Numerical Classification of Bacteria

Part IV [†] Relationships among Some Corynebacteria Based on Serological Similarity Alone

By Eitarō MASUO and Toshio NAKAGAWA

Shionogi Research Laboratory, Shionogi & Co. Ltd., Osaka

Received February 18, 1970

It has been suggested by Sokal and Sneath that the mathematical techniques used in numerical taxonomy would effectively be applied to elucidate serological (and perhaps phylogenetic) relations among organisms. This excellent idea, however, has hitherto never come into operation. The present paper describes a mathematical method for analyzing the intensities of antigen-antibody reactions between test organisms to make clear their serological relations. Eighteen corynebacteria were examined as a test case, and a brief discussion is presented on the results obtained.

It has been recognized that a protein (or protein fraction) of an organism reacts most strongly with the antiserum produced against itself, less strongly with the antiserum against the corresponding protein fraction of a similar organism, and still less strongly with that of a dissimilar organism. This rule is generally true although noticeable exceptions are sometimes encountered. Thus, the intensity of antigen-antibody reaction can be used as a measure of similarity between organisms. The real significance of the above-mentioned serological similarity remains ambiguous until the nature of antigen-antibody reaction is fully elucidated. It is likely that large numbers of small differences and resemblances in amino acid sequence, and perhaps also, in the folding of polypeptide chain contribute to the reaction intensity. In this sense, the serological similarity may be regarded as an estimate of the overall structural resemblance between the proteins under investigation.

In modern genetics, the protein structure

is considered to be determined by the fine structure of genes. In this context, Crick¹⁾ has suggested that the detailed information of protein structure would be utilized to establish "protein taxonomy." Such a study may bring a significant development into taxonomy and possibly into phylogeny. Sokal and Sneath²⁾ published in their recent book an excellent idea that the mathematical technique used in numerical taxonomy would effectively be applied to investigate phylogenetic relations among various organisms. This suggestion, however, has hitherto never come into operation. The present paper deals with the mathematical procedure, by which the mutual correlations of some corynebacteria were analyzed on the basis of serological similarities. A brief discussion is presented on the results obtained.

[†] Part I. E. Masuo and T. Nakagawa, *Nippon Nogeikagaku Kaishi*, **42**, 627 (1968); Part II and III, E. Masuo and T. Nakagawa, *Agr. Biol. Chem.*, **33**, 1124, 1570 (1969).

1) F. H. C. Crick, "The Biological Replication of Macromolecules, XIIth Symposium of the Society for Experimental Biology," Cambridge University Press, London, 1958, p. 138.

2) R. R. Sokal and P. H. A. Sneath, "Principles of Numerical Taxonomy," Freeman & Co., San Francisco and London, 1963, p. 40.

(a) Reaction Intensities Measured by Tube Precipitation.

(b) Reaction Intensities Measured by Double Gel Diffusion.

[illegible]

MATERIALS AND METHODS

Test organisms and serological data. Eighteen organisms of genus *Corynebacterium* were examined. The species names of these bacteria (Operational Taxonomic Units; OTU's) are listed in Table I along with the related serological data cited from an original literature of Lazar.³⁾ In his work, nucleoprotein extracts were used as antigens and antigen-antibody reactions were examined both by tube precipitation and by double gel diffusion. The results were coded with six code symbols "—", "±", "+", "##", "###" and "####" which indicated reaction intensities in the order of increasing intensity.

Algorithm for computing serological similarity. Algorithm for computing the serological similarity is explained below by processing a hypothetical matrix shown in the upper left of Fig. 1 (Matrix I), where antigen-antibody reaction intensities are expressed by the six code symbols stated above. First, each of these

symbols is converted into one of the six integers ranging from 0 to 5; — into 0, ± into 1, + into 2, and so on. In Matrix II obtained through this conversion, the values of diagonal elements, *i.e.* homologous reaction values, are not equal to unity nor even equal to one another. It seems rational to stipulate that the similarity of any antigen to itself should be equated to unity. Second, therefore, we divide every matrix element by the diagonal element in the same column to derive Matrix III, in which all diagonal elements are equal to unity. The square matrix obtained in this way is usually asymmetric with respect to the diagonal. For example, two similarity values, 0.5000 and 0.6000, are found for the pair of OTU's A and B. To express the similarity between A and B, some value between 0.5000 and 0.6000 would appear adequate. Thus, in the third and fourth steps, we calculate arithmetical means and geometrical means to obtain Matrices IVa and IVb, respectively. Either of these can be used as the source matrix,

FIG. 1. Explanatory Diagram for Algorithm.

3) I. Lazar, *J. Gen. Microbiol.*, **52**, 77 (1968).

upon which the clustering of OTU's is carried out. It is not clear *a priori* which matrix is preferable to the other.

Computer and program. Computer analysis was performed on an IBM 1620 digital computer at the authors' research laboratory by using a computer program NUTAX,* about which detailed explanation has been given elsewhere.⁴⁾

RESULTS

Two sets of serological data shown in Table I were transformed into source matrices by the above-mentioned conversion. Thus, four

source matrices were obtained. On each of them, three different clusterings were carried out by single-, average-, and complete-linkage procedures based upon the pair group method.^{4,5)} Twelve dendrograms were framed up in this way, and twelve phenon-value matrices were written out from these.

Although these dendrograms display quite similar patterns, it seems desirable, or even necessary, to select the most representative dendrogram and to discuss the serological relations of test organisms based on this one. The selection was done by comparing correlation coefficients among the sixteen matrices

FIG. 2. Correlation Coefficients among Various Similarity Matrices.

[S: Single-, A: Average-, C: Complete-Linkage Clustering].

* Program description and Input-and-Output Specifications can be obtained as a set of microfilms on request to T. Nakagawa.

4) T. Nakagawa, *Ann. Rept. Shionogi Res. Lab.*, **19**,

107 (1969).

5) R. R. Sokal and P. H. A. Sneath, "Principles of Numerical Taxonomy," Freeman & Co., San Francisco and London, 1963, p. 180~183, 309.

FIG. 3. The Most Representative Dendrogram and Shaded Similarity Matrix Depicting the Serological Relationships among 18 *Corynebacteria*.

illustrated in Fig. 2.* Each of the four source matrices has three correlation coefficients between the other three source matrices. The sum of coefficients divided by three is the mean coefficient of that matrix. The source matrix depicted at the centre of the lower left triangle—"0" in (III)—may be regarded as the most representative source matrix because it has the highest mean coefficient. Compared among three phenon-value matrices derived from this source matrix, the one obtained by the average-linkage method—"A" in (III)—shows the highest correlation coefficient to

wards the source matrix.

The most representative dendrogram shown in Fig. 3 has thus been obtained by applying the average-linkage clustering on the source matrix that was derived by processing the double gel diffusion data with the method of arithmetical mean. In the same figure, the most representative source matrix is illustrated in a form of shaded similarity matrix, in which all OTU's are arranged in the same order as that of the dendrogram. The latter presentation makes visible the undistorted similarity relations among the test organisms; the distortion brought about in the course of constructing the dendrogram has been eliminated in the shaded similarity matrix.

* Correlation coefficient between any two matrices, $[L]$ and $[M]$, was calculated by the formula

$$\frac{\sum(L \times M) - \frac{2(\sum L) \times (\sum M)}{t(t-1)}}{\left[\sum L^2 - \frac{2(\sum L)^2}{t(t-1)} \right]^{1/2} \times \left[\sum M^2 - \frac{2(\sum M)^2}{t(t-1)} \right]^{1/2}}$$

where L and M stand for two corresponding matrix elements and t represents the number of OTU's under study.

DISCUSSION

Since the serological data used in the present work have been cited from Lazar's paper,³¹ his view will be introduced first before pre-

FIG. 4. Serological Relationships Depicted in the Lazar's Original Literature.³⁾

—, Strong reaction; ---, reaction; ---, weak reaction; very weak or doubtful reaction. [I]~[VII]; Serological groups.

senting the authors' own. He made up a schematic diagram illustrating the serological relationships among the eighteen corynebacteria examined. It is reproduced in Fig. 4. In this diagram, six groups (I~VI) are arranged around *C. fascians* (group VII consisting of only one member). Reaction intensities (or degrees of serological resemblance) between these groups are shown by four kinds of tie-lines.

Now, the serological relations will be discussed by comparing Figs. 3 and 4. Good correspondence between both figures is recognized at a glance. For example, there are five distinct stems in the dendrogram of Fig. 3, which correspond to groups I, III, IV, V, and VI in Fig. 4. First of all, we shall discuss these groups. In the dendrogram shown in Fig. 3, four strains of plant pathogenes (*C. flaccumfaciens*, *C. flaccumfaciens* var. *aurantiacum*, and two strains of *C. poinsettiae*) crowd together at high phenon levels and *C. betae* joins to them at a somewhat lower level, thus resulting in a stem whose constituents are the same as those of Lazar's group III. Close

inspection at the shaded similarity matrix reveals that *C. betae* shows comparatively low similarities to the other members. These relations are consistent with the conventional classification; Bergey's Manual⁶⁾ and its companion volume "Index Bergeyana"⁷⁾ describe that *C. flaccumfaciens* and *C. poinsettiae* are identical or of high resemblance to each other. Similar descriptions are found also for groups I (two types of *C. diphtheriae*), IV (*C. insidiosum* and two strains of *C. michiganense*), VI (two strains of *C. sepedonicum*) and V (*C. tritici* and *C. rathayi*). For the last group, *C. rathayi* and *C. tritici* are actually used as a synonym in the conventional taxonomy. The identity or close resemblance between the members within a group has also been supported by other workers. In a work on chemical compositions and antigenic

6) "Bergey's Manual of Determinative Bacteriology," 7th ed., Williams and Wilkins Co., Baltimore, 1957.

7) "Index Bergeyana; A Companion Volume to Bergey's Manual of Determinative Bacteriology," ed. by R. E. Buchanan, J. G. Holt, and E. F. Lessel, Williams and Wilkins Co., Baltimore, 1966.

features of cell walls, Cummins⁸⁾ stated that *C. poinsettiae* and *C. flaccumfaciens* appeared to be identical. Lelliot⁹⁾ consented to this comment from serological viewpoint, and pointed out further that *C. insidiosum* and *C. michiganense* were related to each other as closely as the former pair of corynebacteria were.

In Fig. 4, *C. equi*, *C. renale*, and *C. kutscheri* are clustered to form group II. This grouping has a serious disadvantage; it makes one imagine that these three animal corynebacteria were closely related to one another. As clearly shown in the shaded matrix of Fig. 3, the serological similarity between any two of these three bacteria is very low.* In the dendrogram of Fig. 3, these organisms connect one another at low phenon levels. It would be better to posit them as separate single-items rather than to cluster them into a group. Mutual remoteness of these organisms is supported on evidence other than serological. On the basis of morphological and physiological characters, Ramamurthi¹⁰⁾ proposed that *C. renale*, alike *C. fascians*, should be included in genus *Nocardia*. On the other hand, Jensen¹¹⁾ stated that *C. equi* might probably with as much right be placed in *Mycobacterium*, based on the finding that this organism was markedly oxidative in metabolism and was acid-fast *in vivo*. Jensen's view does not seem peculiar since it has been claimed by many other authors that *Corynebacterium* and *Mycobacterium* border closely upon each other.^{9,12)}

Although a decisive conclusion could not be derived, the taxonomic position of *C. fascians* was the centre of interest in this work. According to Fig. 3, this plant pathogene is

considered far more akin to typical animal parasites (two types of *C. diphtheriae*) than to the other fifteen bacteria (three animal diphtheroids in Lazar's group II and twelve plant pathogens in group III~VI). In Fig. 4, on the other hand, *C. fascians* is indicated as an intermediate between groups I and III which have hitherto been considered as far distant groups. In contradiction to the fact, group VII (*C. fascians*) is connected to group III by a thick tie-line which means that the serological reaction between these groups is as strong as the reaction between groups VII and I. The present authors consider that the shaded similarity matrix of Fig. 3 is a better presentation for illustrating the serological relations.

In the preceding paragraph, we have discussed the serological or taxonomic position occupied by *C. fascians* in the assembly of eighteen corynebacteria examined. We now refer to the taxonomic position of *C. fascians* in a wider variety of microorganisms. Jensen¹¹⁾ and afterward Lacey¹²⁾ reported that cultures of this bacterium exhibited acid-fastness when grown on special media. Some strains tentatively labeled as *Mycobacterium* (e.g. *Mycobact. rhodochrous*) were reported to be acid-fast when first isolated, but not so when examined afterward.¹³⁾ Lanelle *et al.*¹⁴⁾ stated recently that the acid-fastness of these bacteria was a variable characteristic. Thus, it became more and more doubtful whether acid-fastness was a decisive key character to discriminate between *Corynebacterium*, *Mycobacterium*, and *Nocardia* in the conventional taxonomy or in Bergey's Manual. In recent years, the problematic position of *C. fascians* and closely related *Mycobacterium* species were repeatedly discussed by Gordon,¹⁵⁾ who suggested that there was no clear delimitation between the two genera.

8) C. S. Cummins, *J. Gen. Microbiol.*, **28**, 35 (1962).

9) R. A. Lelliot, *J. Appl. Bacteriol.*, **29**, 114 (1966).

* Lazar himself, who designed Fig. 4, was aware of this fact.

10) C. S. Ramamurthi, *Mem. Cornell Univ. Agr. Exp. Stn.*, No. 336 (1959).

11) H. L. Jensen, *Proc. Linnean Soc. N.S.W.*, **59**, 19 (1934) [*Ann. Rev. Microbiol.*, **6**, 77 (1952)].

12) M. S. Lacey, *Trans. Br. Mycol. Soc.*, **38**, 49 (1955) [R. A. Lelliot, *J. Appl. Bacteriol.*, **29**, 114 (1966)].

13) R. E. Gordon and L. M. Mihm, *J. Gen. Microbiol.*, **21**, 735 (1959).

14) M. A. Lanelle, J. Asselineau and G. Castellnuovo, *Ann. Inst. Pasteur*, **108**, 69 (1965).

15) R. E. Gordon, *J. Gen. Microbiol.*, **43**, 329 (1966).

Pebble¹⁶⁾ found that the carotenoid composition of a strain of *C. fascians* was of much resemblance with those of some *Mycobacterium* strains. These findings, together with the results of Adansonian taxonomic studies of Harrington¹⁷⁾ and later of the authors¹⁸⁾ (based on the overall similarity of a large number of phenetic characters) suggest that *C. fascians* may be regarded as a boundary form between

genera *Corynebacterium*, *Mycobacterium*, and further *Nocardia*. Any proposal in generic assignment of such wandering species should be made only on the basis of further studies and on the establishment of clearer differentiation between neighboring genera. Serological study on organisms currently posited in different genera might furnish us with valuable information.

Although the above discussion has been presented on the basis of the most representative dendrogram and matrix, we arrived at substantially the same conclusion from any of the other dendrograms and matrices.

16) J. Pebble, *J. Gen. Microbiol.*, **52**, 15 (1968).

17) B. H. Harrington, *ibid.*, **45**, 31 (1966).

18) E. Masuo and T. Nakagawa, *Nippon Nogeikagaku Kaishi*, **42**, 627 (1968); *Agr. Biol. Chem.*, **33**, 1124 (1969); *Ann. Rept. Shionogi Res. Lab.*, **19**, 105 (1969).