Letter to the Editor

Difference and Diversity of Plant Pathogen Populations: A New Approach for Measuring

E. Kosman

School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Ramat Aviv, Tel-Aviv 69978, Israel.

I thank the Colton Foundation, the Gileady Programme, and the Immigrant Absorption Fund of Tel-Aviv University for their support. I thank D. Steinberg (School of Mathematical Sciences, Tel-Aviv University), Z. Eyal (Institute for Cereal Crops Improvement, Tel-Aviv University), and A. Eshel (Faculty of Life Sciences, Tel-Aviv University) for critically reading the manuscript. I thank the anonymous reviewers for their many valuable comments and helpful suggestions.

Accepted for publication 31 July 1996.

The diversity of pathogen populations is one of their important intrinsic characteristics. According to Groth and Roelfs (3), an optimal diversity index should satisfy several conditions. A pathogen population is more diverse (diversity index is higher) if (i) it consists of a larger number of phenotypes for a given number of isolates; (ii) it is characterized by an even distribution of phenotypes; and (iii) the number of differences in virulence between phenotypes is larger. Shannon's entropy and Simpson's index are two diversity indices that are commonly used in plant pathology studies. They are based on the relative frequencies of different races. However, they consider nonidentical phenotypes as equally distinct and ignore similarities of different races. Nevertheless, some isolates exhibit distinct virulence patterns, while others are quite similar. Therefore, a diversity index that also reflects this relationship between isolates would provide more information. The hierarchical version of Shannon's index (3,8) and the "weighted mean proportion" (6) were proposed to incorporate a degree of phenotypic similarity into a measurement of diversity of pathogen populations. However, grouping of phenotypes in the case of the hierarchical index of Shannon is rather subjective, which makes comparative analysis of results difficult. On the other hand, the "weighted mean proportion" is based on the "all possible comparisons" of all isolates and, in fact, it is calculated without regard to a phenotypic structure of population.

Rogers' index and Nei's standard genetic distance are two indices that are commonly used for comparison between two populations. Rogers' index is based on the frequencies of the phenotypes that occur in the populations, regardless of how many virulences these phenotypes share. Thus, Rogers' index does not account for close similarities of pathogen phenotypes that may differ from one another only by one, or a few, mutations. On the other hand, Nei's distance, as applied to plant pathogens, measures frequencies of individual virulences without regard to phenotype. Thus, Nei's distance is better suited for randomly mating sexual populations than for most plant pathogens that undergo asexual reproduction in addition to, or in place of, sexual reproduction. A more appropriate index of genetic distance for plant pathogens would take into account both phenotypic frequencies and degrees of similarity among distinct phenotypes. The "index of mean differences" (6) was proposed to provide a measure of phenotypic dissimilarity as a component of distance between populations. However, it is based on "all possible comparisons" for all pairs of isolates between two populations, and it does not consider how many similar phenotypes these populations share.

The object of this paper is to describe a new index of genetic distance between populations and a related index of genetic diversity within populations that take into account both phenotypic frequencies and degrees of similarity among distinct phenotypes. These new indices are particularly suited to species with populations made up of clonal lineages resulting from widespread asexual reproduction.

INDEX OF GENETIC DISTANCE BETWEEN POPULATIONS

The data obtained from testing the virulence of a plant pathogen to a special set of k differentials are considered. The distance between two isolates is defined as the number of differentials on which the isolates respond differently. The distance between two pathogen populations, A and B, is then defined as follows. To each isolate sampled from one population an isolate sampled from the second population is matched so as to minimize the sum of distances between corresponding pairs of isolates. Finding the best matches is known as the "assignment problem" in operations research (1), and a number of algorithms have been developed to minimize the sum of distances (2; Appendix).

Thus, the idea is to take an equal number of isolates, n, from the two populations and match up the samples of phenotypes so that there are as few discordant loci (across pairs) as possible. The distance between the populations is then calculated as the sum of distances between the matched pairs of isolates. If the samples of isolates from the two populations are phenotypically identical, then the distance between them is 0. If the samples are different, then the distance between the populations is greater than 0. The distance reaches its maximum value if, and only if, each population has a single virulence phenotype, and the phenotypes of the two populations are absolutely different, i.e., the isolates respond differently on the entire set of differentials. It is possible to normalize the distance, so that it ranges from 0 to 1, by dividing the obtained minimum value of the sum of distances between matched pairs of isolates, $Ass_{\min}(A,B)$, by the product of the number of differentials, k, and the number of matched pairs, n_p:

$$K = \frac{Ass_{\min}(A,B)}{n_p k} \quad (1)$$

in which $n_p = n$, because both samples are of equal size.

Rogers' index and Nei's distance, taken separately, provide information concerning only one aspect of diversity and similarity between the two populations, and both have to be used jointly to

Corresponding author: E. Kosman; E-mail address: kosman@math.tau.ac.il

Publication no. P-1996-0930-01O

obtain relatively reliable results of comparison (4,5). In contrast, the newly proposed K distance provides a comparison between the two populations on the basis of "optimal" matching phenotypes with similar virulence patterns. Therefore, K distance enables the detection of nuances in the comparison between populations that cannot be revealed even when Rogers' index and Nei's distance are applied together.

I will compare properties of Rogers' index, Nei's distance, and K distance. Rogers' index (9) is calculated by the formula

$$R = \frac{1}{2}\sum_{i=1}^{n}|p_{Ai} - p_{Bi}| \quad (2)$$

in which p_{Ai} and p_{Bi} are the frequencies of the ith race in populations A and B, respectively, and n is the total number of different phenotypes in both populations. Rogers' index varies from 0, for two populations with identical race structure, to 1, for populations with no phenotypes in common.

Nei's standard genetic distance (7) is aimed at measuring the difference between two populations on the basis of the frequencies of alleles at a number of genetic loci. For dimorphic loci, let p_{Ai} and $1 - p_{Ai}$ and p_{Bi} and $1 - p_{Bi}$ denote the frequencies of the two alleles at the ith locus in populations A and B, respectively. Nei's standard genetic distance is defined by the formula

$$N = -\ln\frac{J_{AB}}{\sqrt{J_A J_B}} \quad (3)$$

in which, in the case of L loci with just two alleles per locus,

$$J_{AB} = \frac{1}{L}\sum_{i=1}^{L}\left[p_{Ai}p_{Bi} + (1-p_{Ai})(1-p_{Bi})\right]$$

$$J_A = \frac{1}{L}\sum_{i=1}^{L}\left[p^2{}_{Ai} + (1-p_{Ai})^2\right]$$

$$J_B = \frac{1}{L}\sum_{i=1}^{L}\left[p^2{}_{Bi} + (1-p_{Bi})^2\right]$$

Nei's distance values may vary between 0, when the two populations have identical frequencies of alleles over all the loci tested, and infinity, if the two populations share no alleles.

Rogers' index, Nei's distance, and K distance fulfill the following relations. (i) Rogers' index for two populations reaches its minimum value ($R = 0$) if, and only if, K distance between them is also minimal ($K = 0$). (ii) The minimum value of Rogers' index ($R = 0$) or K distance ($K = 0$) implies the minimum of Nei's distance ($N = 0$), whereas the opposite is incorrect (an example is considered below). (iii) Nei's distance between two populations is maximal if, and only if, the K distance also reaches its maximum. (iv) The maximum value of Nei's or K distances always results in the maximum of Rogers' index, but, conversely, the maximum value of Rogers' index does not necessarily imply the maximum of either Nei's or K distances.

The following numerical example distinctly shows the different approaches in the analysis of the data. Consider three isolate populations, A, B, and C, each of which consists of two isolates (a_1, a_2; b_1, b_2; and c_1, c_2) with phenotypes as follows:

A	B	C
a_1: 11111001	b_1: 11111100	c_1: 11110001
a_2: 11110110	b_2: 11110011	c_2: 11111110

The dichotomous structure of the alleles at eight genetic loci is identical for these three populations, even though they have no phenotypes in common. It is apparent that the isolate pairs a_1 and c_1, a_2 and c_2, b_1 and c_2, and b_2 and c_1 differ in just one out of the eight virulence loci. On the other hand, all isolates in population A are different from isolates in population B in two virulence loci.

Nei's distance between pairs of given populations equals 0, $N(A,B) = N(A,C) = N(B,C) = 0$. Therefore, according to the Nei distance, these three populations do not differ. Rogers' index assumes its maximal value in cases in which populations have no phenotypes in common. Hence, $R(A,B) = R(A,C) = R(B,C) = 1$, and according to Rogers' index these three populations are absolutely different. Therefore, if only Rogers' index and Nei's distance are used together in the analysis, it is impossible to distinguish between three pairs of given populations. The K distance differentiates between these populations.

Calculate the K distance between given populations. If the populations A and B are considered, first determine the distances between all pairs of isolates, one of which is taken from A and the other from B:

$d(a_1,b_1) = 2$, because a_1 and b_1 are different at two loci, 6 and 8

$d(a_1,b_2) = 2$, because a_1 and b_2 are different at two loci, 5 and 7

$d(a_2,b_1) = 2$, because a_2 and b_1 are different at two loci, 5 and 7

$d(a_2,b_2) = 2$, because a_2 and b_2 are different at two loci, 6 and 8

According to the definition of K distance, only two possibilities exist for matching isolates from the populations A and B. The two corresponding pairs are (i) a_1 and b_1, a_2 and b_2, or (ii) a_1 and b_2, a_2 and b_1. Hence, $Ass_{\min}(A,B) = 4$ is the minimum of $d(a_1,b_1) + d(a_2,b_2) = 4$ and $d(a_1,b_2) + d(a_2,b_1) = 4$, and the K distance between the two populations, A and B, is $K(A,B) = 4/(2 \times 8) = {}^1/_4$. In the same way, $Ass_{\min}(A,C) = 2$ is the minimum of $d(a_1,c_1) + d(a_2,c_2) = 1 + 1 = 2$ and $d(a_1,c_2) + d(a_2,c_1) = 3 + 3 = 6$, and $Ass_{\min}(B,C) = 2$ is the minimum of $d(b_1,c_1) + d(b_2,c_2) = 3 + 3 = 6$ and $d(b_1,c_2) + d(b_2,c_1) = 1 + 1 = 2$. Thus, K distances between A and C, and B and C are $K(A,C) = 2/(2 \times 8) = {}^1/_8$ and $K(B,C) = 2/(2 \times 8) = {}^1/_8$, respectively. Therefore, the K distance reveals that the populations A and C are more similar than A and B.

The following remark relates to estimating the dissimilarity between populations in the case in which samples A and B, from two populations, are of different sizes. When sample sizes are not equal, one solution is to expand their sizes. For example, sample A is expanded to a larger sample A', in which the number of isolates of each phenotype is multiplied by some number r_A. Thus, if there are n_A isolates in sample A, there will be $r_A n_A$ isolates in sample A'. Similarly, the size of sample B is expanded from n_B isolates to $r_B n_B$ isolates in sample B'. The numbers r_A and r_B are chosen to yield the lowest number divisible by both n_A and n_B (that is, the lowest common multiple, $\mathrm{lcm}(n_A,n_B)$, of n_A and n_B). Therefore, $\mathrm{lcm}(n_A,n_B) = r_A n_A$, $\mathrm{lcm}(n_A,n_B) = r_B n_B$, and the generated samples A' and B' will contain r_A and r_B copies of each isolate from A and B, respectively. For example, for $n_A = 12$ and $n_B = 15$, $r_A = 5$ and $r_B = 4$, so that the size of A' and B' is $\mathrm{lcm}(12,15) = 60$. It is obvious that the original relative frequencies of phenotypes in A and B are preserved in the new samples A' and B'. Hence, formula 1 adjusted for samples A' and B' can be used to calculate the K distance between populations:

$$K = \frac{Ass_{\min}(A',B')}{n_p k} \quad (1')$$

in which the number of matched pairs, n_p, equals the size $\mathrm{lcm}(n_A,n_B)$ of samples A' and B', and k is the number of differentials.

While this approach works, it could be very cumbersome. For example, suppose $n_A = 50$ and $n_B = 51$. This would require expanding samples A and B to 2,550 isolates with 51 copies of each isolate from sample A and 50 copies of each isolate from sample B. It would take a lot of time and computer memory to make the comparisons among so many isolates. The following solution can be proposed in this case. Since each sample is presumably a random sample from the population it represents, one can randomly

select 50 isolates from sample B to compare with the 50 isolates in sample A and calculate the index. Such indices can be calculated several times with different random samples of 50 isolates from sample B. Then the average value of the indices can be used as genetic distance between the two populations.

INDEX OF GENETIC DIVERSITY WITHIN POPULATIONS

Diversity indices that are commonly used in plant pathology studies consider nonidentical phenotypes as equally distinct and ignore the number of differences in virulence between isolates, property 3 of Groth and Roelfs (3). Such indices are Shannon's entropy (*Sh*) (8)

$$Sh = -\sum_{i=1}^{n} p_i \ln p_i \tag{4}$$

and Simpson's index (*Si*) (10)

$$Si = 1 - \sum_{i=1}^{n} p^2{}_i \tag{5}$$

in which p_i is the frequency of the ith phenotype, and n is the total number of distinct phenotypes. The "assignment problem" can be used for generating a diversity index that also considers similarity of response patterns of isolates.

Consider an arbitrary sample from a pathogen population, A, that is presented by n patterns of dichotomous responses of its isolates to some set of k factors. The distance between two isolates equals the number of factors to which the isolates respond differently. To each sampled isolate, match an isolate from the population to make up n pairs so as to maximize the sum of distances between corresponding pairs of isolates. Finding such matches can be realized by the solution of the appropriate "assignment problem." The proximity between two isolates is defined as the number of differentials on which the isolates respond similarly. Thus, proximity = k – distance. The sum of distances between matched pairs of isolates is maximal, so the sum of proximities is minimal. Therefore, if $Ass^p{}_{\min}(A,A)$ is the obtained minimum value of the sum of proximities, then the maximum value of the sum of distances equals $Ass_{\max}(A,A) = nk - Ass^p{}_{\min}(A,A)$.

The new diversity index, *Ko*, of the given pathogen population A can be determined by dividing the obtained maximum value of the sum of distances between matched pairs of isolates, $Ass_{\max}(A,A)$, by the product of the number of differentiating factors, k, and the number of sampled isolates, n:

$$Ko = \frac{Ass_{\max}(A,A)}{nk} \tag{6}$$

The diversity index, defined in such a way, ranges from 0 to 1. The only case in which the diversity score is assigned 0 corresponds to a population that is limited to a single phenotype. The maximal diversity score is assigned to a population with phenotypes occurring in equal frequencies, provided that the sample consists of pairs of isolates with absolutely different response patterns on the entire set of differentiating factors (in particular, for a population with all 2^k possible phenotypes).

Shannon's and Simpson's indices also maintain the following properties: (i) when a given set of phenotypes is prescribed, they assign maximum diversity to a population that is divided equally among them (this is, in fact, property 2 of Groth and Roelfs [3]); and (ii) if the number of phenotypes in two populations is different, and the phenotypes within each population are equally distributed, the higher diversity index must be assigned to the population with more phenotypes (this is property 1 of Groth and Roelfs [3], if evenness of phenotype distribution is fulfilled). Both these properties are not generally compelling when the distance between isolates (property 3 of Groth and Roelfs [3]) is also considered. The extent of the similarity among isolates contributes considerably to the diversity within a population and is not less important than the relative frequencies of different phenotypes. This fact will be further demonstrated by examples.

Thus, the *Ko* index preserves some of the properties of Shannon's and Simpson's diversity indices, but not all of them in their absolute form. It enables one to reveal nuances in structure of populations, which cannot be established by these two indices.

The following examples demonstrate properties of *Ko* diversity index as compared with those of Shannon and Simpson. Three isolate populations, A, B, and C, are considered, each of which consist of two isolates tested on six host differentials with phenotypes as follows:

A	B	C
a_1: 111000	b_1: 111100	c_1: 111110
a_2: 000111	b_2: 001111	c_2: 011111

Calculate the *Ko* diversity index for population B. Four different ordered pairs of isolates are possible: (b_1,b_1), (b_1,b_2), (b_2,b_1), and (b_2,b_2). The distances between the paired isolates are determined as follows:

$d(b_1,b_1) = d(b_2,b_2) = 0$,
because these two pairs consist of identical isolates

$d(b_1,b_2) = d(b_2,b_1) = 4$,
because b_1 and b_2 are different at four loci, 1, 2, 5, and 6

Only two possibilities exist for matching isolates from population B to themselves. The two corresponding pairs are (i) b_1 and b_1, b_2 and b_2, or (ii) b_1 and b_2, b_2 and b_1. The solution of the "assignment problem" (matching isolates for maximum) is the maximum of $d(b_1,b_1) + d(b_2,b_2) = 0$ and $d(b_1,b_2) + d(b_2,b_1) = 8$, i.e., $Ass_{\max}(B,B) = 8$. Dividing this value by $2 \times 6 = 12$, the *Ko* diversity index for population B is $Ko(B) = {}^2/_3$. In the same way, $Ass_{\max}(A,A) = 12$ and $Ass_{\max}(C,C) = 4$ for populations A and C, respectively. Dividing by 12, the *Ko* diversity indices for these populations are $Ko(A) = 1$ and $Ko(C) = {}^1/_3$. Thus, *Ko* index reveals the difference in diversity of the considered populations, which results from the difference in the measure of dissimilarity between isolates within these populations. In contrast, both Shannon's and Simpson's indices are equal for the given populations: $Sh(A) = Sh(B) = Sh(C) = \ln 2$ and $Si(A) = Si(B) = Si(C) = 0.5$.

The following example relates to the assignment of the maximal score of *Ko* diversity index. Consider the response patterns to three host differentials of isolates from four pathogen populations:

D_1	D_2	D_3	D_4
110	110	110	000
110	110	001	100
110	001	100	010
110	001	011	001
001	100	010	110
001	100	101	101
001	011		011
001	011		111

The last population, D_4, consists of all $2^3 = 8$ possible phenotypes occurring in equal frequencies. Both Shannon's and Simpson's indices establish the difference in diversity among the given populations: $Sh(D_1) = \ln 2$, $Sh(D_2) = \ln 4$, $Sh(D_3) = \ln 6$, and $Sh(D_4) = \ln 8$, and $Si(D_1) = {}^1/_2$, $Si(D_2) = {}^3/_4$, $Si(D_3) = {}^5/_6$, and $Si(D_4) = {}^7/_8$. Moreover, according to these results, diversity increases from populations D_1 to D_4. In contrast, *Ko* index is maximal (equals 1) for all these populations, i.e., they are considered as equally diverse.

Additional insight into the comparison can be obtained by considering the collections of pairwise distances in each population. If d_{ij} is the distance between isolates i and j, then the average distance between the isolates of each sample can be calculated by the formula

$$\bar{d}_k = \bar{d}(D_k) = \frac{1}{n_k^2} \sum_{i,j=1}^{n_k} d_{ij} \quad , k = 1, 2, 3, 4$$

Here, n_k is the number of isolates, so that there are n_k^2 entries in the distance matrix. The absolute deviation from the average distance can be expressed by the formula

$$s_k = s(D_k) = \frac{1}{n_k^2} \sum_{i,j=1}^{n_k} \left| d_{ij} - \bar{d}_k \right| \quad , k = 1, 2, 3, 4$$

According to these formulas, $\bar{d}_1 = \bar{d}_2 = \bar{d}_3 = \bar{d}_4 = 1.5$ and $s_1 = 1.5$, $s_2 = 1$, $s_3 = 0.83$, and $s_4 = 0.75$. Thus, the average distance is the same for all these samples, whereas the deviation decreases from populations D_1 to D_4. Therefore, a greater number of different phenotypes in a population increases its level of diversity. However, the measure of variance between distinct phenotypes decreases from population D_1 to population D_4. Therefore, the former effect is compensated by the latter one, which enables one to suppose the equal maximal level of diversity for all these populations.

Consider two populations, A and B, each of which consists of three isolates tested on eight host differentials with phenotypes as follows:

A	B
a_1: 00111111	b_1: 00000001
a_2: 11000000	b_2: 11000000
a_3: 10000000	b_3: 10000000

The relative frequencies of the different phenotypes within each population are equal ($p_i = 1/3$, $i = 1, 2, 3$), and these populations are equally diverse according to the indices of Shannon and Simpson. The *Ko* index is different for these populations: $Ko(A) = 16/(3 \times 8) = 2/3$ and $Ko(B) = 6/(3 \times 8) = 1/4$; and, therefore, population A is more diverse than population B.

Two populations, A_1 and A_2, are made up of population A to which an additional isolate was duplicated ($a_4 = a_3$ in A_1 and $a_4 = a_1$ in A_2):

A_1	A_2
a_1: 00111111	a_4: 00111111
a_2: 11000000	a_1: 00111111
a_3: 10000000	a_2: 11000000
a_4: 10000000	a_3: 10000000

Thus, they consist of three distinct phenotypes with unequal relative frequencies: $p_1 = p_2 = 1/4$, and $p_3 = 1/2$. It is easy to see that populations A_1 and A_2 are equally diverse according to Shannon and Simpson, and they are less diverse than A. *Ko* index gives the following values: $Ko(A_1) = 16/(4 \times 8) = 1/2$ and $Ko(A_2) = 30/(4 \times 8) = 15/16$, i.e., population A_2 is considerably more diverse than A_1. But the most important result is a direct consequence of the inequality $Ko(A_1) < Ko(A) < Ko(A_2)$: when a given set of phenotypes is prescribed, *Ko* index does not necessarily assign maximum diversity to a population that is divided equally among them. The *Ko* index for population A is not maximal, even though the three phenotypes occur at equal frequencies, because $Ko(A_2) = 15/16 > 2/3 = Ko(A)$, and population A_2 consists of the same phenotypes as population A with unequal frequencies. It seems reasonable to regard the two similar phenotypes in A_2 as "virtually the same" and, thus, to regard population A_2 as being composed of equal proportions of two very different phenotypes. Likewise, it seems reasonable to regard population A as being composed essentially of two phenotypes, one at $2/3$ and one at $1/3$ frequency. By this reasoning, one should regard population A_2 as more diverse than population A, even though Shannon's and Simpson's indices lead to the opposite conclusion.

One more characteristic property of Shannon's and Simpson's diversity indices is not generally implemented for *Ko* index: a higher score of *Ko* index is not necessarily assigned to the population with a greater set of phenotypes, if the relative frequencies of the different phenotypes in each of two populations are equal. This can be exemplified in the following populations:

A	B
a_1: 10000000	b_1: 00001111
a_2: 11000000	b_2: 11110000
a_3: 01000000	

According to Shannon's and Simpson's indices, population A is more diverse than B. In contrast, $Ko(A) = 4/(3 \times 8) = 1/6$ is considerably less than $Ko(B) = 16/(2 \times 8) = 1$, which infers a higher level of diversity for population B as compared with A.

APPENDIX

Some computer programs for the "assignment problem" can be found in the "List of Interesting Optimization Codes in Public Domain," which is compiled by Jiefeng Xu of the University of Colorado at Boulder (xu@benji.colorado.edu). This list has a bias towards Unix platforms. The codes can be retrieved via anonymous ftp: (i) Site: ftp://netlib.att.com in /netlib/toms, File: 548.Z, Language: Fortran; and (ii) Site: ftp://ftp.bilkent.edu.tr at /pub/IEOR/Opt/goldberg directory, File: csas.tar.z, Language: C.

The reader needs to prepare a matrix of distances between all pairs of isolates as input to the programs.

LITERATURE CITED

1. Bellman, R., Cooke, K. L., and Lockett, J. A. 1970. Algorithms, Graphs and Computers. Academic Press, Inc., New York.
2. Carpaneto, G., and Toth, P. 1980. Algorithm 548. Solution of the assignment problem [H]. ACM (Assoc. Comput. Mach.) Trans. Math. Software 6:104-111.
3. Groth, J. V., and Roelfs, A. P. 1987. The concept and measurement of phenotypic diversity in *Puccinia graminis* on wheat. Phytopathology 77:1395-1399.
4. Leonard, K. J., Roelfs, A. P., and Long, D. L. 1992. Diversity of virulence within and among populations of *Puccinia recondita* f. sp. *tritici* in different areas of the United States. Plant Dis. 76:500-504.
5. Long, D. L., Roelfs, A. P., Leonard, K. J., and Roberts, J. J. 1993. Virulence and diversity of *Puccinia recondita* f. sp. *tritici* in the United States in 1991. Plant Dis. 77:786-791.
6. McCain, J. W., Groth, J. V., and Roelfs, A. P. 1992. Inter- and intrapopulation isozyme variation in collections from sexually reproducing populations of the bean rust fungus, *Uromyces appendiculatus*. Mycologia 84:329-340.
7. Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283-292.
8. Pielou, E. C. 1975. Ecological Diversity. John Wiley & Sons, Inc., New York.
9. Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. Pages 145-153 in: Studies in Genetics. University of Texas, Austin.
10. Simpson, E. N. 1949. Measurement of diversity. Nature 163:688.