA). Sensitivity analysis of the trained NNs identified which input PLFAs significantly contributed to training the NNs.

MATERIALS AND METHODS

Source of PLFA profiles analyzed. PLFA data used in our analyses were from a previous study of marine sediment microbiota by Steward et al. (37). This study should be consulted for sample site descriptions and significant features of the sediments sampled. Briefly, sediment samples collected for the study were 10 matched sets for each of three worm species, with each set including one worm burrow sediment sample, one surface sediment sample, and one subsurface sediment sample. Surface sediments were collected from approximately the top centimeter of sediment near but not adjacent to the worm burrows. Subsurface sediment samples were taken approximately 10 to 20 cm deep, near but not adjacent to worm burrows. Sediments were scraped from the linings of burrows built by the capitellid polychaete *Notomastus lobatus* and the hemichordate *Balanoglossus aurantiacus* by using sterile spatulas. Segments of tubes constructed by the maldanid polychaete *Branchyoasychus americana* were recovered intact. Large macrofauna were excluded from all samples. The samples were placed in sterile Whirlpack bags and immediately frozen on dry ice and then transferred to −80°C pending PLFA extraction and analysis as described by Steward et al. (37). This sample set included 90 total samples, each yielding a profile of 60 different fatty acids. Forty-two PLFAs accounted for 93 to 97% of the total number of PLFAs in each sample (37). The balance was made up of low-abundance compounds. Characterization of the sample set by PCA showed that monounsaturated fatty acids, branched saturated fatty acids, and PLFA 10me16:0 are important components of the sediment microbial communities (37).

Data for the NN and LDA. The complete input data set consisted of 89 PLFA profiles (one sample was lost during processing) and 23 randomly generated PLFA profiles which served as controls. Each PLFA profile consisted of 60 PLFAs. For NN and sensitivity analysis, the entire data set was normalized, with a minimum value of the data set being 0.1 and a maximum value of 0.9 (26). For LDA, the data were normalized by an arcsine transformation.

Output data for sediment type and worm species were coded as four-digit binary numbers. The first two digits refer to the worm species and the next two digits refer to sediment type. The worm species were coded 00, 10, 01, and 11, corresponding to the profiles of marine worms *B. aurantiacus*, *B. americana*, and *N. lobatus* and to randomly generated profiles, respectively. The sediment types were coded 01, 00, 10, and 11, corresponding to surface, burrow, and subsurface sediment profiles and randomly generated profiles, respectively.

Cross-validation scheme. To evaluate the predictive power of the NN and linear discriminant analyses, we employed the following cross-validation scheme: the order of the data was randomized, 90% of the data were used to train the NN and LDA, and the remaining 10% were used to test NN and LDA performance. Correct and incorrect classifications were recorded. The above scheme was repeated 10 times for each cross-validation.

Sensitivity analysis. The relative importance of each PLFA to predict the target values was calculated by performing sensitivity analysis on the trained NN (26). In this study, the sensitivity of an output parameter, $\mathrm{Out}_{j=1,2,\ldots,nj}$ (there are *nj* output parameters), to an input parameter, $\mathrm{In}_{i=1,2,\ldots,ni}$ (there are *ni* input parameters), was defined as the normalized ratio between variations caused in Out_j by variations introduced in In_j and is represented by the following equations:

$$\mathrm{NS}_{i,jc} = (\mathrm{dOut}_{j,c}/\mathrm{dIn}_{i,c})(\mathrm{In}_{i,c}/\mathrm{Out}_{j,c})$$

$$\mathrm{S}_i = [\Sigma_{j=1,2,\ldots,nj;\, c=1,2,\ldots,nc}(\mathrm{NS}_{i,jc})]/[\Sigma_{i=1,2,\ldots,ni;\, j=1,2,\ldots,nj;\, c=1,2,\ldots,nc}(\mathrm{NS}_{1,jc})] \quad (1)$$

where $i = 1,2,\ldots,ni$ (input index), $j = 1,2,\ldots,nj$ (output index), and $c = 1,2,\ldots,nc$; (sample [case] index). The normalized sensitivity for PLFA profile *c*, $\mathrm{NS}_{i,jc}$ was calculated for every combination of input, *i*, and output parameters, *j*, and for every PLFA profile (there are *nc* profiles). The overall sensitivity to an input, S_i, was determined by taking the average over all 112 PLFA profiles and all four binary outputs used to classify them, which corresponds to a total of 26,880 ($112 \times 4 \times 60$) partial sensitivity values. Finally, the sensitivity values obtained are represented as relative values, calculated as a percent value of the sum of all sensitivities (equation 1, S_i).

Phospholipid fatty acid nomenclature. Fatty acids are designated A:BωC, where A is the total number of carbon atoms, B is the number of double bonds, and C is the position of the double bond from the aliphatic end of the molecule. The geometry of this bond is indicated with a "c" for *cis* and a "t" for *trans*. The prefixes "i" and "a" refer to iso and anteiso terminal methyl-branching, respectively (24). Mid-chain methyl branches are designated by "me" preceded by the position of the methyl group from the acid end of the molecule. Cyclopropyl fatty acids are designated by "cy."

Statistical software. The NN and sensitivity analysis software were implemented in BrainCel version 3.0 (Promised Land Tech.) and Matlab version 5.2 equipped with the NN tool box (MathWorks Inc.). For NN analysis, the logsig equation, $\mathrm{logsig}(x) = 1/(1 + e^{-x})$, was used as a transfer function and error back propagation was used to optimize the connection weights. Full interconnection between the layers was used. Learning and momentum rates are self-adjusting in Matlab and BrainCel. The input and output architecture of the NN consisted of 60 input neurons representing the 60 PLFAs and four-output neurons coding for both sediment types and worm species. In order to minimize the effective number of degrees of freedom in the network, training was stopped when the error measured with independent test data started to increase (9). This criterion was also used to select the optimal number of hidden neurons (9). LDA was conducted by the procedure DISCRIM (prior probabilities equal, covariance matrix pooled), chi-square was conducted by the procedure FREQ, and Kendall's coefficient of concordance was conducted by the procedure CORR KENDALL in the SAS program (release 6.11; SAS Inc., Cary, N.C.). Cross-validation results and the Student *t* tests were tabulated by MS Excel 98 (Microsoft Inc., Redmond, Wash.) on Macintosh 8.6 or MS Windows 98 operating systems.

RESULTS AND DISCUSSION

Optimization of the NN. The effects of architecture on the predictive abilities of NNs were determined by changing the

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries

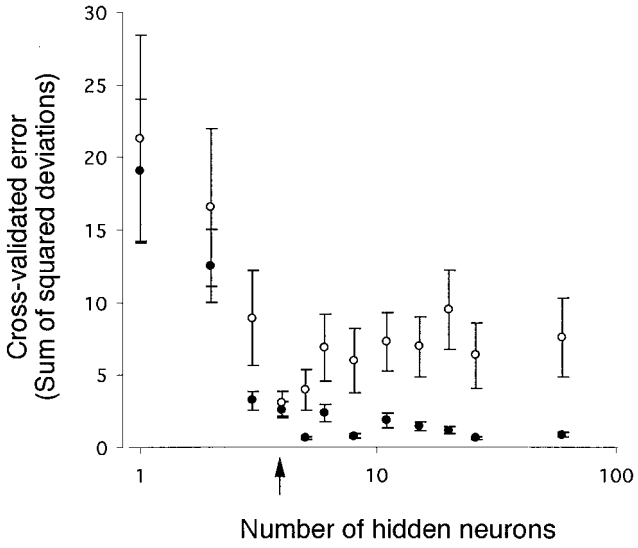

FIG. 1. Effect of the number of hidden neurons on the cross-validation error. The optimum value of neurons was found to be 4 (arrow), where the training (closed circle) and testing (open circle) errors are the lowest and still similar. Each circle is the mean of 10 separately trained NNs, and each error bar represents the standard deviation of the mean.

number of hidden layer neurons in NNs and calculating the cross-validation error (sum of squared deviations) of predicted and actual outputs of training and test data (Fig. 1). The same cross-validation scheme was used as outlined in Materials and Methods. Optimization of the NN was necessary to prevent the possibility of "overfitting" the data (9). Overfitting the data decreases the performance of an NN, making it unable to generalize predictions. The optimal number of hidden neurons for the NN was found to be four (Fig. 1). Hence, the optimal architecture for the NN with this data set was 60 input neurons, four hidden neurons, and four output neurons.

Cross-validation of microbial community structure analysis. The LDA cross-validation results (Table 1) were significantly different from those obtained by using NNs (Table 2; $p < 0.0001$; the Student t test). LDA predictions had a combined error of 8.5% (19 incorrect out of 224; Table 1) while those of the NN had a combined error of 2.7% (6 incorrect out of 224; Table 2), indicating that the NN was significantly more accurate than the LDA. Table 1 shows that PLFAs were more correctly associated with sediment types (93.8%) than with worm species (89.3%), indicating that microbial community compositions were more determinable by whether they were from the surface, burrow, or subsurface than if they were from a particular worm species (Table 1). Similar differences between the associations of PLFAs and sediment types or worms species were not apparent when cross-validation was conducted with NNs (Table 2).

Sensitivity analysis. The analysis of the sensitivity of individual PLFAs to NN outputs was repeated with 10 separately trained NNs to validate the rank order of the 60 PLFAs with respect to their individual sensitivities. Chi-square tests of the ranked order of the 60 PLFAs revealed that they are very similar ($p = 0.98$). Kendall's concordance coefficient, which provides the simultaneous association (relatedness) between samples, was 0.99, indicating that the rank order was very consistent.

Approximately 50% of the variation in the data set was explained by 17 fatty acids (Fig. 2), indicating that certain PLFAs, such as 16:1ω7c and 18:1ω9c (relative sensitivities of 5.3 and 3.8%, respectively), play more important roles in determining the outputs of the NN than other PLFAs, such as 20:4ω3 (sensitivity of 0.5%). Examination of these findings with reference to those of Steward et al. (37) provide clues as to why certain PLFAs explained more variability than others. For example, PLFA 16:1ω7c occurs at much higher concentrations in microbial communities found in burrows and surface sediments (14.3 ± 3.5 mol% [mean ± standard deviation] [n = 60]) than those found in subsurface sediments (7.6 ± 2.6 mol% [n = 29]). Hence, significant differences in the concentrations of PLFAs between the sediment types as well as the low variability of the PLFA in each type are likely key factors used by the NN to recognize patterns. Also, sediments containing high concentrations of PLFA 18:1ω9c are more often associated with microbial communities of *B. aurantiacus* (3.2 ± 1.1 mol% [n = 30]) or *N. lobatus* (3.3 ± 0.6 mol% [n = 30]) than *B. americana* (2.3 ± 0.5 mol% [n = 29]). Presumably, the concentration of 18:1ω9c was used by the NN to distinguish microbial communities of *B. americana* from those of the other species. Conversely, PLFAs with low sensitivities, such as 20:4ω3 (sensitivity of 0.5%), did not significantly contribute to the training of the NN because the concentrations between sedi-

TABLE 1. Cross-validation of microbial communities by LDA of arcsine-transformed PLFA data[a]

Actual sample type	Predicted no. of samples by:								Total no. of samples
	Sediment type				Worm species				
	Burrow	Subsurface	Surface	Random	*B. aurantiacus*	*B. americana*	*N. lobatus*	Random	
Sediment type									
Burrow	**23**	4		1					28
Subsurface	1	**24**							25
Surface	1		**36**						37
Random				**22**					22
Worm species									
B. aurantiacus					**30**		1		31
B. americana					3	**23**	1		27
N. lobatus					2	5	**25**		32
Random								**22**	22

[a] Approximately 90% of the data were used for LDA, and the remaining 10% of the data were used for cross-validation. The cross-validation was repeated 10 times; each time, the order of the records used for LDA and cross-validation was randomized. Numbers in bold are correct predictions; numbers not in bold are incorrect predictions. The total number of correct predictions by sediment type was 105 (93.8%); the total number of incorrect predictions was 7 (6.3%). The total number of correct predictions for worm species was 100 (89.3%); the total number of incorrect predictions was 12 (10.7%).

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries

TABLE 2. Cross-validation of microbial communities by a back-propagating NN of PLFA data[a]

Actual sample type	Predicted no. of samples by:								Total no. of samples
	Sediment type				Worm species				
	Burrow	Subsurface	Surface	Random	*B. aurantiacus*	*B. americana*	*N. lobatus*	Random	
Sediment type									
Burrow	**30**	1	1						32
Subsurface		**28**							28
Surface			**29**	1					30
Random				**22**					22
Worm species									
B. aurantiacus					**29**	1	1		31
B. americana						**28**			28
N. lobatus					1		**30**		31
Random								**22**	22

[a] Approximately 90% of the data were used to train the data and the remaining 10% were used for cross-validation. The cross-validation was repeated 10 times; each time, the order of the records was randomized. Numbers in bold are correct predictions; numbers not in bold are incorrect predictions. The total number of correct predictions for sediment and worm type was 109 (97.3%); the total number of incorrect predictions was 3 (2.7%).

ment types (burrow, 0.3 ± 0.4 mol% [n = 30]; subsurface, 0.3 ± 0.3 mol% [n = 29]; surface, 0.3 ± 0.1 mol% [n = 30]) and worm species (*B. aurantiacus*, 0.4 ± 0.3 mol% [n = 30]; *B. americana*, 0.2 ± 0.3 mol% [n = 29]; *N. lobatus*, 0.3 ± 0.1 mol% [n = 30]) were not significantly different and/or too variable for the NN to recognize patterns in the PLFAs.

These findings are also consistent with the more physiological and organismal interpretations of Steward et al. (37). Both 16:1ω7c and 18:1ω9c are products of aerobic bacteria and microeucaryotes (15) and would be expected to be more abundant in the surface sediments and in the oxidized linings of the burrows of *N. lobatus* and *B. aurantiacus* (37). Reduced subsurface sediments, as well as the more reduced tube sediments of *B. americana* (37), would not be expected to harbor high levels of aerobic organisms. In fact, many PLFAs contributing strongly to the sensitivity of the NN analyses (Table 3) have well-defined origins in the aerobic and anaerobic microbiota (15, 37).

Distinguishing which of the input PLFAs are used to determine the outputs of a trained NN by sensitivity analysis enabled us to identify the major PLFAs associated with microbial communities of worm burrows and sediment types. Steward et al. (37) identified which PLFAs accounted for a significant portion of the variance by using PCA. A significant portion of the variability in both sensitivity analysis and PCA was attributed to PLFAs 16:0, 16:1ω7c, and 18:1ω9c (Table 3), indicating that trained NN yielded some correspondence to the results obtained by PCA (37). However, 18 of the 24 PLFAs were not common to both analysis methods, indicating significant disagreements between the methods. Since NNs are able to use

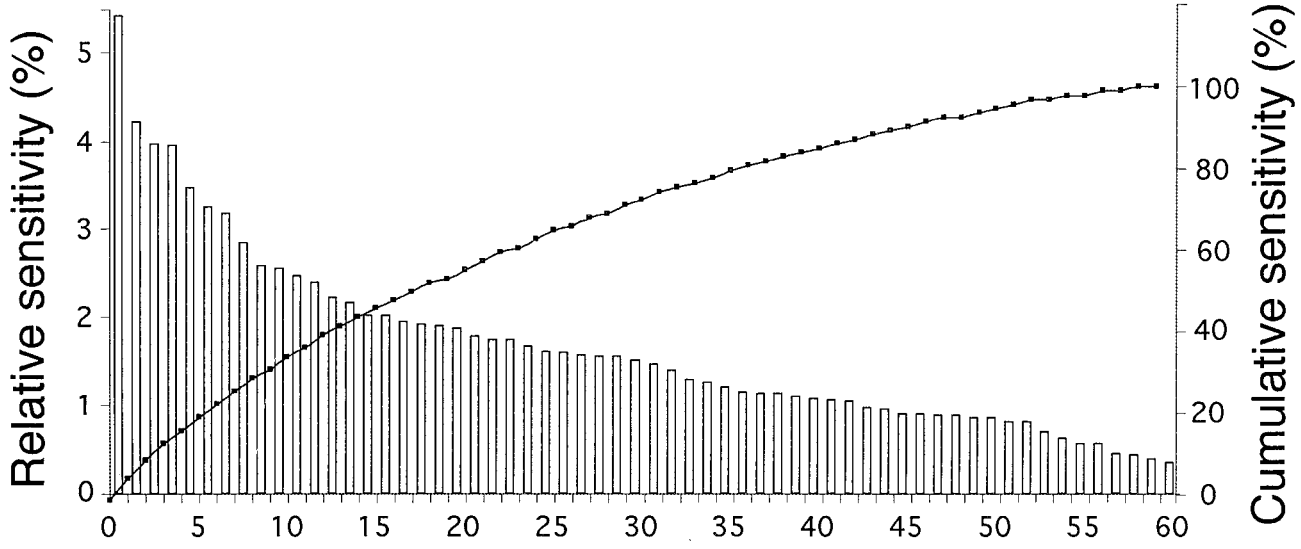

FIG. 2. Relative (bars) and cumulative (solid squares) sensitivities of an optimized NN to specific PLFAs. The graph represents 1 of the 10 sensitivity analyses conducted. PLFAs are arranged in order of their importance to the prediction of target values. These PLFAs are listed in decreasing order of importance: (PLFAs 1 to 5) 16:1ω7c, 16:0, 18:1ω7c, 18:1ω9c, 18:0, (PLFAs 6 to 10) 16:1ω7t, cy17:0, 17:0, cy19:0(ω7,8), 20:0, (PLFAs 11 to 15) 22:4ω6, 10me16:0, 20:3ω3, poly19, i16:0, (PLFAs 16 to 20) 17:1ω8c/a17:0, 22:5ω6, 16:1ω5c, 17:1ω6c, 10me18:0, (PLFAs 21 to 25) 16:1ω13t, 18:3ω3/Br17:1/i18:0, 22:6ω3, 18:1ω7t, i15:0, (PLFAs 26 to 30) 18:3ω6/10me17:0, 15:1ω6c, Poly20, α15:0, 20:4ω6, (PLFAs 31 to 35) i17:0, 22:5ω3, br19:1, a16:0/16:1ω9c, 10me14:0, (PLFAs 36 to 40) br17:1ω7c, 16:2ω6/br15:0, 20:5ω3, 24:0, Poly17, (PLFAs 41 to 45) 14:0, 22:0, 11me18:0, 20:2ω6, 20:3ω6, (PLFAs 46 to 50) 19:1ω12c, Poly22, 20:1ω7c, 19:1ω8c, 18:2ω6, (PLFAs 51 to 55) br17:0, 15:0, i14:0, 19:1ω6c, 18:4v3/12me17:0, (PLFAs 56 to 60) 12me16:0, 20:1ω9c, 20:4ω3, mono F.A., br17:0.

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries

TABLE 3. Major PLFAs contributing to the variability of PCA and sensitivity analysis[a]

PLFA type	Significant PLFA	
	PCA	Sensitivity analysis
Saturated	16:0	16:0
		17:0
		18:0
		20:0
Branched, saturated	i15:0	
	a15:0	
	i16:0	
Monounsaturated		16:1ω7t
	16:1ω7c	16:1ω7c
	18:1ω7c	18:1ω7c
		18:1ω9c
Polyunsaturated	18:3ω3	
	20:4ω6	
	20:5ω3	
		22:4ω6
	22:6ω3	
Methyl-branched, saturated	10me16:0	10me16:0
Cyclopropyl		cy17:0
	cy19:0	
		cy19:0(ω7,8)

[a] PCA results were derived from those of Steward et al. (37). Shown are the 12 most significant PLFAs.

nonlinear associations among PLFAs to relate to target values, such disagreements are to be expected. Another factor which might contribute to these differences is that the PCA conducted by Steward et al. (37) was based on 89 PLFA profiles while the NN and sensitivity analysis were conducted on the same 89 profiles plus an additional 23 randomly generated PLFA profiles which served as controls for NN training.

The application of neural computing approaches to analyze complex data is relatively new in microbiology. For example, NNs have been used to identify the restriction enzyme patterns of *Escherichia coli* O157:H7 (12), the pyrolysis mass spectra of *Mycobacterium tuberculosis* complex species (17), the promoter sites of *E. coli* (22), protein-DNA binding sites (29), fatty acids of marine heterotrophic bacteria (8) and mycobacteria from sputum samples (3), stable low-molecular-weight rRNA profiles (28), and *nifH*-specific binding patterns from denaturing gradient gel electrophoresis analysis (31). The advantage of using neural computing approaches is that they allow pattern recognition that cannot be discerned by conventional statistical methods (26). Sensitivity analysis of trained NNs yields information on which individual input neurons (in this study, PLFAs) significantly contributed to the targeted outputs.

In summary, an optimized neural network was used to associate PLFA data of microbial communities to worm species and sediment type. Cross-validation of the trained NNs and LDA with independent test data revealed that NNs provided better predictions than LDA. The relevant PLFAs for NN prediction were identified by sensitivity analysis.

ACKNOWLEDGMENTS

We thank M. Caroline Roper and Wes R. Johnson for proofreading the manuscript.

This research was supported by the DOD/ONR grant no. N000149610403 to C.R.L.; NSF grant no. MCB-9802342, DOD/ONR grant no. F188-N00014-97-1-0806, and EPA grant no. R826944-01-0 to P.A.N.; and FMRH/BSAB/60/98 Fundação para a Ciência e Tecnologia/MCT to J.S.A. Thanks to two anonymous reviewers for their insights and suggestions.

REFERENCES

1. **Acinas, S. G., F. Rodriguez-Valera, and C. Pedros-Alio.** 1997. Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16s rDNA. FEMS Microbiol. Ecol. **24:** 27–40.
2. **Aleksander, I., and H. Morton.** 1991. Principles and promises, p. 1–20. *In* An introduction to neural computing. Chapman & Hall, Ltd., London, England.
3. **Almeida, J. S., A. Sonesson, D. B. Ringelberg, and D. C. White.** 1995. Application of artificial neural networks (ANN) to the detection of *Mycobacterium tuberculosis*, its antibiotic resistance and prediction of pathogenicity amongst *Mycobacterium* spp. based on signature lipid biomarkers. Bin. Comput. Microbiol. **7:**53–59.
4. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and *in-situ* detection of individual microbial-cells without cultivation. Microbiol. Rev. **59:**143–169.
5. **Baird, B. H., and D. C. White.** 1985. Biomass and community structure of the abyssal microbiota determined from the ester-linked phospholipids recovered from Venezuela Basin and Puerto Rico Trench sediment. Mar. Geol. **68:**217–231.
6. **Balkwill, D. L., E. M. Murphy, D. M. Fair, D. B. Ringelberg, and D. C. White.** 1998. Microbial communities in high and low recharge environments: implications for microbial transport in the vadose zone. Microb. Ecol. **35:**156–171.
7. **Berardesco, G., S. Dyhrman, E. Gallagher, and M. P. Shiaris.** 1998. Spatial and temporal variation of phenanthrene-degrading bacteria in intertidal sediments. Appl. Environ. Microbiol. **64:**2560–2565.
8. **Bertone, S., M. Giacomini, C. Ruggiero, C. Piccarolo, and L. Calegari.** 1996. Automated systems for identification of heterotrophic marine bacteria on the basis of their fatty acid composition. Appl. Environ. Microbiol. **62:** 2122–2132.
9. **Bishop, C. M.** 1995. Neural networks for pattern recognition. Clarendon Press, Oxford, England.
10. **Bogosian, G., P. J. L. Morris, and J. P. O'Neil.** 1998. A mixed culture recovery method indicates that enteric bacteria do not enter the viable but nonculturable state. Appl. Environ. Microbiol. **64:**1736–1742.
11. **Bossio, D. A., K. M. Scow, N. Gunapala, and K. J. Graham.** 1998. Determinants of soil microbial communities: Effects of agricultural management, season, and soil type on phospholipid fatty acid profiles. Microb. Ecol. **36:** 1–12.
12. **Carson, C. A., J. M. Keller, K. K. McAdoo, D. Wang, B. Higgins, C. W. Bailey, J. G. Thorne, B. J. Payne, M. Skala, and A. W. Hahn.** 1995. *Escherichia coli* O157:H7 restriction pattern recognition by artificial neural networks. J. Clin. Microbiol. **33:**2894–2898.
13. **Chandler, D. P., F. J. Brockman, and J. K. Fredrickson.** 1997. Use of 16s rDNA clone libraries to study changes in a microbial community resulting from *ex situ* perturbation of a subsurface sediment. FEMS Microbiol. Rev. **20:**217–230.
14. **Ferris, M. J., and D. M. Ward.** 1997. Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. **63:**1375–1381.
15. **Findlay, R. H., and F. C. Dobbs.** 1993. Quantitative description of microbial communities using lipid analysis, p. 271–284. *In* P. R. Kemp, B. F. Sherr, E. B. Sherr, and J. J. Cole (ed.), Handbook of methods in aquatic microbial ecology. Lewis, Boca Raton, Fla.
16. **Findlay, R. H., and L. Watling.** 1998. Seasonal variation in the structure of a marine benthic microbial community. Microb. Ecol. **36:**23–30.
17. **Findlay, R. H., B. Trexler, J. B. Guckert, and D. C. White.** 1990. Laboratory study of disturbance in marine sediments: response of a microbial community. Mar. Ecol. Prog. Ser. **62:**121–133.
18. **Freeman, R., R. Goodacre, P. R. Sisson, J. G. Magee, A. C. Ward, and N. F. Lightfoot.** 1994. Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra. J. Med. Microbiol. **40:**170–173.
19. **Gray, J. P., and R. P. Herwig.** 1996. Phylogenetic analysis of the bacterial communities in marine sediments. Appl. Environ. Microbiol. **62:**4049–4059.
20. **Guckert, J. B., C. P. Antworth, P. D. Nichols, and D. C. White.** 1985. Phospholipid, ester-linked fatty acid profiles as reproducible assays for changes in prokaryotic community structure of estuarine sediments. FEMS Microbiol. Ecol. **31:**147–158.
21. **Guckert, J. B., D. B. Ringelberg, D. C. White, R. S. Hanson, and B. J. Bratina.** 1991. Membrane fatty-acids as phenotypic markers in the polyphasic taxonomy of methylotrophs within the proteobacteria. J. Gen. Microbiol. **137:**2631–2641.
22. **Horton, P. B., and M. Kanehisa.** 1992. An assessment of neural networks and statistical approaches for prediction of *E. coli* promoter sites. Nucleic Acids Res. **20:**4331–4338.
23. **Ibekwe, A. M., and A. C. Kennedy.** 1998. Phospholipid fatty acid profiles and

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries

carbon utilization patterns for analysis of microbial community structure under field and greenhouse conditions. FEMS Microbiol. Ecol. **26:**151–163.
24. **Kates, M.** 1986. Techniques in lipidology: isolation, analysis and identification of lipids, 2nd ed. Elsevier, Amsterdam, The Netherlands.
25. **Kell, D. B., A. S. Kaprelyants, D. H. Weichart, C. R. Harwood, and M. R. Barer.** 1998. Viability and activity in readily culturable bacteria: a review and discussion of the practical issues. Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol. **73:**169–187.
26. **Masters, T.** 1993. Practical neural network recipes in C++. Academic Press, New York, N.Y.
27. **McDougald, D., S. A. Rice, D. Weichart, and S. Kjelleberg.** 1998. Nonculturability: adaptation or debilitation? FEMS Microbiol. Ecol. **25:**1–9.
28. **Noble, P. A., K. D. Bidle, and M. Fletcher.** 1997. Natural microbial community compositions compared by a back-propagating neural network and cluster analysis of 5S rRNA. Appl. Environ. Microbiol. **63:**1762–1770.
29. **O'Neill, M. C.** 1998. A general procedure for locating and analyzing protein-binding sequence motifs in nucleic acids. Proc. Natl. Acad. Sci. USA **95:** 10710–10715.
30. **Ovreas, L., L. Forney, F. L. Daae, and V. Torsvik.** 1997. Distribution of bacterioplankton in meromictic lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16s rRNA. Appl. Environ. Microbiol. **63:**3367–3373.
31. **Piceno, Y. M., P. A. Noble, and C. R. Lovell.** 1999. Spatial and temporal assessment of diazotroph assemblage composition in vegetated salt marsh sediments using denaturing gradient gel electrophoresis analysis. Microb. Ecol. **38:**157–167.
32. **Rajendran, N., O. Matsuda, N. Imamura, and Y. Urushigawa.** 1992. Variation in microbial biomass and community structure in sediments of eutrophic bays as determined by phospholipid ester-linked fatty acids. Appl. Environ. Microbiol. **58:**562–571.
33. **Rao, V. B., and H. V. Rao.** 1993. C++ neural networks and fuzzy logic. MIS:Press, New York, N.Y.
34. **Ringelberg, D. B., S. Sutton, and D. C. White.** 1997. Biomass, bioactivity and biodiversity: microbial ecology of the deep subsurface: analysis of ester-linked phospholipid fatty acids. FEMS Microbiol. Rev. **20:**371–377.
35. **Rumelhart, D. E., G. E. Hinton, and R. J. Williams.** 1986. Learning internal representation by error back propagation, p. 318–362. *In* D. E. Rumelhart and J. L. McCleland (ed.), Parallel distributed processing. M.I.T. Press, Cambridge, Mass.
36. **Stephen, J. R., G. A. Kowalchuk, M. A. V. Bruns, A. E. McCaig, C. J. Phillips, T. M. Embley, and J. I. Prosser.** 1998. Analysis of beta-subgroup proteobacterial ammonia oxidizer populations in soil by denaturing gradient gel electrophoresis analysis and hierarchical phylogenetic probing. Appl. Environ. Microbiol. **64:**2958–2965.
37. **Steward, C. C., S. C. Nold, D. B. Ringelberg, D. C. White, and C. R. Lovell.** 1996. Microbial biomass and community structures in the burrows of bromophenol producing and non-producing marine worms and surrounding sediments. Mar. Ecol. Prog. Ser. **133:**149–165.
38. **Sundh, I., M. Nilsson, and P. Borga.** 1997. Variation in microbial community structure in two boreal peatlands as determined by analysis of phospholipid fatty acid profiles. Appl. Environ. Microbiol. **63:**1476–1482.
39. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. **56:**782–787.
40. **Torsvik, V., K. Salte, R. Sorheim, and J. Goksoyr.** 1990. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl. Environ. Microbiol. **56:**776–781.
41. **White, D. C.** 1988. Validation of quantitative analysis for microbial biomass, community structure, and metabolic activity. Arch. Hydrobiol. Beih. Ergebn. Limnol. **31:**1–18.
42. **Zweifel, U. L., and Å. Hagström.** 1995. Total counts of marine bacteria include a large fraction of non-nucleoid-containing bacteria (ghosts). Appl. Environ. Microbiol. **64:**2180–2185.

Downloaded from http://aem.asm.org/ on December 28, 2016 by University of North Texas Libraries