APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 1991, p. 3171–3177 Vol. 57, No. 11
0099-2240/91/113171-07$02.00/0
Copyright © 1991, American Society for Microbiology

Biochemical Fingerprinting of Water Coliform Bacteria, a New Method for Measuring Phenotypic Diversity and for Comparing Different Bacterial Populations

I. KÜHN,[1,2]* G. ALLESTAM,[2] T. A. STENSTRÖM,[2] AND R. MÖLLBY[1]

Department of Bacteriology, Karolinska Institute, S-104 01 Stockholm,[1] and Department of Bacteriology, National Bacteriological Laboratory, S-105 21 Stockholm,[2] Sweden

Received 17 May 1991/Accepted 13 August 1991

A simple, automated microplate system for biochemical characterization of water isolates can be used to obtain fingerprints of the bacterial flora from various water samples. Mathematical models for calculating the diversities and similarities between bacterial populations are described for such fingerprints. The diversity may give information on whether an indigenous or allochthonous flora is present, and the similarities between bacterial populations, as calculated by using a population similarity coefficient (S_p), may indicate contaminations between different water samples. The system was demonstrated on coliform bacterial populations from various water samples, with or without suspected intercontamination. For unrelated water samples, the S_ps were close to 0, whereas repeated samples of the same source showed S_ps of 0.64 to 0.74. The S_p values from several water samples were also clustered to form a dendrogram, thus indicating the relative similarities between the bacterial populations to confirm suspected common sources of pollution.

Species of several genera within the family of *Enterobacteriaceae* are arbitrarily grouped as coliform bacteria (6). Bacteria belonging to this group are isolated on selective media at defined incubation temperatures, and the detection of acid, gas, and aldehydes produced from the metabolism of lactose is used as the main identification criterion for the group. Different medium formulations and isolation procedures are currently used, and this may affect the selection of strains (10).

In water microbiology, the main purpose of quantifying coliform bacteria is to detect fecal pollution and thus the possible presence of fecal pathogens. The occurrence of *Escherichia coli* in such investigations is regarded as evidence of fecal contamination, since this species is a common inhabitant of the intestines of humans and animals but very seldom can multiply in the water environment (19). However, *E. coli* may be more susceptible than true fecal pathogens to chlorination or environmental stress factors. Therefore, all species of *Enterobacteriaceae* are often identified as one group, the total coliform bacteria. The amount of such coliform bacteria in drinking water is used as the main criterion to decide whether the water is suitable for human consumption. However, the coliform bacteria other than *E. coli* may originate from a multitude of other sources, including soil, decaying vegetation, industrial processes, and effluents (5). Some coliform bacteria can also multiply on suitable solid materials (3), or their presence may be due to elevated levels of organic matter in raw water or within drinking-water systems. The growth of *Enterobacter cloacae* has even been used as a bioindicator for the assessment of biodegradable organic matter in drinking water (22).

The existing routine methods for the investigation of fecal and total coliform bacteria in drinking water do not allow directly for separation of the bacteria into species. If further identifications are made, this is usually done with the IMViC test (6) or with commercial identification kits. Subtyping below the species level is sometimes required to identify potential sources of contamination, both in epidemiological investigations and in evaluations of the adverse effects of pollution of a water body, a well, or other types of drinking water sources and treatment systems.

A system for simple and automatic identification of biochemical phenotypes of bacteria (the PhenePlate [PhP] system; BioSys inova, Stockholm, Sweden) has been developed (12, 14); this system is based on evaluation of the kinetics of several biochemical reactions, performed in microtiter plates (biochemical fingerprinting). The reactions have been selected to give a high level of discrimination among strains in a specific group of bacteria. Different sets of reagents have earlier been developed for *E. coli* (13), *Klebsiella* and *Enterobacter* spp. (14), and *Salmonella* spp. (11a). Several epidemiological studies of these species have been performed (15, 17, 26). The system was also used to identify the contaminating source of *Klebsiella* strains in a Swedish river (24).

Recently, a set of reagents suitable for typing of bacteria belonging to the family *Enterobacteriaceae* and related bacteria was developed (the PhP-48 plate). The aims of the present investigation were (i) to demonstrate the specificity and reproducibility of this system for subtyping coliform bacteria isolated from water, (ii) to demonstrate the use of biochemical fingerprinting for studying the diversity of coliform bacterial populations in water, and (iii) to develop simple mathematical models that can be used to describe phenotypic relationships between bacterial populations in different water samples.

MATERIALS AND METHODS

The PhP system. The PhP system for coliform bacteria consists of microtiter plates with dehydrated reagents. In the present study, plates with two sets of 48 reagents were used (PhP-48 plate) (Table 1). The reagents have been selected to give a high level of discrimination among various *Enterobac-*

* Corresponding author.

TABLE 1. Tests used in the PhP-48 system

Test no.	Reagent	SD	Test no.	Reagent	SD
1	Mannoic acid-*d*-lactone	3.8	25	Sorbose	6.2
2	L-Arabinose	1.7	26	Deoxyglucose	2.9
3	D-Xylose	2.2	27	Deoxyribose	5.2
4	Galactose	2.6	28	Rhamnose	4.5
5	Maltose	2.2	29	D-Fucose	2.7
6	Cellobiose	6.4	30	L-Fucose	5.2
7	Trehalose	5.0	31	Tagatose	4.1
8	Palatinose	2.7	32	Amygdalin	2.5
9	Sucrose	6.8	33	Arbutin	6.9
10	Lactose	4.0	34	β-methyl-glucoside	5.1
11	Melibiose	3.8	35	5-Ketogluconate	2.8
12	Lactulose	4.7	36	Gluconate	4.1
13	Gentobiose	6.4	37	Melbionate	5.3
14	Melezitose	4.4	38	Galactolactone	5.5
15	Raffinose	6.3	39	Salicine	6.7
16	Inosine	3.5	40	pH 5.5 control[a,b]	1.4
17	Adonitol	5.5	41	Citrate[b]	7.1
18	Inositol	5.3	42	Fumarate[b]	3.0
19	D-Arabitol	6.3	43	Malinate[b]	3.6
20	Glycerol	2.6	44	Malonate[b]	8.1
21	Maltitol	4.2	45	Pyruvate[b]	3.6
22	Sorbitol	3.8	46	L-Tartrate[b]	3.1
23	Dulcitol	4.2	47	Urea[b]	3.2
24	pH 7.4 control[c]	2.1	48	Ornithine[b]	4.8

[a] Negative control for tests 41 through 48.
[b] Reagent giving an alkaline reaction when positive.
[c] Negative control for tests 1 through 39.

teriaceae strains and related species. The general preparation method for plates was described earlier (12).

The bacteria were suspended into a medium containing 0.1% proteose peptone (Difco) and 0.01% bromothymol blue. Aliquots of 0.15 ml of the bacterial suspensions were added to each of 48 wells of a microtiter plate containing the dehydrated reagents. The plates were kept in a refrigerator overnight and then incubated on the following morning at 35°C. The A_{620} of each well was read after 7, 24, and 48 h with a microtiter plate reader (Titertek Multiskan; Flow Laboratories). The values were automatically transferred to a microcomputer, multiplied by 10, and stored in the computer as integer values, yielding a score ranging from 0 to 30 for each test. After the last reading, the mean value from all three readings of each well was calculated, yielding a biochemical fingerprint of each isolate consisting of 48 numbers, each one ranging from 0 (yellow, acid reaction) to 30 (blue, alkaline reaction). Readings after 72 h were also performed, but they did not yield any further information and were therefore excluded from the analysis.

Calculations. The biochemical fingerprints of all isolates were compared pairwise, and the similarity between each pair of strains was calculated as the correlation coefficient (r). The inter-assay reproducibility was calculated from duplicate assays on different occasions of 50 coliform reference isolates obtained from the Swedish National Bacteriological Laboratory, and the intra-assay reproducibility was calculated from duplicate assays on the same occasion of 10 isolates from the same source as above. The identity level was calculated as the mean of the correlation coefficients obtained from multiple assays of the same isolate, subtracted by two standard deviations (SD) of this mean (15), yielding a 95% confidence level. Isolates with coefficients of correlation to each other higher than the identity level were assigned to the same biochemical phenotype (BPT).

The discrimination power of the PhP-48 system was calculated as described previously (13, 14). Two different ways of measuring the discrimination power were used. (i) The resolution index was calculated as 1 minus the mean of all correlation coefficients obtained when all isolates within the studied population were compared. The resolution index is a measure of the discriminating capacity of the used set of biochemical reactions for a population of isolates. It may also be used as a measure of the metabolic diversity of a population of bacterial strains. A low resolution index indicates that the population consists of strains that are similar to each other, whereas a high resolution index means a heterogeneous population containing many different groups of bacteria. (ii) The diversity of a bacterial population was measured by using Simpson's index of diversity (D_i) according to the formula $D_i = 1 - \Sigma[N_i \times (N_i - 1)]/[N \times (N - 1)]$, where N_i equals the number of isolates assigned to the ith BPT and N is the total number of isolates assayed (8). D_i depends on the distribution of isolates into different types. If a population consists of isolates evenly distributed into many different types, the D_i will be high, reaching a maximum value of 1.0 if all isolates are different from each other; if the population contains some dominating types, the D_i will be low, reaching a minimum value of 0 if all isolates are identical.

The similarity between the bacterial populations in two samples (the population similarity coefficient [S_p]) was defined as $(Sx + Sy)/2$, where Sx is the similarity of population x in population y and Sy is the similarity of population y in population x. Sx was calculated in the following way: N_x is the total number of isolates sampled from population x, i is the isolate number in population x (from 1 to N_x), p_{xi} is the proportion of isolates identical to isolate i (including isolate i) in population x, p_{yi} is the proportion of isolates identical to isolate i in population y, and q_{xi} is the proption of isolates identical to isolate i in populations x divided the proportion of isolates identical to isolate i in population y. q_{xi} is calculated for each isolate in population x according to the formula $q_{xi} = p_{xi}/p_{yi}$. However, this quotient should always be less than 1; if q_{xi} is greater than 1, then $q_{xi} = 1/q_{xi}$. The total similarity of population x in population y is then calculated as $Sx = \Sigma q_{xi}/N_x$. Sy is calculated in the same way. An example of the calculation of S_p is shown in Table 2.

S_p thus depends on the proportion of isolates showing identity between the two compared populations. When two bacterial populations with no common phenotypes are compared, S_p will always be 0. When two populations containing some identical isolates are compared, S_p will increase when the number of strains identical between the two populations increases, reaching a maximum value of 1.0. When the S_p values are calculated by the computer, each isolate is compared with all other isolates in the two populations, and the pairwise correlation coefficients are calculated. When the correlation coefficient to any other isolate exceeds the chosen identity level, the isolates are regarded as identical. Thus, when the S_ps for samples from different assays were calculated, the inter-assay identity level (0.965) was used for estimates of identical strains between the two populations, whereas the intra-assay identity level (0.973) was used for calculations of S_p values within a population and for comparisons of populations from the same assay. Comparisons of several bacterial populations yielded a matrix of S_p coefficients that was clustered according to the UPGMA method (23).

All data, including optical readings, calculations of correlation coefficients, diversity indices, S_p values, and cluster-

TABLE 2. Calculation of S_p as examplified by comparisons of two hypothetical populations[a]

Parameter	Calculation of Sx (sample no. in population x)					Parameter	Calculation of Sy (sample no. in population y)				
	1	2	3	4	5		1	2	3	4	5
p_{xi}	2/5	2/5	1/5	1/5	1/5	p_{yi}	1/5	3/5	3/5	3/5	1/5
p_{yi}	1/5	1/5	3/5	0	0	p_{xi}	2/5	1/5	1/5	1/5	0
q_{xi}	1/2	1/2	1/3	0	0	q_{yi}	1/2	1/3	1/3	1/3	0

[a] Contents of population x: AABCD ($N_x = 5$). Contents of population y: ABBBF ($N_y = 5$). $S_p = 17/60 = 0.28$ $(Sx + Sy)/2$. For Sx, $\Sigma q_{xi} = 8/6$ and $Sx = 8/30$ $(\Sigma q_{xi}/N_x)$. For Sy, $\Sigma q_{yi} = 9/6$ and $Sy = 9/30$ $(\Sigma q_{yi}/N_y)$.

ing and printing of dendrograms, were handled with the PhP software (BioSys inova).

Isolation and assay of bacterial strains. A total of 360 coliform bacteria were isolated from 18 different water samples, collected from different sources (drinking water, surface water, and wastewater) (Table 3) by using the membrane filtration method (6), and incubated on LES-Endoagar (Difco Laboratories, Detroit, Mich.) at 35°C for 48 h. Between 10 and 34 oxidase-negative colonies from each water sample where randomly picked from the membrane filters and grown on nonselective nutrient agar. Each isolate was then suspended in the medium, inoculated into the PhP-48 plates, and incubated for 48 h as described above. After the last reading, calculations were automatically performed.

RESULTS

PhP-48 system for coliform water isolates. In Table 1, the resolution capacity for each one of the 48 tests in the PhP-48 system is given as an SD. The SD values were calculated from the test results for the 360 isolates. All tests, except for test 2 (L-arabinose), gave SD values higher than those of the corresponding negative control tests, which means that different isolates differ in their reactions and thus that the tests can discriminate between coliform water isolates.

The mean intra-assay reproducibility, as calculated from simultaneous duplicate assays of 10 strains, was 0.987, and the mean inter-assay reproducibility, as calculated from duplicate assays on different occasions of 50 strains, was 0.983, yielding intra- and inter-assay identity levels of 0.973 and 0.965, respectively (Table 4).

When all isolates from all samples in Table 3 were compared with each other, the total number of BPTs was 185 (Table 3) and the diversity index among 100 isolates randomly sampled from the above 360 isolates was 0.995, indicating that the PhP-48 plates yield high-level discrimination among random coliform water isolates. The resolution index, as calculated from comparisons of all 185 BPTs was 0.512, further indicating that the discrimination capacity of the tests in the PhP-48 plate was high.

Phenotypic diversity of coliform bacterial populations from water. Table 3 shows the number of BPTs identified, the diversity indices, and the resolution indices of 18 bacterial populations. Very high diversities (above 0.97) were found in raw water (samples 8, 9, and 11). High diversities were also found in wastewater infiltration tanks (samples 10A, 10B, 12A, and 12B). As expected, the total bacterial numbers were high in these samples. However, some samples with low bacterial numbers showed a high diversity, e.g., samples 2, 5, and 7A (all from wells). Very low diversities were found in all samples from source 4 (a well and a connected tap).

The total number of BPTs identified within the 18 water samples was 185 (Table 3). The sum of the number of BPTs

TABLE 3. Water samples studied[a]

Sample no. and source	N_{tot}	N	N_{BPT}	D_i	Res
Well water					
1, Dug well	1,270	27	14	0.886	0.247
2, Drilled well	190	34	24	0.966	0.511
3A, Flooded basement	50	12	6	0.818	0.250
3B, Contaminated well	110	12	9	0.939	0.313
4A, Drinking water from tap	700	14	3	0.385	0.203
4B, Drinking water from well	340	14	5	0.593	0.298
4C, Drinking water from well	400	32	5	0.387	0.200
5, Drinking water from well	35	17	14	0.978	0.474
6, Drinking water from well	75	17	12	0.934	0.461
7A, Drinking water from well	82	17	12	0.956	0.212
7B, Drinking water from well	1,440	17	9	0.868	0.188
Raw water and surface water					
8, Lake surface water	1,500	34	30	0.991	0.403
9, Groundwater	1,300	24	23	0.997	0.419
10A, Wastewater treatment step 1	>1,000,000	10	7	0.911	0.235
10B, Wastewater treatment step 2	>1,000,000	19	4	0.696	0.193
11, Lake surface water	3,000	10	10	1.000	0.310
12A, Wastewater infiltration	>1,000,000	28	14	0.920	0.332
12B, Wastewater infiltration	>1,000,000	22	19	0.987	0.343

[a] N_{tot}, total number of coliform bacteria per 100 ml; N, number of isolates typed with the PhP-48 system; N_{BPT}, number of biochemical phenotypes identified in each sample; D_i, diversity index; Res, resolution index within each sample.

TABLE 4. Reproducibility of the PhP-48 system for coliform water isolates

Comparison	No. of assays[a]	Reproducibility[b]	SD	Identity level[c]
Inter-assay	50	0.983	0.009	0.965
Intra-assay	10	0.987	0.007	0.973

[a] Assays were done in pairs.
[b] Calculated as mean correlation coefficient of duplicate assays.
[c] Calculated as reproducibility − 2 standard deviations.

within the water samples was 220, and thus only 35 BPTs in the total material were found in more than one sample. When a certain BPT was found in more than one water sample, these samples were usually suspected to be related. Thus, it seems that the coliform bacterial population in a water sample mainly consists of BPTs that are unique to that particular environment.

Comparisons of isolates from different water samples. The population similarity coefficients (S_p) between unrelated water samples were calculated (Table 5) and were usually found to be low (mean value, 0.01; SD, 0.02; maximum value, 0.13). In contrast, Table 6 shows the data obtained from comparisons between samples where a relationship is known or suspected. Samples 4A and 4B, from a well and from a tap connected to this well (Table 4), contain many identical strains and show an S_p as high as 0.74 to each other. Sample 4C was collected from the same well as samples 4A and 4B but 4 months later, and it still shows the same high S_p when compared with the water sampled earlier. Samples 10A and 10B from different sites in a wastewater treatment plant also contain similar populations. Sample 3B originated from a contaminated well and sample 3A originated from a flooded basement, where well 3B was a possible source for the flooding. The high population similarity coefficient (S_p = 0.28) strongly indicated a spread of bacteria between these two sampling points.

Samples 7A and 7B were derived from two wells in the same area. These wells had earlier shown high numbers of coliform bacteria, and the water had been declared unsuitable for human consumption. Samples 12A and 12B were derived from two different infiltration plants in the same area as the source of samples 7A and 7B, and it was suspected that one or both was the contaminating source. The S_p value between the two wells was high, and water from the infiltration plant 12A showed very high S_p values when compared

TABLE 6. Comparisons of presumably related water samples

Sample	S_p
4A − 4B	0.74[a]
4AB − 4C	0.64
10A − 10B	0.38[a]
3A − 3B	0.28[a]
7A − 7B	0.22[a]
7A − 12A	0.16[a]
7A − 12B	0.00[a]
7B − 12A	0.29[a]
7B − 12B	0.00[a]
12A − 12B	0.00[a]

[a] Calculated from intra-assay data (identity level, 0.973).

with water from both wells (Table 6), indicating that the former plant is the contaminating source.

The bacterial populations from all 18 water samples were compared, and a matrix of S_p values was calculated. This matrix was then clustered according to the UPGMA method. The resulting dendrogram is shown in Fig. 1, which clearly shows that samples 7A and 7B from adjacent wells were similar to a sample from one of the suspected contaminating wastewater infiltration sites (sample 12A) but not to a sample from the other site (sample 12B). Likewise, the similarities between the other related samples described above are shown in the figure.

DISCUSSION

A quantitative measurement of the amount of coliform bacteria in water samples is the basis for most risk assessments in relation to the spread of fecal pathogens. Such a quantification does not normally indicate the possible source of the contamination, even if multiple potential population sources are analyzed. When the same species is dominant both in the water sample and in a potential source of contamination, a relationship may be suspected, but since many coliform bacterial species are normally found in fecal pollution sources a clear connection between a sample and a contamination source is seldom established. When a waterborne outbreak of bacterial disease has occurred and the etiological agent has been isolated, a multitude of different techniques, such as serotyping, phage typing, resistotyping, protein or DNA fingerprinting, etc., may be used for subtyping isolates to confirm the epidemiological relations between the outbreak and the causative agent. However, typing

TABLE 5. Similarity matrix obtained from comparison of independent water samples[a]

Sample no.	S_p with sample:									
	1	2	3AB	4AB	5	7A	8	9	10	11
2	0.01									
3AB	0.00	0.13								
4AB	0.00	0.00	0.00							
5	0.00	0.00	0.00	0.00[b]						
6	0.00	0.00	0.00	0.00[b]	0.07[b]					
7A	0.00	0.00	0.00	0.00	0.00	0.00				
8	0.08	0.05	0.00	0.00	0.00	0.02	0.00			
9	0.04	0.03	0.00	0.00	0.01	0.00	0.00	0.03		
10	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00	
11	0.00	0.01	0.00	0.02[b]	0.00[b]	0.00[b]	0.00	0.00	0.00	0.00

[a] Values are presented as population similarity coefficients (S_p). When not otherwise indicated, samples were assayed on different occasions, and thus the S_p was calculated from the inter-assay identity level (0.965).
[b] Calculated from intra-assay data (identity level, 0.973).

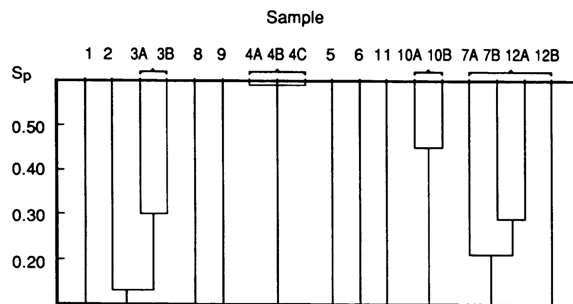

FIG. 1. Dendrogram derived from UPGMA clustering of population similarity coefficients (S_ps) obtained from comparisons of coliform bacterial isolates from 18 different water samples. All values were calculated from the inter-assay identity level.

below the species level is seldom used as a tool to indicate possible pollution sources.

The PhP system for biochemical fingerprinting has been used for subtyping various species belonging to the family *Enterobacteriaceae*. In previous studies, special sets of tests that were specific for the studied species were used (13, 14, 16). In the present study, and in contrast to other diagnostic kits for bacterial identification (e.g., API, MicroID) where the tests are selected to yield a low intraspecies differentiation but a maximum interspecies differentiation, the tests in the PhP-48 plate were selected to give a maximum differentiation among all strains belonging to the family of *Enterobacteriaceae* and related genera, regardless of species. The reproducibility of the PhP-48 system was high for both inter- and intra-assay analyses (Table 4), and the discriminatory power for independent coliform water isolates was also very high with this test set. Therefore the PhP-48 system can be used for identifying biochemical phenotypes among a wide range of coliform species. Our data indicate that it may also be used for simultaneous subtyping and species identification of coliform water isolates (unpublished data). The PhP system for biochemical fingerprinting may therefore be useful both in investigations within species and on a population basis. Stenström and Kühn (24) used it to show a relationship between an industrial outlet and the contamination with *Klebsiella pneumoniae* of the intake water to a drinking water treatment plant 10 miles downstream. It was then shown that the same biochemical phenotype of *K. pneumoniae* that was dominant in the industrial outlet also was dominant in intake water to the drinking water treatment plant and that this phenotype was not found in any of the other possible contaminating sources assayed.

In ecological investigations of microorganisms, the phenotypic or metabolic diversity of a microbial population is often described by quantifying the number of types found in the population or by compiling tables showing the reactions for isolates from various samples (1, 4). The Shannon diversity index has been used as a quantitative diversity measure in several investigations (2, 25). Simpson's diversity index, used in the present study, has been used for comparing the levels of discrimination of various typing systems (8), but it is also an excellent measure of the phenotypic diversity within a microbial population (9). It is simple to calculate and may be used on all kinds of data.

The diversity indices for coliform bacterial populations in some water samples from the present study are shown in Table 3. High total coliform counts were often but not always associated with a high diversity index. This association was found in most potential contaminating sources. Samples with a high bacterial count but with a low diversity index may contain many autochthonous bacteria (bacteria that can multiply in the water source). This was probably the case with samples 4A and 4B and was further verified with a second sample from the same well, collected 4 months later and containing a similar bacterial population. Low bacterial counts with a high diversity (e.g., samples 2 and 5) might also indicate wastewater contamination, if the bacterial population from the wastewater has been subject to a high dilution or partial die-off when reaching the well. Since many coliform bacteria may persist for extended time periods in soil and water (7), these results may also be due to a previous contamination with fecal bacteria in the wells.

The resolution indices may give some information on the metabolic diversity of a population. A low resolution index means that the BPTs from a water sample are similar to each other, possibly indicating the presence of only a few species. Sample 7B, for example, has a resolution index of only 0.188, which indicates that the nine BPTs found in that particular sample were similar to each other (Table 3). On the other hand, sample 2, showing a very high diversity index, also showed a very high resolution index. This indicates that the sample contains a bacterial population that is very heterogeneous, with regard to both metabolic activity and BPTs. An interesting feature is that the resolution index for all 185 BPTs found among the 360 isolates studied was higher than the resolution index within any sample. This means that the BPTs found within one sample were more similar to each other than were randomly sampled coliform isolates. One explanation for this finding is that a certain environment is suitable for bacteria belonging to certain groups, whereas other bacteria cannot multiply in that environment. Another explanation is that a bacterial clone that has been in residence for some time in a particular environment may start to undergo changes and diverge into different but related BPTs. Although the tests included in the PhP system have been carefully selected to give stable results, during repeated subculturing or long-term storage of the bacteria (11) the normal evolution will sooner or later induce changes in bacterial strains and make them diverge into different BPTs, as measured with the PhP system (15). The phenotypic stability during environmental stress, e.g., starvation, irradiation, or other sublethal damages, has not been investigated. However, the fact that identical coliform bacteria may be found in one well at different occasions and at different but related sampling sites indicates that at least a large proportion of the coliform bacteria that have been isolated on LES-Endoagar have sufficiently stable biochemical fingerprints under natural circumstances.

When working with samples containing mixed microbial populations, the aim is often to measure the similarities between different populations or to study changes within a population due to various environmental factors. This is often done by quantifying various groups of bacteria and by measuring changes in the bacterial numbers within the various groups. Regression analysis may then be used to describe possible relations between different groups of microorganisms (27), and factor analysis may be used to describe relations between environmental factors and the various groups of microorganisms (28). The minimum percent similarity analysis (21) measures the similarity between bacterial populations by calculating the sum of the minimum percentages of each species that is shared by two populations. The geometric mean similarity index measures the

mean of the percentage of each species that is shared by two populations (18, 20). The population similarity coefficient (S_p), as described in the present study, measures the total similarity of the proportions of each phenotype that is shared between two compared populations. The S_p was constructed to be automatically computed directly from the raw data obtained from a set of biochemical tests without prior identification of different types.

The population similarity coefficient may also be used for comparisons of data obtained from studies performed by methods other than the PhP system. When numerical identification methods have been used, the S_p can be used to calculate results obtained from any kind of data. With other, nonnumerical typing methods that yield distinct types, such as serotyping, biotyping, etc., and when only a species identification has been performed, the S_p can be used to compare the distribution of the types (or species) obtained in different samples.

In the present study, the S_p between bacterial populations from independent water samples was often 0 or close to 0, which means that the samples contained no or only a few identical strains (Table 5). Our other studies also have shown that S_p values above 0.10 are very seldom obtained from independent bacterial populations (data not shown). In contrast, when samples from sources with known or suspected connections were compared, the S_p values were usually higher, with a maximum of 0.74 for samples 4A and 4B (a well and a connected tap). These samples all contained the same dominating phenotype and thus showed low diversities. Laboratory experiments with repeated samplings over 3 days yielded S_p values of 0.42 to 0.52 from an artificial bacterial population containing a mixture of five different coliform bacterial strains and 0.34 to 0.45 from a natural bacterial population with a high diversity (data not shown). Thus, the S_p seems to be an adequate measure of possible relationships between different bacterial populations. Since S_p is a similarity coefficient, like the correlation coefficient obtained from pairwise comparisons between single isolates, it may also be used to form a similarity matrix that can be clustered in the same way as any other similarity matrix, and the dendrogram thus obtained may be used to visualize possible relationships between bacterial populations from different water samples (Fig. 1).

In conclusion, the PhP-48 plate proved to be a practically useful and valuable tool for typing coliform bacteria from water samples. The diversity index and the population similarity coefficient are very simple and useful methods to describe the phenotypic diversity within and the relationships between various microbial populations, respectively. Since the test is simple to perform with prefabricated microtiter plates, and all results are automatically read and computed by a standard personal computer, the system is well suited for studies involving large numbers of isolates. However, the mathematical models used here may well be applied on any other set of data used to characterize the microorganisms in a water sample.

ACKNOWLEDGMENTS

We thank Birgitta Karlsson and the staff at the BVA section, Swedish National Bacteriological Laboratory, for skillful technical assistance.

This work was supported by grants 584/89 L139 and 0777/89 D326 from the Swedish Council for Forestry and Agricultural Research and by Karolinska Institute funds.

REFERENCES

1. **Balkwill, D. L., J. K. Fredrickson, and J. M. Thomas.** 1989. Vertical and horizontal variations in the physiological diversity of the aerobic chemoheterotrophic bacterial microflora in deep southeast coastal plain subsurface sediments. Appl. Environ. Microbiol. **55:**1058–1065.
2. **Bianchi, M. A. G., and A. J. M. Bianchi.** 1982. Statistical sampling of bacterial strains and its use in bacterial diversity measurement. Microb. Ecol. **8:**61–69.
3. **Colbourne, J.** 1985. Materials usage and their effects on the microbiological quality of water supplies, p. 47–59. *In* W. R. White and S. Passmore (ed.), Society for Applied Bacteriology symposium series no. 14, Microbial aspects of water management. Society for Applied Bacteriology, London.
4. **Fredrickson, J. K., D. L. Balkwill, J. M. Zachara, W. L. Shu-Mei, F. J. Brockmann, and M. A. Simmons.** 1991. Physiological diversity and distributions of heterotrophic bacteria in deep cretaceous sediments of the Atlantic coastal plain. Appl. Environ. Microbiol. **57:**402–411.
5. **Geldreich, E. E.** 1990. Microbiological quality of source waters for water supply, p. 3–31. *In* G. A. McFeters (ed.), Drinking water microbiology. Springer Verlag, Berlin.
6. **Greenberg, A. E., R. R. Trussel, and L. S. Clesceri (ed.).** 1989. Standard methods for the examination of water and wastewater, 17th ed. American Public Health Association, Washington, D.C.
7. **Hazen, T. C., and G. A. Toranzos.** 1990. Tropical source water, p. 32–53. *In* G. A. McFeters (ed.), Drinking water microbiology. Springer Verlag, Berlin.
8. **Hunter, P. R., and M. A. Gaston.** 1988. Numeric index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465–2466.
9. **Hunter, P. R., and J. C. Izsák.** 1990. Diversity studies of salmonella incidents in some domestic livestock and their potential relevance as indicators of niche width. Epidemiol. Infect. **105:**501–510.
10. **Jones, D.** 1988. Composition and properties of the family *Enterobacteriaceae*, p. 1–21. *In* B. M. Lund, M. Sussman, D. Jones, and M. F. Stringer (ed.), Society for Applied Bacteriology symposium series no. 17. Enterobacteriaceae in the environment and as pathogens. Society for Applied Bacteriology, London.
11. **Katouli, M., I. Kühn, and R. Möllby.** 1990. Evaluation of the stability of biochemical phenotypes of *Escherichia coli* upon subculturing and storage. J. Gen. Microbiol. **136:**1681–1688.

11a. **Katouli, M., I. Kühn, A. Brauner, A. A. Farhoudi-Moghaddam, and R. Möllby.** Application of biochemical fingerprinting in investigations of the clonal groups of *Salmonella havana*. J. Med. Microbiol., in press.

12. **Kühn, I.** 1985. Biochemical fingerprinting of *Escherichia coli*: a simple method for epidemiological investigations. J. Microbiol. Methods **3:**159–170.
13. **Kühn, I., A. Brauner, and R. Möllby.** 1990. Evaluation of numerical typing systems for *Escherichia coli* using the API 50 CH and the PhP-EC systems as models. Epidemiol. Infect. **105:**521–531.
14. **Kühn, I., L. G. Burman, L. Eriksson, and R. Möllby.** 1990. Subtyping of *Klebsiella* spp. by biochemical fingerprinting: a simple system for epidemiological investigations. J. Microbiol. Methods **11:**177–185.
15. **Kühn, I., A. Franklin, O. Söderlind, and R. Möllby.** 1985. Phenotypic variations among enterotoxinogenic O-groups of *Escherichia coli* from Swedish piglets with diarrhoea. Med. Microbiol. Immunol. **174:**119–130.
16. **Kühn, I., K. Tullus, and L. G. Burman.** 1991. Epidemiological studies of fecal *Enterobacter cloacae* using biochemical fingerprinting with the PhP-KE system. Epidemiol. Infect., in press.
17. **Kühn, I., K. Tullus, and R. Möllby.** 1986. Colonization and persistence of *Escherichia coli* phenotypes in the intestines of children aged 0 to 18 months. Infection **14:**7–16.
18. **Lacroix, J. M., and M. C. Lavoie.** 1987. Microcomputer package for statistical analysis of microbial populations. Cabios **3:**309–312.

19. **Linton, A. H., and M. H. Hinton.** 1988. Enterobacteriaceae associated with animals in health and disease, p. 71–87. *In* B. M. Lund, M. Sussman, D. Jones, and M. F. Stringer (ed.), Society for Applied Bacteriology symposium series no. 17. Enterobacteriaceae in the environment and as pathogens. Society for Applied Bacteriology, London.
20. **Moore, W. E. C., L. V. Holdeman, E. P. Cato, I. J. Good, E. P. Smith, R. R. Ranney, and K. G. Palcanis.** 1984. Variations in periodontal floras. Infect. Immun. **46:**720–726.
21. **Moore, W. E. C., L. V. Holdeman, R. M. Smibert, D. E. Hash, J. A. Burmeister, and R. R. Ranney.** 1982. Bacteriology of severe periodontitis in young adult humans. Infect. Immun. **38:**1137–1148.
22. **Rice, E. W., P. V. Scarpino, G. S. Logsdon, D. J. Reasoner, P. J. Mason, and J. C. Blannon.** 1990. Bioassay procedure for predicting coliform bacterial growth in drinking water. Environ. Technol. **7:**821–828.
23. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. W. H. Freeman & Co., New York.
24. **Stenström, T. A., and I. Kühn.** 1988. Phenotypic variations within *Klebsiella pneumoniae*. A tool to trace sources of contamination in surface water. Water Sci. Technol. **20:**429–431.
25. **Torsvik, V., K. Salte, R. Sorheim, and J. C. Goksoyr.** 1990. Comparisons of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl. Environ. Microbiol. **56:**776–781.
26. **Tullus, K., B. Berglund, B. Fryklund, I. Kühn, and L. G. Burman.** 1988. Epidemiology of fecal *Enterobacteriaceae* strains in 22 neonatal wards and influence of antibiotic policy. J. Clin. Microbiol. **26:**1166–1170.
27. **Väätänen, P. C.** 1980. Effects of environmental factors on microbial populations in brackish waters off the southern coast of Finland. Appl. Environ. Microbiol. **40:**48–54.
28. **Väätänen, P. C.** 1980. Factor analysis of the impact of the environment on microbial communities in the Tvärminne area, southern coast of Finland. Appl. Environ. Microbiol. **40:**55–61.