JOURNAL OF BACTERIOLOGY, Apr., 1967, p. 1382-1391

Vol. 93, No. 4
Printed in U.S.A.

# Selection of Characters for an Adansonian Analysis of Mycobacterial Taxonomy[1]

LAWRENCE G. WAYNE
*Veterans Administration Hospital, San Fernando, California*

Received for publication 3 January 1967

Tests which are most frequently employed to provide characters for numerical analysis of bacterial taxonomy provide little differentiation among the slow-growing members of the genus *Mycobacterium*. A special scheme was proposed to deal with this problem. Characters were based, to some extent, on quantitative reactions to tests, and were defined, for each test, in terms of natural clustering behavior of a large collection of assorted strains. Inter- and intrataxon analyses were based on comparisons of reaction patterns with tentative hypothetical median strain patterns. This permitted the analysis of larger numbers of cultures than can be practically studied in a conventional $n \times n$ table.

In recent years, there has been an increasing tendency to re-examine existing schemes of bacterial taxonomy in the light of Adansonian principles (8). Among some genera of bacteria, the usefulness and validity of older classifications are being confirmed by numerical analysis based, wholly or in part, on Adansonian principles. Among other genera, including the mycobacteria, this method promises to provide niches for many organisms which have never been adequately classified.

There have been, historically, two general approaches to mycobacterial taxonomy. One approach is an outgrowth of an interest in the position of mycobacteria within the general framework of bacterial taxonomy. Practitioners of this approach have been most successful in developing taxa among rapidly growing mycobacteria (1, 3), but have been less productive in providing subdivisions among slow growers. A second school has evolved among investigators whose primary interest is tuberculosis. A major effort among these investigators was once directed toward evolution of so-called "virulence" tests which were intended to evaluate mycobacteria in terms of their relationships to human, bovine, or avian tubercle bacilli (6), or all of these. This has since branched out into a search for methods for characterizing slow-growing mycobacteria in general (13).

Among mycobacteria, problems of growth rates and limited accumulation of intermediate metabolic products have led to a special situation. Application of a battery of conventional tests which are useful for establishing taxa among bacteria in general tends to result in a lumping of all slowly growing mycobacteria. Thus, an investigator studying a "natural" taxonomy of bacteria might find a fairly uniform "species," consisting of slowly growing mycobacteria.

For practical purposes, it becomes necessary to evolve specialized schemes for investigation of certain groups of organisms, and the mycobacteria appear to represent such a group. Such a decision forces the introduction of monothetic concepts, and thus bias, into the purely Adansonian system. Once this decision has been made (in this case based on acid-fastness and, to a lesser extent, on growth rate), it is desirable to minimize further introduction of bias into the selection of characters. This report presents some solutions to this problem that have proven useful in the analysis of the genus *Mycobacterium*.

## MATERIALS AND METHODS

Between 600 and 700 cultures in our collection of mycobacteria were each examined for the properties described in previous publications (14; Wayne et al., Am. Rev. Respirat. Diseases, *in press*). Of these, distribution of the following properties will be treated in some detail in this paper.

*Microcolonial test (17).* Cultures were recorded as neutral red-positive or -negative, as determined in the microcolonial test.

*Tween agar opacity (16).* Results were recorded as the number of weeks required for first appearance of subsurface opacity in 2.5% Tween-agar butts inoculated on the surface only.

*Aryl sulfatase test (12).* This test was used to deter-

[1] Based, in part, on a paper presented at the Round Table on Mycobacterial Taxonomy, Annual Meeting of the American Society for Microbiology, Los Angeles, Calif., 2 May 1966.

mine the number of weeks required for the first evidence of free phenolphthalein after addition of potassium phenolphthalein sulfate to a liquid culture.

*Nitrate reduction* (*15*). Results were expressed as logarithm of millimicromoles of nitrite per milliliter produced by a culture grown in the presence of nitrate. The cultures were sampled for nitrite determination when the bacillary density reached approximately 1 mg/ml.

*Catalase* (*12*). The reaction with catalase was evaluated in terms of height, in millimeters, of a column of bubbles generated over a 14-day-old culture of bacilli on Loewenstein-Jensen medium in a 16 by 150 mm tube, on addition of a mixture of $H_2O_2$ and Tween 80.

*Tween 80 hydrolysis* (*16*). Results were expressed as number of days of incubation required for a bacillary suspension to produce a characteristic color change in a solution of Tween 80 and neutral red.

*Iron uptake* (*9*). Results were expressed as number of days of incubation required for a culture on Lowenstein-Jensen medium, supplemented with 2% ferric ammonium citrate, to develop a deep rust color in the colonies.

## RESULTS AND DISCUSSION

Lockhart and Koenig (5) have discussed the use of secondary data, which they define as information regarding descriptive or quantitative characters applicable only to organisms yielding a positive result in another primary character. Secondary data appeared to provide more precise definition of groups, when applied to a study of the genus *Erwinia*. In their study, the expansion was achieved by increasing the number of characters through application of quantitative differences. As these authors point out, this does not necessarily cause a great increase in net amount of data available. With slowly growing, acid-fast bacilli, however, additional tests as well as secondary data must also be introduced into a special scheme, and this does indeed increase net data.

Sneath has suggested (8) that additive characters (based on intensity of reaction) exaggerate dissimilarities due to different growth rates. He says ". . . a slow growing strain will usually give delayed reactions in chemical tests." Among mycobacteria we have found reaction rates to be characteristic of groups of strains, without necessarily being strict functions of growth rates.

When dealing with organisms that grow or metabolize slowly, one observes that the range of variation in reactivity in different tests is larger, in absolute terms, than with rapid growers. This expanded range of variability makes it quite difficult to decide how long a test must be observed or how intense a reaction must be to be considered positive. To avoid purely arbitrary decisions in this area, we have plotted frequency distributions of intensities or rates of reactions and, thus, have been able to select natural cutoff points at sites of lowest frequency, in quantitative adaptations of selected tests. By employing this approach we have, in some instances, encountered polymodal distributions of reaction rates (or intensities) and, thus, have been able to define several reaction clusters. On further analysis, it is found that cluster behavior in a given test correlates with patterns derived from other tests.

Examples of the use of reaction rate histograms are illustrated in Fig. 1. The mycobacteria may be broadly divided into rapid growers and slow growers, and members of these two general divisions behave differently in the microcolonial test; only the slow growers exhibit a positive neutral-red reaction in this test (17). Therefore, to demonstrate the relationship of growth rates and reaction rates among these organisms, the histograms were divided into two groups: the solid bars represent organisms which were positive in the microcolonial test, and the open bars represent cultures which were negative in that test.

The reactivity in two of the tests was expressed in terms of the number of weeks of incubation required to achieve an arbitrarily defined reaction. In the test for production of opacity in Tween-agar, most of the rapid growers yielded opacity within the first week, with fewer in the ensuing weeks. A number were negative in this test even after 6 weeks. The slow growers exhibited two major clusters, one yielding opacity between the 2nd and 4th weeks, and another which reacted very slowly or not at all within the 6 weeks of incubation. From this test, two characters were derived: (i) opacity produced within 1 week, and (ii) opacity produced within 4 weeks.

In the aryl sulfatase test, two major clusters were seen, both among slow growers and among rapid growers. From this test, one character was derived: appearance of free phenolphthalein within 3 weeks.

Reactivity in two of the tests was expressed in terms of intensity of reaction at a selected time. Four clusters were seen in the nitrate reduction test. The rapid growers produced over $10^3$ m$\mu$-moles of nitrite per ml with a few scattered exceptions. The slow growers fell into four groups (possibly five, although we elected not to segregate the small cluster occurring between log values of 3.0 and 3.2). Three characters were derived from this test: log millimicromoles of nitrite (i) exceeds 0.1, (ii) exceeds 1.4, and (iii) exceeds 2.9.

In the catalase test, the rapid growers were predominantly highly reactive, producing a

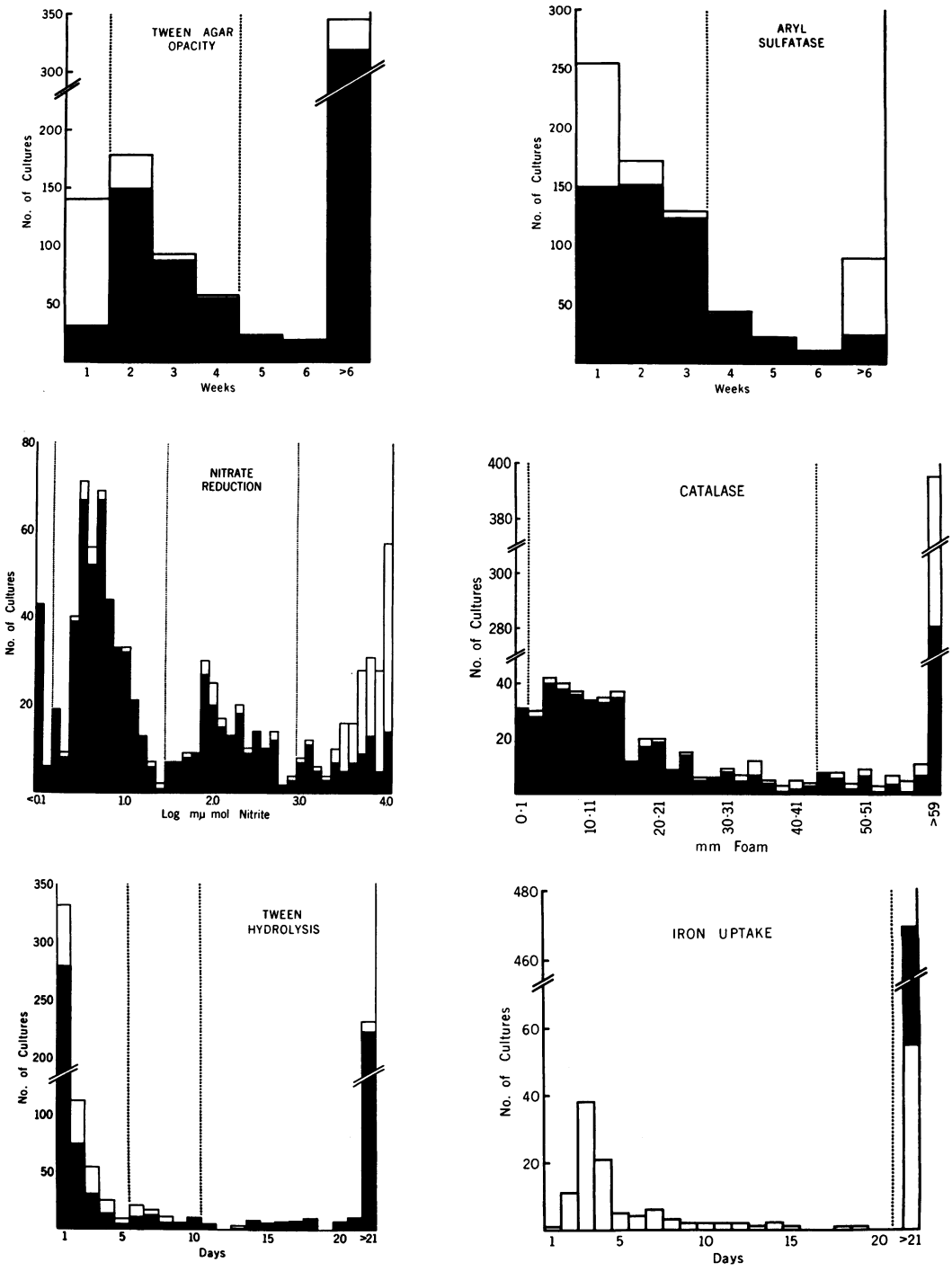


FIG. 1. *Distribution of quantitative responses of a collection of mycobacteria to six biochemical tests. The solid bars represent numbers of slow-growing strains yielding the indicated response, and the open bars represent the rapid growers.*

column of bubbles exceeding 60 mm. Three rather poorly demarcated reaction clusters were noted among slow growers. The catalase-negative strains (i.e., 1 mm of bubbles or less) were treated separately from the moderate and intense reactors. Two characters were derived from this test: (i) foam exceeds 1 mm, and (ii) foam exceeds 45 mm.

Two tests were scored on the basis of the number of days of incubation required to yield a positive reaction. Hydrolysis of Tween 80 occurred within 5 days with most rapid growers. A few nonhydrolyzers were seen in this group. Among slow growers, two large clusters occurred. A third, poorly defined cluster appeared between the 5th and 10th days, and it was treated separately, although the validity of this cluster may be questioned. Two characters were derived: (i) color change within 5 days, and (ii) color change within 10 days.

The iron uptake test yielded essentially positive or negative results, with no clear intermediate clustering. None of the slow growers showed rusting of colonies within 21 days. (Many were kept for 42 days without reacting.) The rapid growers could be divided into reactors and nonreactors. One character was thus derived: rusting of colonies within 21 days.

The six tests described illustrate the method used to select 11 characters for use in an Adansonian analysis. Similar studies were carried out to yield an additional 19 characters based on 13 other tests. Eleven of the tests yielded one character each, six tests yielded two characters each, and one test (nitrate reduction) and another (maximal growth temperature) yielded three and four characters each, respectively. Among the species and groups tested, none was distinguished from any other on the basis of characters derived from fewer than three different tests. These 30 characters were determined for 609 of our cultures. Because of the problems of dealing with an $n \times n$ table of similarities of this many cultures, and the lack of precedent for defining the limits of clusters of cultures described by this specialized set of characters, an alternative method of defining clusters was selected.

As the alternative to the use of $n \times n$ tables, we used judgment (perhaps more elegantly expressed as "educated bias"), based on preliminary visual scan of data, to establish "hypothetical median strain" (4) patterns for certain well-recognized groups, as well as some groups that we postulated. The reaction patterns of the entire collection were then sent to the computers at the Veterans Administration Western Research Support Center, to determine the similarity score ($S$ value) of each strain against reference cards based on each of the hypothetical median strains. For this purpose, each character was recorded on a positive or negative basis, and all matches, positive and negative, were included in each comparison. Histograms were then prepared to determine the distribution of *intra*taxon $S$ values for each pattern. The mean intrataxon $S$ value was also determined. Similarly, when the character patterns of the hypothetical median strain of each group were compared with one another, in a single $n \times n$ table, $S$ values were obtained which represent *inter*taxon similarity scores (Table 1). Thus, the evolution of clusters of organisms with high internal similarities (i.e.,

TABLE 1. *Similarity matrix of 15 hypothetical median strains in the genus Mycobacterium*[a]

| Organism | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. *"M." rhodochrous* | (91)[b] | | | | | | | | | | | | | | |
| 2. *M. phlei* | 80 | (92) | | | | | | | | | | | | | |
| 3. *M. smegmatis* | 70 | 90 | (95) | | | | | | | | | | | | |
| 4. *M. fortuitum* | 73 | 87 | 77 | (91) | | | | | | | | | | | |
| 5. *M. flavescens* | 63 | 77 | 67 | 83 | (91) | | | | | | | | | | |
| 6. *M. scrofulaceum* | 60 | 53 | 50 | 60 | 73 | (87) | | | | | | | | | |
| 7. "Battey" | 63 | 63 | 67 | 57 | 53 | 77 | (93) | | | | | | | | |
| 8. *M. avium* | 53 | 53 | 57 | 53 | 57 | 73 | 83 | (89) | | | | | | | |
| 9. *M. "nonchromogenicum"* | 70 | 70 | 67 | 77 | 73 | 77 | 80 | 77 | (87) | | | | | | |
| 10. *M. "terrae"* | 70 | 63 | 63 | 57 | 53 | 63 | 73 | 63 | 80 | (87) | | | | | |
| 11. *M. aquae* | 57 | 43 | 40 | 43 | 53 | 70 | 60 | 50 | 67 | 73 | (87) | | | | |
| 12. *M. kansasii* | 63 | 63 | 60 | 63 | 67 | 70 | 67 | 57 | 80 | 73 | 80 | (86) | | | |
| 13. *M. gastri* | 50 | 50 | 53 | 57 | 60 | 77 | 67 | 63 | 73 | 67 | 73 | 80 | (89) | | |
| 14. *M. bovis* | 40 | 33 | 37 | 40 | 43 | 73 | 63 | 73 | 57 | 63 | 63 | 63 | 83 | (87) | |
| 15. *M. tuberculosis* | 53 | 47 | 43 | 53 | 57 | 67 | 57 | 67 | 63 | 57 | 50 | 63 | 77 | 80 | (96) |

[a] Similarity scores ($S$) are expressed as percentage.

[b] Numbers in parentheses represent the mean intrataxon $S$ score for each cluster.

homogeneous groups) and low intertaxon similarity scores (i.e., marked differences between groups) is evidence of a consistent and useful classification.

Information on the clustering behavior of mycobacteria may be derived by studying the histograms illustrated in Fig. 2A, B, and C. The distributions of *S* values of the entire collection are plotted for each of the 15 hypothetical median strain reference patterns. The frequencies are plotted on a logarithmic scale for ease of visualization. The total height of each bar indicates the number of cultures with a given *S* value against the given reference. The height of the shaded portion of each bar represents the number of cultures for which the given *S* value is the largest of the 15 determined.

The relationship between *M. avium* and the "Battey" strains of Runyon's Mycobacterial Group III (7) is of considerable interest. In the "Battey" histogram (Fig. 2A), a total of 207 cultures exhibited *S* values of 77% or higher against the Battey reference pattern. Of these, 130 (63%) showed highest similarity to this reference pattern. An additional 21 (10%) showed highest similarity to the *M. avium* pattern. Of the remaining 56 cultures with Battey *S* values in excess of 77%, none exhibited a similarity score to this reference pattern in excess of 87%. Conversely, in *M. avium* histogram, a total of 175 cultures exhibited *S* values of 77% or higher against the *M. avium* reference pattern. Of these, 27 (15%) showed highest similarity to this reference pattern. An additional 116 (66%) showed highest similarity to the Battey pattern. Of the remaining 32 cultures with *M. avium S* values in excess of 77%, none exhibited a similarity score to this reference pattern in excess of 83%. Thus, a large proportion of both Battey and *M. avium* strains exhibit similarity scores against *both* reference patterns of 83% or higher; inasmuch as the similarity of these two hypothetical median strains to one another is 83% (see Table 1), these data do not appear to justify establishment of two distinct species. It is, of course, possible that addition of further characters unrelated to those used in this analysis might yield a clear distinction between these clusters.

Another pair of Group III mycobacterial types which appear to differ from *M. avium* and the Battey bacilli are designated here under the names *M. "terrae"* and *M. "nonchromogenicum."* In a prior publication (14) we had proposed the species name *M. terrae* for a group of nonpathogenic, nonpigmented, slow-growing mycobacteria. Tsukamura (10) had proposed the name *M. nonchromogeniucm* for what has since appeared to be a similar group of organisms. Subsequent to our publication of *M. terrae*, we learned that Tsukamura (11) had previously proposed that *M. nonchromogenicum* be renamed *M. terrae*. For purposes of this presentation, we use the name *M. "nonchromogenicum"* to designate organisms corresponding most closely to hypothetical median strain patterns derived from cultures received from Tsukamura, and *M. "terrae"* to designate organisms corresponding to the description we published. The *S* value obtained between these two reference patterns was 80%. As seen in the histograms (Fig. 2A), the shaded portions of both "species" are erratically polymodal and extend down into the 76% range. Of 80 cultures which gave a maximal *S* score against either or both of these reference patterns, 8 (10%) had tie scores for both, at levels between 76 and 86%. It is not yet clear whether these two groups should be combined as a single species, with a rather broad range of variation, or whether two or more species may be defined among these organisms.

Yet another nonpigmented, slow-growing *Mycobacterium* species has been classified under the name *M. gastri* (14). This species, although fairly distinct from the photochromogen *M. kansasii* (*S* = 80%), was reported to be similar to a low-catalase variety of *M. kansasii* (12). In Fig. 2A, a bimodal distribution of cultures with maximal similarity to the *M. gastri* pattern is seen. Furthermore, the *M. kansasii* histogram is distorted with high frequencies in the 76 to 83% *S* range. A comparison of these two groups demonstrated that 12 of 47 cultures in the *M. kansasii* shaded histogram with *S* values below 90%, and 10 of 33 cultures in the *M. gastri* shaded histogram with *S* scores below 93%, had previously been placed, on monothetic grounds, among the low-catalase group of *M. kansasii* cultures. Most of the remainder of the cultures in the left-hand *M. gastri* peak had been placed, on monothetic grounds, in a variety of other groups. The intermediate position of the photochromogenic strains with low catalase activity is of special interest as they may represent transition organisms in evolution of the pathogenic *M. kansasii* from free-living *M. gastri*.

In Fig. 2B, the histograms of *M. aquae* and *M. scrofulaceum*, although presenting some evidence of bi- or trimodal distribution, do indicate a division between two major subgroups among the scotochromogenic mycobacteria formerly lumped together in Group II. The reference patterns are markedly different from one another (*S* = 70%). Similarly, *M. bovis*, which was once considered to be a variety of *M. tuberculosis*, has an *S* score of only 80% against

this species, and these two groups of mammalian tubercle bacilli demonstrate a rather tight distribution against their respective reference patterns (Fig. 2B).

Histograms of *S* values against the five hypothetical median patterns of rapid growers (Runyon's Group IV) are presented in Fig. 2C. Organisms in this group characteristically exhibit a negative neutral-red reaction in the microcolonial test. Rather tight distributions are seen in the shaded portions of the histograms of *M. phlei*, *M. smegmatis*, and *"M." rhodochrous*. However, some of these tight patterns reflect the fact that most of the tests used to date had

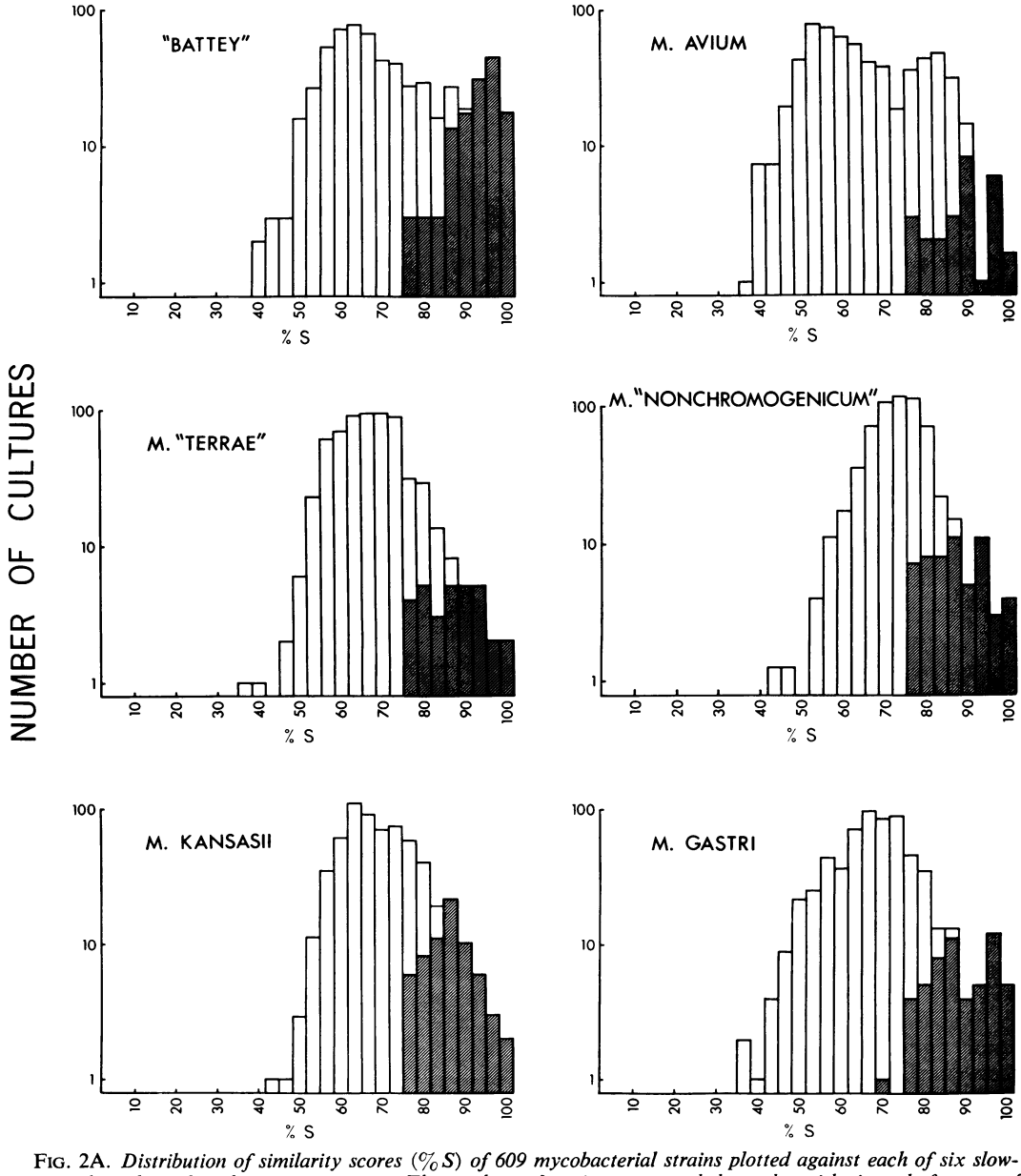


FIG. 2A. *Distribution of similarity scores (% S) of 609 mycobacterial strains plotted against each of six slow-growing hypothetical median strain patterns. The numbers of strains are recorded on a logarithmic scale for ease of visualization. The shaded bars represent strains for which the indicated % S is the highest seen against any of 15 patterns tested. The open bars represent strains which showed a higher % S against some other pattern.*

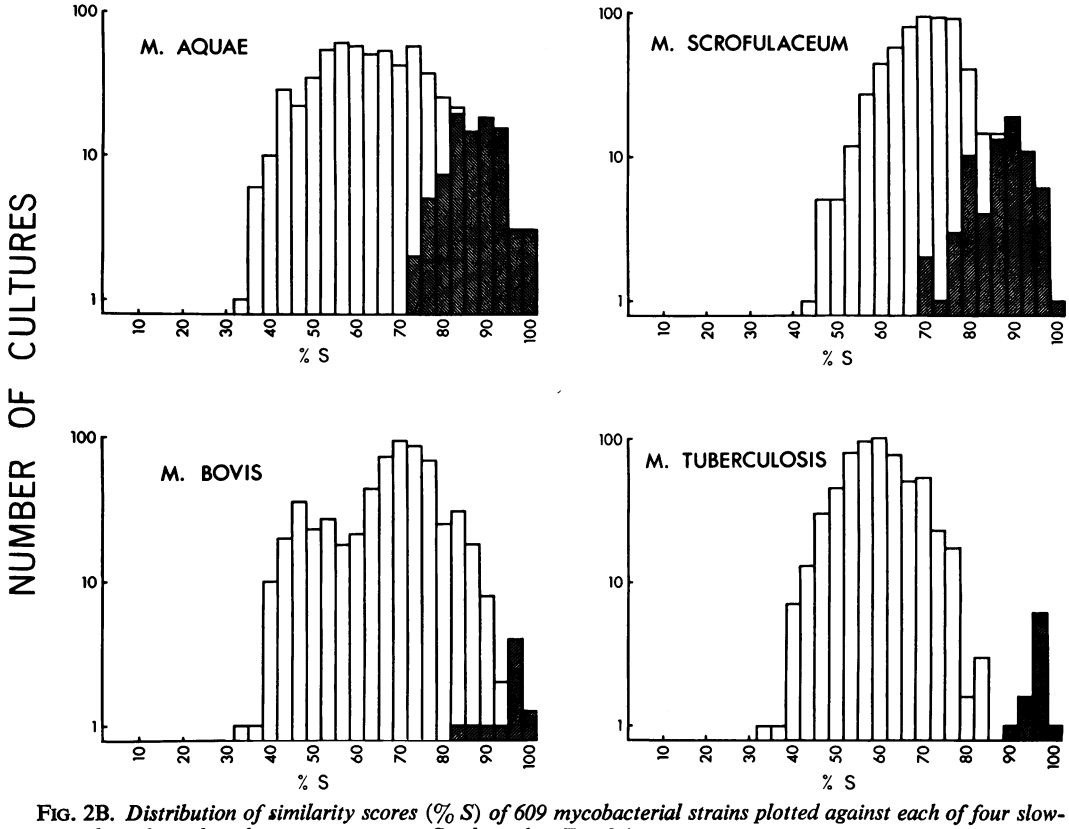


FIG. 2B. *Distribution of similarity scores (% S) of 609 mycobacterial strains plotted against each of four slow-growing hypothetical median strain patterns. See legend to Fig. 2A.*

been selected mainly for usefulness in subdividing the slow growers; thus, the hypothetical median strain of *M. phlei* has *S* scores of 90 and 87%, respectively, against *M. smegmatis* and *M. fortuitum*. If a comprehensive set of special tests is to be used for numerical analysis of the genus *Mycobacterium*, it will be necessary to balance this selection with tests which will account for the rapid growers as well. Of special interest is the histogram for "*M.*" *rhodochrous*. The distribution of *S* scores for these cultures is rather tight, and the hypothetical median strain exhibits *S* values of 80% or less against the rapid growers, and does not exceed 70% against any slow grower. Gordon has described these organisms as "some strains in search of a genus" (2), and the isolated position of the peak in the "*M.*" *rhodochrous* histogram attests to their peculiarity. *M. flavescens*, another pigmented, rapidly growing species, appears to be more at home in the genus, possibly functioning as a link between rapid and slow growers.

The foregoing discussion has dealt with problems remaining in the establishment of, definitions of, and divisions among closely related species. Most of the distinctions based on the proposed hypothetical median strain patterns hold up well. Thus, only 8 of 609 cultures examined have a maximal *S* score against *any* reference pattern of less than 77%. On the other hand, of 105 *inter*taxon *S* scores derived by comparing all 15 hypothetical median strain patterns to one another, only 20 were equal to or greater than 77%, and two thirds of these *S* scores were less than 70%. The mean *intra*taxon *S* scores (i.e., of all cultures showing maximal similarity to a given pattern) ranged from 86 to 96%.

Of 609 cultures analyzed, 68 were involved in ties for high score against two or more reference patterns. Of these, 57 cultures were involved in two-way ties, 8 in three-way ties, and 3 in four-way ties. Only eight cultures tied at the 90 to 93% *S* level, and all of these involved the *M. fortuitum* and *M. phlei* patterns. The distribution of ties among the hypothetical median strains is illustrated in Table 2. This distribution indicates a clustering pattern compatible with intertaxon similarities as seen in the intertaxon *S* scores in Table 1. A given culture will show up more than once if it was involved in a three- or four-way tie.

The sequence of tabulation of *S* scores (Table

1) suggests relationships ranging from "*M.*" *rhodochrous*, which is only grudgingly tolerated in the genus, through a number of rapidly growing saprophytes, to the slow growers, including pathogenic opportunists, and down to *M. tuberculosis* in the penultimate position of phenotypic isolation. (We must, for the time being, award the ultimate position in the genus to *M. leprae*, which has, so far, refused to permit analysis of its in vitro character.) This distribution is compatible with the concept that *M. bovis* and *M. tuberculosis* occur near the terminus of an evolutionary process, existing as they do, in nature, essentially as parasites, even though they are not strict parasites in the laboratory.

The inference of relative homogeneity of a cluster in the shaded bars of Fig. 2A, B, and C may only be made with respect to the right-

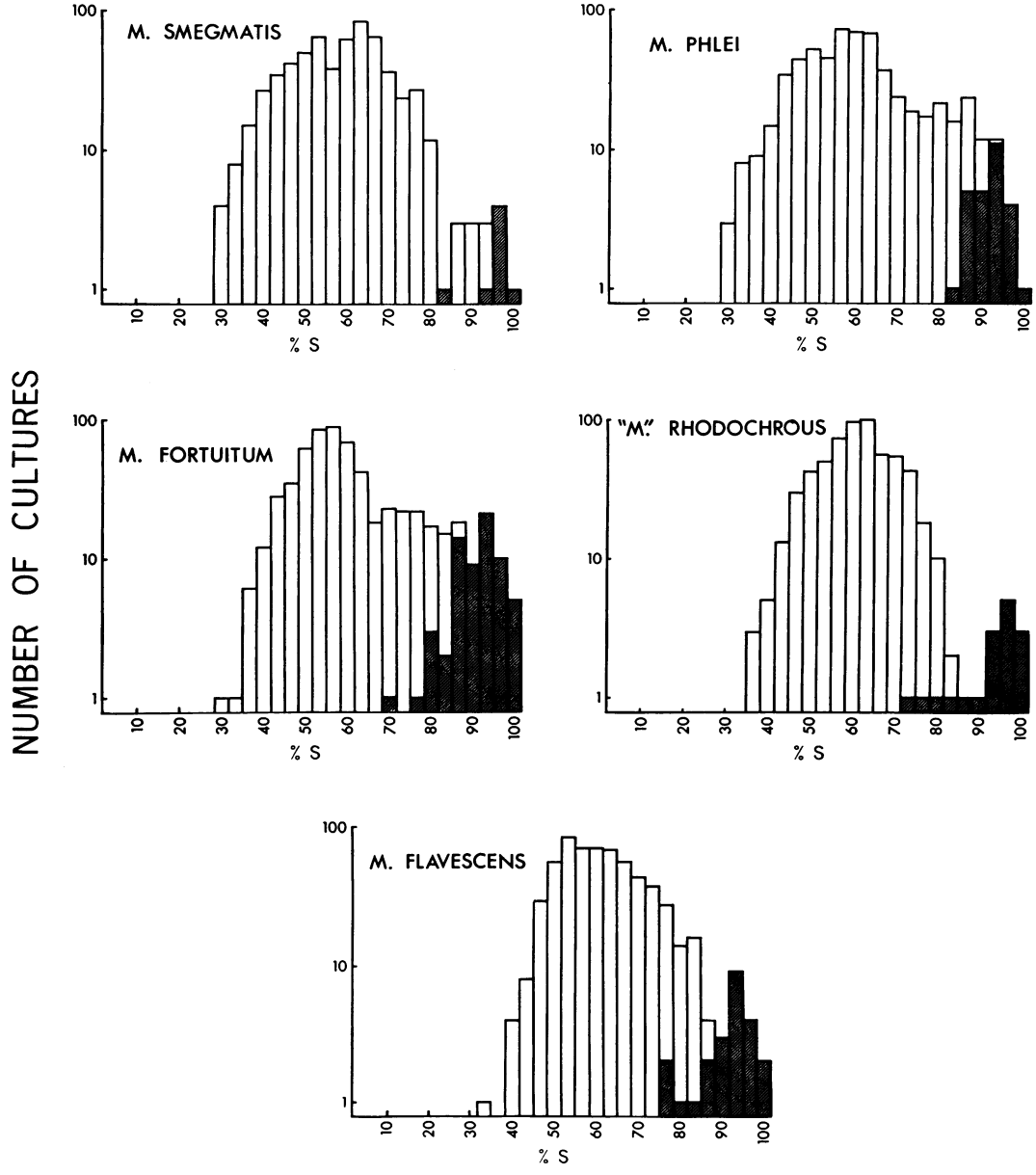


FIG. 2C. *Distribution of similarity scores (% S) of 609 mycobacterial strains plotted against each of five rapid-growing hypothetical median strain patterns. See legend to Fig. 2A.*

TABLE 2. *Distribution of ties among Mycobacterium cultures exhibiting maximal similarity (S) scores to two or more hypothetical median strains (HMS)*

| HMS no. | Reference pattern | Total strains[a] | No. of ties involving HMS no. | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1 | *"M." rhodochrous* | 17 | | | | | | | | | | | | | | | |
| 2 | *M. phlei* | 26 | — | | | | | | | | | | | | | | |
| 3 | *M. smegmatis* | 7 | — | — | | | | | | | | | | | | | |
| 4 | *M. fortuitum* | 69 | — | 13 | — | | | | | | | | | | | | |
| 5 | *M. flavescens* | 24 | — | — | — | — | | | | | | | | | | | |
| 6 | *M. scrofulaceum* | 69 | — | — | — | 1 | — | | | | | | | | | | |
| 7 | "Battey" | 129 | — | — | — | — | — | — | | | | | | | | | |
| 8 | *M. avium* | 27 | — | — | — | — | — | 2 | — | | | | | | | | |
| 9 | *M. "nonchromogenicum"* | 58 | — | — | — | — | 3 | -- | 5 | — | | | | | | | |
| 10 | *M. "terrae"* | 32 | — | — | 1 | — | — | — | 4 | — | 8 | | | | | | |
| 11 | *M. aquae* | 86 | — | — | — | — | — | — | 1 | — | 6 | 5 | | | | | |
| 12 | *M. kansasii* | 67 | — | — | — | — | 1 | — | 2 | — | 8 | 4 | 6 | | | | |
| 13 | *M. gastri* | 55 | — | — | — | — | 1 | — | — | — | 4 | — | 6 | 13 | | | |
| 14 | *M. bovis* | 9 | — | — | — | — | — | 3 | — | 2 | — | — | — | — | — | | |
| 15 | *M. tuberculosis* | 10 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | |

[a] Total strains indicates number included in shaded bars of Fig. 2A, B, and C.

hand peak, i.e., nearest the 100% point. Secondary peaks at lower levels may well be composed of two or more natural clusters, each differing in a different set of characters from the reference pattern.

This has been an interim report, suggesting some approaches that may be made to the numerical classification of mycobacteria. The statistical approaches still need polishing, because there are still questions dealing with tie scores, and decisions about cutoff points at low *S* values. The methods permit the evaluation of a proposed new species by examination of a small collection of representative cultures, establishment of a tentative hypothetical median strain pattern, and determination of *S* scores of an entire large culture collection against the new pattern. The overall pattern of those cultures clustering nearest the 100% point may then be re-examined to determine whether each character was correctly ascribed to the tentative hypothetical median strain. If over half of the cultures in such a cluster disagree with a character ascribed to the tentative hypothetical median strain, that character may be corrected in the description of the tentative median pattern to produce a more representative hypothetical median strain pattern. That is, the use of tentative sorts, and correction of reference patterns when necessary, should minimize the obvious hazards of preselection of a hypothetical median strain rather than definition of such a strain after a full polythetic $n \times n$ analysis has been made. Discovery of such new clusters should diminish the size of the secondary peaks in the shaded portions of the histograms in Fig. 2A, B, and C. Furthermore, it should then be possible to perform a polythetic $n \times n$ analysis on the relatively small number of cultures which occur in these secondary peaks, in order to detect new clusters. The use of the reference pattern analysis thus permits segregation of the large numbers of cultures which do fit well-recognized species, permitting recognition of smaller clusters which might have been obscured in the full collection. Only 30 characters were used in this analysis; more tests are needed which can provide characters which may improve distinction among taxa.

For practical purposes, it is usually possible to select a few characters that have the highest correlations with membership in a specific taxon, and to use these tests for *identification* of a great majority of the acid-fast bacilli usually encountered in diagnostic laboratories. Refinements in serological procedures should be most useful for providing some confirmations of identifications.

Presumably, everyone who finds himself treading the tortuous trail of numerical taxonomy should pause at intervals and ask two intricately related questions. Why are these organisms being classified, and why are these particular tests being used in this classification? These questions are especially crucial for mycobacteriologists. Inherent in the concept of so-called "natural" classification is an assumption that further random addition of characters will not significantly

change the *S* matrix. If, however, the original set of characters used was inappropriate, then this will not be true. A *presumably* random selection of characters from a battery of general bacteriological tests is actually a biased selection, in terms of *internal* analysis within the genus *Mycobacterium*, because so many tests are inappropriate for mycobacteria. For this reason, the general bacterial taxonomist may be satisfied with a lumping of strains which would be unacceptable to someone deeply concerned with subdivisions among slowly growing mycobacteria. Those who are so concerned must evolve their own specialized schemes, while still trying to keep bias within bounds.

Thus, in further extension of the concepts of Lockhart and Koenig (5), it would appear useful to prepare general and special schemes of numerical taxonomic analysis. A cluster evolving from a general analysis would be used, in effect, monothetically to select the special scheme to be employed. The special scheme could consist of secondary data, as defined by Lockhart and Koenig, as well as data from additional tests which are not included in the general scheme. Thus, although definition of limits of "species" would vary within different general clusters, this may be, for practical purposes, essentially a semantic problem, and definitions and labels would still be meaningful for most purposes. The extent of further subdivision of the initially derived cluster would be based on needs, i.e., the reason the classification was undertaken in the first place.

ACKNOWLEDGMENTS

The laboratory tests from which the characters were derived were performed with the excellent technical assistance of Joseph R. Doubek and Gilbert A. Diaz. Allie Broadway coded laboratory data for the numerical analysis. Edward Gocka and Edward Warburton, of the Veterans Administration Western Research Support Center, provided advice and adapted programs for, as well as carried out, numerical analysis of the data.

LITERATURE CITED

1. BOJALIL, L. F., J. CERBON, AND A. TRUJILLO. 1962. Adansonian classification and mycobacteria. J. Gen. Microbiol. **28**:333–346.
2. GORDON, R. E. 1966. Some strains in search of a genus—*Corynebacterium, Mycobacterium, Nocardia* or what? J. Gen. Microbiol. **43**:329–343.
3. GORDON, R. E., AND J. M. MIHM. 1959. A comparison of four species of mycobacteria. J. Gen. Microbiol. **21**:736–748.
4. LISTON, J., W. WIEBE, AND R. R. COLWELL. 1963. Quantitative approach to the study of bacterial species. J. Bacteriol. **85**:1061–1070.
5. LOCKHART, W. R., AND K. KOENIG. 1965. Use of secondary data in numerical taxonomy of the genus *Erwinia*. J. Bacteriol. **90**:1638–1644.
6. RICHMOND, L., AND M. M. CUMMINGS. 1950. An evaluation of methods of testing the virulence of acid-fast bacilli. Am. Rev. Tuberc. **62**:632–637.
7. RUNYON, E. H. 1965. Pathogenic mycobacteria. Advan. Tuberc. Res. **14**:235–287.
8. SNEATH, P. H. A. 1962. The construction of taxonomic groups, p. 304. *In* G. C. Ainsworth and P. H. A. Sneath [ed.], Microbial classification. Cambridge Univ. Press, Cambridge, England.
9. SZABO, I., AND E. VANDRA. 1963. Mycobacterium minetti (Penso *et al.* 1952). Bacteriological and epidemiological observations. Acta Microbiol. Acad. Sci. Hung. **10**:215–233.
10. TSUKAMURA, M. 1965. A group of mycobacteria from soil resources resembling nonphotochromogens (Group 3). Med. Biol. (Tokyo) **71**:110–113.
11. TSUKAMURA, M. 1966. Classification of mycobacteria. 1. Adansonian classification of slowly growing mycobacteria. Med. Biol. (Tokyo) **72**:75–78.
12. WAYNE, L. G. 1962. Two varieties of *Mycobacterium kansasii* with different clinical significance. Am. Rev. Respirat. Diseases **86**:651–656.
13. WAYNE, L. G. 1964. The mycobacterial mystique: Deterrent to taxonomy. Am. Rev. Respirat. Diseases **90**:255–257.
14. WAYNE, L. G. 1966. Classification and identification of mycobacteria. III. Species within Group III. Am. Rev. Respirat. Diseases **93**:919–928.
15. WAYNE, L. G., AND J. R. DOUBEK. 1965. Classification and identification of mycobacteria. II. Tests employing nitrate and nitrite as substrate. Am. Rev. Respirat. Diseases **91**:738–745.
16. WAYNE, L. G., J. R. DOUBEK, AND R. L. RUSSELL. 1964. Classification and identification of mycobacteria. I. Tests employing Tween® 80 as substrate. Am. Rev. Respirat. Diseases **90**:588–597.
17. WAYNE, L. G., I. KRASNOW, AND M. HUPPERT. 1957. Characterization of atypical mycobacteria and of nocardia species isolated from clinical specimens. I. Characterization by means of the microcolonial test. Am. Rev. Tuberc. Pulmonary Diseases **76**:451–467.