Microb. Ecol.7:283–296 (1981)

MICROBIAL ECOLOGY

A Functional Evenness Index for Microbial Ecology

Marc Troussellier and Pierre Legendre

Laboratory of Marine Biology, Université des Sciences et Techniques du Languedoc, F-34060 Montpellier Cedex, France; and Department of Biological Sciences, Université de Montréal, Montreal, Quebec H3C 3J7, Canada

Abstract. Microbial ecologists attempting to describe community structures through the use of synthetic parameters face enormous difficulties. These stem in part from the necessity of using standard taxonomic reference levels in a field where the species level is poorly defined. This paper presents an attempt to obviate this problem. A "functional evenness" index (E) is defined using information measures; it is based directly on the characteristics of the bacteria, as determined, for example, with the API 20B method. Comparisons of this index with classic structure indices, such as taxonomic evenness (Pielou) or systematic dominance (Hulburt), show that it behaves like an evenness index, while bypassing the taxonomic study required before computation of the classic indices. Its use is illustrated with samples of aerobic heterotrophic bacteria obtained from brackish lagoon sediments.

Introduction

In natural ecosystems, plant and animal species are organized in a complex interactive system referred to as the biotic community, or biocoenosis. For various reasons, primarily practical ones, students of ecosystems prefer the use of more homogeneous subunits of the biocoenosis. This has led to the definition of several such subunits, more or less hierarchical and partially redundant, such as community, population, taxocene [6], or guild [25]. Whatever the subunit chosen, a classic methodology exists to study its composition and structure. Composition parameters include abundance and species richness, whereas structural parameters describe the relation between abundance and richness analytically (demographic models) or synthetically (diversity indices).

Although the measurement of such parameters poses no special problem for animals or multicellular plants, it is not so for unicellular organisms and in particular for bacteria. Given an adequate in situ bacterial sample from a natural environment (which is at best a compromise between statistical requirements and the actual constraints of sample handling), the establishment of sound demographic bases for the study of a natural bacterial community is very difficult. The composition of a community is defined, as mentioned above, in terms of the number of species and of their relative abundances. It is then necessary to define a stable and standardized reference taxonomic level, with the

Send reprint requests to P. Legendre at the above address.

0095-3628/81/0007-0283 $02.80

same qualities of constancy and universality as the biological species of botanists and zoologists. To find such a reference level, bacteriologists may use two different taxonomic techniques: classic systematics or numerical taxonomy, with widely diverging conventions of data analysis, each limited by important but differing constraints. The chief advantages of numerical (phenetic) taxonomy include a more objective description of individuals through a standardized body of characters, and the simultaneous use of a large number of unweighted characters. On the other hand, classic systematics provides a universal system of nomenclature. In both cases, the species level is often arbitrarily defined; in any case, some researchers are beginning to doubt the biological reality of bacterial species, due to their lack of genetic isolation [7, 27, 30].

For these reasons, and without advocating the biases due to the bacteriological technique of sample handling—quantitative and qualitative selection due to dilution and culture media, it may seem unjustified to compute even such simple structure parameters as diversity indices. It is, however, possible to try to bypass the delicate step of species recognition and still look for synthetic parameters with which to characterize bacterial communities. Indeed, the small size of bacterial organisms and their lack of morphological differentiation make it necessary to use other characters, physiological or metabolic, to describe individuals and eventually arrive at a classification or an identification. As a result, taxonomy is based on characters that describe the potential functions of bacteria grown on culture media, in tests designed to demonstrate their ability to use certain products in their metabolism (mainly the case in section 4 of Methods, hereafter) or to live under specified physical or chemical conditions, for instance as described by Hauxhurst et al. [15] and Martin and Bianchi [20]. Whey then, not use structural indices, based not on the distribution of individuals in a vector of taxonomic units, but rather on their distribution with reference to the various potential functions investigated? The main objective of the present paper is to define such an index, based on information theory. This new measure, computed on test bacteriological samples, is then compared with like parameters calculated from classic systematic and numerical taxonomic studies in order to ascertain its validity and interpretation.

Materials and Methods

1. A "Functional Evenness" Index

According to Pielou [23], the specific diversity of a biologial collection depends on both "the number of species and the *evenness* with which the individuals are apportioned among them." In order to measure a synthetic structural parameter based directly on the metabolic functions and morphology of the strains, as was called for in the introduction, an *index of functional evenness* appropriate for binary data will be developed based on information theory. It is called *functional* because it is based on the set of potential functions that form the 27 characters listed in section 4: indeed, a positive reaction to a test shows that the bacterium has the enzymatic potential of using a given substrate, if present in the environment.

Since this index does not require a previous determination of species units, it does not contain information as to the "number of species" component of diversity and can be best referred to, for the moment, as a measure of evenness. It is compared with a classic measure of evenness in order to ascertain this analogy.

For a data matrix of binary characters, such as the one describing each bacterial sample of the present study, Lance and Williams [18] have derived a measure of total information, which they used in a quite different context for numerical ecology studies. Their method is actually best suited for taxonomic rather than ecological studies [28].

The computation of total information is as follows (Fig. 1). First, for each presence-absence character i, the

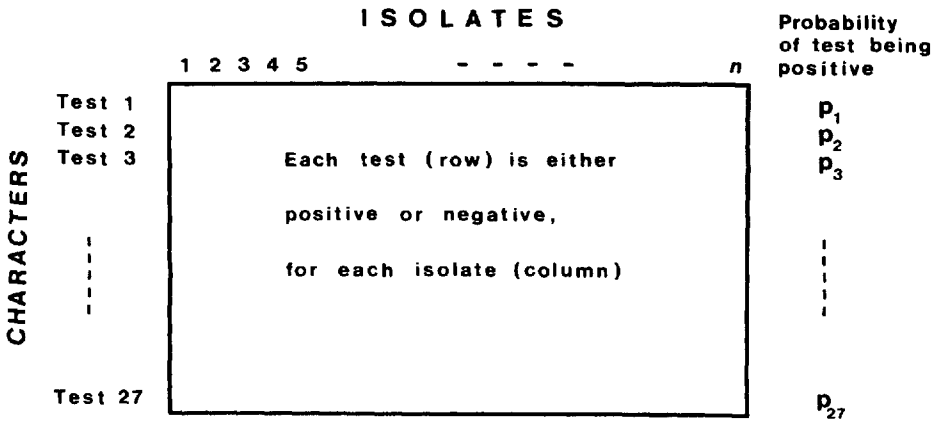

Fig. 1. Data matrix for E.

entrophy $H(i)$ can be computed from its probability (relative frequency) of presence p_i and of absence $(1 - p_i)$ through all the strains of the sample, using the formula

$$H_i = -[p_i \log p_i + (1 - p_i) \log (1 - p_i)] \quad \text{for } 0 < p_i < 1.$$

This is a direct application of Shannon's (26) entropy formula, $H = -\Sigma p_i \log p_i$, for a two-states character with relative frequencies p_i and $(1 - p_i)$. Considering that all the characters have the same weight, the total information of the sample is obtained by summing the various H_i obtained for each character:

$$I = -\sum_{i=1}^{c} [p_i \log p_i + (1 - p_i) \log (1 - p_i)] \quad \text{for } 0 < p_i < 1.$$

Here c, the number of binary characters, is 27 (below). If there are reasons to believe that different weights w_i have to be attributed to the various characters i, the formula could be written as

$$I = -\sum_{i=1}^{c} w_i [p_i \log p_i + (1 - p_i) \log (1 - p_i)] \Bigg/ \frac{1}{c} \sum_{i=1}^{c} w_i$$

$$\text{for } 0 < p_i < 1.$$

The denominator in this equation is such that, when all the w_is are equal to 1, this formula will reduce to the previous one. The w_is could represent different degrees of chemical or biochemical activity. If the index is used in a different context, for instance, with data on presence or absence of various species—as in Williams et al. [32]—pertaining to the same taxocene, the w_is could be made proportional to the biomass of each species or eventually to a measure of biological activity more appropriate to the specifics of the study.

In the original formula of Lance and Williams, this measure I was multiplied by n, the number of objects in the data matrix (corresponding to the number of strains in the data matrices described below); this information was pertinent to their computations. Here, however, we are interested only in measuring the information contained in the two vectors of probabilities p_i and $(1 - p_i)$, so that the number of strains is of no particular relevance. If it were included, it would be present in both the denominator and numerator of the equation of evenness (below), and so would cancel out. This formula will lead to a value of zero when all the strains of the sample are described in exactly the same way by the 27 binary characters; it has no superior limit since its value depends on the number of characters, c.

In order to standardize this coefficient between 0 and 1, that is, to make it independent of the number of

characters, one may divide I by I_{max}, the maximum value which could be taken by I. By studying the limits of function I, or more simply by plotting H_i as a function of p_i, it can be shown that I is maximum when all the relative frequencies p_i are equal to 0,5. Consequently,

$$I_{\max} = -\sum_{i=1}^{c} w_i \log 0,5 \Big/ \frac{1}{c}\sum_{i=1}^{c} w_i = -c \log 0,5$$

The index of functional evenness sought is then

$$E = \frac{I}{I_{\max}} = \frac{1}{c \log 0,5} \sum_{i=1}^{c} [p_i \log p_i + (1-p_i)\log(1-p_i)] \quad \text{for } 0 < p_i < 1,$$

when all the w_is are equal to one, which is the case in the example below. This formula is independent of the logarithmic base, as long as the same base is used in the numerator and the denominator. One cannot overlook the resemblance between this equation and the classic measure of biological evenness, summarized below.

The value provided by this index will be zero when all the strains in the sample are coded in the same way by the c binary characters (maximum functional evenness) and one when the probabilities of all the character states are 0,5. When all the strains in the sample are described by different combinations of the c characters, a situation in which the classic evenness index J (below) is one, E should be close to maximum. The problem of computing the standard error of E is addressed in the Appendix.

2. *Numerical Taxonomy and Diversity Measures*

In order to compare the new index with a classic measure of evenness, taxonomic entities at about the species level had to be defined. Specific diversity and evenness could then be computed from their probability distribution.

As the systematics of environmental bacteria is poorly understood, available identification keys were not adequate in identifying the strains to the species level from the 27 characters available (below). Therefore, a numerical taxonomic procedure was designed and strictly applied to each of the 34 samples (below), thereby identifying the strains to a reference taxonomic level, comparable for all the samples of this study. This level is hereafter referred to as the species-like level.

In each sample, similarity was computed between strains on the basis of the 27 binary characters, using the simple matching coefficient [29]. Although it would perhaps have been interesting to use several hundred characters as in some studies, 27 seem to suffice for the purpose of the present study [9, 14; P. H. A. Sneath, personal communication]. Agglomerative hierarchical clustering was performed on this similarity matrix by the weighted centroid method (WPGMC) [13], which is a space-conserving strategy [28]. Simultaneously, an ordination of the strains in a reduced space of principal coordinates [12] was computed from the same similarity matrix. In order to determine the most appropriate level of synthesis in the dendrogram representing the hierarchical clustering, the various levels of clustering were drawn, for 10 of the 34 samples, on the reduced-space ordination diagram of the strains. The similarity level of 95% was selected to become the species-like level for all 34 samples, since it was the most inclusive level which still respected the relative position of the strains in the reduced-space diagrams. This level consistently lies under the generic level determined by identification keys (below). Furthermore, the study of 55 strains obtained from four type culture collections, with the API 20B apparatus (see section 4) and the same method of numerical taxonomic analysis, showed [31] that the 95% level of similarity is satisfactory, in that it forms taxa very close to the species level of these international reference collections. The 90% level of similarity corresponded closely to the genus level.

From the probability distribution of the species-like entities in each sample, Shannon's [26] diversity index was computed, and from it the measure of taxonomic evenness J was obtained [22] through the formula

$$J = H/H_{\max}$$

$$\text{where } H = -\sum_{i=1}^{m} p_i \log p_i \quad \text{for the } m \text{ species } i$$

$$\text{and } H_{\max} = \log m$$

Diversity studies are usually not based on a census of the total population; this is obviously the case with bacterial samples, mainly because of the selectivity of culture media. This does not invalidate the comparison of samples through such indices, as long as they have been handled in a comparable manner [17, 20].

3. *Dominance Index*

The 27 characters were used to identify each strain to the genus level using dichotomic keys [1, 5]. From the strain's probability distribution in the various genera, Hulburt's dominance index [16] was computed:

$$d_2 = p_1 + p_2$$

where p_1 and p_2 are the relative frequencies (probabilities) of the two most common genera in the sample.

4. *Sampling and Bacteriological Methods*

To illustrate the use of the proposed index, we used samples of heterotrophic aerobic bacteria, obtained from the sediments of neighboring stations in two eutrophic lagoons of the Arcachon basin (01°W, 45°N), 60 km west of Bordeaux, France. Station 1, located near the continental margin of the basin, is open to the influence of tide (average salinity $24^o/_{oo}$, depth varying from a few centimeters to about 1 meter); station 2 stands in the artifically closed Certes lagoon (average salinity $15^o/_{oo}$, depth 30 cm) and obtains its water supply from rainfall. From a sample of about 400 cm^2 of surface sediment, roughly 5 mm thick and homogenized by hand in the laboratory, diluted aliquot fractions were counted via spread plates employing two culture media. Two media were used because these aquatic environments have changing salinity, and because the salinity range was unknown at the beginning of the experiment (Table 1): the first medium is Bacto-nutrient Agar Difco with a salinity of $5^o/_{oo}$, and was expected to provide results complementary to those obtained with the second one, ZoBell medium 2216, with a salinity of $30^o/_{oo}$. The plates were incubated at about 20°C and counted after 4 or 5 days, which provided an estimate of abundances in those aerobic heterotrophic communities. Longer incubation time showed no significant increase in the number of isolates.

Homogenization by hand was performed in the same manner and for the same length of time (1 min) before each dilution or plating, so as to render the results comparable. This technique is currently being used [14] and has been shown to be equivalent to automated homogenization for sediment samples [19].

Thirty-four of the samples, obtained between July 1977 and August 1978, were studied qualitatively. From the dilution plate with the largest number of well-isolated strains, individual strains were purified and subjected to the battery of tests in the API 20B apparatus, which provides 22 biochemical tests [2]: gelatin hydrolysis; presence of β-galactosidase; nitrate reduction; acid production from carbohydrates: sucrose, arabinose, mannitol, fructose, glucose, maltose, starch, rhamnose, galactose, mannose, sorbitol, and glycerol; presence of urease; indole production H_2S production; acetyl methyl carbinol production; Simmons' citrate utilization; oxidase activity; and catalase activity. Each test is analogous to a binary character, as it is coded by the presence or absence of the characteristic tested for. The API 20B system has been developed explicitly for the study of environmental bacteria, as exemplified by Baleux [2]; it provides results comparable to those of conventional microbiological techniques, with the additional advantages of standard procedures, and quick and easily readable results [1, 10, 11]. API 20B is different from API 20E, designed for enteric bacteria.

A 23rd test, the Hugh and Leifson test (oxidation/fermentation of glucose) was performed in the laboratory. Four other binary morphological and staining characters were recorded for each isolated strain: cell shape, Gram stain, motility, and presence of endospore. Table 1 shows the spatio-temporal origin of the 34 samples, in which slightly over 1000 strains were studied. For each sample, the data form a $27 \times n$ matrix (Fig. 1), where n is the number of strains, 27 being the constant number of binary characters.

Results

The values of the structure parameters E, J and d_2, computed for the 34 samples, are shown in Figs. 2 and 3.

Table 1. Characteristics of the bacterial samples.

		Station 1					Station 2				
			Medium 1		Medium 2			Medium 1		Medium 2	
Sampling	Date	Salinity	*N*	*n*	*N*	*n*	Salinity	*N*	*n*	*N*	*n*
1	77-07-13	17	$1{,}0 \cdot 10^5$	52	$4{,}1 \cdot 10^6$	35	15	$6{,}3 \cdot 10^4$	32	$1{,}5 \cdot 10^6$	—
2	77-07-20	21	$1{,}9 \cdot 10^6$	42	$1{,}0 \cdot 10^5$	—	17	$2{,}5 \cdot 10^5$	23	$1{,}0 \cdot 10^6$	33
3	77-08-05	25	$3{,}6 \cdot 10^7$	19	$1{,}2 \cdot 10^5$	18	16	$1{,}2 \cdot 10^6$	21	$1{,}5 \cdot 10^7$	14
4	77-08-19	27	$3{,}2 \cdot 10^4$	—	$2{,}5 \cdot 10^7$	—	18	$1{,}4 \cdot 10^5$	28	$2{,}8 \cdot 10^6$	—
5	77-09-05	27	$1{,}0 \cdot 10^4$	—	$3{,}4 \cdot 10^4$	22	18	$2{,}5 \cdot 10^5$	15	$2{,}0 \cdot 10^6$	13
6	77-12-03	29	$2{,}7 \cdot 10^5$	47	$3{,}8 \cdot 10^6$	52	21	$1{,}7 \cdot 10^5$	25	$3{,}2 \cdot 10^6$	16
7	78-02-16	15	$4{,}5 \cdot 10^5$	48	$1{,}0 \cdot 10^6$	58	8	$1{,}3 \cdot 10^6$	60	$3{,}5 \cdot 10^6$	44
8	78-03-30	13	$1{,}6 \cdot 10^5$	15	$6{,}4 \cdot 10^6$	57	7	$7{,}2 \cdot 10^5$	49	$3{,}5 \cdot 10^6$	44
9	78-05-05	10	$4{,}1 \cdot 10^4$	23	$2{,}6 \cdot 10^6$	29	4	$3{,}0 \cdot 10^5$	26	$4{,}5 \cdot 10^6$	27
10	78-08-17	19	$3{,}3 \cdot 10^5$	30	$8{,}0 \cdot 10^7$	39	20	$3{,}5 \cdot 10^5$	21	$3{,}5 \cdot 10^6$	47

N, number of bacteria per gram of fresh sediment; *n*, number of strains used for detailed bacteriological analysis;—, samples discarded because the number of strains isolated was too small to authorize an interpretable statistical analysis. Salinity is expressed in $^o/_{oo}$.

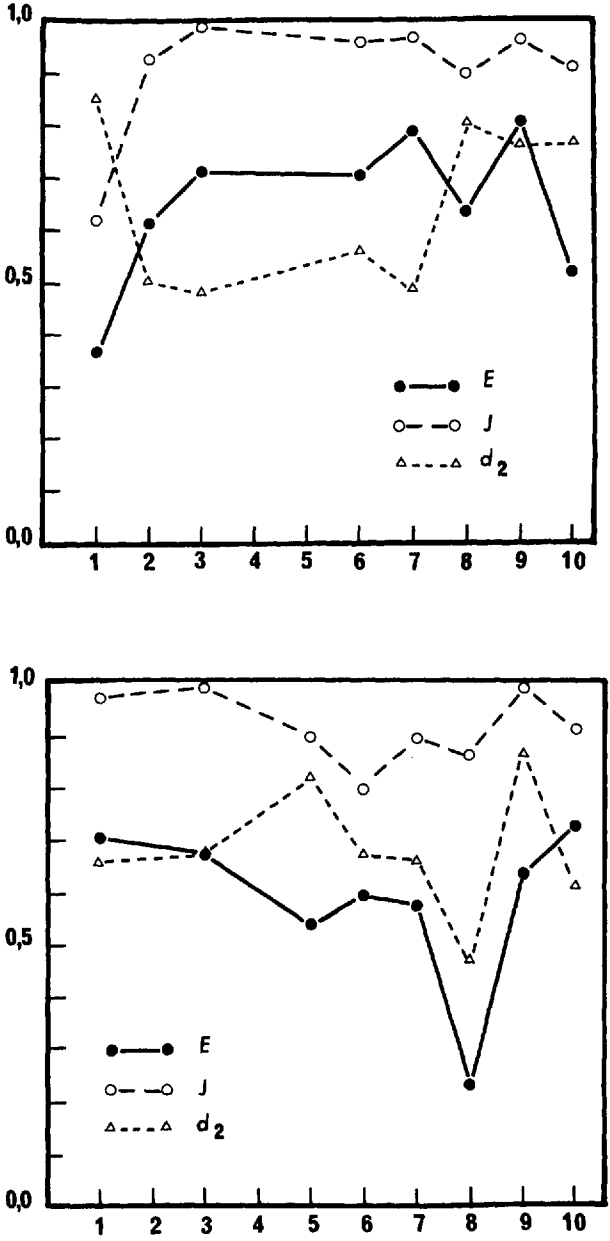

Fig. 2. Structural parameters at station 1. Top: culture medium 1. Bottom: medium 2. Other characteristics of these samples are found in Table 1. Abscissa: sample number. Ordinate: dimensionless measures of evenness (for E and J), or relative frequencies (for d_2).

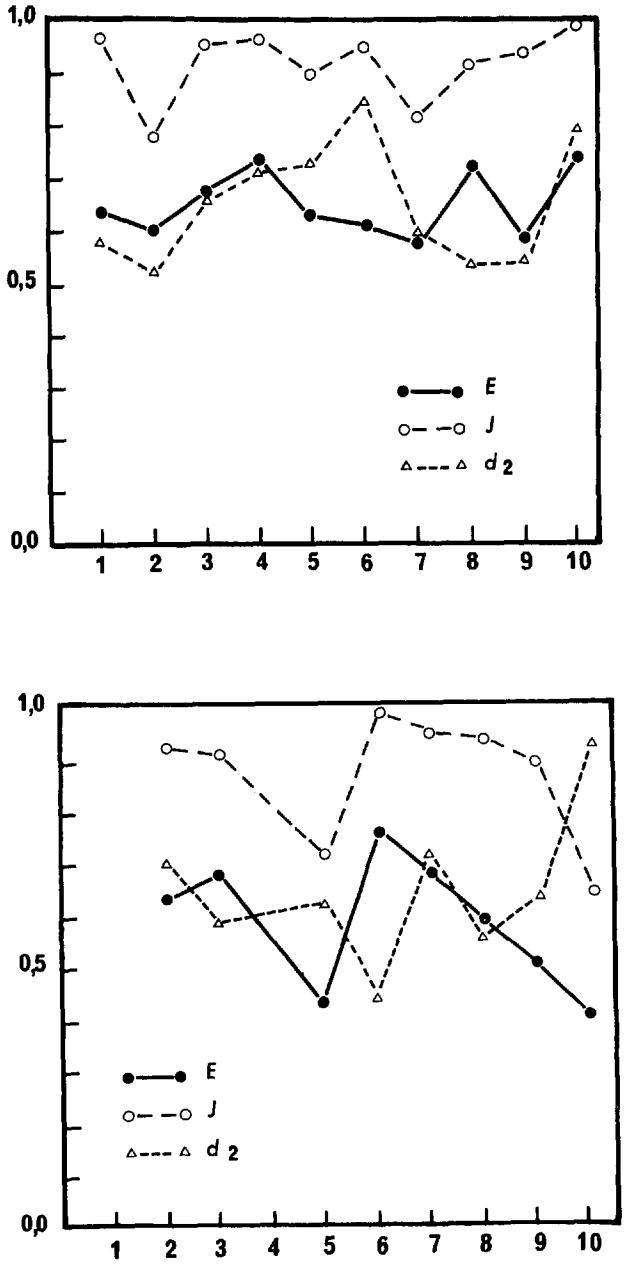

Fig. 3. Structural parameters at station 2. Top: culture medium 1. Bottom: medium 2. Other characteristics of these samples are found in Table 1. Abscissa: sample number. Ordinate: dimensionless measures of evenness (for E and J), or relative frequencies (for d_2).

Considering the long time intervals between samples, relative to the short generation time of bacteria, even the samples from the same series can be considered as statistically independent [24]. This makes it possible to use correlation coefficients to test the significance of the relations between parameters.

The Spearman rank correlation coefficient r was used to measure the relation between the three structural parameters of each of the four data series (two stations, two culture media). In all cases, a significant relation was found between functional evenness (E) and taxonomic evenness (J), at the 5% probability level.[1] On the other hand, the dominance index (d_2) is significantly related with E and J in only one case, station 1 medium 1, whereas in all other cases it is significantly related to neither E nor J, thus emphasizing the commonness of behavior of E and J.

Spearman's r was also used to study the relation between the number of species-like entities m and both evenness indices, since in classic diversity studies, the number of species and the evenness of their frequency distribution are considered as independent components of deversity [23]. Indeed, no significant relation was found between either m and E or m and J, thus showing that E shares with J this additional important property of independence from m.

One last remark should be made about the independence of the characters used in computing E. Indeed, bacteriological characters seldom measure completely different aspects of the strains. However, the use of mostly redundant characters may well cause important distortions of E, this being equivalent, at the extreme, to repeating certain characters in the data matrix corresponding to each sample. Inasmuch as the information duplicated has a marked effect in increasing or decreasing E, its repetition may have the described deleterious effect on E. This problem will also occur in numerical taxonomy, following which the eveness J was computed. In order to identify redundant characters, a correlation matrix was computed between the 27 characters, using as data base the percentages of presence in the 34 samples. Clustering this matrix showed groups of interrelated characters and made it possible to identify 17 primarily independent characters among the 27 originally available. Functional evenness E was computed again from these 17 characters, and showed no significant difference from E shown in Figs. 2 and 3, computed on the 27 characters. This demonstrates that in this case, the effect of redundant characters, if any, was negligible.

Discussion and Conclusions

Relation Between E and J

The correlations, always significant between functional evenness E and taxonomic evenness J, in the samples studies here, show that E is indeed a measure of evenness, even though it does not require a taxonomic treatment prior to its computation. Consequently, one can choose between the two measures of evenness, which lead to comparable results.

Two reasons favor the choice of E over J when the data are of the type described above

[1] This same relation between E and J was found recently in three experiments on the biodegradation of organic compounds over time, to be published elsewhere.

and determination of species is not required. The first is that E varies more than J (larger coefficient of variation in all cases), thus making it potentially more sensitive to environmental influences; this is not surprising since there is a loss of information implied by the numerical taxonomic procedure, when passing from a metric distance matrix containing the complements of the simple matching coefficients (themselves representing the original data matrix), to the ultrametric matrix [28] which could represent clustering at the 0,95 similarity level. Secondly, E is much easier to compute than J which requires a prior numerical (or classic) taxonomic study of each sample, implying in bacteriology the choice of a more or less arbitrary level of clustering.

Relation Between Evenness and Dominance

The results presented above show few significant relations between the dominance index d_2 and either one of the evenness indices E or J. This is not surprising as d_2 was computed on dominant genera while the evenness indices describe the species-like level: Pielou [23] argues that the diversity of the various taxonomic categories is largely independent, to the point of being additive. On the other hand, the identifications with keys were made only to the genus level because the identification of environmental bacteria to the species level, with present-day knowledge, is costly if not impossible (see Introduction).

Other Uses of the Functional Evenness Index

It was briefly mentioned in the Methods section that the functional evenness index E could prove useful in other types of ecological studies. Indeed, ecologists must often deal with qualitative information, or with information that may seem quantitative or semiquantitative but in reality is qualitative. For instance, dosages of heavy metals or of other toxic components in animals are relevant only with relation to a threshold of toxicity for the species under study, and could be noted by binary coding: above or below this threshold. Other examples have been mentioned in connection with the weighting of characters, in the definition of E. In these and other cases, the use of synthetic indices which, like E, can be computed directly from the information available, could prove valuable in the characterization of community structure and behavior.

Finally, with data on the presence of selected biochemical compounds within individuals (or colonies in bacteriology), the functional evenness index makes it possible to measure the increase in biochemical diversity over time within populations, a subject which Odum [21] claims to be understudied.

Further Work

One of our reviewers has been kind enough to suggest further experiments with the functional evenness index, and we take the liberty to report these valuable suggestions here. The problem is to follow the evolution of bacterial communities with time. In the case of natural communities, we can hope to study only a small fraction of the total number of bacteria in a sample, selected (by the bacteriological methodology) among the

agar-cultivable strains. The two approaches suggested hereafter are experimental and consist of a follow-up of a synthetic bacterial community:

Repeat the present paper's study in a big gnotobiotic aquarium with a mixture of agar-cultivable species, and follow changes first with none, then with one environmental parameter varying at a time. This would explain an entire system and 100% of the species present. Proceed to the functional evenness data analysis, using not API 20B characters (which measure capabilities and not activities), but rather data from either of the following methods:

1. Use high-pressure liquid chromatography/gas chromatography/mass spectrometry analyses of specific compounds, described by White et al. [3, 4, 8] Find the concentration of each of 10 or 20 different hydrocarbons. Each compound becomes a character, and from its concentration we derive its proportion. Thus one treats chemical diversity.

2. Incubate replicate samples with trace amounts of 10 or 20 different ^{3}H-labeled organic substrates; do microautoradiography and record the percentage of cells incorporating each. This treats real in situ activities.

Acknowledgements. The field work was supported by contract CNEXO: ECOTRON No. 78-1843, and computer time was provided by contract DGRST "Etangs méditerranéens" No. 79-7-0416, both contracts to the Laboratoire d'Hydrobiologie Marine, Université des Sciences et Techniques du Languedoc, Montpellier. The presence of P. Legendre in Montpellier was supported in part by NSERC Grant T 1735. We thank Prof. M. Amanieu and Dr. B. Baleux for their support of this work and critical review of the manuscript. We are also grateful to Lynne Hosley and Philippe Ross for revision of the English manuscript.

References

1. Baleux, B.: Dégradation biologique des agents de surface—Etude chimique et microbiologique. Thèse de doctorat d'Etat ès Sciences naturelles, Université de Montpellier. Réf. CNRS 12384 (1976)
2. Baleux, B.: A computer study of the evolution of aerobic heterotrophic bacterial populations in sewage and river waters. Microb. Ecol. **4**, 53–65 (1977)
3. Bobbie, R. J., and D. C. White: Characterization of benthic microbial community structure by high-resolution gas chromatography of fatty acid methyl esters. Appl. Environ. Microbiol. **39**, 1212–1222 (1980)
4. Bobbie, R. J., J. S. Nickels, G. A. Smith, S. D. Fozio, R. H. Findlay, W. M. Davis, and D. C. White: Effect of light on biomass and community structure of estuarine detrital microbiota. Appl. Environ. Microbiol. **42**, 150–158 (1981)
5. Buchanan, R. E., and N. E. Gibbons (eds.): Bergey's Manual of Determinative Bacteriology, 8th Ed. Williams & Wilkins Co., Baltimore (1974)
6. Chodorowski, A.: Ecological differentiation of turbellarians in Harsz-Lake. Pol. Arch. Hydrobiol. **6**, 33–73 (1959)
7. Cowan, S. T.: Heretical taxonomy for bacteriologists. J. Gen. Microbiol. **61**, 145–154 (1970)
8. Davis, W. M., and D. C. White: Fluorometric determination of adenosine nucleotide derivatives as measures of the microfouling, detrital, and sedimentary microbial biomass and physiological status. Appl. Environ. Microbiol. **40**, 539–548 (1980)
9. Debette, J., J. Losfeld, and R. Blondeau: Taxonomie numérique de bactéries telluriques non fermentantes à Gram-négatif. Can. J. Microbiol. **21**, 1322–1334 (1975)
10. Feltham, R. K. A.: A taxonomic study of the Micrococcaceae. J. Appl. Bacteriol. **47**, 243–254 (1979)
11. Gauthier, M. J., and R. Clément: Essais d'application de microméthodes à l'étude taxonomique de bactéries hétérotrophes aérobies. Rev. Inst. Pasteur (Lyon) **11**, 35–53 (1978)
12. Gower, J. C.: Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika **53**, 325–338 (1966)

13. Gower, J. C.: A comparison of some methods of cluster analysis. Biometrics **23**, 623–637 (1967)
14. Griffiths, A. J., and R. Lovitt: Use of numerical profiles for studying bacterial diversity. Microb. Ecol. **6**, 35–43 (1980)
15. Hauxhurst, J. D., T. Kaneko, and R. M. Atlas: Characteristics of bacterial communities in the Gulf of Alaska. Microb. Ecol. **7**, 167–182 (1981)
16. Hulburt, E. M.: The diversity of phytoplanktonic populations in oceanic coastal and estuarine regions. J. Mar. Res. **21**, 81–93 (1963)
17. Kaneko, T., R. M. Atlas, and M. Krichevsky: Diversity of bacterial populations in the Beaufort Sea. Nature **270**, 596–599 (1977)
18. Lance, G. N., and W. T. Williams: Computer programs for hierarchical polythetic classification ("similarity analyses"). Comput. J. **9**, 60–64 (1966)
19. Litchfield, C. D., J. B. Rake, J. Zindulis, R. T. Watanabe, and D. J. Stein: Optimization of procedures for the recovery of heterotrophic bacteria from marine sediments. Microb. Ecol. **1**, 219–233 (1975)
20. Martin, Y. P., and M. A. Bianchi: Structure, diversity and catabolic potentialities of aerobic heterotrophic bacterial populations associated with continuous cultures of natural marine phytoplakton. Microb. Ecol. **5**, 265–279 (1980)
21. Odum, E. P.: Fundamentals of Ecology, 3rd Ed. W. B. Saunders, Philadelphia (1971)
22. Pielou, E. C.: The measurement of diversity in different types of biological collections. J. Theor. Biol. **13**, 131–144 (1966)
23. Pielou, E. C.: An Introduction to Mathematical Ecology. Wiley-Intersciences,New York (1969)
24. Rheinheimer, G.: Microbial ecology of a brackish water environment. Ecological Studies 25. Springer-Verlag, New York (1977)
25. Root, R. B.: The niche exploitation pattern of the blue-gray gnatcatcher. Ecol. Monogr. **37**, 317–319; 331–349 (1967)
26. Shannon, C. E.: A mathematical theory of communications. Bell Syst. Tech. J. **27**, 379–423; 623–656 (1948)
27. Sneath, P. H. A.: Phylogeny of microorganisms. Symp. Soc. Gen. Microbiol. **24**, 1–39 (1974)
28. Sneath, P. H. A., and R. R. Sokal: Numerical Taxonomy—The Principles and Practice of Numerical Classification. W. H. Freeman, San Francisco (1973)
29. Sokal, R. R., and C. D. Michener: A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **38**, 1409–1438 (1958)
30. Sonea, S., and M. Panisset: Introduction à la nouvelle bactériologie. Les Presses de l'Université de Montréal and Masson, Paris (1980)
31. Troussellier, M.: Approche expérimentale et méthodologique pour l'étude des peuplements bactériens d'un milieu lagunaire atlantique (domaine de Certes-Arcachon). Thèse de doctorat, Université des Sciences et Techniques du Languedoc, Montpellier (1981)
32. Williams, W. T., J. M. Lambert, and G. N. Lance: Multivariate methods in plant ecology. V. Similarity analyses and information-analysis. J. Ecol. **54**, 427–445 (1966)

Appendix—The Standard Error of E

The values of E were used without standard error in our example, since they did not have to be looked at as estimates of properties from larger universes, but merely as descriptions of the bacterial samplings themselves. This did not prevent us from comparing them, through correlation, with other properties of the same series of samplings, such as J and d_2.

In other situations, a statistical sample (not to be confused with an ecologist's sample of biological material, called "sampling" here) of element measurements is meant to represent a statistical population. Several samples can be compared to see if they are likely to have been drawn from the same population, through t-tests for two samples, or one-way analysis of variance for several samples. In this case, one needs to know the samples' means as well as the standard errors of these means.

Several methods are available in the literature for computing the standard error of Shannon's diversity statistic H. One of these requires several samplings and looks at each one as representing an element in the population of all possible samplings; other methods make use of a single sampling and manage to compute from it a standard error for H.

If a large number of independent samplings are available for a given area, it has been shown [4, 6, 7] how to cumulate the samplings until H reaches stability, and then to obtain independent estimates of H. It is then possible to compute their mean and variance, from which the standard error of H can be calculated.

Mills and Wassel [5] agree that the variance of H must be computed on the basis of several independent samples from the same community, as above. When this cannot be done (as in surveys involving several sampling times or several stations), they advocate the use of rarefaction instead of H, since a measure of the estimation error can be attached to it. These authors have also tried to divide the sampling into a fixed number of (small) subsamplings, then to compute the mean and the variance of H throughout the subsamplings and finally the standard error of H. However, they rightfully note that the subsamplings' H are always an underestimation of the sampling's H, due to impoverishment in species, H being sensitive to the proportion of rare species present in the samplings but likely to be absent from most of the subsamplings. They conclude that this is not a proper way to compute a variance for H from a single sampling.

In bacterial ecological studies of natural environments, however, even if one could cumulate samplings until H stabilizes, as in Pielou [6], we suspect that the heterogeneity of a natural bacterial environment may be such that H may never stabilize as samples are added, but continues to increase, showing thus a gradual shift in our ecological scale of perception as the biocoenosis becomes more and more encompassing.

Basharin gives a formula, reported by Lloyd et al. [3], for the standard error of a single H value, computed from a large number of individuals sampled at random from a whole community. Complementing this formula, Hutcheson [2] derived a t-test for comparing two H estimates.

A final and more realistic method uses the so-called jackknife procedure, a statistical technique designed specifically for analyzing novel statistics. Chiefly, it consists of obtaining N different estimates of the statistic (here H) after removing in turn each of the N individuals or isolates identified in a single sampling; even when a single-isolate species is removed, the error in H is not very large. From each estimate of H, a so-called pseudovalue is calculated; the mean and standard error of the statistic, H, are then computed from the pseudovalues. See Sokal and Rohlf [8] for the details. This method has been applied to H by Adams and McCune [1].

As for evenness, Pielou [7] has shown that the standard error of J cannot be estimated directly from the standard error of H when s, the true number of species in the population, is simply estimated by the number of species m in the sampling, as is usually the case in routine survey work. The reason is that J is then a ratio of two random variables, each with a sampling variance.

This problem does not hold for E, the index of functional evenness, since the divisor used, I_{max}, is not a random variable but rather a constant times a predetermined number of tests. Any one of the methods above—including the jackknife procedure— could be used, depending on the situation, to obtain an estimate of the standard error of I. From that point, one simply divides it by the constant I_{max} to get the standard error of E.

We are grateful to our statistician colleagues, Dr. Yves Lepage from Université de

Montréal and Dr. André Plante from Université du Québec à Montréal. Both have contributed important elements to the discussion presented in this Appendix.

References

1. Adams, J. E., and E. D. McCune: Application of the generalized jackknife to Shannon's measure of information used as an index of diversity. in J. F. Grassle, G. Pattil, W. Smith, and C. Tallie (eds): Ecological Diversity. Statistical Ecological Series, Vol. 6, pp. 117–132. International Cooperative Publishing House, Burtonsville, Md. (1979)
2. Hutcheson, K.: A test for comparing diversities based on the Shannon formula. J. Theor. Biol. **29**, 151–154 (1970)
3. Lloyd, M., J. H. Zar, and J. R. Karr: On the calculation of information-theoretical measures of diversity. Am. Midl. Nat. **79**, 257–272 (1968)
4. Lloyd, M., R. F. Inger, and F. W. King: On the diversity of reptile and amphibian species in a Bornean rain forest. Am. Nat. **102**, 497–515 (1968)
5. Mills, A. L., and R. A. Wassel: Aspects of diversity measurement for microbial communities. Appl. Environ. Microbiol. **40**, 578–586 (1980)
6. Pielou, E. C.: The measurement of diversity in different types of biological collections. J. Theor. Biol. **13**, 131–144 (1966)
7. Pielou, E. C.: Ecological Diversity. John Wiley & Sons, New York (1975)
8. Sokal, R. R., and F. J. Rohlf: Biometry—The Principles and Practice of Statistics in Biological Research, 2nd Ed. W. H. Freeman, San Francisco (1981)