Binary (Presence-Absence) Similarity Coefficients
Author(s): Alan H. Cheetham and Joseph E. Hazel
Reviewed work(s):
Source: *Journal of Paleontology*, Vol. 43, No. 5 (Sep., 1969), pp. 1130-1136
Published by: SEPM Society for Sedimentary Geology
Stable URL: http://www.jstor.org/stable/1302424
Accessed: 14/04/2012 21:54

BINARY (PRESENCE-ABSENCE) SIMILARITY COEFFICIENTS

ALAN H. CHEETHAM AND JOSEPH E. HAZEL

Smithsonian Institution and U.S. Geological Survey, Washington, D.C.

ABSTRACT—Various numerical coefficients have been employed in comparisons of taxa or bioassociational units, especially in studies involving large arrays of multivariate data. Nomenclatural and conceptual difficulties have arisen from the translation of coefficients from variable-oriented (R-mode) to case-oriented (Q-mode) use. This paper points out the conceptual relationships and some of the properties of coefficients expressing similarity or difference between units defined by binary data.

INTRODUCTION

IN RECENT years there has been increased use in systematic biology of large arrays of multivariate data for both taxonomic and bioassociational (including biogeographic, ecologic, paleoecologic, and biostratigraphic) interpretations. Many taxa and bioassociational units are polythetic in concept, i.e., distinctive in the aggregate of their attributes, one or more of which they share with other such units. Where either the number of examples or cases composing the units or the number of attributes describing them is large, adequate comparison among units becomes more and more dependent upon numerical coefficients of similarity or difference. Various numerical indices have been employed in systematic analyses of quantitative, multistate qualitative, and binary (presence-absence) data. The increasing availability of digital computers, and of programs for them which calculate these indices, doubtless will result in continued acceleration of the use of numerical methods of comparison.

Coefficients of association in studies of relationships between variables or attributes (R-mode) are a well-known part of standard statistical procedure. Translation of these measures to express relationships between examples or cases (Q-mode) of taxa or bioassociational units has resulted in both conceptual and nomenclatural complications. Our purpose is not to justify or criticize the use of any particular coefficient or coefficients, but rather to point out some of the conceptual and nomenclatural relationships among them.

Coefficients which express relationships of either similarity or difference between units defined by binary data are especially important in systematic biology because of the wide availability of data of this type, such as features present or absent in specimens or species present or absent in samples. This discussion is limited to binary coefficients. Some binary coefficients, however, are simply special cases of those, such as the Correlation Coefficient, that express similarity or difference between units based upon quantitative or multistate qualitative data. These relationships are included in the discussions of the binary coefficients to which they are applicable.

Most of the binary coefficients discussed here appear to have been developed intuitively and tested empirically. Those that are special cases of quantitative or multistate coefficients, however, share the mathematical derivation of their non-binary equivalents. Used in R-mode studies, some of these mathematically derived binary coefficients, such as the Phi Coefficient, have known probability distributions on which statistical inferences can be based. In Q-mode analysis however, coefficients do not estimate population parameters in an exactly analogous way, because the defining attributes are more than just an unbiased sample of a universe of possible attributes. Consequently, numerical comparison of taxa or bioassociational units is not, strictly speaking, a statistical procedure.

The properties of some of the coefficients listed here have been discussed in greater detail by Sokal & Sneath (1963, p. 121–157) for taxonomic applications and by Simpson (1960) for bioassociations. Not all coefficients are logically applied to both areas. Those incorporating negative matches (mutual absences) in the numerator are used commonly in taxonomic studies and rarely in bioassociational ones. Those excluding some mismatches (presences in one sample but not the other) from the denominator are common in bioassociational studies, but, as far as we know, have not been used in taxonomy.

Because of separate development of coefficients in the two areas of application, similar or identical coefficients have been proposed under different names, and the same name has been used for different coefficients. Unlike statistical coefficients, these similarity or difference measures have not been assigned standard symbols, and even the symbols used for their computation

are different in the two areas (see, for example, Simpson, 1960; Sokal & Sneath, 1963, p. 126). The following list presents these computational symbols, bioassociational on the left, taxonomic on the right. Each symbol refers to a number of taxa in bioassociational application or a number of characters in taxonomic application. To emphasize the equivalency of coefficients in the two areas of application, we have used both sets of symbols in the expressions that follow the list.

called the Dice Coefficient by Sokal & Sneath (1963, p. 128) and the Burt Coefficient or Pirlot Index by Peters (1968).

Other coefficients proposed for use in both taxonomy and bioassociational studies include the First Kulczynski Coefficient (Sokal & Sneath, 1963, p. 129; Peters, 1968),

$$\frac{C}{N_1 + N_2 - 2C} = \frac{n_{JK}}{n_{Jk} - n_{jK}} ; \qquad (4)$$

C	Present in both units compared	n_{JK}
E_1	Present in first but not second	n_{Jk}
E_2	Present in second but not first	n_{jK}
A	Absent in both (but present in others)	n_{jk}
$N_1(=C+E_1)$	Total present in first	$n_J(=n_{JK}+n_{Jk})$
$N_2(=C+E_2)$	Total present in second	$n_K(=n_{JK}+n_{jK})$
$N_t(=N_1+N_2-C)$	Total present in both units	(not used)
(N_t+A)	Total present in all units (where there are more than two)	n

In the following discussion we do not differentiate between coefficients in which the numerator is multiplied by 100 and those in which it is not. The former, expressed as percentages, are generally used for bioassociational indices, but not in taxonomic work.

COEFFICIENTS

The most widely used measure in both taxonomic and bioassociational studies is

$$\frac{C}{N_1 + N_2 - C} = \frac{n_{JK}}{n_{JK} + n_{Jk} + n_{jK}} , \qquad (1)$$

which was introduced by Jaccard (1908, p. 226–227) as the Coefficient of Community (see also Hagmeier & Stults, 1964; Peters, 1968). In numerical taxonomy, this coefficient is well known as the Jaccard Coefficient (Sokal & Sneath, 1963, p. 129), and under that name has also been used in bioassociational studies such as those of Kaesler (1966), Maddocks (1966), Valentine (1966), and Mello & Buzas (1968).

The Simple Matching Coefficient,

$$\frac{C + A}{N_t + A} = \frac{n_{JK} + n_{jk}}{n} , \qquad (2)$$

was proposed by Sokal & Michener (1958) for numerical taxonomy, but it has also been used by Kaesler (1966) in comparing bioassemblages. It can be used, of course, only where three or more assemblages are being compared; otherwise, A $(n_{jk}) = 0$, and expression (2) becomes identical with (1).

Another commonly used index in both taxonomy and bioassociational studies is

$$\frac{2C}{N_1 + N_2} = \frac{2n_{JK}}{2n_{JK} + n_{Jk} + n_{jK}} \qquad (3)$$

the Second Kulczynski Coefficient (Sokal & Sneath, 1963, p. 130; Peters 1968),

$$\frac{C(N_1 + N_2)}{2(N_1 N_2)} = \frac{n_{JK}}{2n_J} + \frac{n_{JK}}{2n_K} ; \qquad (5)$$

and the measure attributed to Otsuka by Ochiai (1957) and called the Ochiai Coefficient by Sokal & Sneath (1963, p. 130) and the Otsuka Coefficient by Peters (1968),

$$\frac{C}{\sqrt{N_1 N_2}} = \frac{n_{JK}}{\sqrt{n_J n_K}} . \qquad (6)$$

Expression (6), the Otsuka Coefficient, is a special case of the Coefficient of Proportional Similarity (Imbrie & Purdy, 1962, p. 355),

$$\cos \theta = \frac{\Sigma(X_{ij})(X_{ik})}{\sqrt{\Sigma(X_{ij})^2 \Sigma(X_{ik})^2}} , \qquad (7)$$

in which X_i (the character state in numerical taxonomy) may take on only the binary values 0 and 1. In the binary case, the product $(X_{ij})(X_{ik})$ is equal to 1 for positive matches (mutual presences) and to 0 for mismatches and negative matches (mutual absences), and thus the numerator is equal to the sum of the positive $(C = n_{JK})$. Each of the quantities $\Sigma (X_{ij})^2$ and $\Sigma (X_{ij})^2$ is in like manner equal to the total number of presences in each unit compared, that is, to $N_1 (= n_J)$ and $N_2 (= n_K)$, respectively; the denominator thus becomes $\sqrt{N_1 N_2} (= \sqrt{n_J n_K})$.

Measures of similarity proposed for use in bioassociational but not taxonomic studies include the Correlation Ratio (Sorgenfrei, 1959, p. 403),

$$\frac{C^2}{N_1 N_2} = \frac{(n_{JK})^2}{n_J n_K} , \qquad (8)$$

which is the square of expression (6); the Simp-

TABLE 1—Characteristics of coefficients of similarity and difference based on binary (presence-absence) data. Bioassociational symbols are used; see text for numerical taxonomic equivalents. The expressions in column 1 result when two samples have the same number of positive features (equal diversity); those in column 2 when they share no positive features; those in column 3 when they become identical in positive features; and those in column 4 when there are no shared negative features (mutual absences). Column 5 offers a comparison of the coefficients when one sample has twice as many positive features (presences) as the other and the number of features in common (mutual presences) is one half the number in the less positive (less diverse) sample.

Coefficient	1. As $N_1\to N_2$	2. As $C\to 0$	3. As $C\to N_t$	4. As $A\to 0$	5. If $\dfrac{N_1}{N_2}=\dfrac{1}{2}$ and $\dfrac{C}{N_1}=\dfrac{1}{2}$
1 Jaccard $\dfrac{C}{N_1+N_2-C}$	$\to \dfrac{C}{2E_1+C}$	$\to 0$	$\to 1$	unaffected	$1/5$
2 Simple M. $\dfrac{C+A}{N_t+A}$	$\to \dfrac{C+A}{2E_1+C+A}$	$\to \dfrac{A}{E_1+E_2+A}$	$\to 1$	$\to \dfrac{C}{N_t}$ [= Jaccard (1)]	$\dfrac{C+A}{5C+A}$ [if A=C, then =1/3]
3 Dice $\dfrac{2C}{N_1+N_2}$	$\to \dfrac{C}{N_2}$ [= Simpson (8)]	$\to 0$	$\to 1$	unaffected	$1/3$
4 1st Kulcz. $\dfrac{C}{N_1+N_2-2C}$	$\to \dfrac{C}{2E_1}$	$\to 0$	$\to \infty$	unaffected	$1/4$
5 2nd Kulcz. $\dfrac{C(N_1+N_2)}{2(N_1 N_2)}$	$\to \dfrac{C}{N_1}$ [= Simpson (8)]	$\to 0$	$\to 1$	unaffected	$3/8$
6 Otsuka $\dfrac{C}{\sqrt{N_1 N_2}}$	$\to \dfrac{C}{N_1}$ [= Simpson (8)]	$\to 0$	$\to 1$	unaffected	$1/\sqrt{8}$
8 Corr. Ratio $\dfrac{C^2}{N_1 N_2}$	$\to \dfrac{C^2}{(N_1)^2}$ [= (Simpson (8))²]	$\to 0$	$\to 1$	unaffected	$1/8$
9 Simpson $\dfrac{C}{N_1}$	$= \dfrac{C}{N_1}$	$\to 0$	$\to 1$	unaffected	$1/2$
10 Braun-Bl. $\dfrac{C}{N_2}$	$\to \dfrac{C}{N_1}$ [= Simpson (8)]	$\to 0$	$\to 1$	unaffected	$1/4$
11 Fager $\dfrac{C}{\sqrt{N_1 N_2}} - \dfrac{1}{2\sqrt{N_2}}$	$\to \dfrac{C}{N_1} - \dfrac{1}{2\sqrt{N_1}}$ [= Simpson (8) $- \dfrac{1}{2\sqrt{N_1}}$]	$\to -\dfrac{1}{2\sqrt{N_2}}$	$\to 1 - \dfrac{1}{2\sqrt{N_2}}$	unaffected	$\dfrac{1}{\sqrt{8}} - \dfrac{1}{4\sqrt{C}}$

12 Rogers & Tan. $\dfrac{C+A}{N_t+A+E_1+E_2}$	$\rightarrow\dfrac{C+A}{C+A+4E_1}$	$\rightarrow 1$	$\rightarrow\dfrac{A}{2(E_1+E_2)+A}$	$\rightarrow\dfrac{C}{N_t+E_1+E_2}$	$\dfrac{C+A}{9C+A}$ [If A=C, then $=1/5$]
13 Hamann $\dfrac{(C+A)-(E_1+E_2)}{N_t+A}$	$\rightarrow\dfrac{C+A-2E_1}{C+A+2E_1}$	$\rightarrow 1$	$\rightarrow\dfrac{A-(E_1+E_2)}{A+(E_1+E_2)}$	$\rightarrow\dfrac{C-(E_1+E_2)}{N_t}$	$\dfrac{A-3C}{A+5C}$ [If A=C, then $=-1/3$]
14 Yule $\dfrac{CA-E_1E_2}{CA+E_1E_2}$	$\rightarrow\dfrac{CA-(E_1)^2}{CA+(E_1)^2}$	$\rightarrow 1$	$\rightarrow -1$	$\rightarrow -1$	$\dfrac{A-3C}{A+3C}$ [if A=C, then $=-1/2$]
15 Phi $\dfrac{CA-E_1E_2}{\sqrt{N_1N_2(E_1+A)(E_2+A)}}$	$\rightarrow\dfrac{CA-(E_1)^2}{N_1(E_1+A)}$	$\rightarrow 1$	$\rightarrow\dfrac{-E_1E_2}{\sqrt{(E_1{}^2+A)(E_2{}^2+A)}}$	$\rightarrow\dfrac{-E_1E_2}{\sqrt{N_1N_2E_1E_2}}$	$\dfrac{A-3C}{C\sqrt{8(C+A)(3C+A)}}$ $\left[\text{if A=C, then }=-\dfrac{1}{4C}\right]$
17 Resem. Eq. $\left(\dfrac{N_1}{N_t}\right)^{1/z}+\left(\dfrac{N_2}{N_t}\right)^{1/z}=1$	$\rightarrow 2\left(\dfrac{N_1}{N_t}\right)^{1/z}=1$	$\rightarrow (1)^{1/z}+(1)^{1/z}=1*$	$\rightarrow\left(\dfrac{E_1}{E_1+E_2}\right)^{1/z}+\left(\dfrac{E_2}{E_1+E_2}\right)^{1/z}=1$	unaffected	$(\tfrac{2}{5})^{1/z}+(\tfrac{4}{5})^{1/z}=1$
18 Coeff. Diff. $1-\dfrac{C}{N_2}$	$\rightarrow 1-\dfrac{C}{N_1}$ [=1−Simpson (8)]	$\rightarrow 1$		unaffected	3/4
19 No. Feat. Diff. E_1+E_2	$\rightarrow 2E_1$	$\rightarrow N_1+N_2$		unaffected	4C
22 Sokal Dist. (Binary Case) $\sqrt{\dfrac{E_1+E_2}{N_t+A}}$	$\rightarrow\sqrt{\dfrac{2E_1}{C+A+2E_1}}$	$\rightarrow 0$	$\rightarrow\sqrt{\dfrac{E_1+E_2}{E_1+E_2+A}}$	$\rightarrow\sqrt{\dfrac{E_1+E_2}{N_t}}$	$\sqrt{\dfrac{4C}{5C+A}}$ [if A=C, then $=\sqrt{\tfrac{2}{3}}$]
22a Sokal Dist. (Binary Case) $\sqrt{1-\dfrac{C+A}{N_t+A}}$	$\rightarrow\sqrt{1-\dfrac{C+A}{C+A+2E_1}}$	$\rightarrow 0$	$\rightarrow\sqrt{1-\dfrac{A}{E_1+E_2+A}}$	$\rightarrow\sqrt{1-\dfrac{C}{N_t}}$	$\sqrt{1-\dfrac{C+A}{5C+A}}$ [if A=C, then $=\sqrt{\tfrac{2}{3}}$]
	$=\sqrt{1-\text{Simple M. (2)}}$			$=\sqrt{1-\text{Jaccard (1)}}$	

* When the two samples being compared are identical the equation is insoluble.

son Coefficient (Simpson, 1943; 1947; 1960; Cheetham & Deboo, 1963; Hagmeier & Stults, 1964; Peters, 1968),

$$\frac{C}{N_1} = \frac{n_{JK}}{n_J}, \quad \text{where} \quad N_1 \leqq N_2, \quad n_J \leqq n_K; \quad (9)$$

the coefficient introduced by Braun-Blanquet (1932, p. 632) and referred to by Peters (1968) as the Jaccard Coefficient,

$$\frac{C}{N_2} = \frac{n_{JK}}{n_K}, \quad \text{where} \quad N_1 \leqq N_2, \quad n_J \leqq n_K; \quad (10)$$

and the Fager Coefficient,

$$\frac{C}{\sqrt{N_1 N_2}} - \frac{1}{2\sqrt{N_2}} = \frac{n_{JK}}{\sqrt{n_J n_K}} - \frac{1}{2\sqrt{n_K}},$$
$$\text{where } N_1 \leqq N_2, \quad n_J \leqq n_K, \quad (11)$$

which was introduced by Fager in Fager & McGowan (1963) and subsequently has been used by Elliot (1963) and Valentine & Peddicord (1967).

Measures of similarity which have been proposed for use in taxonomy but not in bioassociational studies include the Rogers and Tanimoto Coefficient (Sokal & Sneath, 1963, p. 129),

$$\frac{C + A}{N_t + A + E_1 + E_2} = \frac{n_{JK} + n_{jk}}{n + n_{Jk} + n_{jK}}; \quad (12)$$

the Hamann Coefficient (Sokal & Sneath, 1963, p. 132),

$$\frac{(C + A) - (E_1 + E_2)}{N_t + A} = \frac{(n_{JK} + n_{jk}) - (n_{Jk} + n_{jK})}{n}; \quad (13)$$

the Yule Coefficient (Sokal & Sneath, 1963, p. 132),

$$\frac{CA - E_1 E_2}{CA + E_1 E_2} = \frac{n_{JK} n_{jk} - n_{Jk} n_{jK}}{n_{JK} n_{jk} + n_{Jk} n_{jK}}; \quad (14)$$

and the Phi Coefficient (Sokal & Sneath, 1963, p. 132),

$$\frac{CA - E_1 E_2}{\sqrt{N_1 N_2 (E_1 + A)(E_2 + A)}}$$
$$= \frac{n_{JK} n_{jk} - n_{Jk} n_{jK}}{\sqrt{n_J n_K (n_{Jk} + n_{jk})(n_{jK} + n_{jk})}}. \quad (15)$$

Expression (15), the Phi Coefficient, is a special case of the Pearson Product-Moment Correlation Coefficient, computed by the formula (Sokal & Sneath, 1963, p. 296)

pression (7), demonstrates the equivalence of the Phi and Correlation Coefficients.

In addition to the foregoing measures of similarity, three coefficients of difference have been proposed for computation from binary data. In bioassociational studies, the coefficient z is calculated from the Resemblance Equation,

$$\left(\frac{N_1}{N_T}\right)^{1/z} + \left(\frac{N_2}{N_T}\right)^{1/z} = 1, \quad (17)$$

$$\left(\frac{n_{Jk}}{n - n_{jk}}\right)^{1/z} + \left(\frac{n_{jK}}{n - n_{jk}}\right)^{1/z} = 1. \quad (17a)$$

The Resemblance Equation was proposed by Preston (1962), and Hagmeier & Stults (1964) used $1 - z$ as a measure of similarity. The Coefficient of Difference (Savage, 1960; Peters, 1968),

$$1 - \frac{C}{N_2} = 1 - \frac{n_{JK}}{n_K}, \quad \text{where} \quad N_1 \leqq N_2, \ n_J = n_K \quad (18)$$

is simply one minus expression (10).

The Number of Features of Difference (NFD) used by Stephenson, Williams, & Lance (1968, p. 4, 5),

$$E_1 + E_2 = n_{Jk} + n_{jK}, \quad (19)$$

is a special case of the squared Taxonomic Distance of Sokal (1961) and Sokal & Sneath (1963, p. 147),

$$\sum_{i=1}^{n} (X_{ij} - X_{ik})^2, \quad (20)$$

in which Xi may take on only the binary values 0 and 1; in this case, the value $(X_{ij}-X_{ik})^2$ is equal to 0 for matches and 1 for mismatches, and the squared Taxonomic Distance is equal to the sum of the mismatches, as given in expression (19). As pointed out by Sokal & Sneath (1963, p. 147), this value increases with the number of characters used in comparison, so that in practice, an average Taxonomic Distance (commonly called the Sokal Distance),

$$\sqrt{\frac{\sum_{i=1}^{n} (X_{ij} - X_{ik})^2}{n}}, \quad (21)$$

$$r = \frac{\Sigma(X_{ij})(X_{ik}) - \dfrac{1}{n} (\Sigma X_{ij})(\Sigma X_{ik})}{\sqrt{\left[\Sigma(X_{ij})^2 - \dfrac{1}{n} (\Sigma X_{ij})^2\right]\left[\Sigma(X_{ik})^2 - \dfrac{1}{n} (\Sigma X_{ik})^2\right]}}. \quad (16)$$

In the binary case, substitution of the quantities 0 and 1, as used above in the discussion of ex-

is used. In the binary case, the Sokal Distance becomes

$$\sqrt{\frac{E_1 + E_2}{N_t + A}} = \sqrt{\frac{n_{Jk} + n_{jK}}{n}}, \qquad (22)$$

which is equivalent to

$$\sqrt{1 - \frac{C + A}{N_t + A}} = \sqrt{1 - \frac{n_{JK} + n_{jk}}{n}} \qquad (22a)$$

that is, the square root of one minus expression (2).

PROPERTIES OF COEFFICIENTS

Table 1 lists the coefficients treated above under the names which appear to be the most appropriate for them. Some properties of the coefficients are indicated by variation of certain terms in the numerator and denominator. The expressions in column 1 result when two samples being compared have the same number of positive features (equal diversity). Under this condition, for example, the Dice, 2nd Kulczynski, Otsuka, and Braun-Blanquet Coefficients become identical with the Simpson Coefficient, and the Coefficient of Difference becomes one minus the Simpson Coefficient.

When there are no positive features (mutual presences) shared by two samples being compared (column 2), all similarity coefficients except those that use shared negative features (mutual absences) take on their minimum values. When two samples are identical in positive features (mutual presences), the similarity coefficients assume maximum (and difference coefficients, minimum) values, except for the Fager Coefficient which still has a negative constant, and the Resemblance Equation which becomes insoluble (column 3).

When two samples share no negative features (no mutual absences), the Simple Matching Coefficient becomes identical with the Jaccard Coefficient, and the Sokal Distance becomes the square-root of one minus the Jaccard Coefficient (column 4). Most other coefficients are unaffected.

Column 5 offers a comparison of the coefficients when one sample has twice as many positive features (presences) as the other, and the number of positive features in common (mutual presences) is one half the number in the less positive (less diverse) sample. Under these conditions, for example, the Jaccard Coefficient assumes a value of one fifth, and the Simpson Coefficient becomes one half. For purposes of comparison, we have considered coefficients closer to the Jaccard (that is, less than one third) than to the Simpson Coefficient to emphasize difference. On this basis, the 1st Kulczynski,

Fager, Braun-Blanquet, Rogers & Tanimoto, Hammann, Yule, and Phi Coefficients, the Correlation Ratio, the Coefficient of Difference, and the Sokal Distance, all are like the Jaccard Coefficient in emphasizing difference. The Dice, 2nd Kulczynski, and Otsuka Coefficients, like the Simpson Coefficient, emphasize similarity. The Simple Matching Coefficient behaves differently as the number of shared negative features (mutual absences) varies.

REFERENCES

BRAUN-BLANQUET, J. (transl. by G. D. Fuller and H. S. Conrad), 1932, Plant sociology: the study of plant communities: McGraw-Hill, New York, xvii+439 p., 180 text-figs.

CHEETHAM, A. H., & DEBOO, P. B., 1963, A numerical index for biostratigraphic zonation in the mid-Tertiary of the eastern Gulf: Trans. Gulf Coast Assoc. Geol. Soc. v. 13, p. 139–147.

ELLIOT, R. L., 1963, Faunas of the Mahantango Formation in south-central Pennsylvania, *in* Shepps, V. C. (editor), Symposium on Middle and Upper Devonian stratigraphy of Pennsylvania and adjacent states: Pennsylvania Geol. Survey, 4th ser., Bull. G 39, p. 201–212.

FAGER, E. W., & McGOWAN, J. A., 1963, Zooplankton species groups in the North Pacific: Science, v. 140, p. 453–460.

HAGMEIER, E. M., & STULTS, C. D., 1964, A numerical analysis of the distributional patterns of North American mammals: Syst. Zoology, v. 13, p. 125–155.

IMBRIE, JOHN, & PURDY, E. G., 1962, Classification of modern Bahamian carbonate sediments: Am. Assoc. Petroleum Geologists, Mem. 1, p. 253–272.

JACCARD, P., 1908, Nouvelles recherches sur la distribution florale: Bull. Soc. Vaud. Sci. Nat., v. 44, p. 223–270.

KAESLER, R. L., 1966, A quantitative re-evaluation of the ecology and distribution of Recent Foraminifera and Ostracoda of Todos Santos Bay, Baja California, Mexico: Univ. Kansas Paleont. Contr., Paper 10, 50 p.

MADDOCKS, R. F., 1966, Distribution patterns of living and subfossil podocopid ostracodes in the Nosy Bé area, northern Madagascar: Univ. Kansas Paleont. Contr., Paper 12, 72 p.

MELLO, J. F., & BUZAS, M. A., 1968, An application of cluster analysis as a method of determining biofacies: Jour. Paleont., v. 42, p. 747–758.

OCHIAI, A., 1957, Zoogeographical studies on the Soleoid fishes found in Japan and its neighbouring regions—II: Bull. Japan Soc. Sci. Fisheries, v. 22, p. 526–530.

PETERS, J. A., 1968, A computer program for calculating degree of biogeographical resemblance between areas: Syst. Zoology, v. 17, p. 64–69.

PRESTON, F. W., 1962, The canonical distribution of commonness and rarity: part II: Ecology, v. 43, p. 410–432.

SAVAGE, J. M., 1960, Evolution of a peninsular herpetofauna: Syst. Zoology, v. 9, p. 184–212.

SIMPSON, G. G., 1943, Mammals and the nature of continents: Am. Jour. Sci., v. 241, p. 1–31.

——, 1947, Holarctic mammalian faunas and continental relationships during the Cenozoic: Bull. Geol; Soc. America, v. 58, p. 613–688.

——, 1960, Notes on the measurement of faunal resemblance: Am. Jour. Sci., v. 258a, p. 300–311.

SOKAL, R. R., 1961, Distance as a measure of taxonomic similarity: Syst. Zoology, v. 10, p. 70–79.

SOKAL, R. R., & MICHENER, C. D., 1958, A statistical method for evaluating systematic relationships: Univ. Kansas Sci. Bull., v. 38, p. 1409–1438.

SOKAL, R. R., & SNEATH, P. H. A., 1963, Principles of numerical taxonomy: Freeman, San Francisco, 359 p.

SORGENFREI, T., 1959, Molluscan assemblages from the marine middle Miocene of south Jutland and their environments: Denmark Geol. Unders., ser. 2, no. 79, 2 vols., 503 p.

STEPHENSON, W., WILLIAMS, W. T., & LANCE, G. N., 1968, Numerical approaches to the relationships of certain American swimming crabs (Crustacea: Portunidea): Proc. U. S. Natl. Mus., v. 124, p. 1–25.

VALENTINE, J. W., 1966, Numerical analysis of marine molluscan ranges on the extratropical Northeastern Pacific Shelf: Limnology and Oceanography, v. 11, p. 198–211.

VALENTINE, J. W., & PEDDICORD, R. G., 1967, Evaluation of fossil assemblages by cluster analysis: Jour. Paleont., v. 41, p. 502–507.