APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 1997, p. 1762–1770 Vol. 63, No. 5
0099-2240/97/$04.00+0

Natural Microbial Community Compositions Compared by a Back-Propagating Neural Network and Cluster Analysis of 5S rRNA†

PETER A. NOBLE,‡ KAY D. BIDLE,§ AND MADILYN FLETCHER*

Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, Maryland 21202

Received 18 December 1996/Accepted 7 March 1997

The community compositions of free-living and particle-associated bacteria in the Chesapeake Bay estuary were analyzed by comparing banding patterns of stable low-molecular-weight RNA (SLMW RNA) which include 5S rRNA and tRNA molecules. By analyzing images of autoradiographs of SLMW RNAs on polyacrylamide gels, band intensities of 5S rRNA were converted to binary format for transmission to a back-propagating neural network (NN). The NN was trained to relate binary input to sample stations, collection times, positions in the water column, and sample types (e.g., particle-associated versus free-living communities). Dendrograms produced by using Euclidean distance and average and Ward's linkage methods on data of three independently trained NNs yielded the following results. (i) Community compositions of Chesapeake Bay water samples varied both seasonally and spatially. (ii) Although there was no difference in the compositions of free-living and particle-associated bacteria in the summer, these community types differed significantly in the winter. (iii) In the summer, most bay samples had a common 121-nucleotide 5S rRNA molecule. Although this band occurred in the top water of midbay samples, it did not occur in particle-associated communities of bottom-water samples. (iv) Regardless of the season, midbay samples had the greatest variety of 5S rRNA sizes. The utility of NNs for interpreting complex banding patterns in electrophoresis gels was demonstrated.

Most aquatic environments contain bacteria freely suspended in the water column and bacteria associated with particulate material. These two community types provide different metabolic contributions to the total bacterial community depending on environmental conditions. In oligotrophic ocean waters, free-living bacteria can dominate in numbers and biomass and account for most of the secondary production (13). However, in environments that are characterized by substantial particle loads, attached bacteria make more significant contributions in cell numbers and total productivity (13, 20). In addition, they can exhibit morphological and metabolic characteristics that are distinct from those of freely suspended bacteria (7, 15, 29).

The Chesapeake Bay is a eutrophic estuary characterized by significant amounts of organic and inorganic particulate material. Typically, particle loads are greater in the upper bay due to the large influx of fresh water. The bay exhibits seasonal circulation patterns, with a highly stratified water column in the summer followed by vertical mixing in the cooler winter months. It contains extensive, abundant, and active bacterial communities, which degrade phytodetritus and contribute to the formation of anoxic zones in bottom waters of the midbay region during the summer. Tritiated thymidine uptake experiments have shown that attached bacteria make a relatively minor contribution to total bacterial secondary production in the Chesapeake Bay (12). However, attached bacteria were found to be more active on a per-cell basis in summer, and they made significant contributions to the hydrolysis of a model dipeptide, MCA-leucine (14).

To determine the community compositions of attached and free-living bacteria, Bidle and Fletcher (6) used filtration to separate particle-associated from free-living bacteria. Compositions of these communities were compared by using stable low-molecular-weight RNA (SLMW RNA) analysis, which provides a first-order analysis of communities and demonstrates significant differences in community composition. In this approach, extracted SLMW RNA was labeled with ^{32}P, separated on a high-resolution polyacrylamide gel by electrophoresis, and exposed to X-ray film. By examining the banding patterns in the 107-to-131-nucleotide (nt) range, they established that the community compositions of free-living and particle-associated bacteria were discernibly different for samples collected in the Chesapeake Bay in the winter. Whether or not these communities were truly representative of those of the bay, or whether these communities change with seasonal circulation of the bay waters, was not determined.

In comparing the genetic composition of bacterial communities by SLMW RNA analysis, it is difficult to interpret the complex banding patterns of these molecules. Decoding the patterns is hindered by inherent errors of the electrophoresis method, such as those due to gel warping, under- and overloading of the gel lanes, and under- and overexposure of the X-ray film. Recently, molecular biology software has enabled researchers to convert autoradiographs, photographs, and/or gels to digitized images so that the images can be modified and enhanced. Moreover, software can convert images to numerical data so that banding patterns in gel lanes can be statistically analyzed. This study uses numerical data and cluster analysis methods to compare SLMW RNA molecules extracted from water samples of the Chesapeake Bay. We assumed that samples having a low degree of similarity in the 5S RNA region of

* Corresponding author. Mailing address: Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC 29208. E-mail: fletcher@biol.sc.edu.

† Contribution No. 271 from the Center of Marine Biotechnology.

‡ Present address: Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, SC 29208.

§ Present address: Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0208.

the gel have dissimilar community compositions. The validity of the cluster analysis was assessed by determining the robustness of group memberships by different linkage methods. However, band intensities could not be used for cluster analysis, because group memberships of replicate samples and molecular-weight standards were inconsistent (data not shown). Therefore, a neural computing approach was used to produce numerical data for cluster analysis.

Neural computing deals with the study of artificial neural networks (NNs), which through the process of learning, storing, and supplying information model natural neural systems (1, 3). Artificial NNs are constructed by using computer software and consist of layers of neurons which make independent computations and pass on their outputs to other neurons (24). 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 (1). In a training technique called error back propagation (27), a pattern is presented to an input layer of a network, and the network produces output based on the sum of the weighted inputs (33). 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 (3). This iterative process continues until the errors fall below a designated tolerance level (24). The adjusted weights form a trained network which can be used to recognize patterns.

The focus of this study was to compare the community compositions of free-living and particle-associated bacteria collected from the Chesapeake Bay by using SLMW RNA analysis. Community comparisons were made at different depths of the water column, as well as at stations separated by considerable distances. To determine whether or not the community compositions of Chesapeake Bay water samples were affected by season, we compared samples taken in the summer to those taken in the winter (6). Here, we report on a novel approach for comparing SLMW RNA bands of bacterial communities by using data from trained NNs. Cluster analysis using different linkage methods on these data revealed robust community memberships which could not be obtained otherwise.

(This work was presented in part at the 96th and 97th General Meetings [1996 and 1997] of the American Society for Microbiology.)

MATERIALS AND METHODS

Sample collection. Environmental samples were collected at three locations along the middle axis of the Chesapeake Bay at latitudes 39°20.78′N, 38°34.00′N, and 37°16.21′N on 13 to 15 June 1994 aboard the R/V *Cape Henlopen* (Fig. 1). Collecting stations were designated N3, M3, and S3, representing the upper, middle, and lower bay, respectively. At each station, water was collected at two different depths by using 10-liter Niskin bottles attached to a General Oceanics (Miami, Fla.) rosette. Water collection depths were 1 m below the surface and 1 m above the bottom and were designated top and bottom water, respectively. Total depths of bottom-water samples were 10.4, 12.5, and 11.0 m for stations N3, M3, and S3, respectively.

All water samples were processed immediately after collection. To separate the particle-associated and free-living communities, each water sample was filtered under a vacuum (200 to 500 mm of Hg) onto a 3.0-μm-pore-size polycarbonate filter (47 mm; Millipore, Bedford, Mass.) until saturation. Microorganisms retained on the 3-μm-pore-size filters were operationally defined as the particle-associated community. The filtrate (<3.0-μm pore size) was then collected onto 0.22-μm-pore-size Sterivex filters (Millipore) until saturation, by using a peristaltic pump (30). These filters were prepared in duplicate and contained the free-living community. Total bacteria at all stations were harvested by filtering water onto 0.22-μm-pore-size Sterivex filters until saturation. All filters were frozen on dry ice immediately after processing until the return to the laboratory the next morning. At this point, they were stored at −20 or −80°C.

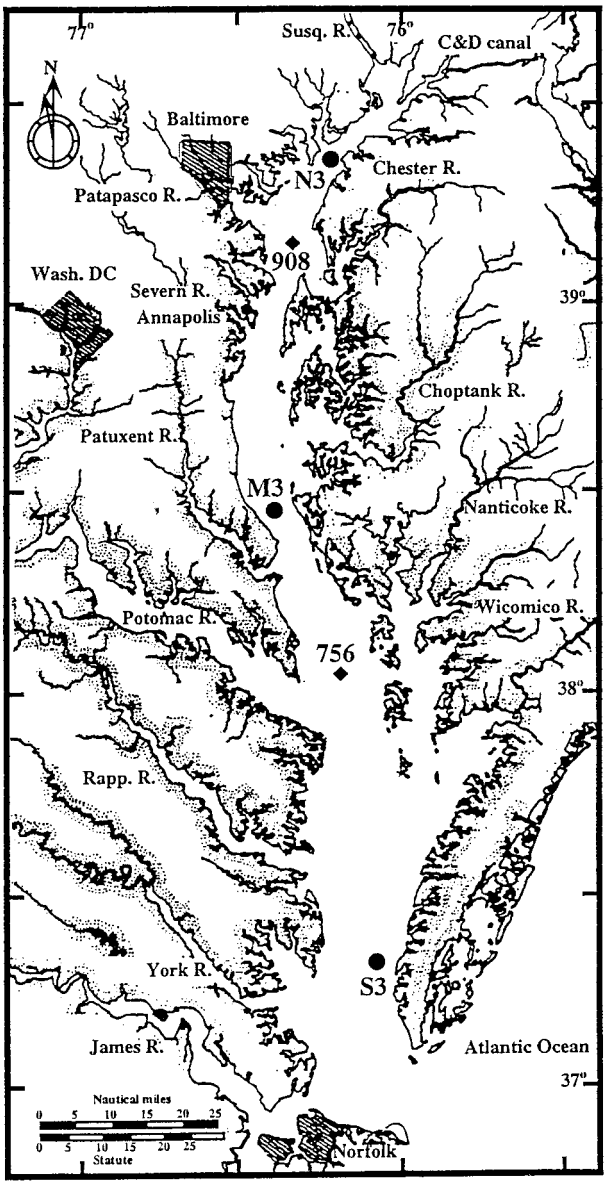

FIG. 1. Map of the Chesapeake Bay showing summer (solid circle) and winter (solid diamond) sampling stations. Samplings from stations N3, M3, and S3 were performed in June 1994. Stations 908 and 756, representing a December 1992 study (6), are shown for comparison.

Determination of cell numbers. Total and free-living bacteria numbers for each sample were determined directly by microscopy as acridine orange direct counts (AODC) (6).

Extraction, labeling, and electrophoresis of SLMW RNA. Extraction of total RNA from filters was based on the procedure of Hoefle (16) and has been previously described (6). All solutions, glassware, and other working materials were rendered RNAse free according to standard procedures (28).

For the particle-associated population, one 3-μl-pore-size filter was cut into small pieces with a clean, sterile scalpel and a glass plate and put into a 15-ml Corex centrifuge tube containing 3.0 ml of extraction buffer (50 mM sodium acetate [NaOAc]–10 mM EDTA–1% [wt/vol] sodium dodecyl sulfate [pH 5.1]). Filter pieces were heated at 100°C for 5 min in extraction buffer and removed with sterile forceps. Cell extracts were immediately extracted with an equal volume of 60°C phenol (containing 0.1% [wt/vol] 8-hydroxyquinoline and equilibrated with 50 mM NaOAc [pH 5.2] until the phenol pH reached <5.2) for 10 min at 60°C. The extract was chilled on ice for at least 2 min, and samples were centrifuged at 12,000 × *g* for 2 min. Aqueous material was extracted in a fresh tube with an equal volume of 60°C phenol-chloroform (4:1) for 5 min and subsequently put on ice for 2 min. Phases were separated by centrifugation (at

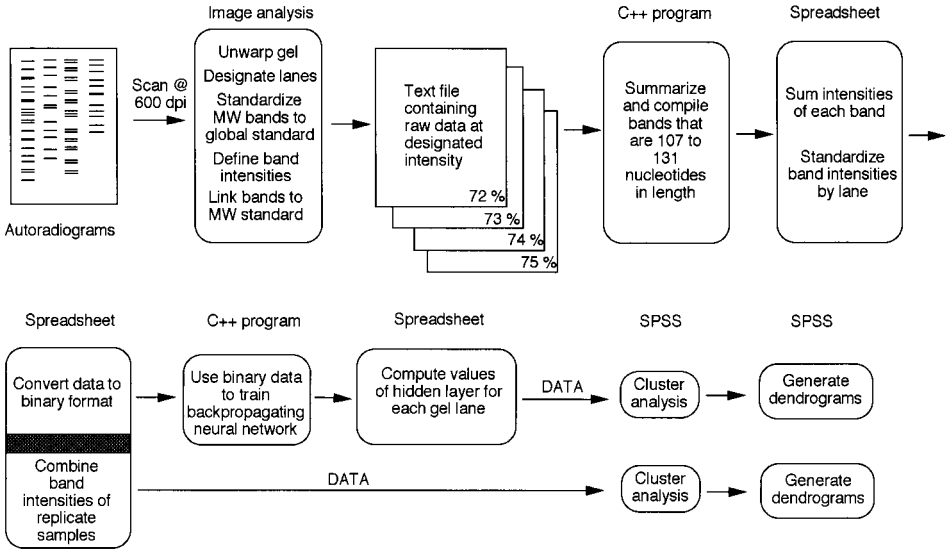

FIG. 2. Strategy for community analysis of 5S rRNA molecules collected from the Chesapeake Bay. Separate cluster analyses were conducted to compare data obtained from the NN with data obtained by combining band intensities of replicate samples. Data obtained from the trained NN were based on the values of the hidden layer (e.g., j_1 to j_{75} in Fig. 3) for each sample.

12,000 × *g* for 2 min). Aqueous material was extracted twice more, once with 200 μl of 2.0 M NaOAc and 2 ml of $CHCl_3$, and once with 2 ml of $CHCl_3$. Final extracts were precipitated with 2.8 volumes of precipitation mixture (100% ethanol-NaOAc-$MgCl_2$ in a ratio of 100:10:1) at −20°C. Precipitated RNAs were collected by centrifugation (at 12,000 × *g* for 20 min) and washed with 70% ethanol. Clean RNAs were dried in a SpeedyVac Concentrator (Savant Instruments, Inc., Farmingdale, N.Y.), resuspended in diethyl pyrocarbonate-treated distilled, deionized water, and quantified with a spectrophotometer. The extracted RNAs were stored at −80°C.

Total RNAs from free-living and total communities, harvested onto Sterivex filters, were extracted by using the same procedures as for the particle-associated bacteria, except for the following. With a 3.0-ml syringe (Becton Dickinson, Cockeysville, Md.), 2.0 ml of extraction buffer was added to the Sterivex housing and recapped with a luer-lock mechanism (30). Filters were heated to 100°C for 5 min. Lysate was withdrawn with a clean syringe and put into a 15-ml Corex tube. Filters were rinsed with 1.0 ml of 50 mM NaOAc–10 mM EDTA (pH 5.2). All subsequent steps were identical to those used for the particle-associated community.

Total RNAs precipitated from all environmental samples were 3′-end labeled with cytidine-3′,5′-[5′-^{32}P]bisphosphate (Amersham, Arlington Heights, Ill.) by using T4 RNA ligase (United States Biochemical, Cleveland, Ohio) as previously described (6). The amount of incorporated label (labeling efficiency) was determined according to standard procedures (28). Five-microliter aliquots of each labeling reaction were pipetted onto two separate Whatman GF/C filters. One of the filters was rinsed (three times) under a vacuum with 5% trichloroacetic acid–20 mM sodium pyrophosphate and represented incorporated label. The other filter, representing total label, was not rinsed. Filters were counted in a Beckman LS 1801 scintillation counter to determine labeling efficiency and specific activities of labeled total RNA. Radioactively labeled RNA was separated from unincorporated label by using Sephadex G-50 spin columns (28).

Equal counts of labeled RNA were subjected to denaturing, high-power polyacrylamide gel electrophoresis. Forty thousand counts per minute was loaded per lane for all samples, except for molecular-weight markers, for which 500 cpm was loaded. In cases where there was less than 40,000 total cpm, all the sample was loaded. The weight corresponding to 40,000 cpm was ca. 0.1 to 0.2 μg, while 5,000 cpm was ca. 0.02 μg of RNA. Polyacrylamide gels (14% polyacrylamide; size, 550 by 170 by 0.4 mm; acrylamide-*N*′,*N*-methylene bisacrylamide ratio, 29:1 [wt/wt]; 7 M urea in TBE buffer [100 mM Tris–83 mM boric acid–1 mM EDTA, pH 8.5]) were run by using a Sequi-sen sequencing gel apparatus (Bio-Rad, Hercules, Calif.) according to the method of Bidle and Fletcher (6).

Commercially available 5S rRNA, $tRNA^{Tyr}$, and $tRNA^{Phe}$ from *Escherichia coli* MRE600 (Boehringer Mannheim, Indianapolis, Ind.) were used as molecular-weight markers. They represent 120, 85, and 76 nt, respectively, and correspond to the three different size classes of SLMW RNA seen, namely, 5S rRNA, tRNA class 2, and tRNA class 1. The 5S rRNA hydrosylate, $tRNA^{Tyr}$, and $tRNA^{Phe}$ were mixed together in the proportions 1:0.2:0.2, at a total of 1.0 ng/lane.

Data for the NN. Autoradiographs were scanned with an image densitometer at a resolution of 600 dpi (Bio-Rad GS-7000) (Fig. 2), and the data were transmitted to a computer (Power Macintosh 7200/120). These images were imported and unwarped by using fingerprinting software (PPC MA Fingerprinting 1.0; Bio-Rad). Once the molecular-weight standards were identified and calibrated, the defined intensities of all bands (which are expressed as percentages by the software and ranged from 72 to 75% in our analysis) were computed. These bands were linked to molecular-weight standards and exported as text files at each intensity. A C++ software program was developed to extract band intensity values in the range of 107 and 131 nt. To account for lane-to-lane and gel-to-gel variations, the summed intensities of each band were normalized to local minima and maxima by using a spreadsheet program (Excel; Microsoft Corp.). Normalized gel data also included control data, which consisted of 24 randomly generated gel lanes. The limits of the random numbers were set by the maximum and minimum values of the normalized band intensity data.

Input data for training the NN were generated by converting the normalized gel data to 6-digit binary numbers. For example, if the summed and normalized intensity of the 107-nt band was 1.31, this number was rounded to an integer and converted to its binary form, 000001. Input data for an entire lane consisted of 25 6-digit binary numbers ordered by size, e.g., from 131 to 107 nt.

Output data for training the NN related input data to collection stations and times, position in the water column, and sample type and were coded as a 10-digit binary number. The first 3 digits refer to the collection station, the next 2 digits refer to time, the next 3 digits refer to position, and the final 3 digits refer to sample type. The collection stations were coded 000, 010, 011, 100, 101, 111, and 001, corresponding to stations 908, 756, M3, N3, and S3, randomly generated data, and molecular-weight-standard data, respectively. Time was coded 00, 01, 11, and 00, corresponding to winter, summer, randomly generated data, and molecular-weight-standard data, respectively. Position was coded 01, 10, 11, and 00, corresponding to top and bottom sampling positions, randomly generated data, and molecular-weight-standard data, respectively. Sample type was coded 001, 011, 010, 100, and 000, corresponding to particle-associated, free-living, and total bacterial samples, randomly generated data, and molecular-weight-standard data, respectively.

Input and output data used for training the NN consisted of the 25 6-digit input binary numbers and the 10-digit binary output data. Data used for testing the NN consisted of the 25 6-digit binary numbers only. To conform with the formatting requirements of the NN program, spaces were inserted between each digit.

NN software. A back-propagating NN, adapted from the work of Rao and Rao (24), was designed as a stand-alone application for the Macintosh by using C++ software (Symantec, Release 5, Cupertino, Calif.). A binhex and binary version of this application is available through an anonymous ftp at inlet.geol.sc.edu. The NN repository directory (/pub/Neural Network) also contains a "neural.readme" file, which describes how to extract the application and what files are needed to make the back-propagating network work (e.g., training.dat and test.dat).

A schematic model of a three-layer NN is shown in Fig. 3. The input layer was defined by the number of binary inputs. For this study, 150 input neurons were needed for the 150 binary digits, representing the band intensities of SLMW RNA molecules from the analysis of one gel lane. The number of neurons in the output layer was defined by the binary digits representing the output data. In this study, 10 binary digits coded for station, position, sample type, and collection time of each sample. The number of neurons in the hidden layer is defined by the

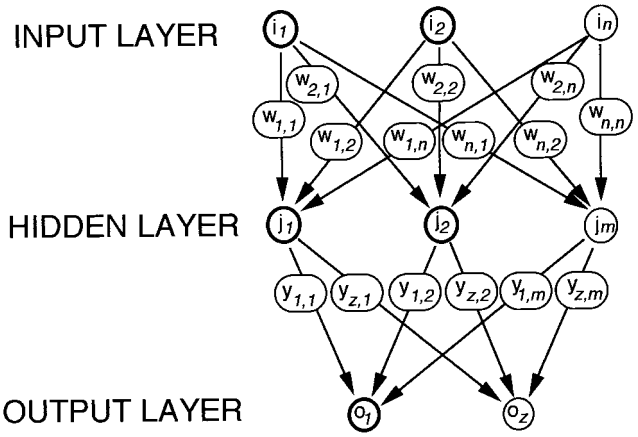

FIG. 3. Schematic model of three-layer neural NN used. i_1 to i_n, j_1 to j_m, and o_1 to o_z represent the neurons of the input, hidden, and output layers, respectively. $w_{1,1}$ to $w_{n,n}$ and $y_{1,1}$ to $y_{z,m}$ represent the weights between the input and hidden layers and between the hidden layer and output layer, respectively.

TABLE 1. Bacterial abundances at stations N3, M3, and S3 in the Chesapeake Bay during June 1994

Station and location in water column	10^5 Cells/ml (mean ± SD)		% of total	
	FL	PA	FL	PA
N3				
Top	12 ± 1.9	14 ± 6.2	46.2	53.8
Bottom	12 ± 8.3	11 ± 4.9	53.3	46.7
M3				
Top	56 ± 12	2.7 ± 1.3	95.4	4.6
Bottom	27 ± 11	2.2 ± 1.0	92.5	7.5
S3				
Top	41 ± 6.6	0.9 ± 0.8	97.9	2.1
Bottom	29 ± 3.7	1.8 ± 0.6	94.2	5.8

[a] Data are given for different depths and geographic locations. Bacterial numbers were determined by acridine orange direct counts (AODC), as previously described (6). FL, free-living bacteria; PA, particle-associated bacteria. Stations N3, M3, and S3 are in the upper, mid-, and lower bay, respectively.

user when the NN is trained. In this study, 75 neurons were used to represent the hidden layer, based on the recommendations of Masters (22).

The value of any neuron in the network can be computed by using the equation

$$f(x) = 1/(1 + e^{-(x)}) \quad (1)$$

where x is the sum of the weights times the inputs (22, 24). For example, in Fig. 3, the value of neuron j_1 is as follows:

$$j_1 = 1/(1 + (e^{(-((w_{1,1}*i_1)+(w_{1,2}*i_2)+(w_{1,n}*i_n)))}))$$

The weights of the hidden layer are generated randomly during the first cycle of training. For subsequent cycles, the weights are adjusted and the values of the neurons are recalculated for the input and output data.The cycles end when the mean square error of the input and out put data is less than the user-defined threshold, or when the maximum number of cycles has been reached. The adjusted sets of weights for matrained network which can be used to recognize SLMW RNA molecules. The hidden layer was used as data for cluster analysis. This was accomplished by importing the adjusted weights to a spreadsheet and calculating the numerical values of the 75 neurons (equation 1) for each gellane.

Numerical analysis. To evaluate the similarity among samples collected from different stations, times, and positions in the water columns, and under different processing conditions (e.g., free-living, particle-associated, and total communities), cluster analysis was performed by using the Euclidean distance coefficient and Ward's and average linkage clustering methods (SPSS, Inc.). Different clustering methods were used to determine the robustness of group memberships (10). To determine which SLMW RNA molecules were common to the clusters, normalized gel data were imported into Transform 3.01 software (Spyglass, Inc., Savoy, Ill.).

RESULTS

Bacterial numbers. Total bacterial numbers ranged from 2.3×10^6 to 5.8×10^6 cells/ml. Bacterial numbers were dominated by free-living bacteria at all stations and times examined, except in the upper bay (station N3) (Table 1). At stations M3 and S3, free-living bacteria accounted for >90% of the total bacteria for both top and bottom water. However, bacteria attached to particles made a greater contribution at station N3, accounting for 55 and 46% of the total bacteria for top and bottom water, respectively. Typically, particle loads were concentrated in the upper-bay regions, primarily due to the high degree of freshwater influx from the Susquehanna River.

RNA analysis and labeling. RNA extracted from natural samples was of high quality as determined by UV spectrophotometry. All sample types yielded A_{260}/A_{280} values of 1.6 to 2.0, indicating that contaminating material was minimal. Total RNA yields varied depending on the sample type and the total number of cells collected per filter. Total and free-living community samples yielded the most RNA per filter, 1.9 to 18.3 and 4.7 to 17.4 μg, respectively. This was presumably due to high cell numbers on the Sterivex filters (10^9/filter). Normalized RNA amounts per cell for total and free-living bacteria were 2.0 to 4.1 and 0.8 to 4.1 fg per cell, respectively. Particle-associated samples had much lower numbers of cells per filter (10^7 to 10^8) and, therefore, the total RNA recovered was less (0.5 to 2.2 μg). Normalized RNA per cell for bacteria associated with particles was 8.7 to 81.5 fg/cell.

Up to 2.0 μg of total extracted RNA was 3′-end labeled with cytidine-3′,5′-[5′-^{32}P]bisphosphate. Labeling efficiencies were much higher for samples collected in the midbay and lower-bay stations, probably due to lack of humic material in the water column. In general, 1 order of magnitude more label was incorporated into corresponding samples from stations M3 and S3 than from station N3. The highest labeling efficiency seen for environmental samples was ca. 20 to 26%. For comparison, 5S rRNA marker from *E. coli* MRE600 labeled at ca. 89%. All labeled RNA from environmental sources had specific activities of 10^5 cpm/μg of RNA.

Autoradiographs of SLMW RNA molecules extracted from the Chesapeake Bay in the summer are shown in Fig. 4. The SLMW RNA molecules are clearly separated into three major classes of molecules: 5S rRNAs, and both class 2 and class 1 tRNAs. Molecular size ranges of bacterial 5S rRNA are 107 to 131 nt, while class 2 and class 1 tRNAs range from 83 to 96 and 72 to 79 nt, respectively (31, 32).

NN computation. The NN was independently trained on three occasions to evaluate its performance and to obtain the adjusted weights of the hidden layer. The training mode took between 30 and 35 min with an error threshold, learning rate, momentum parameter, and noise factor of 0.001, 0.5, 0, and 0, respectively. Details concerning the operating parameters of the NN are available elsewhere (24, 33). The NN converged to the error threshold after 25 to 30 cycles, examining 1,875 to 2,250 patterns. After the NN converged to the error threshold, the NN was tested with data consisting of the 25 6-digit binary numbers. On all occasions, the NN correctly related collection stations and times, positions in the water column, and sample types (data not shown).

Cluster analysis. Dendrograms showing the relationship of microbial communities and their corresponding 5S rRNA gel images are shown in Fig. 5. The dendrograms represent the analysis of the data from one hidden layer. Most of the samples

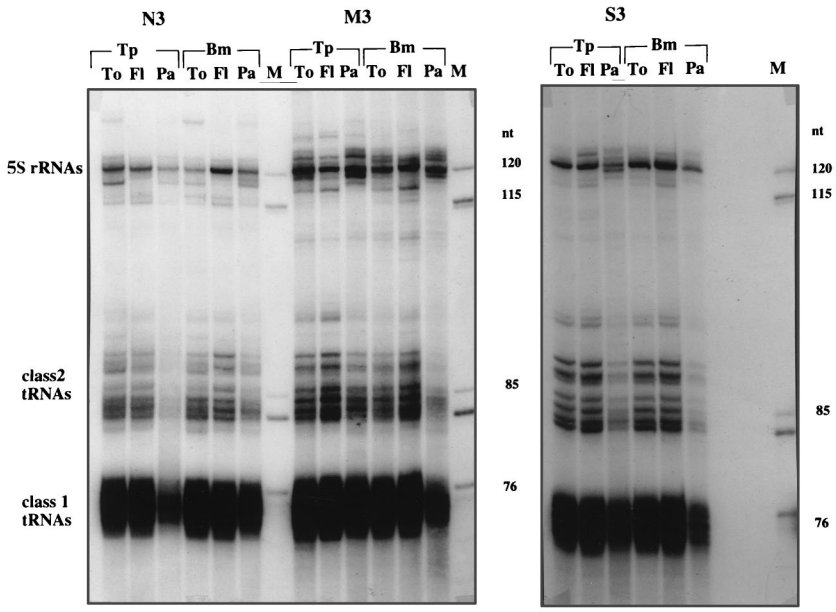

FIG. 4. Autoradiographs showing SLMW RNA banding patterns of summer water samples. Total RNAs were extracted, 3′-end labeled with T4 RNA ligase, and run on denaturing, high-resolution one-dimensional 14% polyacrylamide gels. Results are shown for upper (N3)-, mid (M3)-, and lower (S3)-bay locations, top (Tp) and bottom (Bm) water. To, total bacterial community; Fl, free-living community; Pa, particle-associated community; M, molecular-weight markers (5S rRNA, $tRNA^{Tyr}$, and $tRNA^{Phe}$ from *E. coli* at 120, 85, and 76 nt, respectively). The three classes of SLMW RNAs are indicated.

had robust group memberships. Duplicate samples, samples from the same station which had been independently processed, molecular-weight standards, and randomly generated samples clustered as separate and distinct groups. Based on the composition of the clusters, group memberships appeared to be determined by station rather than by time, position, and type. However, in some cases, station and type or station and position influenced the clustering (see below). Comparison of the clusters to their corresponding gel lanes revealed which bands, or group of bands, were common to the clusters (Fig. 5).

We compared these dendrograms (Fig. 5) to those obtained by analyzing the data from two independently trained NNs (data not shown). Although weights of the hidden layer changed every time the NN was trained, these data yielded similar but not identical dendrograms, indicating that group memberships for most samples were robust. To investigate the relationship between clusters in the dendrograms and sample membership, letter values were assigned to predominant clusters and the community characteristics describing each cluster were examined. A summary of this analysis is shown in Table 2.

Winter samples were represented by clusters A, B, and D (Table 2). Cluster A was composed of free-living and total communities extracted from top and bottom waters of the upper bay and midbay. All samples in this cluster contained 5S rRNA bands of 115 and 109 nt. Secondary 5S rRNA bands, defined as bands that occurred frequently, included 5S rRNA bands of 110, 117, 120, and 122 nt. Clusters B and D consisted of particle-associated communities from the midbay and upper bay, respectively. Cluster B differed from cluster D in that the former was composed of top-water samples only, while the latter was composed of both top- and bottom-water samples. All samples in cluster B had a 5S rRNA band of 121 nt. Secondary 5S rRNA bands for cluster B were 117, 119, and 120 nt. All samples of cluster D contained a dominant 5S rRNA band of 122 nt. Although two of four samples of cluster B contained this band, there was little overlap in community compositions with cluster D.

Summer samples were represented by clusters C, E, and F (Table 2). Community types had little influence on group memberships for these samples, indicating that there was no difference in the composition of free-living or particle-associated communities. With the exception of cluster C, which consisted of midbay samples, position in the water column had no effect on group membership, indicating that the community composition was homogeneous in the water column of the upper- and lower-bay regions. In dendrograms based on the data of two of the three trained NNs, cluster C occurred as two independent clusters, C_1 and C_2 (data not shown). Samples from cluster C_1 were from the top of the water column, while samples from cluster C_2 were from the bottom of the water column. Particle-associated bacteria from cluster C_2 were the only samples in the summer that did not contain a 121-nt 5S rRNA band. Differences in the community composition of these subclusters were presumably due to the presence of anoxic zones in the bottom waters of the Chesapeake Bay during the summer. Subtle discrepancies in band intensities of the rRNA molecules may also explain these differences. The community compositions of subclusters C_1 and C_2 were similar in that they contained 120- and 122-nt 5S rRNA bands, with secondary bands of 110, 121, and 123 nt. Samples of clusters E and F represent the upper and lower bay, respectively, and often clustered together. All samples in clusters E and F contained a common 121-nt 5S rRNA band. Clusters E and F differed from one another by having secondary rRNA bands of 116 and 119 nt, and of 120 nt, respectively.

Comparison of dendrograms using different clustering data. To ensure that cluster analyses using NN data corresponded to cluster analyses using normalized band intensity data, dendrograms were compared. Since clustering of band intensity data was inconsistent, replicate samples were combined. Combining band intensity data resulted in considerable loss of informa-

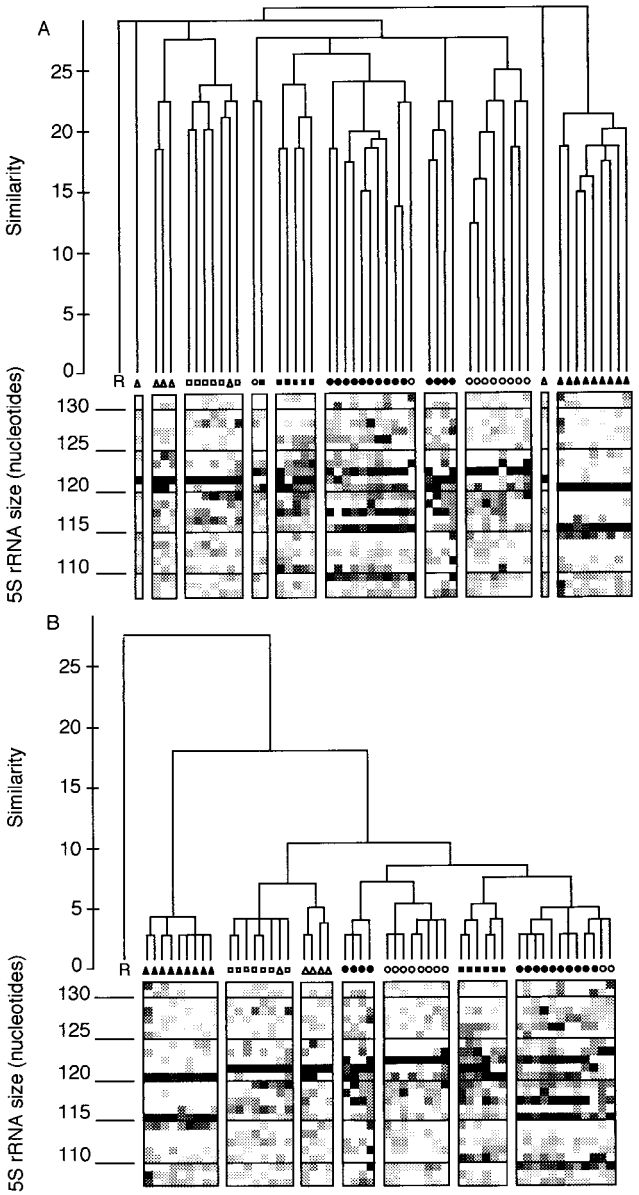

FIG. 5. Dendrograms showing the relationship of microbial communities from the Chesapeake Bay and their corresponding 5S rRNA digitized gel images. The similarity scores of the 75 samples were determined by calculating the squared Euclidean distance and clustering the data by using average (A) and Ward's (B) distance linkage methods. The 24 randomly generated samples are represented by the letter R. Each 5S rRNA molecule is represented by a box in the gel image, whose shade is determined by the intensity of the band; open and solid boxes represent low and high intensities, respectively. ○, station 908; ●, station 756; □, station N3; ■, station M3; △, station S3; ▲, molecular weight standards.

tion, as can be seen in the gel images (Fig. 6). Overall, group memberships were similar to those obtained by using NN data (Fig. 5). For example, samples representing summer stations from the upper and lower bay grouped together, having similar 5S rRNA bands to those observed in clusters E and F (Table 2). Particle-associated communities in the winter formed a cluster corresponding to clusters B and D from the mid- and lower-bay regions, respectively. Winter water samples from the midbay formed a cluster corresponding to cluster A (Table 2). Summer samples from the midbay grouped together, corresponding to cluster C (Table 2). Based on these comparisons, NN data yielded results similar to those obtained by combining replicate normalized band intensity data, but without loss of information pertaining to 5S rRNA molecules.

DISCUSSION

SLMW RNA analysis focuses on stable RNA molecules involved in protein synthesis and cell maintenance. By using this analysis method, we were able to monitor active cells in the community, avoiding the bias associated with conventional culturing methods (26). SLMW RNA analysis has additional advantages: (i) SLMW RNA molecules in bacterial cells remain chemically unchanged regardless of the physiological state of the bacteria (16); (ii) independently extracted environmental samples have identical SLMW RNA banding patterns (6, 18); (iii) SLMW RNA molecules can be extracted from phylogenetically diverse bacterial organisms (6, 17); and (iv) SLMW RNA analysis targets the whole community without the biases created by PCR or sequence-specific probes.

It is also important to note that SLMW RNA banding patterns may not be an absolute representation of abundances of all bacterial types within a community for several reasons. First, RNA concentrations will be highest in the most active community members. Second, extraction efficiencies of RNA vary in different types of bacterial cells. Third, RNA concentration may also be a function of changes in the level of rRNA pools during starvation and recovery (21). For example, organisms under high-nutrient conditions or possessing strong adaptational responses to starvation may contain more ribosomes and may appear to be the dominant 5S rRNA band of a sample. Nonetheless, SLMW RNA analysis provides a method to compare the compositions of physiologically active microbial communities and demonstrate significant differences in community composition.

For the NN analysis, we focused on the 5S rRNA bands to provide a first-order comparison of complex communities. 5S rRNA is highly conserved and similar for closely related organisms, but significant differences in the molecular weight of 5S rRNA molecules indicate that organisms are distantly related. NN and cluster analyses based on 5S rRNA alone clearly demonstrated differences in complex communities, which was the objective of this study. Analysis of tRNA may be used to differentiate among more closely related organisms and can be included in future studies. Although adequate resolution of tRNA in gels becomes very difficult in diverse, complex communities, image analysis techniques may enable second-order analysis of communities containing more closely related organisms.

RNA extracted from water samples was of high quality based on UV spectrophotometric analyses. Normalizing of the RNA yields on a per-cell basis was based on the assumption that RNA was from eubacterial sources. Higher RNA levels per cell were observed for particle-associated communities than for total and free-living counterparts (total bacteria, 2.0 to 4.1 fg/cell; free-living bacteria, 0.8 to 4.1 fg/cell; particle-associated bacteria, 8.7 to 81.5 fg/cell), with values resembling those seen for bacteria with high nutritional status. For comparison, *E. coli* B/r in balanced growth in glucose minimal medium at 37°C has ca. 60 fg of RNA per cell (23). This observation supported the previous finding (12) that bacteria associated with particulate material are more active on a per-cell basis than their free-living counterparts. However, it is also possible that RNA concentrations per cell appeared to be higher for attached cells because their numbers were underes-

TABLE 2. Microbial community characteristics of water samples collected from different positions and locations in the Chesapeake Bay during the winter (1992) and summer (1994)[a]

Cluster[b]	Season		Community type			Position		Location[c]		
	Winter	Summer	Free-living	Particle-associated	Total	Top	Bottom	Upper	Mid	Lower[d]
A	+	−	+	−	+	+	+	+	+	ND
B	+	−	−	+	−	+	−	−	+	ND
C_1	−	+	+	+	+	−	+	−	+	−
C_2	−	+	+	+	+	+	−	−	+	−
D	+	−	−	+	−	+	+	+	−	ND
E	−	+	+	+	+	+	+	+	−	+
F	−	+	+	+	−	+	+	−	−	+

[a] The data were based on dendrograms generated by using the Euclidean distance coefficient with average and Ward's linkage methods on three independently trained NNs.
[b] Clusters A to F refer to samples that have consistent group membership (see text).
[c] Upper, stations 908 and N3; mid, stations 756 and M3; lower, station S3.
[d] ND, not determined.

timated; microscopic counts of particle-associated cells may be low if cells are obscured by particulate material or aggregates.

To confirm that the extracted RNA was derived essentially from bacteria, we probed our total RNA samples with domain-specific oligonucleotide probes specific to the 16S rRNAs and tried to evaluate the percentage contribution of eubacterial, eucaryotic, and archaeal domains. Total RNA hybridized to the eubacterial and eucaryotic probes, but quantification was not possible due to the nature of the membrane used (25). The most probable eucaryotic source of nucleic acids would be phytoplankton due to spring blooms. However, our extraction protocols did not appear to release algal RNA. RNA could be extracted from pure algal cultures only when rigorous mechanical manipulation was used to break the cell walls.

This study confirms previous observations (6, 12) that free-living bacteria dominate Chesapeake Bay bacterial communities in terms of numbers and that they make up large percentages of the total bacterial population. For example, in mid- and lower-bay stations, free-living bacteria comprised >90% of the total bacteria per sample, while in upper-bay samples, ca. 50% of the total community were free living. Similar observations were made in the winter, with free-living bacteria in the upper bay making up 79 and 65% of the total bacteria in top and bottom waters, respectively (6). The high percentages of particle-associated bacteria in the upper-bay samples were probably due, at least in part, to higher particle loads brought in by the Susquehanna River and other large freshwater tributaries emptying into the bay.

A primary objective was to determine whether or not free-living and particle-associated bacterial communities differed in their compositions in summer, as had been observed in samples taken in the winter (6). Cluster analyses of free-living and particle-associated bacteria in the summer suggested that there was no difference between community types. These results are in contrast to those obtained with winter water samples (6). In the winter, free-living bacterial communities exhibited considerable similarities in composition from the mid- to the upper bay, while those communities associated with particulate material were discernibly different, as indicated by their independent and robust clusters (i.e., clusters B and D in Table 2). Cluster analyses revealed that there was little overlap in community compositions of winter and summer samples. These findings suggest that the compositions of free-living communities in the Bay change seasonally, in response to hydrographic conditions. Particle-associated bacteria are dependent not only on hydrographic conditions but also on the nature of particulate matter they attach to. The composition of particle-associated bacterial communities in the upper- and lower-bay samples are different (i.e., clusters B and D in Table 2) because the nature of particulate matter presumably selects for the growth of specific bacterial communities. Free-living communities are homogeneous because vertical mixing in the winter provides stable hydrographic conditions and dissolved organic material (DOM) which promotes bacterial proliferation.

In the summer, the Chesapeake Bay becomes highly stratified, resulting in active bacterial communities that degrade phytodetritus. Degradation of the phytodetritus results in the formation of anoxic zones in the bottom waters of the midbay. Since attached bacteria degrade particulate organic matter (POM) (9, 29), decaying algal materials select for organisms that can break down POM via exohydrosylases. It is possible that the organisms breaking down POM are also assimilating DOM and releasing progeny into the water column as the

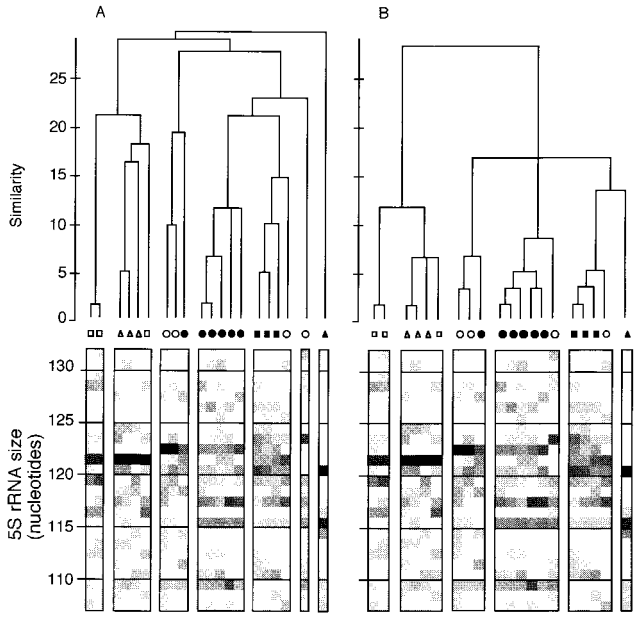

FIG. 6. Dendrograms showing the relationships of microbial communities from the Chesapeake Bay and their corresponding 5S rRNA gel images. Each sample represents the sum of combined replicates. The similarity scores of the 20 samples were determined by calculating the squared Euclidean distance and clustering the data by using average (A) and Ward's (B) distance linkage methods. Labeling of the dendrograms and gel images is the same as for Fig. 5.

particulate material is degraded (2). Based on this hypothesis, there should be little difference between free-living and particle-associated bacteria because bacteria move freely between these two habitats. Since the composition of POM changes with successive phytoplankton blooms, it is possible that the composition of these summer communities fluctuates with changing amounts and types of POM and/or DOM. Further studies elaborating on the effects of POMs and DOMs on the compositions of free-living and particle-associated communities are needed for full understanding of the dynamics of bacterial communities in the Chesapeake Bay in the summer.

The community composition of summer midbay samples was different from those of summer upper- or lower-bay samples. Moreover, summer midbay samples had more similarity to winter samples taken from the same region (i.e., station 756 in Bidle and Fletcher [6]) than to samples taken at the same time. Comparisons of the 5S rRNA bands (Fig. 5) suggest that the community composition in the midbay was quite stable, varying slightly between seasons, possibly due to minor influxes of freshwater, changing particle loads, and/or the establishment or dissolution of an anoxic zone in the summer. Regardless of collection time, midbay samples tended to have a greater variety of 5S rRNA bands than the other stations, indicating that no one group of bacteria dominated the community. The apparent stability of the community composition may have been due to a variety of different sources of POM and/or DOM or another factor(s) not known at this time.

Most summer samples contained a 5S RNA band of 121 nt (Fig. 5) which occurred at a high intensity. This finding suggests that a bacterium or groups of phylogenetically related bacteria are dominant members of the summer communities, although sequencing of molecules is required to confirm whether they are derived from the same or closely related organisms. The presence of specific, preferred carbon sources may account for certain bacteria becoming dominant. It has been established that monosaccharides and dissolved free amino acids, derived from living and decaying phytoplankton populations, contribute to bacterial productivity in the spring and summer months (4). The fact that some bacteria incorporate monosaccharides ca. 16-fold more than dissolved free amino acids suggests that monosaccharides may be a preferred carbon source in the Chesapeake Bay (4). Therefore, it is possible that the 121-nt band represents a community that rapidly utilized preferred carbon sources such as monosaccharides, outcompeting other bacteria.

NNs have traditionally been used to recognize patterns in data that are impossible to substantiate by standard statistical methods (22). In particular, NNs are useful for analyzing fuzzy, noisy, chaotic, and/or unpredictably nonlinear data (22). The application of NNs to the recognition of patterns in molecular biology is relatively new. For example, NNs have been used to identify the restriction enzyme patterns of *Escherichia coli* O157:H7 (8), the pyrolysis mass spectra of *Mycobacterium tuberculosis* complex species (11), the promoter sites of *E. coli* (19), and the fatty acids of marine heterotrophic bacteria (5).

In this study, NNs were trained to recognize SLMW RNA banding patterns produced by polyacrylamide gel electrophoresis. To our knowledge, this is the first study to use data from the hidden layer for examining the banding patterns of molecules produced by electrophoresis. We used three independently trained NNs to determine if data from the hidden layers would produce similar clusters when statistically analyzed. Cluster memberships of the samples were robust and similar regardless of which hidden layer was used for cluster analysis, indicating that our approach was valid. The dendrograms were similar, but not identical, for different hidden-layer data because the weights changed every time the NN was trained. Moreover, the weights were randomized in the first training cycle. We calculated the hidden layer data for each gel lane by using equation 1. By performing cluster analyses and determining group members of the clusters, we determined which SLMW RNA bands were common to particular clusters. Most clusters had one to four bands in common, indicating similarities with respect to bacterial composition of the samples. The described method for analyzing the complex banding patterns of SLMW RNA molecules may be applied to a wide variety of molecular biology problems where interpretation of the data is confused by noise and uncertainty is a factor.

In summary, the application of NNs allowed for interpretation of complex gel data and determination of microbial community changes in the Chesapeake Bay. The compositions of microbial communities of the Chesapeake Bay vary both seasonally and spatially. In winter, the community compositions of particle-associated bacteria are dissimilar, presumably due to differences in the nature of particulate debris from the tributaries. Conversely, the community compositions of free-living bacteria were similar, presumably because vertical mixing of the bay waters distributed autochthonous sources of organic material into the water column. In the summer, there was no apparent difference between community types, presumably because the same organisms degrading POM were also assimilating DOM and producing free-living and/or attached progeny. It is possible that certain bacteria become dominant members of natural communities in the presence of preferred carbon sources.

ACKNOWLEDGMENTS

The crews of the research vessel *Cape Henlopen*, as well as the educators from the Living Classrooms Foundation (Baltimore, Maryland) provided valuable assistance during cruises on the Chesapeake Bay. Gratitude is also extended to Allison Caalim and Tim Potter for their help in sample collecting and laboratory work and to Michael O'Neill (University of Maryland—Baltimore County) for his discussions on NN computing.

REFERENCES

1. **Aleksander, I., and H. Morton.** 1991. An introduction to neural computing, p. 1–20. Chapman & Hall, Ltd., London, England.
2. **Azam, F., and B. C. Cho.** 1987. Bacterial utilization of organic matter in the sea, p. 261–281. *In* M. Fletcher, T. R. G. Gray, and J. G. Jones (ed.), Ecology of microbial communities. Cambridge University Press, Cambridge, England.
3. **Beale, R., and T. Jackson.** 1990. Neural computing: an introduction. IOP Publishing Ltd., New York, N.Y.
4. **Bell, J. T.** 1990. Carbon flow through bacterioplankton in the mesohaline Chesapeake Bay. M.S. thesis. University of Maryland, College Park.
5. **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.
6. **Bidle, K. D., and M. Fletcher.** 1995. Comparison of free-living and particle-associated bacterial communities in the Chesapeake Bay by stable low-molecular-weight RNA analysis. Appl. Environ. Microbiol. **61:**944–952.
7. **Caron, D. A., P. G. Davis, L. P. Madin, and J. M. Sieburth.** 1982. Heterotrophic bacteria and bacteriovorous protozoa in oceanic macroaggregates. Science **218:**795–797.
8. **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.
9. **Cho, B. C., and F. Azam.** 1988. Major role of bacteria in biogeochemical fluxes in the ocean's interior. Nature **332:**441–443.
10. **Everitt, B. S.** 1974. Cluster analysis. Wiley and Sons, New York, N.Y.
11. **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.
12. **Griffith, P., F. Shiah, K. Gloersen, H. W. Ducklow, and M. Fletcher.** 1994.

Activity and distribution of attached bacteria in Chesapeake Bay. Mar. Ecol. Prog. Ser. **108:**1–10.
13. **Griffith, P. C., D. J. Douglas, and S. C. Wainright.** 1990. Metabolic activity of size-fractionated microbial plankton in estuarine, nearshore, and continental shelf waters of Georgia. Mar. Ecol. Prog. Ser. **59:**263–270.
14. **Griffith, P. C., and M. Fletcher.** 1991. Hydrolysis of protein and model dipeptide substrates by attached and nonattached marine *Pseudomonas* sp. strain NCIMB 2021. Appl. Environ. Microbiol. **57:**2186–2191.
15. **Hodson, R. E., A. E. Maccubbin, and L. R. Pomeroy.** 1981. Dissolved adenosine triphosphate utilization by free-living and attached bacterioplankton. Mar. Biol. **64:**43–51.
16. **Hoefle, M. G.** 1988. Identification of bacteria by low molecular weight RNA profiles: a new chemotaxonomic approach. J. Microbiol. Methods **8:**235–248.
17. **Hoefle, M. G.** 1990. RNA chemotaxonomy of bacterial isolates and natural bacterial communities, p. 129–159. *In* J. Overbeck and R. J. Chrost (ed.), Aquatic microbial ecology—biochemical and molecular approaches. Springer-Verlag, Berlin, Germany.
18. **Hoefle, M. G.** 1992. Bacterioplankton community structure and dynamics after large-scale release of nonindigenous bacteria as revealed by low-molecular-weight-RNA analysis. Appl. Environ. Microbiol. **58:**3387–3394.
19. **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.
20. **Kirchman, D., and R. Mitchell.** 1982. Contribution of particle-bound bacteria to total microheterotrophic activity in five ponds and two marshes. Appl. Environ. Microbiol. **43:**200–209.
21. **Kramer, J. G., and F. L. Singleton.** 1992. Variations in rRNA content of marine *Vibrio* spp. during starvation-survival and recovery. Appl. Environ. Microbiol. **58:**201–207.
22. **Masters, T.** 1993. Practical neural network recipes in C++. Academic Press, New York, N.Y.
23. **Neidhardt, F. C., J. L. Ingraham, and M. Schaechter.** 1990. Physiology of the bacterial cell: a molecular approach, Sinauer Associates, Inc., Sunderland, Mass.
24. **Rao, V. B., and H. V. Rao.** 1993. C++ neural networks and fuzzy logic. MIS Press, New York, N.Y.
25. **Raskin, L., R. I. Amann, L. K. Poulsen, B. E. Rittman, and D. A. Stahl.** 1995. Use of ribosomal RNA-based molecular probes for characterization of complex microbial communities in anaerobic biofilms. Water Sci. Technol. **31:** 261–272.
26. **Roszak, D. B., and R. R. Colwell.** 1987. Survival strategies of bacteria in the natural environment. Microbiol. Rev. **51:**365–379.
27. **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.
28. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
29. **Smith, D. C., M. Simon, A. L. Alldredge, and F. Azam.** 1992. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. Nature **359:**139–142.
30. **Somerville, C. C., I. T. Knight, W. L. Straube, and R. R. Colwell.** 1989. Simple, rapid method for direct isolation of nucleic acids from aquatic environments. Appl. Environ. Microbiol. **55:**548–554.
31. **Specht, T., J. Wolters, and V. A. Erdmann.** 1990. Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Res. **18**(Suppl.)**:**2215–2235.
32. **Sprinzel, M., J. Moll, F. Meissner, and T. Hartmann.** 1985. Compilation of tRNA sequences. Nucleic Acids Res. **13:**r1–r49.
33. **Tollenaere, T.** 1990. SuperSAB: fast adaptive back propagation with good scaling properties. Neural Networks **3:**561–573.