COENOSES 1: 21-28, 1986

THE COMMUNITY CONCEPT AND THE PROBLEM OF NON-TRIVIAL CHARACTERIZATION OF YEAST COMMUNITIES

Marc-André Lachance Department of Plant Sciences, University of Western Ontario, London Ontario N6A 5B7
and
William T. Starmer Department of Biology, Syracuse University, Syracuse New York 13210

Keywords: Yeasts, Communities, Angiosperms, Specificity, Habitats, Probability.

Abstract: The problem of non-trivial characterization of ecological specificity in yeast communities as a function of their physiological responses was examined. It was found that when overall specificity is expressed as the mean number of responses which deviate significantly from those of a theoretical random assemblage, the probability vectors associated with the responses can be defined and transformed to obtain multivariate predictors of habitat quality. Significantly, the distribution of the transformed vectors is continuous, and their relationships are linear. To explore the usefulness of the method, data from yeast communities associated with angiosperm trees were subjected to principal components analysis, and responses which depart significantly from expectations were subjected to cross-tabulation. These provided a means of visually elucidating community profiles.

The community concept in microbial ecology

Contemporary microbial ecology is influenced strongly by its success in identifying specific biochemical interactions between microorganisms and their environment. Geomicrobial transformations, methanogenesis, fixation of atmospheric nitrogen, or the hydrolysis of cellulose are examples of bacterial or fungal processes which are well documented (Atlas and Bartha 1981). The physiological diversity of bacteria is so extreme that autecological studies, often of one-to-one interactions, have been of sufficient interest in themselves. Community studies are addressed involving overall processes (e.g. productivity) resulting from the collective action of many bacterial species. Bacterial communities are usually viewed as collections of specialized individuals, each of which catalyses a step in a complex network of chemical reactions.

Yeast community ecology

Compared to bacteria, yeasts have few physiological capabilities. Very few yeasts have highly specialized nutritional requirements or high degrees of tolerance to extreme physical-chemical conditions. The great majority are simply members of the food chain: they colonize substrates rich in organic carbon, and they serve as food for a number of invertebrates. Yeasts are therefore comparable to plants and animals in that the range of their physiological abilities is relatively narrow. They are always organotrophic, usually mesophilic (growth extremes from ca. 0 to 45°C: Phaff et al. 1976), and somewhat tolerant to acidic conditions. Such physiological homogeneity normally precludes the description of yeast communities as complex networks of interacting specialists.

The environment of a yeast community may be viewed as a multidimensional geometric space, following Hutchinson's (1958, 1959) model. The dimensions include physical factors of the environment, as well as biotic factors such as predation, competition, or food. This view has been useful in understanding why various types of organisms occupy certain habitats (McNaugton and Wolf 1979). The yeast community itself can be described in the same manner, as a multidimensional array, the elements of which are its component species, or the characteristics of those species. Its accurate characterization requires that all efforts be made to eliminate, as much as possible, ambiguity and noise from yeast community data. Provided that these goals are met, the yeast ecologist should benefit more than other microbial ecologists from the vast array of methodologies available and exploited by researchers concerned with plant and animal communities.

Taxonomic ambiguity

The principal problem arising when trying to apply measurements of diversity or multivariate analytical

methods to yeast communities is that of taxonomic ambiguity. This point has been well made by Trousselier and Legendre (1981), who proposed a new measure of diversity based on phenotypic properties of bacteria, instead of their nomenclature. Bacterial taxonomy receives much of its bias from differences in the views of individual taxonomists on the species concept, and from inequalities in the intensity of investigations addressed to various bacterial taxa. For example, medically important species are described in great abundance by contrast to species with less relevance to human welfare. This bias is also present in yeast taxonomy, since some species are of considerable industrial interest. In addition, yeast species often bear duplicate or equivocal taxonomic designations, partly because fungal anamorphs and teleomorphs (asexual and sexual forms), although members of the same natural taxa, are classified separately. The alternative, as pointed out by Trousselier and Legendre (1981), and also by some yeast ecologists (Heed et al. 1976, Lachance, Metcalf and Starmer 1982, Lachance and Starmer 1982, Starmer 1981), is to describe microbial communities as collections of average properties of their individual components.

Numerical properties of yeast characteristics

Physiological profiles (vectors of the proportions of yeasts, in a particular community, which give a positive response on each of a series of physiological tests) are in some cases more informative, as multivariate descriptors, than vectors of yeast taxa frequencies (Lachance et al. 1982). Some numerical properties of yeast physiological vectors impose constraints on their best possible use by multivariate methods. First, different physiological properties of yeasts are distributed differently. Second, many of their distributions are asymmetrical. For example, there is nothing unusual in finding few lactose-utilizing yeasts in a collection, since only a minority of yeasts exhibit β-galactosidase activity. On the other hand, the majority of yeasts utilize ethanol, and so a random collection of yeasts is expected to contain many ethanol utilizers. Such asymmetries not only obscure the significance of physiological responses in a community, but also result in non-linear joint distributions of properties with rare versus frequent positive responses. Common standardization techniques may rectify some of these difficulties, but they usually range data in relation to an internal mean, and consequently they distort data structure.

In this paper, we propose an approach to the study of yeast communities based on the differences between the physiological responses in each community and the responses expected in random yeast assemblages. Significant responses are identified based on statistical methods, and are used to illustrate the relationships between physiology and habitat specificity. After suitable transformations, the descriptors are symmetrical, continuous, and linearly related. On this basis, a single coefficient can be derived to quantify the physiological uniqueness of a yeast community.

Probabilities associated with physiological responses

A list of yeast species and their properties was compiled from that in Barnett and co-workers (1979). The physiological traits considered are listed in Table 1. Responses were scored as 0.0 (negative or no growth), and 1.0 (positive growth). Weak or variable responses were treated as half negative and half positive responses. The growth responses (y_j) are distributed as binomials each with expectation p_j. The estimated response for each variable Y is $\hat{p}_j$ (listed in Table 1) which is the probability that a randomly collected isolate should be positive for the jth attribute. The binomial probability law which applies to the jth physiological response in a collection of size n may thus be expressed as

$$P_j(i)=\frac{n!}{i!(n-i)!}\cdot p_j^i\cdot(1-p_j^{n-1})$$

where $P_j(i)$ is the probability that exactly i individuals out of n be positive for the jth character.

Table 1. Mean responses of all described yeasts to standard physiological tests.

Physiological test	$\hat{p}_j$
Galactose	0.58
Sorbose	0.41
Maltose	0.55
Sucrose	.0.62
Cellobiose	0.60
Trehalose	0.69
Lactose	0.17
Melibiose	0.13
Raffinose	0.37
Melezitose	0.43
Inulin	0.11
Soluble starch	0.23
Xylose	0.62
L-arabinose	0.39
D-arabinose	0.21
Ribose	0.35
Rhamnose	0.26
Ethanol	0.73
Glycerol	0.75
Erythritol	0.25
Ribitol	0.51
Galactitol	0.13
Mannitol	0.75
Glucitol	0.73
α-methyl-D-glucoside	0.39
Salicin	0.58
Lactic acid	0.39
Succinic acid	0.66
Citric acid	0.45
Inositol	0.12
Glucose fermentation	0.55
Growth at 37°C	0.43
Sodium nitrate	0.26

If the observed frequency X_j of positive individuals is larger than the expected frequency np_j, then the probability of a more extreme X than X_j is

$$P(X_j > np_j) = \sum_{i=X_j}^{n} P_j(i)$$

Conversely, when X_j is smaller than np_j, the probability is

$$P(X_j < np_j) = \sum_{i=0}^{X_j} P_j(i)$$

Collection size n, in this context, is defined as the sum of the number of different yeast species isolated in each sample (equivalent to a single primary isolation agar plate), in a number of samples from a common habitat (Lachance and Starmer 1982).

Computation of probabilities

Fortran 77 algorithm SPEC was designed to apply the formulae above to data sets. The calculation of $P_j(i)$ for large collection sizes ($n > 40$) causes numerical overflow, but the problem is circumvented by transitory scaling. For very large n values, processor time cost becomes significant. Fortunately, the probabilities associated with large collection sizes may be calculated by reference to a table of normal probabilities. In program SPEC, the criterion for using the normal approximation is a variance ($np[1-p]$) of 10 or less. This choice is based on the results of simulations discussed in the next section.

Three characteristic quantities based on the probability vectors associated with communities were studied. The first is a measure of specificity (S), defined as the mean number of rejections of the null hypothesis that physiological scores do not depart significantly from expectations, for a given significance level α. The second quantity is a transform of the probabilities themselves. PD is the probability associated with the alternative hypothesis that a response departs from expectations, or $1-P(X_j)$. PD is given a sign corresponding to the polarity of the observed deviation. The third quantity is the standard normal deviate (Z) associated with each calculated probability, and symbolized by Z_e, since it is defined in reference to an external mean.

Illustrative data

The origins of the data on yeast communities from angiosperm trees and the taxonomic position of those trees have been described elsewhere (Lachance and Starmer 1982, in their Table 2).

Simulated data were generated with algorithm RANPRO. Random assemblages were obtained by selecting yeast physiological profiles from the complete list of species, with replacement, as a function of a uniform random number generator (Fortran-supplied function RAN), and averaging their responses. Assemblage sizes were varied linearly. Simulated specific communities were obtained in a similar way, but in this case, the number of isolates of each species also was made to vary as a function of n_tT, where n_t is the number of profiles left to be selected, and T the 5th power of a continuous uniform random number from 0 to 1. The n_tT distribution was chosen empirically to give a convenient range of heterogeneities. Program RANPRO also calculated, for each vector, other specificity related quantities, such as the effective number of species ENS (Starmer 1982), and the functional evenness index E (Trousselier and Legendre 1981).

Measurement of linearity

Algorithm LINEAR was designed to assess the bivariate linearity between all pairs of the 33 variables examined in this study. Thornby's (1972) linearity coefficient H was used and interpreted in relation to that author's third assumption, which appears to be the least restrictive, and which defines the maximum variance of H as $3.39/n$. Critical values were obtained by multiplying the standard deviation by the standard normal deviate for $\alpha/2$ in a two-tailed test.

Exploratory methods

Algorithms LOSIDE (agglomerative hierarchical clustering), CONTIN (cross-tabulation based on clustering), and PCA (principal components analysis) were used in this study. Elementary principles of distributions were taken from Sokal and Rohlf (1981). The use and interpretation of multivariate statistics is as discussed by Orlóci (1978).

Natural distribution of physiological responses

The mean responses of 485 described yeast species on 33 selected physiological tests are listed in Table 1. A simple examination of the mean responses will reveal that some physiological attributes are characterized by more than average high responses (*e.g.* trehalose, ethanol, glycerol, mannitol, glucitol), while others are comparatively low (*e.g.* lactose, melibiose, galactitol, inositol).

Detection of unexpected responses

Each proportion in Table 1 is an estimate of the probability p_j that each yeast in a random collection will be positive for the jth attribute, when each attribute is considered separately. In a collection of size n, the expected number of yeasts positive for attribute Y_j is np_j. If, in an actual yeast collection, the number of isolates positive for Y_j departs substantially from the expected mean, then ecological specificity may be suspected. This kind of comparison was made empirically by Heed et al. (1976) in a study of yeasts associated with *Drosophila*.

It is desirable to quantify the degree of significance of departures observed in yeast collections isolated in the course of ecological studies, and thus obtain a measure of specificity. Physiological responses are binomially dis-

tributed, and the probability that a community response is within the range of that of an «average yeast» may be derived. A simple example follows.

A hypothetical physiological attribute Y_j is known to be positive in 60% of all yeasts known ($\hat{p}_j = 0.60$). A collection of 4 isolates is examined. For collection size n = 4, the probability function associated with expected frequencies of yeasts positive for Y_j is shown in Fig. 1 A. Random assemblages of 4 yeasts are expected to contain an average of 2.4 Y_j—positive yeasts. If only one isolate out of 4 is found positive for Y_j, the probability $P(X_j < np_j)$ that this frequency is the result of chance is the sum of all individual probabilities of obtaining one or less Y_j—positive yeast. This is area P_j, in Fig. 1 A, in this case 0.18. Setting a significance level $\alpha = 0.05$, the hypothesis that these yeasts are from the hypothetical distribution cannot be rejected, and the response y_j of these yeasts may not be viewed as specific.

If approximately the same proportion (1/4) of Y_j—positive yeasts is recovered in a larger collection, say n = 15 (Fig. 1 B), the probability that this collection is from the hypothetical distribution with regard to attribute Y_j is area P_j for $i < 4$, which is less than 0.01. One should then conclude that the recovery of 4 or less Y_j—positive isolates in a collection of 15 indicates significant (α=0.05) specificity.

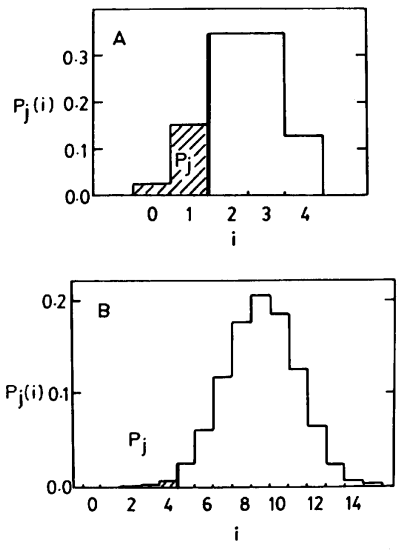

Fig. 1. Probability distributions of expected frequencies (i) of an attribute whose expected response (p_j) is 0.60. In a collection of 4 isolates (A), the probability of obtaining 1 or 0 isolate positive for this attribute in P_j (0.18). If collection size is 15 (B), the presence of 4 (or less) individuals positive for the same attribute would indicate a lower probability ($P_j < 0.01$) that the collection is from a random assemblage.

S, a measure of unexpectedness

In describing yeast communities, it is desirable not only to derive a probability P_j associated with each variable, but also a characteristic function compounding probabilities over the entire physiological profile. Since the probabilities discussed here are attached neither to mutually exclusive, nor to independent events, a true compounded probability may not be calculated by addition or multiplication. Mutual exclusion would also be a prerequisite for the use of χ^2 as a meaningful method of assessing overall non-randomness (Cochran 1954). As an alternative, the mean number of attributes which deviate significantly ($P_j < \alpha$) from expectations, a quantity defined as «S», is proposed. Since the interdependence of the growth variables is not known, S is neither a parameter nor an estimator. It is an empirical measure of the likelihood that, on the average, the physiological responses associated with a collection from a community depart from overall expectation.

Computation of S

The numerical determination of S was studied in a set of artificial communities in which the responses (y_j) were all defined as 0.5, simulating a situation where the entropy is high. The resulting S values ($\alpha = 0.05$) are shown, in Fig. 2, as a function of collection size, computed both with the binomial probability function and with its normal approximation. The two plots merge when collection size reaches 80. This corresponds to a minimum variance near 8 (for highly asymmetrical attributes), which is somewhat larger than the cautionary limit of 5 normally recommended. To be on the safe side, the criterion for using the normal approximation was set at $np(1-p) > 10$. The overestimation of S by the normal approximation is clearly evident for low collection sizes.

Significance of S

To establish the distribution of S in random simulation, $S_{0.05}$ probability points were calculated for several sets of 20 simulated random communities of increasing sizes. The results, shown in Fig. 2, indicate that the mean S values stabilize to a value near α, the significance level initially set for each individual response.

Influence of collection size

As shown in Fig. 2, collection size plays an important role in determining the value of S. We have determined that, in this particular example, S is asymptotic to 0.97, a value reached only when n exceeds 300. This reflects the fact that S is not an estimator of a population parameter. Instead, it must be viewed as a measure of the likelihood, having recovered n yeasts from a particular habitat, that the interaction between those yeasts and that habitat is specific. As collection size increases, the degree of certainty with which one may conclude that specificity prevails also increases.

The usefulness of S determinations for very small collections was investigated by examining the upper theoretical limits of S for n values as low as 2. In this simulation, each vector element (y_j) was set so as to maximize its departure

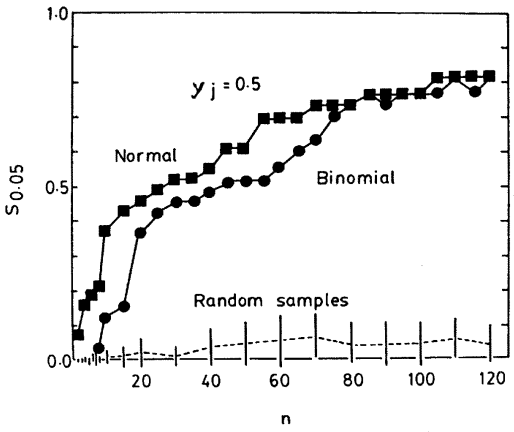

Fig. 2. Computation and significance of S. S values are plotted as a function of collection size, for collections in which all values of Y_j were set to 0.5. Calculations are based on the binomial probability functions (circles) and its normal approximation (squares). The means of twenty S determinations on simulated random collections (probability points) are also plotted as a function of collection size (dotted line). The vertical bars represent two standard deviations for each set of determinations.

from the expected mean np. The results are shown, in Fig. 3, for different α values. The minimum collection sizes above which S has any significance at all are 3 or 4, depending on the stringency of the selected significance level. The examination and interpretation of S for $n < 10$ must therefore be done judiciously, and in the light of theoretical maxima appropriate of the descriptive vectors studied.

Comparison with other measurements of specificity

Several quantities used by ecologists to measure specificity are entropy related, and based on taxon numbers. An example is the effective number of species ENS

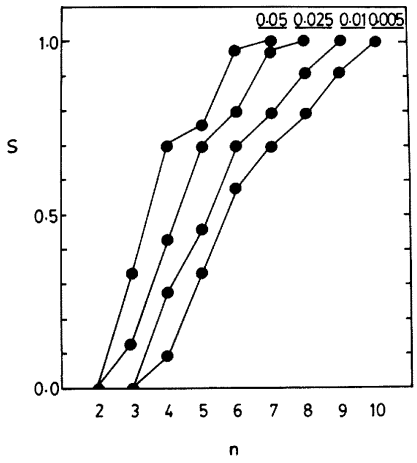

Fig. 3. Upper limits of S values for low collection sizes at different significance levels. Values of Y_j were set to maximize departure from expectation for each variable.

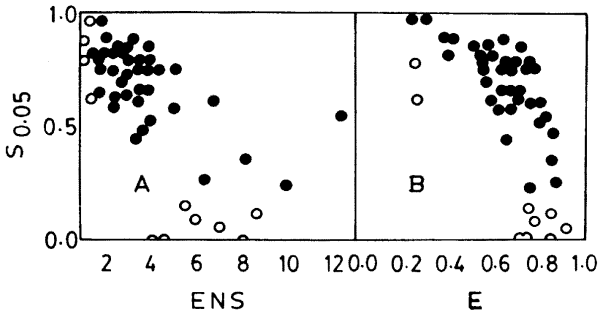

Fig. 4. Dispersion diagrams of S *versus* ENS and E in simulated communities of varying specificity. Collection sizes range from 4 to 20 (open circles), and 22 to 100 (solid circles).

(Starmer 1982) which is the inverse of the probability, given a certain collection, that at least one of the species in that collection will be recovered again in another collection from the same habitat (Simpson 1949). Beside the component of taxonomic ambiguity inherent in the use of ENS (or related entropy measures) in microbiology, we have found that yeast communities isolated from widely diverse habitats exhibit very little, if any, significant variation with respect to quantitative species diversity (Lachance and Starmer 1982). Nonetheless, a comparison between a taxonomic entropy measure and quantity S is of interest. Simulated communities in which ENS was made to vary randomly were examined, as shown in Fig. 4. The joint distribution of S and ENS (Fig. 4A) appears more or less linear, except for communities with smaller sizes, in which ENS overestimates specificity. ENS, by definition, has n as its upper limit. For that reason, and since it is an estimate, ENS requires relatively intense sampling ($n > 15$-30, Lachance and Starmer 1982).

The index of functional evenness (E) proposed by Trousselier and Legendre (1981) is also an entropy measure. It bypasses taxonomic bias by addressing itself directly to descriptive vectors. It relates approximately linearly with S (Fig. 4B) for moderate to large collection sizes. At lower n values, the joint distribution becomes markedly curvilinear, which reflects the fact that E is a parametric estimate, subject to sampling error at low n values.

S is one-sided

It is conceivable that S, in certain cases, may not identify high degrees of specificity which would be detected by species diversity measures. For example, a very selective habitat from which one and only one yeast species is consistently recovered would translate its specificity by an ENS value of 1.0 (maximum specificity). The magnitude of S in such a case would depend on the difference between the physiological profile of that yeast and that of the «average yeast». This very situation has been observed in a yeast community associated with an oak exudate (Bowles and Lachance 1983). In such an extreme case, E is mathematically undefined. In another situation where a

community contains numerous phenotypically homogeneous species, a high ENS value would suggest low specificity, while E and S would lead to the opposite conclusion. The value of S would also depend on the overall phenotypic difference between such a community and a purely random assemblage.

In summary, a high S value defines a community as specific on the basis of its unusual physiology. A low S value leaves specificity undemonstrated, but it does not preclude taxonomic or phenotypic uniqueness.

Probabilities as multivariate descriptors

Derivation of PD and Z_e

The calculation of probabilities necessary to obtain values of S gives rise, for each community, to a multivariate vector. By the simple transformation of P_j to its one complement, one obtains the probability underlying the alternative hypothesis that a particular response is unusual ($1-P_j$). This quantity was the basis for defining variable PD as a symmetrical and externally standardized descriptor. PD has the shortcoming that, except in very small collections, its distribution is sharply discontinuous, since density functions approaching normality range in absolute value between 0.5 and 1.0. Analyses based on PD have been and should be restricted to methods which do not assume data continuity, such as certain types of clustering (Bowles and Lachance 1983).

The standard normal deviate associated with each probability (Z_e) is continuous and has no theoretical boundaries. In practice, its values range from –4.9 to 4.9. In simulated random communities, Z_e follows a normal distribution (with mean 0.0 and standard deviation 1.0).

Linearity in Z_e

The central step of many multivariate methods which involve eigenanalysis is the rotation of variables. It is generally assumed, for this operation, that the original data are free of curvilinear relationships. By virtue of their asymmetry, mean physiological vectors describing yeast communities often do not satisfy those linearity constraints. Their joint distributions are intrinsically hyperbolic.

Thornby's (1972) linearity coefficients (H) were calculated pairwise among all 33 variables in 30 randomly selected communities ranging in size from 5 to 100. A considerable proportion of H values (over 30%) fell outside the confidence interval of the mean H value ($\alpha = 0.05$), suggesting abundant curvilinear relationships. By contrast, the great majority (over 96%) of Z_e vectors obtained from the same data may be considered linearly related.

Application to yeast communities from angiosperm trees

Multivariate ordination

The usefulness of S and Z_e as ecologically significant quantities is obviously dependent on their actual application to real yeast collections. As an example, 27 yeast communities from angiosperm trees (Lachance and Starmer 1982) were examined by principal components analysis (Fig. 5), comparing internal (Z, Fig. 5A) and external (Z_e, Fig. 5B) standardizations. The different patterns in Fig. 5 are comparable, but external standardization did result in a reduction of overlap between certain communities, and it tightened the scattering observed in communities from the Cactaceae, Araliaceae, Ulmaceae, Leguminosae, and, to a lesser extent, the Salicacae.

One major effect of the Z_e transformation data was seen in the loadings of each variable on the components. In both cases (Fig. 5A and B), the first component was a reflection of the utilization of glycosides, alditols, and D-arabinose, and fermentative ability. The component was negatively correlated with growth at 37°C, and the utilization of lactic acid and ethanol. External standardization added nitrate utilization as a major element of this component, a feature overlooked when the data are standardized internally. In addition, external standardization appears to have concentrated more information in the first component, as suggested by an increase from 30 to 43% of the variation explained.

The assimilation of inositol, galactose, and erythritol, in that order, were extracted as important features of the second component illustrated in Fig. 5A. After external standardization (Fig. 5B), the order was reversed, and inositol assimilation could no longer be considered significant. This is an example of how a physiological trait present in a relatively low proportion of yeasts may receive undue weight when data are ranged internally.

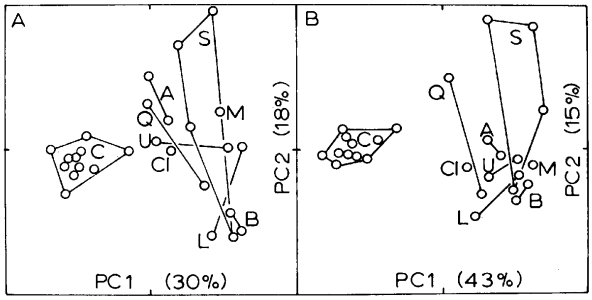

Fig. 5. Ordinations of yeast communities associated with angiosperm trees. Principal components analyses were performed on variance-covariance matrices of internally (A) and externally (B) standardized data. The percent variation extracted in each component is shown in parentheses. Abbreviations: A = Araliaceae; B = Betulaceae; C = Cactaceae; Cl = *Clermontia*; L = Leguminosae; M = *Myoporum*; Q = *Quercus*; S = Salicaceae; U = *Ulmus*.

Cross-tabulation

Physiological responses which significantly deviate from expectation may also be examined directly by means of cross-tabulation. The data may be rearranged according to some extrinsic criteria, or, as in the following example, on

the basis of cluster structure. Table 2 shows the yeast communities of different tree taxa cross-tabulated against their physiological attributes. Partitions and gaps in the data delimit clusters or individuals between which the average cosine was less than 0.5. An analogous table based purely on internal variation had been constructed previously (see Lachance and Starmer 1982, Table 3). A certain polarity opposing carbohydrates with simpler compounds had been detected, but sharp patterns had not emerged. In the present case, it is easily seen that physiological responses in tree yeast communities are highly structured. The following observations may be made:

a – Most yeast communities associated with angiosperm trees show restricted physiological capabilities. Communities associated with the Salicaceae and the Betulaceae are exceptions.

b – The utilization of ethanol and of certain organic acids is generally high in these yeast communities from angiosperms.

c – Certain yeast habitats, the Betulaceae for example, tend to harbor more polytrophic communities over all. The Salicaceae favor yeasts which utilize a less extensive range of compounds, including β-glucosides and some alditols.

d – Growth at 37°C is preponderant in yeasts associated with the Cactaceae.

e – High specificity coefficients, observed in most cases, do not necessarily imply nutritional specialization. They are equally able to detect ecological specificity linked to a broader nutritional potential (e.g. Betulaceae).

Cross-tabulation of significant responses therefore appears to be a simple tool on the basis of which ecologically meaningful hypotheses may be formulated. Unlike multivariate ordinations, the method requires very few assumptions.

Conclusions

We have shown that yeast community physiological vectors may be expressed advantageously as a function of their probabilities of departure from expected responses. The mean number of observed attributes which depart significantly from expectations (S) may be regarded as an overall indicator of ecological specificity. While a meaningful interpretation of S values is mathematically possible over a wide range of collection sizes (down to 3), its significance for small collections ($n < 6$-8) must be evaluated in the light of its upper theoretical boundaries, as established with theoretical extremes (i.e. Fig. 2). In that case, low specificity values may make it advisable to intensify sampling if at all possible. Quantity S is influenced only indirectly by taxonomic diversity, since the polarization of physiological responses in a yeast community is to

Table 2. Significant physiological responses in yeast communities from angiosperm trees. Responses which are lower than expected are identified as «-», and higher responses as «+». Sample size (n) and the specificity coefficient (S) are also shown.

Community	Significant responses																																			n	$S_{.05}$
	Salicin	Cellobiose	Citric acid	Ribose	Melibiose	D-arabinose	Lactose	Trehalose	L-arabinose	Glucitol	Mannitol	Inositol	Nitrate	Glycerol	Rhamnose	Fermentation	Ribitol	Maltose	Melezitose	Galactose	Sorbose	Erythritol	Galactitol	Inulin	Starch	α-methyl-D-glucoside	Raffinose	Sucrose	Xylose	37°C growth	Succinic acid	Lactic acid	Ethanol		
Cactaceae	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	679	0.97
Clermontia			-	-	-	-	-	-	-	-			-				-	-	-	-	-	-		-	-	-	-	-		-		+		32	0.70
Quercus	-	-	-	-	-	-	-	-	-			-					-	-	-	-		-	-	-	-	-	-	-	-	-		+	+	100	0.76
Araliaceae				-	-	-	-	-	-				-		+		-	-	-	-	-	-	-	-	-	-		+	-	+	+	+	+	44	0.70
Ulmus	-	-		-				-					-		+		-	-	-	-	-							-	-	-				26	0.42
Myoporum	-	-	-	-	-	-	-		-			-				-	+	-	-	-	-	-	-	-			+		-	-	-		-	+	32	0.73
Leguminosae	-	-	-	-	-	-	-	-				+			-	-	-	-				-	-	-	-	-				-	-	-			69	0.70
Salicaceae	+	+	+					+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-		+	+	+	249	0.85
Betulaceae	+	+	+	+	+			+	+	+	+	+	-		+		+	+	+		+						+	+	+		+		+	71	0.64

some degree also a function of the number of different species isolated from that community.

The significant departures from expectations (Z_e) in a single yeast community are free of influences from other communities in the same data set. They are dependant to some extent on the number and types of yeast species recognized at a given time, since the proportions in Table 1 are based on taxonomic views which evolve rapidly. Such changes account for the few discrepancies noticeable between the numbers in Table 1, and those reported by others (see Heed et al. 1976, Table 7). Aside from these reservations, the departures and the quantities derived from their associated probabilities are clear indicators of the physiological structure of yeast communities. Whether they are analysed by elaborate multivariate methods or by simpler techniques which facilitate inspection of data, they can provide further elements to our understanding of natural yeast distribution patterns.

Acknowledgments. This work was supported by operating grants from the Natural Science and Engineering Research Council, Canada, to M.A. Lachance, and from the National Science Foundation, USA, to W.T. Starmer (DEB-8108679). Thanks are extended to J.M. Bowles and R.C. Jancey for helpful discussions, and in particular to L. Orlóci for his suggestions on improving the manuscript.

REFERENCES

ATLAS, R.M., and R. BARTHA. 1981. *Microbial Ecology, Fundamentals and Applications.* Addison-Wesley Publ., Reading Massachussetts.

BARNETT, J.A., R.W. PAYNE, and D. YARROW. 1979. *A Guide to Identifying and Classifying Yeasts.* Cambridge University Press, Cambridge.

BOWLES, J.M., and M.A. LACHANCE. 1983. Patterns of variation in the yeast florae of exudates in oak community. Can. J. Bot. **61**: 2984-2995.

COCHRAN, W.G. 1954. Some methods of strengthening the common chi^2 tests. Biometrics **10**: 417-451.

HEED, W.B., W.T. STARMER, M. MIRANDA, M.W. MILLER, and H.J. PHAFF. 1976. An analysis of the yeast flora associated with cactiphilic *Drosophila* and their host plants in the Sonoran Desert and its relation to temperate and tropical associations. Ecology **57**: 151-160.

HUTCHINSON, G.E. 1958. Concluding remarks. Cold Spring Harbor Symposium Quantitative Biology **22**: 415-427.

HUTCHINSON, G.E. 1959. Homage to Santa Rosalia, or why are there so many kinds of animals? American Naturalist **93**: 145-159.

LACHANCE, M.A., B.J. METCALF, and W.T. STARMER. 1982. Yeasts from exudates of *Quercus, Populus, Pseudotsuga,* and *Ulmus*: new isolations and elucidation of some factors affecting ecological specificity. Microbial Ecology **8**: 191-198.

LACHANCE, M.A., and W.T. STARMER. 1982. Evolutionary significance of physiological relationships among yeast communities associated with trees. Can. J. Bot. **60**: 285-293.

MCNAUGHTON, S.J., and L.L. WOLF. 1979. *General Ecology.* Holt, Rinehart and Winston, New York.

ORLÓCI, L. 1978. *Multivariate Analysis in Vegetation Research.* W. Junk, Publishers, The Hague.

PHAFF, H.J., MILLER, and E.M. MRAK. 1976. *The Life of Yeasts.* Harvard University Press, Cambridge, Massachusetts.

SIMPSON, E.H. 1949. Measurement of diversity. Nature (London) **163**: 668.

STARMER, W.T. 1981. A comparison of *Drosophila* habitats according to the physiological attributes of the associated yeast communities. Evolution **35**: 38-52.

STARMER, W.T. 1982. An analysis of the fundamental and realized niche of cactophilic yeasts. In: D.T. Wicklow and G.C. Carroll (eds.), *the fungal community,* pp. 129-155. Marcel Dekker, New York.

SOKAL, R.R., and F.J. ROHLF. 1981. *Biometry.* W.H. Freeman. San Francisco.

THORNBY, J.L. 1972. A robust test for linear regression. Biometrics **28**: 533-543.

TROUSSELIER, M. and P. LEGENDRE. 1981. A functional evenness index for microbial ecology. Microb. Ecol.**7**: 283-296.

Manuscript received: Dec. 1984