Computer Program Designed to Follow Fluctuations in Microbial Populations and Its Application in a Study of Chesapeake Bay Microflora

JAMES D. OLIVER¹ AND R. R. COLWELL²

Department of Biology, Georgetown University, Washington, D.C. 20007

Received for publication 24 August 1973

A computer program has been developed which performs cluster analysis of microorganisms using methods of numerical taxonomy. The program is designed to group related strains, identify the groups by reference to known strains, and calculate a hypothetical median organism (HMO) for each group. The HMO serves to condense taxonomic information and provides a tag for each strain cluster. Every strain in a group is compared with the HMO established for that group. A representative strain for the group is obtained by selection of the strain showing highest similarity to the HMO. New data sets can be compared with data sets of previous analyses. Hence, the occurrence of the same taxonomic groups within separate data sets can be determined. Quantitative or qualitative differences in distribution of taxonomic groups within or between data sets can be measured. The output from the computer is a graphical display, using an on-line plotter; thus, the investigator is provided with visual comparison of data sets. Results obtained from a study applying the computer program in an analysis of taxonomic data obtained for 43 bacterial strains isolated from Chesapeake Bay indicate the usefulness of this method of taxonomic analysis in microbial ecology.

The concept of utilizing computers for analysis of bacterial taxonomic data was first introduced by Sneath in 1957 (11). In the 17 years since, numerical taxonomy has become a powerful tool in bacterial taxonomy (2-5, 7, 10). Results of numerical taxonomy studies, in conjunction with other data, such as deoxyribonucleic acid (DNA) base composition, DNA/DNA reassociation, serology, isozyme analysis, gas-liquid chromatography, and bacteriophage sensitivity, provide a useful approach to bacterial taxonomy which has been termed polyphasic (1). In our laboratory, numerical taxonomy analyses are accomplished using a program (GTP-2), written in PL-1 computer language, which was developed by one of us (RRC). The GTP-2 program has been implemented on an IBM 360-40 computer equipped with both tape and disk storage and an on-line plotter.

Ongoing studies in ecology and taxonomy of estuarine and marine microorganisms (1, 6, 9) have necessitated a number of additions to the original program in order to allow comparison of data obtained for newly isolated microbial strains with those obtained in previous field surveys and to identify taxonomic groups represented among the freshly isolated microbial strains. One set of programs developed and linked to GTP-2 allows identification of taxonomic groups, calculation of a hypothetical median organism for each group, and selection of representative strains for the groups based on calculated relationship with hypothetical median organisms for the groups. Thus, when a given bacterial population is examined, strain comparisons can be made with strains comprising previously studied populations and correlations or differences can be detected, as, for example, those arising from environmental influences.

A preliminary report of this work was presented previously (J. D. Oliver and R. R. Colwell, Abstr. Annu. Meet. Amer. Soc. Microbiol. 1972, G49, p. 38).

MATERIALS AND METHODS

The computer program used and described here was written in computer language PL-1 and can be used on any computer with PL-1 capabilities. Execu-
tion of the program involves less processing time if tape or disk storage is utilized, although punch cards can be used if necessary. The program, as written, permits calculation of data for 1,000 strains, with each strain providing a maximum of 500 coded characters.

A brief description of the major segments of the program is as follows. GTP2A is the main body of the program, which calls up subroutines, depending on specifications provided on a data control card supplied by the operator. Each of the subroutines is optional, with the user maintaining control of all the subroutines via card entry or the computer console, as desired. OVER is the first of the subroutines and reads in the data control card which designates the number of bacterial strains to be examined, the number of phenotypic characteristics coded for each strain, the similarity percentage level at which strains are to be grouped, and the remaining subroutines and their options to be executed. The data cards follow. A card sequence check is performed (if desired) and the coded data, in the "0", "1", or "3" format, i.e., indicating negative, positive, or "not tested" result, respectively, for each test, are transferred to disk storage. Strain identifications, or labels, are printed as soon as they are read into the computer. If any cards are out of sequence, or any incorrect data entries are made, i.e., alphabetic or numbers other than 0, 1, or 3, error messages are produced. The program is designed to correct such errors by substituting a "3" for incorrect punches entering the computer.

PROCESS is that part of the program which compares the coded phenotypic data for each strain with those of every other strain in the data set and calculates percent similarity (%S) for each strain comparison. In this context, either the Jaccard or matching similarity index can be used (12), with selection made by the operator at the time of the computing operation.

OUTPUT comprises the subroutine which groups strain pairs by decreasing %S, after which the strains are rearranged so that strains sharing highest %S are adjacent, or closely located, as accomplished in GTP-2 (4). The computer sequentially assigns numbers to the rearranged strains and prints a table listing original strain sequence numbers, newly assigned numbers, and corresponding strain labels. The %S for strain pairs is assigned a symbol, i.e., an asterisk means 91 to 95% S, a dollar sign means 81 to 85% S, etc., to enhance the visual scanning of the output which is produced in tabular form as a triangle (see Fig. 1).

JDO is a multipurpose subroutine. The first step in the subroutine is an examination by computer of the triangle obtained from OUTPUT to determine if groups of related strains exist. The operator has the option, at this step, of indicating %S necessary for two or more organisms to be considered a group. If it is decided that two or more organisms must possess >70% S, the program will examine each strain in the triangle sequentially, moving along the diagonal of the triangle. For those strains >70% S, a group is demarked. When a value <70% S occurs in the step-by-step scan, additions to the group cease to be made and the search for a new group is initiated. Output from this program consists of an identifying number for the group and for each strain number and the original label or identification of each strain in the groups.

A second procedure carried out by this subroutine is a comparison of the coded phenotype of all strains of each group. From this calculation is obtained a hypothetical median organism (HMO) for each group, with the following criteria used to establish the HMO. If no strain within a group has been examined for a given phenotypic character, the HMO will show a "3" for that character. Similarly, if >66% of the strains in the group are positive for a given character, the HMO is coded "1" for that character and if <34% of the strains in a group are positive the HMO is coded "0" for that character. In cases where the character is present in >34% but <66% of the strains, the HMO is coded "5." The HMO (an HMO is calculated for each group of strains in the analysis) is listed with the data for each group and cards for each HMO are punched so that the HMO can be included in the subroutine MEDIANS. A card is also prepared which designates the number of groups obtained in a given analysis, the number of strains comprising each group, and the maximum number of strains comprising any of the groups (required for plotting routines). Selection of 66% S and 34% S was an arbitrary decision and it should be stated that 75% S and 25% S or 80% S and 20% S could also have been used. For the purposes of our study, we chose to use the %S values as given above.

A third component of one subroutine, CHKMEDS, is an option, which, if utilized, performs the following steps. Each coded phenotypic character, i.e., in the format "0", "1", or "3," for each strain in each of the groups is compared with corresponding coded character of the HMO for that group. For any given strain in a group, the test code must match that of the HMO of the group in order to be considered positive. The "3" and "5" codes are not included in this step of the analysis. A table is printed by the computer, listing group identification number, the number of organisms comprising the group, strain identifications or labels, and %S for each strain with the HMO for the group. Thus, strains which are highly representative of each group can be selected to serve as type or neotype strains in subsequent studies.

MEDIANS is the subroutine which enters the HMO strain data set obtained in earlier studies, actual strain data, data from duplicate or repeated sampling, or data for reference or culture collection strains. The newly entered strain data are compared with corresponding strain data for those HMO computed in subroutine JDO. Output consists of a printed table listing %S for groups from one data set compared with those of the second. The number of characters involved in the calculation of %S are also listed (those characters coded "3," however, are not used in the computation). Comparisons made are with the HMO of the groups with the strain data added to the data set and not with the individual strains of the groups themselves. Up to 99 of these
external, i.e., reference or known, strains can be compared using this subroutine.

PLOTS is the subroutine consisting of two procedures for obtaining data output in graph form. Both graph subroutines are optional. The first, PLOT 1, produces a bar graph of groups identified from a scan of the similarity triangle as the abscissa and the number of strains in each group as the ordinate. Data presented in this manner provide a clear examination of the major groups found in a given sample taken in, for example, ecological studies. The second subroutine, PLOT 2, produces a double bar graph of the groups identified by scanning the similarity triangle, which, again, is the abscissa, and the number of organisms which is the ordinate, with the second bar representing the data introduced for comparison. However, > 7% S between a group of one data set and any group of the other data set is required for the groups to be graphed. For situations where %S = < 70, the computer prints a message, without graphing the data.

The latter subroutine has an additional option. That data set containing the larger number of groups usually comprises the abscissa, but the option is available to specify which data set is to be the abscissa. In either case, X- and Y-axis labels, title, and legend are printed by the computer so that the opportunity for confusion or error on the part of the operator is minimized. Options also exist for cross-hatching the second bar graph set for printing appropriate legends.

The bacterial strains used in the study were isolated from three areas of Chesapeake Bay in Maryland. Surface waters and sediments were collected and, from the samples taken, a total of 43 cultures was purified and subjected to taxonomic analysis. Fifteen strains were from Eastern Bay samples collected 1 July 1970, 18 were from the Rhode River samples collected 21 July 1970, and 10 were from Marumsco Bar samples collected 20 August 1970. The strains are listed in Table 1.

In addition to the 43 fresh isolates, known strains were also included in the analyses. The strain labels used in this study are given in parentheses: Escherichia coli (EC); Pseudomonas sp. (OS); Achromobacter fischeri (AF); Achromobacter sp. (263); Vibrio marinus strain PS-207 (207); Vibrio parahaemolyticus (S3); and Micrococcus sp. (234). Four additional cultures, presumptively identified as V. parahaemolyticus strains and isolated in the course of this study, are also included, and designated VP1 to VP4. All of these reference strains have been used in other, earlier taxonomic studies (4, 6).

Media employed for initial isolation and routine maintenance of cultures have been described elsewhere (8). The details of morphological, biochemical, and physiological tests used in the taxonomic analyses have been published (8). Methods for coding and computing the taxonomic data have also been published (1-3, 7, 10).

RESULTS AND DISCUSSION

In Table 1, the 54 strains examined in this study and for which a taxonomic analysis was done are listed as arranged by computer, according to %S values obtained from the GTP-2 data processing. The computer-assigned strain number (GRAPH IDEN), original strain number (CLUSTER IDEN), and strain labels (COLS 71-78) are given in the table.

| GRAPH IDEN | CLUSTER IDEN | COLS 71-78 | GRAPH IDEN | CLUSTER IDEN | COLS 71-78 |
|------------|--------------|------------|------------|--------------|------------|
| 1          | 37           | M4         | 28         | 33           | R15        |
| 2          | 24           | R1         | 29         | 28           | R16        |
| 3          | 1            | A2         | 30         | 16           | R1         |
| 4          | 15           | A1G1       | 31         | 31           | R19        |
| 5          | 42           | M9         | 32         | 32           | R20        |
| 6          | 3            | A4         | 33         | 13           | L1         |
| 7          | 22           | R8         | 34         | 2            | A3         |
| 8          | 4            | A7         | 35         | 46           | AF         |
| 9          | 5            | A8         | 36         | 6            | A10        |
| 10         | 10           | B3C        | 37         | 7            | A11        |
| 11         | 26           | R13        | 38         | 14           | L4         |
| 12         | 27           | R14        | 39         | 29           | R17        |
| 13         | 23           | R9         | 40         | 30           | R18        |
| 14         | 8            | A16        | 41         | 53           | VP3        |
| 15         | 49           | 234        | 42         | 54           | VP4        |
| 16         | 40           | M7         | 43         | 21           | R6         |
| 17         | 50           | 263        | 44         | 17           | R2         |
| 18         | 44           | EC         | 45         | 9            | A17        |
| 19         | 38           | M5         | 46         | 41           | M8         |
| 20         | 20           | R5         | 47         | 52           | VP2        |
| 21         | 18           | R3         | 48         | 51           | VP1        |
| 22         | 43           | M10        | 49         | 47           | S5         |
| 23         | 39           | M6         | 50         | 48           | 207        |
| 24         | 35           | M2         | 51         | 12           | BEY        |
| 25         | 36           | M3         | 52         | 11           | B5         |
| 26         | 25           | R12        | 53         | 19           | R4         |
| 27         | 45           | OS         | 54         | 34           | M1         |

The 54 strains studied are listed in the order arranged by the computer, i.e., those strains sharing high similarity are adjacent, or placed close to one another. New strain numbers assigned by the GRAPH IDEN subroutine, the original strain sequence numbers, i.e., CLUSTER IDEN, and strain labels, COLS 71-78, are given as produced by the computer. Strain numbers prefixed by M were isolated from samples of water and sediment collected in Marumsco Bar on the Maryland Eastern Shore on 20 August 1970, water temperature 28 to 29°C. Strain numbers prefixed by R were isolated from samples of water and sediment collected in the Rhode River in Chesapeake Bay on 21 July 1970, water temperature 26 to 27°C. Strain numbers prefixed by A, B, or L were isolated from water and sediment samples collected in Eastern Bay on the Eastern Shore of Chesapeake Bay on 1 July 1970, water temperature 25°C. EC, Escherichia coli; OS, Pseudomonas sp.; AF, Achromobacter fischeri; 263, Achromobacter sp.; 207, Vibrio marinus strain PS-207; S3, Vibrio parahaemolyticus; 234, Micrococcus sp.; and VP1, VP2, VP3, VP4, Vibrio parahaemolyticus isolated in this study.
The computer S values, in the form of a similarity triangle produced by the computer, are given in Fig. 1. The %S value data for all the strains tested are thus presented for examination. From the similarity triangle, a subroutine, JDO, notes the groups of strains according to a pre-set of %S value. Eleven groups, as indicated by brackets in Fig. 1, fulfill the %S = >70 requirement.

Table 2 provides a portion of the computer printout, which shows four of the eleven groups observed in the analysis. Computer-assigned strain identification numbers, as indicated on the axes of the similarity triangle in Fig. 1 and

| Table 2. Four bacterial groups identified by scanning the similarity triangle shown in Fig. 1a |
| Group 1 |
| Strain # | Organism |
| 6 | A4 |
| 7 | R8 |
| 8 | A7 |
| 9 | A8 |
| 10 | B3C |
| 11 | R13 |
| 12 | R14 |
| 13 | R9 |
| 14 | A16 |
| Median organism for group 1: |
| 00005001000000100033333100013331300500001000001050500000000005500100033 |
| 1111133111111111111111010333333333331111050000103333333315333359113333333 |
| 33310103333333333333333333333333 |
| Group 5 |
| Strain # | Organism |
| 27 | OS |
| 28 | R15 |
| 29 | R16 |
| 30 | R1 |
| 31 | R19 |
| 32 | R20 |
| Median organism for group 5: |
| 00100100550000010003333305010333331100000010000101000010050000000000500551005033 |
| 11111331111111111111151010333333303311510000150533333333115333313513333333 |
| 33331010033333333333333333333333 |
| Group 6 |
| Strain # | Organism |
| 33 | L1 |
| 34 | A3 |
| 35 | AF |
| 36 | A10 |
| 37 | A11 |
| 38 | L4 |
| Median organism for group 6: |
| 001001005500000100033333100133333100100000100001001000010000000000000500505333 |
| 331333311111111111111510505050103333333003311510001133333333115333313513333333 |
| 33333501000333333333333333333333333 |
| Group 10 |
| Strain # | Organism |
| 47 | VP2 |
| 48 | VP1 |
| 49 | S3 |
| Median organism for group 10: |
| 333333333333333333333333333333333333333333333333333333333333333333333333333333333 |
| 333333333333333333333333333333333333333333333333333333333333333333333333333333333 |
| a The strains comprising the groups demonstrated > 70% intragroup S-values. The strain labels are given, as well as the computer-assigned strain identification numbers. The coded taxonomic data arrayed to represent the hypothetical median organism (HMO) for each of the four groups are also shown.

FIG. 1. Computer output consisting of a triangular array of %S values computed using 210 characters scored for each of the 54 strains examined. The symbols shown represent %S between each of the 2,862 possible comparisons as follows: 0 = < 50%, 1 = 51 to 55%, 2 = 56 to 60%, 3 = 61 to 65%, 4 = 66 to 70%, 5 = 71 to 75%, 6 = 76 to 80%, 7 = 81 to 85%, 8 = 86 to 90%, 9 = 91 to 95%, * = 96 to 100%. Numbers at the left and bottom of the triangle indicate the strain numbers assigned by the computer (see Table 1).
in Table 1, as well as strain labels, are given for each of the strains comprising the four groups.

Nine strains were found to comprise group 1, including organisms isolated from two locations in Chesapeake Bay, the Rhode River (R strains), and Eastern Bay (A and B strains). All nine strains were from sediment samples. The strains making up group 5 included five Rhode River strains, and these were found to be similar to the Pseudomonas sp. (OS) included as a reference strain. The group, therefore, is considered to be Pseudomonas. Group 6 included the luminescent organism Achromobacter fischeri (AF), as well as two isolates (L1 and L4) from Eastern Bay, also luminescent. Strains A3, A10, and A11, from Eastern Bay samples, did not demonstrate luminescence under laboratory conditions. However, they may prove to be dark mutants or nonluminescent. Nevertheless, they are related at a significant %S level to other members of this group. Group 10 included two strains (VP1 and VP2) identified as V. parahaemolyticus, and the group also includes a known strain of V. parahaemolyticus (S3). The identification of the two isolates as V. parahaemolyticus was thereby confirmed by this computer analysis of the taxonomic data when the known strain of V. parahaemolyticus was added for purposes of comparison. Thus, reference strains are valuable in classification of the microbial groups obtained from natural samples. The inclusion of known strains is also useful in demonstrating the absence of given species in a sample. Only one of the 53 cultures (M6), a member of cluster 4, showed >65%S with Escherichia coli (see strain 18 in Fig. 1) which confirmed the low incidence of coliforms in the samples examined.

The HMOs calculated for each of four groups are given in Table 2. The code numbers, 0, 1, 3, and 5, indicate the absence, presence, not tested, or not compared test result for the 210 characters used in the study and arranged as HMO. From the list given, those characters describing the strains comprising each of the groups can be identified, and a description of the groups can be prepared. The characters describing each of the groups can be readily tabulated by computer but, for reasons of brevity, they have not been included in this paper.

A portion of the printout of comparisons calculated for each strain of the groups with the HMO for that group is indicated as %S in Table 3. From the list given in Table 3, a representative strain for each group is selected. It is obvious that, for Group 1, strains A4, A8, and R13 are equally suitable for use as representa-

| Medium organism of group no. | No. organisms in group | Group organisms | Percent similarity |
|-----------------------------|------------------------|-----------------|-------------------|
| 1                           | 9                      | A4              | 91                |
|                             |                        | R8              | 88                |
|                             |                        | A7              | 88                |
|                             |                        | A8              | 91                |
|                             |                        | B3C             | 90                |
|                             |                        | R13             | 91                |
|                             |                        | R14             | 89                |
|                             |                        | R9              | 90                |
|                             |                        | A16             | 88                |
| 5                           | 6                      | OS              | 97                |
|                             |                        | R15             | 90                |
|                             |                        | R16             | 90                |
|                             |                        | R1              | 85                |
|                             |                        | R19             | 89                |
|                             |                        | R20             | 84                |
| 6                           | 6                      | L1              | 91                |
|                             |                        | A3              | 93                |
|                             |                        | AF              | 100               |
| 10                          | 3                      | VP2             | 100               |
|                             |                        | VP1             | 100               |
|                             |                        | S3              | 100               |

Table 3. Portion of the printout showing comparisons of each group HMO with member strains of the group. The %S values are used to select actual strains to represent the groups.

The graph obtained from data given in Table 2 is shown in Fig. 2. It is evident from the graph that group 1 was predominant, representing nine strains. Groups 2, 3, 8, 9, and 11, on the other hand, contained only two strains each and are present in relatively lower numbers in the samples examined. Thus, a rapid visual comparison of the major groups present in the samples is permitted (Fig. 2), as well as a quick determination of the number of groups occurring in the samples tested.

A printout of comparisons of the seven HMOs called up from the data bank, with the HMOs from the 11 groups identified from the data in Fig. 1, is given in Table 4. In this table, set 1 represents the 11 group HMOs identified in Fig. 1, and set 2 represents those introduced into the
analysis for comparison. Set 2 represents HMOs for seven groups from a computer analysis of 35 cultures isolated from the hemolymph of blue crabs (*Callinectes sapidus*) captured in Marum sco Bar at the same time and at the same location as the water and sediment samples examined in this study were collected.

It is evident (Table 4) that the nine strains comprising group 1 were related (S = 74%) to the group 2 strains isolated from crab hemolymph, as well as to the group 4 strains. Similarly, group 5 was found to be related to groups 2 and 4 (S = 78%). The luminescent strains comprising group 6 were found to be most closely related to the strains in group 4 (S = 82%). Thus, the same taxonomic groups of bacteria were present in water, sediment, and crabs and may represent a significant portion of the normal flora. Strains of group 10 (*V. parahaemolyticus*) were most closely related (S = 79%) to the group 7 strains (79.9%) from crab hemolymph. Group 10 included no strains actually isolated from the three areas in Chesapeake Bay sampled in this study but were strains identified and classified as *V. parahaemolyticus* in the course of other analyses. *V. parahaemolyticus* can be isolated from blue crabs and, in fact, have been the causative agent in food-poisoning cases in which contaminated crab meat was implicated.

The figures given in parentheses in Table 4 represent the number of coded tests involved in the computations whereby the several groups of the data sets have been compared. This information is helpful in assessing the statistical validity of the results.

Figure 3, a double bar graph prepared from data presented in Table 4, compares taxonomic groups from Chesapeake Bay water and sediment (set 1) with the seven taxonomic groups present in Chesapeake Bay blue crab hemolymph (set 2). The groups of data, set 2, are indicated on the X-axis, with double bars drawn only if groups from Fig. 1 correlated with the set 2 groups at S = >70%. The number of strains in each of the crab hemolymph bacterial groups are represented in the graph by the outer bars, with those identified in Fig. 1 by the inner bars. It is evident from Fig. 3 that a number of bacterial groups isolated from water and sediment samples were similar, on the basis of the HMO comparisons, with the crab bacterial groups. Striking, however, was the lack of correlation, at S = >70%, of any of the set 1 groups

![Figure 2](image)

**Fig. 2.** Bar graph illustrating the group and number of strains comprising each group. These are identified by computer from the similarity triangle given in Fig. 1. The height of the bar indicates the number of strains in each group.

| Set 1 groups | % Similarity between medians of set 1 and set 2 (with tests compared) |
|--------------|-------------------------------------------------------------------|
| Set 1 groups | 1*                  | 2       | 3       | 4       | 5       | 6       | 7       |
| 1            | 68 (97)             | 74 (97) | 51 (97) | 73 (97) | 66 (97) | 70 (97) | 67 (97) |
| 2            | 64 (33)             | 73 (33) | 64 (33) | 85 (33) | 61 (33) | 58 (33) | 67 (33) |
| 3            | 71 (95)             | 73 (96) | 54 (96) | 78 (95) | 73 (95) | 71 (95) | 65 (95) |
| 4            | 68 (97)             | 70 (97) | 51 (97) | 69 (97) | 65 (97) | 67 (97) | 61 (97) |
| 5            | 73 (97)             | 78 (97) | 56 (97) | 78 (97) | 75 (97) | 74 (97) | 63 (97) |
| 6            | 74 (93)             | 74 (93) | 57 (93) | 82 (93) | 72 (93) | 72 (93) | 65 (93) |
| 7            | 74 (100)            | 71 (100) | 53 (100) | 74 (100) | 73 (100) | 76 (100) | 67 (100) |
| 8            | 72 (92)             | 74 (92) | 49 (92) | 75 (92) | 74 (92) | 77 (92) | 67 (92) |
| 9            | 70 (93)             | 72 (93) | 55 (93) | 75 (93) | 69 (93) | 70 (93) | 63 (93) |
| 10           | 73 (33)             | 64 (33) | 64 (33) | 73 (33) | 76 (33) | 76 (33) | 79 (33) |
| 11           | 71 (89)             | 69 (89) | 55 (89) | 81 (89) | 69 (89) | 67 (89) | 70 (89) |

* Set 2 groups.
with group 3 of set 2, i.e., the crab isolates. Group 3 of the crab isolates was the largest group, comprising 17 strains, isolated from crab hemolymph in this study. Similarly, group 1 of the water and sediment isolates, the largest group of that population, was not found to be the most closely related group to any of the crab bacterial groups. Thus, the logical conclusion to be drawn from this analysis is that the predominant bacterial group occurring in Chesapeake Bay water and sediment differs from that found in Chesapeake Bay blue crab hemolymph.

Clearly, the computer program described in this paper has numerous applications in microbial taxonomy and ecology, with potential value in medical, food, and industrial microbiology. Changes occurring in bacterial populations in a given environment can be monitored over any time period, or different environments can be compared to determine whether the microbial populations are similar or different. Seasonal fluctuations in bacterial populations can be monitored, i.e., data for samples collected in an estuary or river in the winter months can be analyzed by computer, HMOs can be generated, and the information can be maintained in the computer data bank or on punched cards in a file. Sample analyses carried out at a later time, namely, in spring, summer, and fall, would be similarly processed, with successive data sets and HMO files added to the data bank. If any of the taxonomic groups from successive seasonal samples are related to groups of other seasonal samplings, the quantitative and qualitative population changes in the microflora can be assessed. Also, indicator organisms, such as Escherichia coli, can be included to monitor fluctuations in these populations with time, season, temperature, etc. Thus, identification and classification of strains can be readily achieved, with the accumulated data forming a valuable strain data library.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grant No. GB-35281X, Sea Grant Project No. 04-3-158-7, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and Contract No. N00014-67-A-0229-0027 between the Office of Naval Research and the University of Maryland. We wish to acknowledge the helpful guidance of Giuliano Gnugnioli, Georgetown University Computation Center, in writing the computer programs and Thomas Wicks for use of the taxonomic data obtained by him in his studies of crab hemolymph.

LITERATURE CITED

1. Colwell, R. R. 1968. Polyphasic taxonomy of bacteria. p. 421-436. In H. Iizuka and T. Hasegawa (ed.), Culture collections of microorganisms. Proc. Int. Conf. Culture Collections, Tokyo. University of Tokyo Press.
2. Colwell, R. R. 1969. Numerical taxonomy of the Flexibacter. J. Gen. Microbiol. 58:207-215.
3. Colwell, R. R. 1970. Numerical analysis in microbial identification and classification. Develop. Indust. Microbiol. 11:154-160.
4. Colwell, R. R. 1970. Polyphasic taxonomy of the genus Vibrio: Numerical taxonomy of Vibrio cholerae, Vibrio parahaemolyticus, and related Vibrio species. J. Bacteriol. 104:410-433.
5. Colwell, R. R., and J. Liston. 1960. Taxonomic relationships among the pseudomonads. J. Bacteriol. 82:1-14.
6. Colwell, R. R., T. E. Lovelace, L. Wan, T. Kaneko, T. Staley, P. K. Chen, and H. Tubiash. 1973. Vibrio parahaemolyticus—Isolation, identification, classification, and ecology. J. Milk Food Technol. 36:202-213.
7. Colwell, R. R., M. L. Moffett, and M. D. Sutton. 1968. Computer analysis of relationships among phytopathogenic bacteria. Phytopathology 58:1207-1215.
8. Colwell, R. R., and W. J. Wiebe. 1970. "Core" characteristics for use in classifying aerobic, heterotrophic bacteria by numerical taxonomy. Bull. Georgia Acad. Sci. 28:165-185.
9. Kaneko, T., and R. R. Colwell. 1973. Ecology of Vibrio parahaemolyticus in Chesapeake Bay. J. Bacteriol. 113:24-32.
10. Lockhart, W. R., and J. Liston (ed.). 1970. Methods for numerical taxonomy. American Society for Microbiology, Washington, D.C.
11. Sneath, P. H. A. 1967. Application of computers to bacterial taxonomy. J. Gen. Microbiol. 17:201-226.
12. Sokal, R. R., and P. H. A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman Co., San Francisco.