Numerical Classification of Salmonella Serotypes

By W. R. LOCKHART AND J. G. HOLT

*Department of Bacteriology, Iowa State University,
Ames, Iowa, U.S.A.*

(Received 21 October 1963)

SUMMARY

Numerical methods were applied to an analysis of the relationships among *Salmonella* serotypes listed in the Kauffmann–White schema. Although the result suggested a possible new basis for schematic arrangement of these serotypes, a complete and satisfactory classification could not be derived entirely from the computer results. Examination of this outcome suggests some cautions to be observed in the design and interpretation of experiments in numerical taxonomy.

INTRODUCTION

In most numerical studies of bacterial classification, the ‘operational taxonomic units’ (or OTU’s) have been actual cultures. Although numerical analysis might equally well be applied to descriptions of taxa or to pooled data taken from the literature (Sneath, 1962), this approach has been used infrequently (e.g. Lysenko & Sneath, 1959). This paper reports an attempt to develop a numerical classification of the genus *Salmonella*, using Kauffmann’s (1961) descriptions of *Salmonella* serotypes as the OTU’s.

METHODS

Descriptions of each of the 700 *Salmonella* serotypes listed in the Kauffmann–White schema (Kauffmann, 1961) were recorded in form suitable for computer analysis, using a method previously described (Lockhart & Hartman, 1963; Lockhart, 1963). Entries for which Kauffmann lists two biochemical types (e.g. *Salmonella arechavaleta* 4, 5, 12 and *S. arechavaleta* 4, 12) were treated as two separate serotypes. For each listed property of a given serotype, a code symbol (A, B, C or D) was assigned for each alternative state in which the property might occur. A total of 155 properties was used, scored as follows: *Somatic antigens 1 through 56* (56 features); A = present, B = absent, D = incomplete or variant (listed in parentheses by Kauffmann, 1961). The Vi antigen was not scored. *Flagellar antigens—phase variation* (1 feature); A = antigens listed in only one phase, B = antigens listed for two phases, C = no H antigens listed. *Individual flagellar antigens a through z* (26 features), *z₁ through z₄₇* (47 features) and *1 through 11* (11 features); A = antigen listed in first phase, B = antigen listed in second phase, C = antigen not listed in either phase, D = antigen sometimes present in either phase (listed in parentheses by Kauffmann, 1961). *Biochemical characteristics* (14 features); A = +, B = ++ or + with a superscript, C = – or x, D = d (symbols as defined by Kauffmann, 1961). A number of the antigens listed here (e.g. somatic antigens 26, 29, 31, 32, 33, 36, 37 and 49) are not actually found in the 1961 schema. Such anti-

gens, forty-five in all, were scored as 'absent' for all OTU's, reducing the total of effective features to 110. This did not affect the outcome; the question of whether or not one should score a similarity when both of two OTU's *lack* a particular antigen (or when both give a negative response to a particular biochemical test) does not arise, since the computer program used was one which enumerates differences only.

During computations, two serotypes were considered 'different' with respect to a given property whenever different symbols were recorded for each in that property. For somatic antigens, then, a difference was recorded for any antigen possessed by one serotype but not by the other. Two serotypes score a difference with respect to any flagellar antigen unless it is possessed by both in the same phase. No attempt was made, in scoring the biochemical features, to discriminate among varying degrees of 'lateness' in positive reactions ($+^{1-3}$ vs. $+^{2-4}$, etc.), although these were distinguished from strong positives (+) and from negatives (-). Kauffmann's symbol (++) is not really equivalent to (+) with a superscript, but he never uses both in describing the same reaction. His symbol (*x*), meaning 'late, irregularly positive, often negative', seems nearly equivalent to an outright negative, and it is so scored. The symbol (*d*) apparently indicates that individual strains of the serotype in question may give almost any response from strong positive to outright negative, and is scored so that it records a difference from any serotype showing constancy with regard to the property in question.

Groupings were obtained according to the monothetic criterion of Lockhart & Hartman (1963), using the Cyclone computer (Iowa State University). In this method, the operator designates a starting OTU, and the machine then locates the second individual which (of all those under study) is most like the first. All the features in which these two are not identical are then eliminated from consideration, and the computer locates the third OTU which is most like the original pair with respect to the remaining features. This process is repeated many times, with the computer output designating the new OTU added at each step and the total number of features eliminated from consideration thus far (this quantity is designated the 'cumulative difference', *d_c*). All members of the groups obtained in this way are thus identical with respect to the common core of remaining properties, presumably producing a classification in harmony with the hierarchical character of most currently accepted taxonomic schemes. Since previous work (Lockhart & Hartman, 1963) had indicated that little useful information is obtained after a group under formation has coalesced with a taxonomically adjacent group, the computer program has an instruction which makes it possible to stop the calculation after the group has reached a predetermined size. All the present experiments were terminated after 100 serotypes had been grouped around the original pair. The program also has an optional instruction permitting certain properties to be eliminated from consideration during formation of a group, and this was done in some instances specified later.

RESULTS

A preliminary examination of the Kauffmann (1961) data shows that the Kauffmann-White schema is essentially a monothetic classification based on somatic antigens. Within the major groups thus designated, subdivisions are based on the flagellar antigens found in the first phase. Antigenic composition in the second

phase and a number of biochemical properties are listed for each type, but neither shows any obvious correlations with the somatic or first-phase antigens. Certain rather characteristic combinations of antigens may be found in the second phase, however. Frequently there is no second phase, but when one exists antigen 1 seems to occur most frequently—nearly always paired with antigens 2, 5, 6 or 7 (sometimes a combination like 1, 5, 7 may be seen). If antigen 1 is not present, the most likely alternatives are the combinations e, n, x or e, n, z₁₅ (and, occasionally, such combinations as d, e, n, z₁₅ or e, n, x, z₁₅). Less often, the combination l, w or the single antigen z₆ may be found. Quite rarely, a few other antigens are reported for the second phase. Although most of these antigens are sometimes found also in the first phase, the characteristic combinations (and their mutual exclusiveness) in second phase are rather striking.

Table 1. *Composition of a monothetic group formed around Salmonella cairo, considering all features*

Cumulative difference	No. of individuals (serotypes) added	New Kauffmann groups represented	New antigens represented		
			Somatic	Phase 1	Phase 2
0	1	B	1, 4, 12, 27	d	1, 2
4	2	.	.	.	5, 7
8	2	.	.	b	.
11	1	.	.	y	.
13	1	.	.	a	.
14	2	.	.	r	.
16	1	.	.	.	6
17	3	.	.	c	.
19	1	.	.	i	.
20	7	.	5	.	.
24	3	.	.	z	.
25	5	.	.	z ₁₀	.
29	8	.	.	.	z ₆
33	1	.	.	.	absent
34	2	.	.	z ₂₉	.
35	1	.	.	z ₃₈	.
37	2	A	2	.	.
38	21	D ₁	9	.	.
39	1	.	.	.	z ₃₉
40	2	.	.	.	z ₃₅
41	1	.	.	k	.
42	9	D ₂	46	.	.
44	3	C ₃	8	.	.
45	18	C ₂	6	.	.
46	4	C ₁	7	.	.

When the 14 biochemical features were subjected to a conventional card-sorting routine, it was found that many of the serotypes were alike in all or most of their biochemical characters, but there was no discernible pattern that could be correlated with antigenic structure. More than a third of the 700 serotypes, including all the species listed in *Bergey's Manual* (1957), were unique; their patterns of physiological characters resembled none of the other types.

The first attempt at formation of groups using the entire body of data was made with *Salmonella cairo* (1, 4, 12, 27; d; 1, 2) specified as the original OTU. *Salmonella*

cairo appears to be a 'typical' member of Kauffmann's Group B, and it had been found in our preliminary study to be biochemically similar to a number of other serotypes. The results of this computation are presented in Table 1, which shows the number of serotypes added at each level of cumulative difference (d_c) and the Kauffmann groups to which they belong. In an earlier trial of this computer program

Table 2. *Composition of a monothetic group formed around Salmonella cairo, with biochemical features omitted from consideration*

Cumulative difference	No. of individuals (serotypes) added*	New Kauffmann groups represented	New antigens represented		
			Somatic	Phase 1	Phase 2
0	1	B	1, 4, 12, 27	d	1, 2
2	1	.	.	.	5
3	1	.	.	.	7
5	2	.	.	b	.
7	9	.	5	a	.
11	1	.	.	.	6
12	1	.	.	.	absent
13	4	.	.	c	.
14	2	.	.	.	z_6
15	2	.	.	i	.
16	3	.	.	r	.
17	3	.	.	y	.
18	4	.	.	z	.
19	4	.	.	z_{10}	.
20	2	.	.	z_{29}	.
21	1	.	.	z_{38}	.
23	2	A	2	.	.
24	21	D ₁	9	.	.
25	1	.	.	.	z_{39}
26	2	.	.	.	z_{35}
27	1	.	.	k	.
29	9	D ₂	46	.	.
30	3	C ₃	8	.	.
31	18	C ₂	6	.	.
32	4	C ₁	7	.	.

* The Table is condensed in those instances when an increase in d_c occurred as a result of admission of a serotype that lacked an antigen possessed by all previous members of the group, rather than one having a 'new' antigen. The specific serotypes added at various stages of the computation were as follows: $d_c = 0$, *S. cairo*; $d_c = 2$, *S. eppendorf*; $d_c = 3$, *S. schwarzegrund*; $d_c = 5$, *S. limete*; $d_c = 6$, *S. uppsala*; $d_c = 7$, *S. 'hessarek'* 4, 12, 27; $d_c = 8$, *S. hessarek* 4, 12, *S. 'arechavaleta'* 4, 12; $d_c = 9$, *S. kisangani*, *S. fulica*, *S. arechavaleta* 4, 5, 12, *S. paratyphi-B*, *S. stanley*; $d_c = 10$, *S. java*; $d_c = 11$, *Salmonella* 4, 12: b: 1, 6; $d_c = 12$, *S. schleissheim*; $d_c = 13$, *S. legion*, *S. abortus-ovis*, *S. altendorf*, *S. womba*; $d_c = 14$, *S. bury*, *S. ayinde*; $d_c = 15$, *S. typhimurium*, *S. agama*; $d_c = 16$, *S. heidelberg*, *S. bradford*, *S. remo*; $d_c = 17$, *S. coein*, *S. teddington*, *S. kamoru*; $d_c = 18$, *S. shubra*, *S. kiambu*, *S. indiana*, *S. entebbe*; $d_c = 19$, *S. haifa*, *S. ituri*, *S. tudu*, *S. fortune*; $d_c = 20$, *S. brancaster* 1, 4, 12, *S. 'brancaster'* 1, 4, 12, 27; $d_c = 21$, *S. Wilhelmsburg*; $d_c = 23$, *S. paratyphi-A*, *S. paratyphi-A* var. *durazzo*; $d_c = 24$, *S. sendai*, *S. miami*, *S. os*, *S. saarbruecken*, *S. onarimon*, *S. frintop*, *S. blankenese*, *S. goeteborg*, *S. ipeko*, *S. ridge*, *S. typhi*, *S. ndolo*, *S. zega*, *S. seremban*, *S. jamaica*, *S. lawndale*, *S. stellenbosch*, *S. angola*, *S. portland*, *S. canastel*, *S. gallinarum-pullorum*.

(Lockhart & Hartman, 1963), a substantial increase in cumulative difference (d_c) was noted when a new subgroup fused with the group under formation. The peculiar nature of the present data prevented this; addition of a new subgroup might occur

with only a slight increase in d_c . The existence of the subgroup could be detected however, by the fact that an appreciable number of new individuals would join the main group at the same time. In Table 1, for example, 21 types were added as a group at $d_c = 38$, when somatic antigen 9 was admitted. For simplicity, no lists of individual serotype designations are given in the tables; instead, the new antigens possessed by serotypes added at each successive level of d_c are indicated. These last three columns of the tables thus show which features have been eliminated during formation of the group. This information was not included in the computer output, but could easily be deduced by comparing the results with Kauffmann's (1961) tables. As we will discuss later, members of the group may or may not have any of the 'eliminated' features, but cannot possess a feature which has not yet been listed.

Table 3. *Composition of a monothetic group formed around Salmonella saarbruecken, considering all features*

Cumulative difference	No. of individuals (serotypes) added	New Kauffmann groups represented	New antigens represented		
			Somatic	Phase 1	Phase 2
0	1	D ₁	1, 9, 12	a	1, 7
6	2	B	4, 5	.	.
12	1	.	.	c	.
13	1	.	.	z	.
15	2	.	.	.	5
16	6	.	.	.	2
19	1	.	.	b	.
20	3	.	.	.	6
21	4	.	27	.	.
22	4	.	.	d	.
23	2	.	.	i	.
24	4	.	.	r	.
25	2	.	.	y	.
26	8	.	.	z ₁₀	.
32	15	.	.	.	z ₆
36	3	.	.	.	absent
37	2	A	2	.	.
38	3	.	.	z ₂₉	.
39	1	.	.	z ₃₈	.
40	1	.	.	.	z ₃₉
41	1	.	.	z ₃₅	.
42	1	.	.	k	.
44	9	D ₂	46	.	.
45	3	C ₃	8	.	.
46	18	C ₂	6	.	.
47	4	C ₁	7	.	.

The fate of biochemical features is not indicated in the tables, for reasons that become obvious on examination of Table 2, which shows the result when the computation was repeated with the 14 biochemical features eliminated. The groups formed in Tables 1 and 2 are identical. The order in which individual serotypes were added to the groups was slightly different in the early stages of the two computations, but the eventual composition of the groups was the same. The process of group formation may appear more clear after examination of the footnote to Table 2, wherein is indicated the order in which specific serotypes were added to the group.

It will be noted that the d_c level at which the second computation ended (32) is just 14 less than that at which the first had ended (46). The only essential difference is in the elimination of the 14 biochemical features; that is, *any* combination of biochemical characteristics is acceptable for admission to the group.

Table 4. *Composition of a monothetic group formed around Salmonella dar-es-salaam, considering all features*

Cumulative difference	No. of individuals (serotypes) added	New Kauffmann groups represented	New antigens represented		
			Somatic	Phase 1	Phase 2
0	1	D ₁	1, 9, 12	1, w	e, n, x
3	1	.	.	b	.
7	1	B	4	.	.
10	1	.	.	z ₁₀	.
12	1	.	27	d	.
13	1	.	.	y	.
15	1	.	.	.	z ₁₅
16	3	.	.	a	.
19	3	.	.	c	.
21	3	.	5	.	.
22	1	.	.	k	.
23	4	.	.	v	.
28	2	.	.	z	.
30	1	.	.	z ₁₃	.
31	1	.	.	z ₂₈	.
33	3	.	.	.	1, w
34	1	.	.	.	absent
35	9	.	.	.	z ₆
36	4	.	.	e	.
37	1	.	.	i	.
38	1	.	.	r	.
39	2	.	.	z ₂₉	.
40	1	.	.	z ₃₈	.
41	1	.	.	.	z ₃₉
42	3	.	.	.	z ₃₅
44	8	D ₂	46	.	.
46	2	A	2	.	.
47	3	.	.	absent	.
49	12	F	11	.	.
50	11	I	16	.	.
51	6	J	17	.	.
52	4	L	21	.	.
53	4	M	28	.	.

It was concluded that the primary basis for group formation had been possession of the 1, 2; 1, 5; 1, 6 or 1, 7 antigenic grouping in the second phase, with a secondary subgrouping based on somatic antigens. Although the machine eventually began to accept such single second-phase antigens as z_6 , z_{39} , and z_{35} , it rigorously excluded the second phase complexes, e, n, x or e, n, z_{15} . The antigenic composition in first phase, and the biochemical characteristics, appear to have no influence on the composition of the group. *Salmonella cairo* gathered about it neither the rest of Group B nor a biochemical group, but one based on second-phase antigens.

When the calculation was repeated using *Salmonella typhimurium* (1, 4, 5, 12;

i; 1, 2) as the starting OTU, the resulting group was again the same as that shown in Tables 1 and 2. Thus the group formed around one of the classical species also is based on second-phase antigenic composition rather than on the Kauffmann groupings. Essentially similar results were obtained in several further experiments. By way of illustration, two of these are shown in Tables 3 and 4. The starting types for these groupings were *S. saarbruecken* (1, 9, 12; a; 1, 7) and *S. dar-es-salaam* (1, 9, 12; 1, w; e, n, x), both members of Kauffmann's Group D₁. In both cases the initial groups formed are mixtures of Kauffmann Groups D₁ and B, and are based primarily on the characteristic second-phase antigenic complexes 1, 2-5-6 or 7 and e, n, x or z₁₅, respectively. In these and other computations, a few single second-phase antigens were eventually included in the group under formation (as was the lack of a second phase), but the two principal kinds of groups (1, 2, etc., *vs.* e, n, x) were never mixed. In no case was there any evidence that biochemical features had an influence on group formation, or that more distinctive groups were formed around species listed in *Bergey's Manual* (1957) than around serotypes chosen at random as a starting point.

DISCUSSION

Unfortunately, the physiological characters listed by Kauffmann (1961) for each serotype were of no help in achieving a numerical classification. Relationships based entirely on the biochemical features did not include all the serotypes, and could not be correlated either with the Kauffmann-White serological groups or with any of the groupings obtained by computer analysis. The fact that exclusion of the biochemical features did not effectively alter the result of computations (cf. Tables 1 and 2) indicates that not even a selected few of the physiological features were involved in group formation. This conclusion applies primarily to attempts to rearrange the Kauffmann data, however, and not necessarily to all efforts to achieve a classification of salmonellas. The physiological data given by Kauffmann for each 'type' appear not to be based on the characters of a single type strain, but to represent a summary of the reactions shown by an indeterminate number of isolates having the requisite antigenic composition. The result of this, for any 'type' of which a number of individual isolates had been studied, is that a great many of the biochemical results are recorded as 'variable'. This seems to be especially true of the frequently encountered species, so that attempts to form groups around these serotypes, as they are characterized by Kauffmann, become almost hopeless. A better result might be obtained if data were available for the reactions obtained either from a great many individual isolates or from single, type strains (using 'type' in its nomenclatural sense). Since our study was restricted to Kauffmann's (1961) data, we could hope—at best—only to produce a classification of his classification.

In our results, the primary criterion for group formation seems to have been antigenic composition in second phase, with somatic antigens furnishing a basis for further subdivisions. Members of a Kauffmann-White somatic group were rejected by the computer if they did not fit the established second-phase pattern of the numerical grouping. In Tables 1 and 2, for example, all members of Group B which showed the e, n, x or e, n, z₁₅ pattern were excluded, while members of Kauffmann's Groups A, D₁, D₂, etc. were admitted if they possessed the prevailing 1, 2-5-6 or 7 combination of antigens in second phase. This trend is even more strikingly demon-

strated in Tables 3 and 4, where the numerical groups (though consistent for second phase antigens) are composed from the outset of a mixture of Kauffmann (somatic) groups. The illustrations shown in the Tables were selected because the 'starting' serotypes were those with complex somatic antigens. Other computations, in which the original serotype was one listed by Kauffmann as possessing only a single somatic antigen, produced groups of even more mixed somatic composition—though still consistent for second-phase antigens.

First-phase antigens seem to be so diverse that they contribute little or nothing to the pattern of a group; the Tables indicate that first-phase components occur more or less at random among the groups obtained. However, serotypes with a complex first phase, such as d, e, h or g, m, s, t, were consistently excluded.

Members of these numerical 'groups' did not necessarily possess common properties, but all members of any group were commonly lacking in certain properties (antigens). As a group was formed, it was defined in terms of a list of the properties a member might possess without being excluded from the group. The groupings, therefore, are not really monothetic (Sneath, 1962). They are partially polythetic or polythetic, but restricted; that is, a given individual may be considered a member of a group if the properties it possesses are restricted to within the range of those characteristic of the group. But any given member does not necessarily have all the properties that are included in this definition of the group.

Whatever the actual nature of groups, even if they are altogether polythetic, it may be necessary to effect simplifications of this sort for purposes of diagnosis. Practical definition of groups may require application of some such monothetic criterion as accepting the possession or lack of a very few correlated properties as sufficient evidence for assigning an individual to a group or excluding it therefrom. Perhaps any really useful classification is necessarily somewhat arbitrary. The definition of taxa, and particularly the diagnosis of unknown isolates as members thereof, might otherwise prove operationally impossible—or at least so cumbersome as to be hardly worth the bother.

In this light, a computer in some cases is less useful for a complete analysis of groupings than for obtaining hints as to possible criteria for group formation (as we have done in this case). With some sets of data, the multidimensional nature of relationships within and between polythetic groups is such that even lengthy computation produces numerical results that confuse rather than illuminate. One can propose a revision of the Salmonella scheme, based partly on second-phase antigens (Holt & Lockhart, unpublished), but the scheme will be 'better' than Kauffmann's only if it is more practical—not simply because it is derived by numerical methods. In any case, such an arrangement is no more than suggested by these results. The computer certainly did not produce this or any other classification; it merely indicated, in a vague and incomplete manner, some trends in the data. The machine, in fact, was severely handicapped both by the nature of the data and by the inadequacy of the directions we provided. The result is instructive, for it provides some insight into the limitations as well as the advantages of numerical approaches to classification.

We have seen that an OTU may be omitted from a group simply because it has an 'excluded' property—i.e. one that has not yet been included in the definition of the group. This seems to have occurred a number of times in our experiments, when

possession of an obscure antigen (usually in the first phase) caused rejection of individual serotypes that might otherwise have been included. An investigator may simply ignore such anomalies, but the computer, of course, cannot. The particular properties which become 'accepted' as characteristic of a group depend not only on the properties of the OTU designated as a starting point, but also on the order in which similar individuals happen to be listed in the data. Thus, if two serotypes have the same number of differences, though in different specific properties, from the common group pattern, the computer will 'accept' whichever one is encountered first, and eliminate from further consideration any of its properties which are not also possessed by the rest of the group. The properties thus eliminated are now, in effect, admitted to the definition of the group, and the computer has reversed the investigator's intended criterion for group formation! This is illustrated in Tables 2 and 3, where although the predominant grouping in both cases is around the 1, 2-5-6 or 7 complex of second-phase antigens, the secondary subdivision in one case is composed of Kauffmann Group B (somatic antigens 1, 4, 5, 12, 27) individuals only, and in the other case is a mixture of Group B and Group D₁ (somatic antigens 1, 9, 12) serotypes. In the latter case the Group D₁ antigens were accepted because the starting OTU possessed them, but the search for similar individuals then began in Group B (serotypes were stored in the computer 'memory' in the order in which they are listed by Kauffmann), and somatic antigens 4 and 5 were accepted at once. Similarly, all the results reflected a tendency for acceptance of somatic antigens listed near the start of Kauffmann's data (i.e. groups A, B, C, D) rather than equally appropriate representatives of groups listed later and thus less likely to be encountered during searches of the stored data. These 'position effects' are a serious defect in the computer program, for they lead sometimes to a quite trivial property of a single individual becoming one of the criteria for group formation.

Further, the program necessarily shows a bias against complex antigens. A 'group' will eventually include in its definition a variety of antigens that can be added individually with small increments in d_c , but will exclude antigen complexes whose admission would involve a larger increase in d_c . A group formed about a starting OTU possessing antigens 1 and 5 in its second phase, for example, will accept, singly, antigens 2, 6 and 7, and eventually other single antigens such as z_6 , but it will reject a complex like e, n, x.

Although these particular difficulties are characteristic of the 'monothetic' method we have used, other techniques for group-formation have deficiencies of their own (Lockhart & Hartman, 1963). Numerical taxonomy has enormous potential, but depending on the nature and complexity of the data, and the adequacy of the programs used, computer analysis may or may not provide complete and satisfactory solutions for specific problems in classification. The machine will not automatically transform confusion into order; on the contrary, it will perpetuate faithfully any errors or misconceptions the investigator has unwittingly included in his program. One must therefore examine carefully all the assumptions, intended or otherwise, that are implied in the way the data were scored or in the directions given the machine for calculation of similarities and formation of groups. The computer will follow all such instructions quite literally, and will exercise neither discrimination nor good judgement in any instance—however obvious—unless these are explicitly specified beforehand by the investigator.

We are grateful to Dr R. E. Buchanan, who first suggested this investigation and whose advice and encouragement have assisted in its progress. This work was supported by Research Grants AI-03124 and AI-05219 from the National Institutes of Health, U.S. Public Health Service.

REFERENCES

- Bergey's Manual of Determinative Bacteriology* (1957). Ed. by R. S. Breed, E. G. D. Murray & N. R. Smith. Baltimore: Williams and Wilkins.
- KAUFFMANN, F. (1961). *Die Bakteriologie der Salmonella-Species*. Kopenhagen: Munksgaard.
- LOCKHART, W. R. (1963). Scoring of data and group-formation in quantitative taxonomy. *Dev. ind. Microbiol.* 5 (in the Press).
- LOCKHART, W. R. & HARTMAN, P. A. (1963). Formation of monothetic groups in quantitative bacterial taxonomy. *J. Bact.* 85, 68.
- LYSENKO, O. & SNEATH, P. H. A. (1959). The use of models in bacterial classification. *J. gen. Microbiol.* 20, 284.
- SNEATH, P. H. A. (1962). The construction of taxonomic groups. *Symp. Soc. gen. Microbiol.* 12, 289.