Oecologia (Berl.) 18, 359—367 (1975)

Components of Diversity

J. David Allan

Department of Zoology, University of Maryland, College Park, Maryland 20742

Received August 1, 1974

Summary. The information theory measure $H' = -\Sigma[p_i \log p_i]$ is partitioned into components to allow evaluation of various contributions to total diversity. If a species collection is sampled at several microhabitats within each of several sites, we may ask whether the niche breadth of a particular species, and the diversity of the entire collection, are greater with respect to microhabitats or sites. The usefulness of these measures is discussed in the context of within-habitat and between-habitat contributions to diversity.

Introduction

The concept of species diversity and the adoption of information theory measures (Margalef, 1957; Pielou, 1969) have been useful developments in ecological theory. On the other hand, assemblages typically exhibit a species-abundance distribution that has a few categories very common and many categories rather rare; hence this measure is not particularly sensitive, and comparisons must be chosen with care to be of value. One potentially profitable approach has been to partition a total diversity index into component parts associated with several dimensions over which some variation occurs.

This use, analogous to an analysis of variance, allows some evaluation of which dimensions make more major contributions to total diversity. Pielou (1967) has evaluated taxonomic diversity at the species, generic, and familial levels, Levins (1968) partitioned species diversity into factors such as season, time of day, microhabitat etc., and Lewontin (1972) examined variation in blood group systems in man. The purpose of this paper is to outline several methods of apportioning diversity and to examine the relative merits of each.

I will assume the investigator wishes to apportion diversity among three components. Suppose we collect a series of identical small-scale samples in a particular area. Examples might include stream insects in a square foot of substratum, *Drosophila* attracted to each of several different types of fruit placed as baits, or protozoans in each pitcher in a clump of pitcher plants. We may repeat this small-scale sampling at several different sites, or times of day, or seasons. Pursuing the distributional example, suppose we determine the abundances of species present at each of several microhabitats in each of several sites, and wish to enquire into the contributions of microhabitats and sites to total diversity.

Two basic and very similar diversity equations are available: the Brillouin (1962) expression

$$H = \frac{1}{N} \log \left[\frac{N!}{N_1! N_2! \ldots N_s!} \right] \tag{1}$$

and the Shannon and Weaver (1949) expression

$$H' = -\sum_{i=1}^{s}\left[\frac{N_i}{N}\log\frac{N_i}{N}\right] \tag{2}$$

where N = total number of individuals in the collection;
N_i = number of individuals in the i^{th} species;
s = number of species.

Pielou (1966, 1969) has argued that H as defined by Eq. (1) is appropriate for most biological assemblages where we are unable to treat the data collected as a true sample of some definable universe. However, H is affected by the total number of items in the collection, as well as by number of categories and their proportional representation, while H' is affected only by the number of categories and their proportional representation. Since apportioning diversity results in a progressive reduction in the number of items associated with each component, I choose to utilize H' and hence emphasize proportions. This seems most reasonable when the number of categories and items associated with a particular component are determined arbitrarily by the investigator. However, where the size of the categories has more of a biological basis, for example in evaluating the diversity associated within specific, generic and familial levels of biological classification (Pielou, 1967), then H is more appropriate.

The Niche Breadth Approach

Levins (1968) introduced the use of a frequency distribution over habitats as a measure of the breadth of a species' niche, reasoning that species which are distributed over more habitats and more evenly ought to have greater niche breadths. Colwell and Futuyma (1971) and Pielou (1972) have added to this approach. Suppose we collect up to s species from each of r microhabitats in q sites. Then each observation is an N_{ijk} for the abundance of the k^{th} species located in the j^{th} microhabitat of the i^{th} site. Since some species will not be represented in every sample, N_{ijk} may be zero and the number of species found in each sample will typically be less than s. Now each species has a niche breadth over microhabitats, over sites, and a total niche breadth.

$$B_t(k) = -\sum_{i=1}^{q}\sum_{j=1}^{r}\left[\frac{N_{ijk}}{N_{..k}}\log\frac{N_{ijk}}{N_{..k}}\right] \tag{3}$$

where $B_t(k)$ = the total niche breadth of the k^{th} species;
$N_{..k}$ = total number of individuals of the k^{th} species.

$$B_m(k) = \sum_{i=1}^{q}\frac{N_{i.k}}{N_{..k}}\left\{-\sum_{j=1}^{r}\left[\frac{N_{ijk}}{N_{i.k}}\log\frac{N_{ijk}}{N_{i.k}}\right]\right\} \tag{4}$$

where $B_m(k)$ = average microhabitat niche breadth of the k^{th} species;
$N_{i.k}$ = number of individuals of k^{th} species in i^{th} site, summed over r microhabitats.

In essence, the term in brackets is the microhabitat niche breadth of species k at site i. There are q such observations for each of the sites, and their sum is weighted by the proportion of individuals at each site ($N_{i.k}/N_{..k}$) for the species in question.

$$B_s(k) = -\sum_{i=1}^{q}\left[\frac{N_{i.k}}{N_{..k}}\log\frac{N_{i.k}}{N_{..k}}\right] \tag{5}$$

where $B_s(k)$ = site niche breadth for the k^{th} species.

Finally, since all three terms are calculated independently, a convenient check is

$$B_t(k) = B_s(k) + B_m(k). \tag{6}$$

Hierarchical Diversity

Pielou (1967) developed the following approach to take into account the hierarchical nature of biological classification. She utilized H to treat specific, generic and familial levels; I will present formulae for H' and the corresponding microhabitat, site and species example.

$$H'_T = -\sum_{i=1}^{q}\sum_{j=1}^{r}\sum_{k=1}^{s}\left[\frac{N_{ijk}}{N_{...}}\log\frac{N_{ijk}}{N_{...}}\right] \tag{7}$$

where H'_T = total diversity of the assemblage;
$N_{...}$ = total individuals in the assemblage.

$$H'_{sp} = -\sum_{k=1}^{s}\left[\frac{N_{..k}}{N_{...}}\log\frac{N_{..k}}{N_{...}}\right] \tag{8}$$

where H'_{sp} = diversity associated with the average distribution of individuals over all microhabitats and sites.

$$H'_{si} = \sum_{k=1}^{s}\frac{N_{..k}}{N_{...}}\left\{-\sum_{i=1}^{q}\left[\frac{N_{i.k}}{N_{..k}}\log\frac{N_{i.k}}{N_{..k}}\right]\right\} \tag{9}$$

where H'_{si} = diversity associated with the average distribution of individuals over sites calculated for each species and then summed with weights proportional to species abundances;

$\frac{N_{..k}}{N_{...}}$ = relative abundance (weight) of the k^{th} species;

$-\sum_{i=1}^{q}\left[\frac{N_{i.k}}{N_{..k}}\log\frac{N_{i.k}}{N_{..k}}\right]$ = diversity of k^{th} species over $i = 1 \rightarrow q$ sites;

$$H'_{mi} = \sum_{k=1}^{s}\sum_{i=1}^{q}\frac{N_{i.k}}{N_{...}}\left\{-\sum_{j=1}^{r}\left[\frac{N_{ijk}}{N_{i.k}}\log\frac{N_{ijk}}{N_{i.k}}\right]\right\} \tag{10}$$

where H'_{mi} = diversity associated with the average distribution of individuals over microhabitats, calculated at a particular site, for a particular species, then summed with weights proportional to the abundance of individuals at that site and for that species;

$\frac{N_{i.k}}{N_{...}}$ = relative abundance (weight) of the k^{th} species at the i^{th} site;

$-\sum_{j=1}^{r}\left[\frac{N_{ijk}}{N_{i.k}}\log\frac{N_{ijk}}{N_{i.k}}\right]$ = diversity of k^{th} species at i^{th} site over $j = 1 \rightarrow r$ microhabitats.

Finally, since the sum of microhabitat diversity, site diversity and species diversity should equal total diversity, a convenient computational check is:

$$H'_T = H'_{sp} + H'_{si} + H'_{mi}. \tag{11}$$

"Pooling Samples" Diversity

Lewontin (1972) developed the following approach in his analysis of blood group systems in man. The method relies on the choice of H' rather than H particularly strongly, since it uses the fact that if all (say) microhabitats are identical in their species proportions at a particular site, then the proportional representation of species at that site will be equal to the weighted sum of distributions over microhabitats. Hence the difference between the two measures would be zero, indicating an absence of between microhabitat diversity. Note an important difference in interpretation here, as the hierarchical diversity approach would view as high uncertainty what the pooling approach would view as no additional diversity. I will return to this important distinction later with examples.

$$H'_{\text{mic},\,i} = \sum_{j=1}^{r} \frac{N_{ij.}}{N_{i..}} \left\{ -\sum_{k=1}^{s} \left[\frac{N_{ijk}}{N_{ij.}} \log \frac{N_{ijk}}{N_{ij.}} \right] \right\} \tag{12}$$

where $H'_{\text{mic},\,i}$ = within microhabitat diversity of the i^{th} microhabitat, calculated as a species diversity measure in each microhabitat of that site, then summed with weights proportional to the total number of individuals in each microhabitat;

$\frac{N_{ij.}}{N_{i..}}$ = relative abundance (weight) of individuals in the j^{th} microhabitat of the i^{th} site summed over $k = 1 \rightarrow s$ species, to total individuals in the i^{th} site;

$-\sum_{k=1}^{s} \left[\frac{N_{ijk}}{N_{ij.}} \log \frac{N_{ijk}}{N_{ij.}} \right]$ = species diversity in j^{th} microhabitat of i^{th} site.

$$\bar{H}'_{\text{mic}} = \sum_{i=1}^{q} \frac{N_{i..}}{N_{...}} \left[H'_{\text{mic},\,i} \right] \tag{13}$$

where $\bar{H}'_{\text{mic}}$ = the average within microhabitat species diversity, averaged over sites;

$\frac{N_{i..}}{N_{...}}$ = the relative abundance (weight) of individuals in the i^{th} site summed over $j = 1 \rightarrow r$ microhabitats and $k = 1 \rightarrow s$ species, to total individuals.

$$H'_{\text{site},\,i} = -\sum_{k=1}^{s} \left[\frac{N_{i.k}}{N_{i..}} \log \frac{N_{i.k}}{N_{i..}} \right] \tag{14}$$

where $H'_{\text{site},\,i}$ = the species diversity at the i^{th} site, calculated by pooling over microhabitats at that site.

$$\bar{H}'_{\text{site}} = \sum_{i=1}^{q} \frac{N_{i..}}{N_{...}} \left[H'_{\text{site},\,i} \right] \tag{15}$$

where $\bar{H}'_{\text{site}}$ = the average species diversity at sites, averaged over sites.

Now, if all microhabitats are identical, the average site diversity ($\bar{H}'_{\text{site}}$) will be equal to the average microhabitat diversity ($\bar{H}'_{\text{mic}}$), since

$$\frac{N_{ijk}}{N_{ij.}} \times r = \frac{N_{i.k}}{N_{i..}} \tag{16}$$

hence, any differences between microhabitats will be reflected in the degree to which $\bar{H}'_{\text{site}}$ exceeds $\bar{H}'_{\text{mic}}$.

$$WMC = \bar{H}'_{\text{mic}} \tag{17}$$

where WMC is now defined as the within microhabitat component of diversity.

$$BMC = \bar{H}'_{\text{site}} - \bar{H}'_{\text{mic}} = (15) - (13) \tag{18}$$

where BMC is now defined as the between microhabitat component of diversity.

Similarly,

$$BSC = H'_{sp} - \bar{H}'_{\text{site}} = (8) - (15) \tag{19}$$

where BSC is now defined as the between site component of diversity and equals the total species diversity obtained by pooling observations over microhabitats and sites, less the average site diversity.

By definition,

$$H'_{sp} = BSC + BMC + WMC \tag{20}$$

Eq. (20) does not constitute a computational check as the four terms are derived from three equations.

Sample Computations and Discussion

Several sets of dummy data were computed to gain some insight into the behavior of the various indices (Tables 1–4). In the first example (Table 1), species are identical and sites are identical; only the distribution over microhabitats is non-uniform. Note that niche breadth is maximal with respect to sites ($=\log$ [number of sites]) and less for microhabitats. Hierarchical diversity reveals that diversity associated with the species distribution is maximal ($=\log$ [number of species]) as is diversity associated with distributions over sites, but the discovery of five species coupled with the choosing of only three sites will typically result in H'_{sp} being greater than H'_{si}. Since each microhabitat has all five species evenly distributed, all the information about species abundances is contained at the within microhabitat level (WMC). The fact that absolute abundance varies over microhabitats, while relative species distribution does not, is revealed by the value for microhabitat distribution ($H'_{mi} = 0.900$) while the between microhabitat component is zero.

The next example (Table 2) consists of collections where each microhabitat has only one species, sites are identical, and diversity stems from differences between microhabitats.

Two more complex examples provide uneven species distributions within microhabitats, with differences between microhabitats but not between sites (Table 3) or the converse (Table 4). The only point I wish to make here is that as long as each microhabitat contains about the same number of species and the individuals are distributed among species according to about the same distribution (*i.e.* a few common, more rare), then most of the diversity will partition out at the within microhabitat level. This, along with potential misinterpretations

Table 1. Example of components of diversity

a) Dummy data for diversity computations

Species	Site 1 Microhabitats			Site 2 Microhabitats			Site 3 Microhabitats		
	1	2	3	1	2	3	1	2	3
1	2	5	1	2	5	1	2	5	1
2	2	5	1	2	5	1	2	5	1
3	2	5	1	2	5	1	2	5	1
4	2	5	1	2	5	1	2	5	1
5	2	5	1	2	5	1	2	5	1

b) Results

Species (k)	$B_t(k)$	$B_s(k)$	$B_m(k)$	$B_s(k)+B_m(k)$
1	1.999	1.099	0.900	1.999
2	1.999	1.999	0.900	1.999
3	1.999	1.099	0.900	1.999
4	1.999	1.099	0.900	1.999
5	1.999	1.099	0.900	1.999

H'_T	H'_{sp}	H'_{si}	H'_{mi}	$H'_{sp}+H'_{si}+H'_{mi}$
3.608	1.609	1.099	0.900	3.608

	WMC	BMC	BSC	$WMC+BMC+BSC$
	1.609	0.000	0.000	1.609

Table 2. Example of components of diversity

a) Dummy data for diversity computations

Species	Site 1 Microhabitats			Site 2 Microhabitats			Site 3 Microhabitats		
	1	2	3	1	2	3	1	2	3
1	10	0	0	10	0	0	10	0	0
2	0	10	0	0	10	0	0	10	0
3	0	0	10	0	0	10	0	0	10
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0

owing to having one component represented by many more subdivisions than another (*e.g.* choosing 2 sites each with 20 microhabitats), are the principal pitfalls one may encounter. For the first reason mentioned above, Lewontin's (1972) interpretation of racial differences based on blood groups in man probably

Table 2 (continued)

b) Results

Species (k)	$B_t(k)$	$B_s(k)$	$B_m(k)$	$B_s(k)+B_m(k)$
1	1.097	1.097	0	1.097
2	1.097	1.097	0	1.097
3	1.097	1.097	0	1.097
4	0	0	0	0
5	0	0	0	0

H'_T	H'_{sp}	H'_{si}	H'_{mi}	$H'_{sp}+H'_{si}+H'_{mi}$
2.197	1.099	1.099	0	2.197

	WMC	BMC	BSC	$WMC+BMC+BSC$
	0.000	1.099	0.000	1.099

Table 3. Example of components of diversity

a) Dummy data for diversity computations

Species	Site 1 Microhabitats				Site 2 Microhabitats				Site 3 Microhabitats			
	1	2	3	4	1	2	3	4	1	2	3	4
1	3	2	1	5	3	2	1	5	3	2	1	5
2	4	16	0	9	4	16	0	9	4	16	0	9
3	21	2	2	2	21	2	2	2	21	2	2	2
4	12	12	12	12	2	12	12	12	12	12	12	12
5	0	1	0	2	0	1	0	2	0	1	0	2

b) Results

Species (k)	$B_t(k)$	$B_s(k)$	$B_m(k)$	$B_s(k)+B_m(k)$
1	2.339	1.099	1.241	2.339
2	2.063	1.099	0.965	2.063
3	1.872	1.099	0.774	1.872
4	2.485	1.099	1.386	2.485
5	1.735	1.099	0.637	1.735

H'_T	H'_{sp}	H'_{si}	H'_{mi}	$H'_{sp}+H'_{si}+H'_{mi}$
3.571	1.363	1.099	1.110	3.571

	WMC	BMC	BSC	$WMC+BMC+BSC$
	1.139	0.224	0.000	1.363

Table 4. Example of components of diversity

a) Dummy data for diversity computations

Species	Site 1 Microhabitats				Site 2 Microhabitats				Site 3 Microhabitats			
	1	2	3	4	1	2	3	4	1	2	3	4
1	2	2	2	2	6	6	6	6	4	4	4	4
2	8	8	8	8	3	3	3	3	5	5	5	5
3	1	1	1	1	0	0	0	0	3	3	3	3
4	9	9	9	9	6	6	6	6	4	4	4	4
5	12	12	12	12	3	3	3	3	0	0	0	0

b) Results

Species *(k)*	$B_t(k)$	$B_s(k)$	$B_m(k)$	$B_s(k)+B_m(k)$
1	2.398	1.011	1.386	2.398
2	2.410	1.024	1.386	2.410
3	1.949	0.562	1.386	1.949
4	2.432	1.046	1.386	2.432
5	1.887	0.500	1.386	1.887

H'_T	H'_{sp}	H'_{si}	H'_{mi}	$H'_{sp}+H'_{si}+H'_{mi}$
3.786	1.519	0.881	1.386	3.786

	WMC	*BMC*	*BSC*	*WMC+BMC+BSC*
	1.351	0.000	0.168	1.519

overestimates the portion of diversity within races and underestimates that between races.

In conclusion, a variety of approaches exist for apportioning diversity into component parts. Each species' niche breadth may be partitioned into a component associated with microhabitats, and a component associated with sites. The entire species × site × microhabitat matrix may be partitioned into an average distribution of individuals across species (species diversity), across sites (a sort of collective site niche breadth), and across microhabitats (a sort of collective microhabitat niche breadth). However, which is largest may be strongly influenced by logistical decisions concerning number of sites and microhabitats chosen, and in many cases it will be more instructive to examine the site and microhabitat niche breadths in the context of particular species. The average distribution of individuals across species may be partitioned into a within microhabitat component of species diversity, a between microhabitat, within site component, and a between sites component. This latter approach is particularly useful for asking, when one assemblage (*e.g.* the tropics) is more diverse than another (*e.g.* temperate zone), whether each microhabitat of the more diverse assemblage is itself more diverse, or whether the added diversity derives from greater between-habitat differences (MacArthur, 1965).

Acknowledgements. I thank R. C. Lewontin and E. C. Pielou for useful advice. The manuscript benefitted from readings by D. E. Gill, D. H. Morse and S. A. Woodin.

References

Brillouin, L.: Science and information theory, 2nd ed. New York: Academic Press 1962

Colwell, R. K., Futuyma, D. J.: On the measurement of niche breadth and niche overlap. Ecology **52**, 567–576 (1971)

MacArthur, R. H.: Patterns of species diversity. Biol. Rev. **40**, 510–533 (1965)

Margalef, A. R.: Information theory in ecology. General Systems **3**, 36–71 (1958)

Levins, R.: Evolution in changing environments. Monographs in population biology. Princeton: Princeton University Press 1968

Lewontin, R. C.: The apportionment of human diversity. In: Evolutionary biology, vol. 6, p. 381–398, Th. Dobzhansky, ed. New York: Appleton-Century-Crofts 1972

Pielou, E. C.: The measurement of diversity in different types of biological collections. J. theoret. Biol. **13**, 131–144 (1966)

Pielou, E. C.: The use of information theory in the study of the diversity of biological populations. Proc. 5th Berkeley Symposium on Math. Stat. and Prob. **4**, 163–177 (1967)

Pielou, E. C.: An introduction to mathematical ecology. New York: Wiley 1969

Pielou, E. C.: Niche width and niche overlap: a method for measuring them. Ecology **53**, 687–692 (1972)

Shannon, C. E., Weaver, W.: The mathematical theory of communication. Urbana, Ill.: University of Illinois Press 1949

Dr. J. David Allan
Department of Zoology
University of Maryland
College Park, Maryland 20742, USA