Diagnostic Probability Matrix for Identification of Slowly Growing Mycobacteria in Clinical Laboratories

LAWRENCE G. WAYNE,^{1,2*} MICAH I. KRICHEVSKY,³ DAVID PORTYRATA,⁴ AND CLAUDE K. JACKSON⁴

Tuberculosis Research Laboratory, Veterans Administration Medical Center, Long Beach, California 90822¹; Department of Medical Microbiology, College of Medicine, University of California, Irvine, California 92664²; Microbial Systematics Section, National Institute of Dental Research, Bethesda, Maryland 20205³; and ORI, Inc., Silver Spring, Maryland 20910⁴

Received 30 April 1984/Accepted 6 July 1984

A probability matrix is presented for identification of slowly growing mycobacteria that are likely to be encountered in clinical laboratories. The matrix includes 23 features that are useful for identifying members of 14 species or species complexes. The computer program identifies strains as a function of the ID (identification) score, which measures the discrimination among possible alternative identifications, and the R (ratio) score, which measures the degree of fit to the most likely taxa. It is not necessary to employ all 23 tests when initiating an identification; the program will suggest additional tests to perform when a partial data set fails to yield a definitive identification. Two independent sets of cultures comprising a total of 1,212 strains were used to test the matrix. Correct diagnoses were based on clustering behavior in numerical taxonomic analysis with larger numbers of features. The probable efficiencies with the two sets were 94.2 and 83.4%, respectively, and the accuracy of the definitive identifications for both sets exceeded 95%. A discussion is presented of situations when it may be appropriate to override an R score that has caused the rejection of an identification and to thereby enhance the efficiency.

The variety of species of mycobacteria that may be encountered in clinical specimens and the highly specialized nature of the tests used to identify them present formidable problems to clinical microbiologists. Fifteen years ago a perfect-fit strategy was presented; it was based on a battery of only nine tests and was useful for routine identification of members of 11 taxonomic groupings other than *Mycobacterium tuberculosis* that were commonly seen in clinical laboratories (28). Since that time more taxa have been recognized, necessitating the addition of more tests to permit resolution among the taxa. The chances of a strain deviating in one or more features from the norm for its species increases with the number of features examined, thus making it impractical to continue to rely on a perfect-fit strategy for identification. Instead, it becomes necessary to use a probabilistic approach, in which tests are weighted on the basis of the expected frequency of each feature for each taxon in the matrix.

A taxonomic probability matrix has been described for identifying slowly growing mycobacteria to the species level (34). It incorporated 34 features and omitted some species that were not yet sufficiently characterized. However, it is not necessary to identify all strains to the species level; members of closely related *Mycobacterium* species with comparable clinical significance may be combined into a species complex without compromising decisions on the clinical status of a given patient (28). Thus, *M. terrae*, *M. nonchromogenicum*, and *M. triviale* are almost never the cause of disease (37) and are so similar to one another as to be combined conveniently into an *M. terrae* complex. Likewise, the clinical management of infections caused by *M. avium* and by *M. intracellulare* is similar, so these two, closely related species may be combined as the *M. avium* complex. Such consolidations permitted the construction of a clinical probability matrix based on only 23 features for

identification of members of 14 species or complexes of slowly growing mycobacteria. Furthermore, most strains can be identified even when some of these features are omitted. For purposes of this study, slowly growing mycobacteria were defined according to the eighth edition of *Bergey's Manual* (16), i.e., "Grossly visible colonies apparent only after 7 or more days on media seeded with fresh inocula, diluted sufficiently to yield well isolated colonies."

MATERIALS AND METHODS

Construction of the matrix. The principles of numerical taxonomic (N.T.) analysis and construction of the diagnostic matrix are the same as used for the previously described taxonomic matrix (34). The species and species complexes selected for inclusion are shown in Table 1. Eight of the taxa are presented as discrete species in both the prior matrix and in the present one. Two of the species from the prior matrix, *M. avium* and *M. intracellulare*, were combined into a single group, the *M. avium* complex, for the present study. The feature frequencies for the complex were computed from pooled data of the strains examined in earlier cooperative studies of the International Working Group on Mycobacterial Taxonomy (IWGMT) (13). Similarly, the *M. terrae* complex represents a consolidation of the species *M. terrae* and *M. nonchromogenicum*, which appeared separately in the prior matrix; in addition, an adjustment was made in feature 13 (see below) to permit inclusion of *M. triviale* in this complex, even though it had not appeared in the earlier matrix. *M. microti*, which had appeared in the larger taxonomic matrix, is rarely if ever seen in human clinical specimens, is very similar to *M. tuberculosis* (34), and was omitted from the present matrix. Similarly, *M. bovis* BCG is so similar in behavior to the parent species, *M. bovis*, that a separate category was not included for this organism. A separate category for *M. africanum* was omitted for reasons discussed previously (2, 34). Four species that were not included in the prior matrix were incorporated into the new

* Corresponding author.

TABLE 1. Feature frequencies of properties used in probability matrix

Species or complex	Frequency (%) growth with the following features ^a :																						
	Urease	Catalase, quant.	Tween hydr.	Phosphatase	Nitrate	Scotochromogen	Photochromogen	Niacin	Isoniazid 1	Isoniazid 10	Thiacetazone 10	Pyrazinamidase	NaCl 5%	Galactosidase	Esterase	Catalase 68°	25°	45°	Oleate 250	TCH 1	HA 500	PNB 500	Picrate 0.2%
<i>M. gastri</i>	99	1	99	99	1	1	1	1	1	8	1	9	63	1	1	1	99	1	1	99	1	1	1
<i>M. marinum</i>	99	33	99	99	1	1	99	1	83	83	67	99	17	1	1	99	99	1	17	99	99	67	1
<i>M. kansasii</i>	99	99	99	99	97	3	94	1	15	15	9	78	1	1	9	99	99	1	9	99	11	89	1
<i>M. scrofulaceum</i>	91	99	1	1	18	99	1	1	99	86	30	99	1	1	25	99	99	1	91	99	85	86	1
<i>M. gordonae</i>	8	96	91	35	1	99	1	1	87	5	99	78	1	1	99	99	99	1	60	99	35	88	1
<i>M. xenopi</i>	1	1	1	1	1	86	1	1	1	1	99	50	1	1	99	99	1	99	1	99	1	50	1
<i>M. avium complex</i>	1	19	1	10	1	2	1	1	98	72	98	99	5	3	98	87	50	47	91	99	71	90	1
<i>M. terrae complex</i>	1	99	99	99	45	1	1	1	99	99	99	89	50	96	29	99	99	1	59	99	96	82	50
<i>M. tuberculosis</i>	99	1	70	99	98	1	1	99	4	1	10	94	1	50	99	2	5	1	15	96	1	2	1
<i>M. bovis</i>	99	1	1	99	8	1	1	1	1	1	1	1	1	50	99	1	1	1	1	1	1	5	1
<i>M. flavescens</i>	99	99	99	20	99	99	1	1	1	1	99	50	99	50	60	99	99	17	33	99	1	92	99
<i>M. simiae</i>	99	99	1	1	1	1	90	46	99	70	99	82	1	1	89	99	99	1	50	99	99	99	18
<i>M. szulgai</i>	80	99	80	99	99	99	1	1	50	1	99	99	1	1	75	99	99	1	50	99	1	99	1
<i>M. malmoense</i>	71	1	99	1	1	1	1	1	99	1	99	80	1	1	67	50	86	1	1	99	71	99	1

^a Observed frequency values of 100% are set at 99%, and those of 0% are set at 1% in this matrix. When frequency values are not known, they are set at 50%. A more complete description of features, definitions of the abbreviations, and references for the tests used are found in Table 2. TCH, Thiophene-2-carboxylic acid hydrazide; HA, hydroxylamine hydrochloride; PNB, *p*-nitrobenzoic acid.

one. *M. flavescens*, a nonpathogen which is usually treated as a rapid grower but exhibits an intermediate growth rate, bears some resemblance to the pathogen *M. szulgai* and was included for that reason. Three pathogens, *M. szulgai*, *M. simiae*, and *M. malmoense*, were not sufficiently characterized to permit inclusion in the earlier matrix but can now be entered by inclusion of data from recent cooperative studies of the IWGMT (32, 33).

Features used in the matrix. The 23 features used in the diagnostic matrix were coded in binary form by the RKC system (15); those features not yet entered into the permanent RKC code were assigned numbers in the 98,000 series. The features and references to descriptions of the techniques are listed in Table 2.

Twenty of the features in the present diagnostic matrix (see Table 2) were selected from the 34 used in the taxonomic matrix (34). Three more features were added, with frequencies derived from the pooled N.T. data of IWGMT studies (13, 23, 25, 32, 33). These features, numbered 10, 13, and 23 in Table 2, corresponded to growth in the presence of (per milliliter): 10 µg of isoniazid, 50 mg of NaCl, and 2 mg of picric acid, respectively. In a few instances, it was necessary to fill in missing values in Table 1 with feature frequencies derived from an independent data set based on N.T. analyses of the culture collection of L.G.W. or other published data (8, 21). Coding for colony pigment was modified from the earlier matrix by substituting "Colonies are scotochromogenic" (RKC 20087) for "Pigment other than chlorophyll produced" (RKC 20018); "Photochromogenicity" (RKC 20008) was coded as before.

As described previously (34), all feature frequencies above 99% were set at 99% and all frequencies below 1% were set at 1% in the matrix to eliminate multiplication by zero and undue influence from extreme values in the computation of probability products. Where frequency data were not available, values were set at 50%, which eliminated the influence of that test on the diagnosis of the taxon for which the data were lacking. Furthermore, the frequency for feature 13 (resistance to 50 mg of NaCl per ml) was set at 50% for the

M. terrae complex, thus permitting *M. triviale*, the only NaCl-resistant slow-grower, to be identified as a member of that complex.

Computation of identification scores. The probability matrix was incorporated into an identification program named IDDNEW (6), which is part of a group of computer programs that make up a microbial information system (9). IDDNEW is based on a modification of the Bayes theorem of probabilities. The logic of the procedure lies in the fact that, given independent tests, the probability of obtaining a particular combination of results is the product of the probabilities for obtaining the desired results in each of the tests taken singly. This statistical approach has been used widely in medical diagnosis and more recently in bacteriological studies (3, 4, 5, 11, 12, 35, 36).

The likelihood that an unknown strain belongs to any given taxon is defined as the probability of obtaining the observed test results with a strain of this taxon. This probability is found for each taxon by multiplying the probabilities of the individual test results.

The probability values actually used in the computations are decimal scores obtained by dividing the percent positive feature frequencies by 100. The results returned by the program also are displayed as the decimal probabilities.

When an unknown strain gives a positive result, the probability of the result is taken directly from the matrix. When the result given by an unknown is negative, the probability is found by subtracting the matrix figure from 1.00 (e.g., a 0.01 probability of being positive is equivalent to a 0.99 probability of being negative).

There are two cases in which the probabilities of single features interact with other features. A series of mutually exclusive conditions (termed multistate) share probabilities as if they were a single property. For example, colony pigmentation in which a colony could exhibit one of a number of possible colors would be a multistate feature. The combined probabilities of all possible pigmentation states should add up to 1.0. The usual scheme for coding consists of scoring as positive whichever color is found. The remain-

TABLE 2. Tests used to determine features employed in diagnostic matrix (Table 1)

Seq. no.	Feature	Feature code no. (RKC)	Feature key word(s) for Table 1	Reference(s)
1.	Urease is produced	34143	Urease	30, 14
2.	Catalase, over 45 mm of foam	98009	Catalase, quant.	31
3.	Tween 80 is hydrolyzed within 10 days	98030	Tween hydr.	30
4.	Acid phosphatase is produced	34136	Phosphatase	8
5.	Nitrate reduction, moderate ^a	98007	Nitrate	1, 28
6.	Colonies are scotochromogenic	20087	Scotochromogen	30
7.	Colonies are photochromogenic	20008	Photochromogen	30, 26
8.	Niacin accumulates in medium	24424	Niacin	31
9.	Grows in medium containing 1 µg of isoniazid per ml	16250	Isoniazid 1	31
10.	Grows in medium containing 10 µg of isoniazid per ml	16251	Isoniazid 10	31
11.	Grows in medium containing 10 µg of thiacetazone per ml ^b	16276	Thiacetazone 10	31
12.	Pyrazinamidase demonstrated by pyrazinoic acid assay in agar	98282	Pyrazinamidase	21
13.	Grows in medium containing 5% NaCl	18006	NaCl 5%	30
14.	β-Galactosidase is produced	34123	Galactosidase	30
15.	α-Esterase is produced	34146	Esterase	7
16.	Catalase resists inactivation at 68°, 20 min	24425	Catalase 68°	10
17.	Growth at 25°C	17014	25°	31
18.	Growth at 45°C	17017	45°	31
19.	Grows in medium containing 250 µg of oleic acid per ml	16089	Oleate 250	29
20.	Grows in medium containing 1 µg of thiophene-2-carboxylic acid hydrazide per ml	16315	TCH 1	31
21.	Grows in medium containing 500 µg of hydroxylamine hydrochloride per ml ^b	16263	HA 500	31
22.	Grows in medium containing 500 µg of <i>p</i> -nitrobenzoic acid per ml	16264	PNB 500	31
23.	Grows in medium containing 0.2% picric acid	16259	Picrate 0.2%	17

^a Moderate nitrate reduction was coded when results were reported as positive for nitrate reduction within 4 h by a modified Virtanen technique (1) or as the logarithm of nanomoles of nitrate is greater than 1.4 by the technique of Wayne and Doubek (27, 28).

^b This specific drug itself or this concentration of the drug was not included in the cited reference, but the technique described for the other drugs in that reference are applicable to the agent or concentration shown here.

der of the possible colors are scored as negative. A colorless colony is denoted by scoring all colors as negative. A probability matrix of positive feature frequencies would contain no entry for the probability of occurrence of colorless strains. The combined probabilities for the multistate feature would not add up to the required 1.0. To overcome this lack of a frequency corresponding to the complement of the positive states of the multistate feature, we have added a "pseudostate" which compensates for the missing probability. That is, if all the multistate features are scored as negative, the probability for the pseudostate is used as the probability for the multistate condition. The pseudostate is transparent to the user of the matrix but is used in the calculation of probabilities to ensure that the complement of all positive multistate features is accurately represented.

The second case is that of "linked" features, usually in pairs. The most common type of features that are linked are those in which growth at two concentrations of a material is assessed (e.g., antibiotics and chemical inhibitors). The features are linked rather than independent in that the probability of occurrence of the second feature (usually measured as growth) is necessarily no more than, and is normally less than, the probability of occurrence of the first feature. The probability for linked features is calculated differently for the four possible combinations of results. If the lower concentration is negative and the second is positive (−, +), the test is discarded as an error. If the two tests are positive (+, +), the probability of the first is used in the calculation. If both tests are negative (−, −), the probability of the second is used. The combination of the first test being positive and the second being negative (+, −) is calculated by subtracting the second probability from the first. It is this

last case that can give problems (35, 36). The matrix described in this communication has two species for which the probabilities of growth in 1 and 10 µg of isoniazid per ml were identical and of intermediate values (*M. marinum*, 0.83 and *M. kansasii*, 0.15). In such cases, the algorithm causes a (+, −) result to yield a probability of zero which reduces the calculation to nonsense. The usual strategy of substituting 0.01 for zero still prevents an isolate from being judged as identified by the criteria we use. (Extreme probabilities of 0.99 and 0.01 cause no problem since the combined results of [+ , −] are almost never encountered.) To prevent this skewing of results, we adopted the following strategy. When a (+, −) result was encountered, the difference between probabilities was determined. If the difference was greater than 0.05, the difference in probabilities was used as the combined probability in the usual manner. If it was 0 to 0.05, the probability of the first feature alone was used. Two reasons led us to this adjustment in the identification program. The first is that in those species having nearly identical tolerances for inhibitors, the actual threshold is above both concentrations. This effectively reduces the two features to one for such species. Second, the threshold difference of 0.05 was chosen because only differences greater than this value were deemed of usable significance in conditional probability calculations such as are performed in identification matrices (12).

To display clearly the relative values of the likelihoods of the taxa, they were normalized by dividing the likelihood of each taxon by the sum of the likelihoods of all of the taxa. The normalized values are referred to as identification (ID) scores. The best likelihood is defined as the maximum likelihood that each taxon could achieve by using only those

test features recorded for the unknown. It is assumed that the results obtained match their individual matrix probabilities; i.e., the features are set to positive when the values in the probability matrix are ≥ 0.50 , and they are set to negative for values of < 0.50 , and the best likelihood is calculated.

A second score is calculated as the ratio (R score) obtained by dividing the observed likelihood by the best possible likelihood for each taxon, when only those features recorded for a given strain were used (34).

Interpretation of the scores. As the program was written, a level of 0.99 was the threshold identification level; i.e., for a strain to be labeled as identified, its ID score had to equal or exceed 0.99. An ID score of 0.99 or greater to a given taxon means that the likelihood of the feature pattern of the strain being found in that taxon is at least 99-fold greater than in all other possible taxa combined. The ID score is thus a measure of discrimination.

However, the possibility remains that a strain may fit one taxon far better than all other taxa in the matrix, and still not belong to that taxon, simply because taxa exist that are not described in the matrix. Therefore, we introduced a measure of fit, the R score (34). The R score is the ratio of the observed likelihood score of the strain to the best possible likelihood for each taxon with the tests employed. An R score less than 0.01 is highly associated with an erroneous diagnosis in terms of actual N.T. cluster location of a strain (34) and was established as the threshold below which an identification would be rejected.

When the ID score was below 0.99, the computer scanned the matrix to determine whether additional tests could be performed which might shift the likelihood values sufficiently to yield an ID score of > 0.99 . However, when the R score for a given taxon was below the threshold of 0.01, the addition of more tests could not contribute to an identification; new data could not raise the R score since even a correct result would contribute equally to the observed likelihood and the best likelihood values.

"Special-case" interpretations. In preliminary evaluations of the diagnostic matrix, it became evident that strict adherence to the threshold criteria for ID and R scores could lead to some diagnostic failures which could be overridden safely by applying special interpretations. In the clinical diagnostic process, the most frequently required differentiation is that of members of the *M. tuberculosis* complex (i.e., *M. tuberculosis*, *M. bovis*, and intermediate forms such as *M. africanum*) from all other mycobacteria (22). Furthermore, the chemotherapeutic use of some drugs that are included in the matrix leads to resistant strains in this complex; the frequency of such resistances varies geographically from region to region, so it is not possible to provide consistent values for these features. Therefore, the decision was made to consider a strain that yielded an ID score above 0.990 for either species or a sum of the two ID scores greater than 0.990 for the two species *M. tuberculosis* and *M. bovis*, regardless of the value of the R score, as identified within the *M. tuberculosis* complex. The computer was not programmed to make this decision, but it is the option of the person interpreting the report.

Data sets for testing the matrix. To evaluate the efficiency and accuracy of the diagnostic matrix, data sets were needed for collections of cultures that had been identified by N.T. analysis based on a larger number of features than was included in the matrix. Two sources of such data were available. A series of IWGMT cooperative studies on slowly growing mycobacteria were based on N.T. study of 342 strains, with between 81 and 140 features per strain. From

studies on slowly growing mycobacteria other than members of the *M. tuberculosis* complex (13, 23, 25, 32, 33), an average of 20 features included in the diagnostic matrix were available per strain. From the study of the tuberculosis complex, an average of only 12.5 matrix features per strain were available.

Since the matrix was developed predominantly from IWGMT data, it was desirable to use data from a separate collection as an additional test of the matrix. W series strains, comprising a collection of 870 cultures of slowly growing mycobacteria in the collection of L.G.W., had been subjected to N.T. analysis based on 42 features (19). Seventeen of these features are in the diagnostic matrix, permitting further independent analysis of the matrix from incomplete data.

RESULTS

The results of application of the diagnostic matrix to the two data sets were interpreted in terms of their agreement with clustering behaviors in the full N.T. analysis. Of 342 strains from the series of IWGMT studies, 279 (81.6%) met or exceeded both threshold criteria ($ID \geq 0.99$; $R \geq 0.01$) for identification (Table 3). In addition, 14 met the special-case criteria cited above, i.e., $ID (M. tuberculosis) + ID (M. bovis) \geq 0.99$, regardless of R score. Of the total 293 strains so identified, all but 10 of the identifications agreed with N.T. clustering behavior, corresponding to 96.6% accuracy. A total of 15 species or complexes are represented in this list, but errors in diagnosis, i.e., disagreements with N.T. clustering, were restricted to four of these groups, *M. scrofulaceum*, *M. gordonae*, the *M. avium* complex, and *M. malmoense*. Nine of the disagreements involved strains that fit no recognized N.T. cluster, and one strain misidentified as *M. gordonae* fell into the *M. avium* N.T. cluster. Especially noteworthy is the observation that application of the special-case criteria permitted correct recognition of 14 members of the *M. tuberculosis* complex that would have been rejected by strict application of the ID and R score thresholds. Included among these 14 strains were 1 that clustered with *M. tuberculosis*, 1 with *M. bovis*, 2 with *M. microti*, 4 with *M. bovis* BCG, and 6 with *M. africanum*.

An additional 16 strains had ID scores for the most likely species or complex that were below the threshold but had R scores that did not reject the identification, i.e., $R \geq 0.01$ (Table 3). The best-fit identification for 11 of these agreed with their N.T. clustering, and the ID scores of all but 2 of them fell in the range of 0.949 to 0.987. Some matrix data were missing for all strains in this series, and it became evident on individual review that most of these strains would probably have reached the identification thresholds if missing tests had been performed.

A total of 43 strains in the IWGMT series exhibited best R scores below the threshold; 9 of these were among the 14 covered in the special-case criteria for the *M. tuberculosis* complex. Further testing to fill in missing data for the remaining 34 strains (Table 4) (corresponding to 9.9% of the strains examined) could not have facilitated a diagnosis, since adding data can only improve the ID score, not the R score (34). Based on N.T. cluster distributions of these strains, 11 of them appear to be atypical representatives of their taxa, and the remaining 23 either represent taxa that were not included in the matrix (*M. shimoidae* and *M. asiaticum*) or are strains which fell into none of the recognized N.T. clusters. In that sense, the results obtained with the matrix for these 23 strains (6.7% of the strains examined) are correct.

TABLE 3. IWGMT series strains with R scores equal to or greater than 0.01 for their most likely identities, grouped by most likely identity

Most likely identity ^a	ID scores ≥ 0.99			ID scores < 0.99 (no. of strains)		
	No. of strains that:		% Correct	Same cluster	Other cluster	No cluster
	Meet criteria	Disagree with N.T. cluster				
<i>M. tuberculosis</i>	52	0	100	0	0	0
<i>M. bovis</i>	24	0	100	0	0	0
<i>M. tuberculosis</i> complex ^a	14	0	100	0	0	0
<i>M. scrofulaceum</i>	21	3	85.7	2	0	0
<i>M. gordonae</i>	24	4	83.3	4	1	3
<i>M. avium</i> complex	44	2	95.5	4	0	0
<i>M. xenopi</i>	7	0	100	0	0	0
<i>M. terrae</i> complex	23	0	100	0	0	0
<i>M. gastri</i>	12	0	100	0	0	0
<i>M. kansasii</i>	37	0	100	0	0	0
<i>M. marinum</i>	6	0	100	0	0	0
<i>M. flavescens</i>	4	0	100	0	0	1
<i>M. simiae</i>	12	0	100	0	0	0
<i>M. szulgai</i>	5	0	100	1	0	0
<i>M. malmoense</i>	8	1	87.5	0	0	0

^a For the *M. tuberculosis* complex special-case criteria, ID (*M. tuberculosis*) + ID (*M. bovis*) ≥ 0.99 , regardless of R score, were used. Of a total of 342 strains tested, 293 (85.7%) met the criteria. Of these, 10 (3.4%) disagreed with N.T. clustering. The percent correct was 96.6%. For strains with ID scores of < 0.99 , 11 (3.2%) were in the same N.T. cluster, 1 (0.3%) was in another cluster, and 4 (1.2%) were in no cluster.

When this matrix was designed, the first 13 features (Table 1) were considered the most critical ones, and the remaining tests were included for further support. This consideration was tested by analyzing the IWGMT series of strain data in a short matrix, comprising only the features numbered 1 through 13 in Table 1, but including all taxonomic groups listed in that table. The use of only 13 features in the matrix had the expected result of reducing the proportion of strains that met the criteria for identification from 85.7% achieved with the full panel (Table 3) to only 75.6% for the short panel; the proportion of erroneous identifications among these did not change.

An analysis based on the full matrix was also made on the independent data set from the W series of strains. Whereas each of these 23 features had been available for some or all of the strains in the IWGMT set, features 4, 12, 14, 15, 21, and 26 from Table 2 were lacking for all strains of the W series. Data for the remaining 17 features in the matrix were available for each of the strains in this series. Of 870 strains in this series, 519 met or exceeded both standard threshold criteria for identification (Table 5), and nine more met the special-case criteria for the tuberculosis complex. Of the total of 528 strains so identified, all but 19 of the identifications agreed with N.T. clustering, corresponding to 96.4% accuracy, the same as the accuracy seen with the IWGMT series (Table 3). The disagreements between matrix diagnosis and N.T. cluster were distributed among seven different species or complexes in the W series, as compared with only four in the IWGMT series.

An additional 112 W series strains had ID scores for the most likely species or complex that were below the threshold but had R scores that did not reject the identification. The best fit for 79 of these strains agreed with the N.T. clustering, so most of these would probably have been identified if missing data had been available. Another 24 strains fit no recognized N.T. cluster, and additional data would probably have led to depression of their R scores below threshold for any species or complex.

A total of 240 strains in the W series exhibited best R

scores below the threshold; 10 of these were covered under the special-case criteria for the *M. tuberculosis* complex. Of the remainder, 118 appeared to be atypical representatives of the taxa with which they clustered in N.T. analysis, and 64 strains were unclustered or fell into an N.T. cluster (*M. shimoidei*?) that was not included in the matrix. Eleven strains in this group clustered with *M. kansasii* but were recognized as low-catalase environmental strains (18) of the species, which occupy a transitional position between *M. kansasii* and *M. gastri*.

TABLE 4. IWGMT series strains not identified because R scores were below the rejection threshold of 0.01, grouped by N.T. cluster

N.T. cluster	No. of strains showing best fit to ^a :	
	Same species or complex	Other species or complex
<i>M. tuberculosis</i> ^b	0	0
<i>M. bovis</i> ^b	0	0
<i>M. scrofulaceum</i>	1	0
<i>M. gordonae</i>	2	0
<i>M. avium</i> complex	5	0
<i>M. xenopi</i>	0	0
<i>M. terrae</i> complex	0	0
<i>M. gastri</i>	0	0
<i>M. kansasii</i>	2	0
<i>M. marinum</i>	0	0
<i>M. flavescens</i>	0	1
<i>M. simiae</i>	0	0
<i>M. szulgai</i>	0	0
<i>M. malmoense</i>	0	0
<i>M. shimoidei</i> ^c	0	4
<i>M. asiaticum</i> ^c	0	3
Fall in no named cluster		16

^a Of 342 strains tested, 34 (9.9%) had R scores below the rejection level, 10 (2.9%) showed best fit to the same species or complex, and 24 (7.0%) showed best fit to another species or complex.

^b Special-case interpretation for *M. tuberculosis* and *M. bovis* (see Table 3).

^c *M. shimoidei* and *M. asiaticum* not included in matrix because of insufficient data to calculate feature frequencies.

TABLE 5. W series strains with R scores equal to or greater than 0.01 for their most likely identities, grouped by most likely identity

Most likely identity	ID scores ≥ 0.99			ID scores < 0.99 (no. of strains)		
	No. of strains that		% Correct	Same cluster	Other cluster	No cluster
	Meet criteria	Disagree with N.T. cluster				
<i>M. tuberculosis</i>	12	0	100	0	0	0
<i>M. bovis</i>	5	0	100	0	1	0
<i>M. tuberculosis</i> complex	9	1	88.9	0	0	0
<i>M. scrofulaceum</i>	77	8	89.6	7	1	6
<i>M. gordonae</i>	40	0	100	38	0	7
<i>M. avium</i> complex	183	5	97.3	10	1	5
<i>M. xenopi</i>	11	0	100	0	0	1
<i>M. terrae</i> complex	96	1	99.0	3	0	3
<i>M. gastri</i>	23	0	100	7	0	0
<i>M. kansasii</i>	47	1	97.9	9	0	1
<i>M. marinum</i>	5	2	60.0	4	2	0
<i>M. flavescens</i>	17	0	100	0	0	0
<i>M. simiae</i>	3	1	66.7	1	2	0
<i>M. szulgai</i>	0	0		0	2	0
<i>M. malmoeense</i>	0	0		0	0	1?

^a For the *M. tuberculosis* complex, special-case criteria were used (as in footnote a of Table 3). Of a total of 870 strains tested, 528 (60.7%) of strains met the criteria. Of these, 19 (3.6%) disagreed with N.T. clustering. The percent correct was 96.4%. For strains with ID scores < 0.99 , 79 (9.1%) were in the same cluster, 9 (1%) were in other clusters, and 24 (2.8%) were in no cluster.

DISCUSSION

Although there was a marked difference in the proportion of strains that were identified (i.e., exceeded threshold values for both the ID and R scores) in the two data sets tested, it is striking that the disagreement with N.T. clustering among those so identified was only 3.4% for the IWGMT series and 3.6% for the W series (Tables 3 and 5). This agreement indicates that identifications can be accepted with a high degree of confidence when the defined criteria for matrix identification are met.

Of the combined strains that had R scores that exceeded threshold for one or more species or complex in the matrix (Tables 3 and 5) and were thus not rejected from all of the possible groups, 16 of 309 (5.4%) in the IWGMT series and 112 of 640 (17.5%) in the W series had ID scores that were below the threshold. This difference in apparent efficiency of the matrix with the two data sets can be explained by the fact that an average of 20 matrix features were available for the IWGMT strains (other than those in the *M. tuberculosis* complex), but only 17 of these features were available for the W strains. Unfortunately, it was not possible to retrieve and test these large collections of strains to fill in the missing data, but extrapolation of probable results suggests that the discriminating ID scores would have been pushed over the threshold for most of them, permitting definitive identification to be made.

The apparent failures of the identification matrix are represented in largest part by strains with an R score that was below the threshold for all possible taxa (Tables 4 and 6). For these strains, addition of missing data would not have permitted identification, since adding data can improve only an ID score, not an R score (34). Only 9.9% of the IWGMT series had R scores below threshold, but 26.5% of the W series were rejected on this basis. This discrepancy can be explained in part by considering N.T. clustering behavior. Only 16 of 342 (4.7%) of the IWGMT strains in Table 4 fell into no recognized N.T. cluster, and thus should not have been identified. That is, these are not failures, but rather correct answers from the matrix. On the other hand, 64 (7.4%) of the 870 W strains in Table 6 did not belong to

recognized N.T. clusters and, in this sense, also are not failures.

The most notable difference in apparent efficiency of the matrix when applied to the two test populations is represented by the distribution of strains that had appeared in recognized N.T. clusters, but were rejected in the diagnostic matrix by virtue of low R scores for all of the species or complexes that were included in the matrix. This occurred in only 11 of 342 (3.2%) of the IWGMT strains (Table 4), but in 158 (18.2%) of the 870 W strains (Table 6). We can only

TABLE 6. W series strains not identified because R scores were below the rejection threshold of 0.01, grouped by N.T. cluster

N.T. cluster ^a	No. strains showing best fit to ^b :	
	Same species or complex	Other species or complex
<i>M. tuberculosis</i>	0	0
<i>M. bovis</i>	0	0
<i>M. scrofulaceum</i>	9	3
<i>M. gordonae</i>	13	5
<i>M. avium</i> complex	43	7
<i>M. xenopi</i>	1	0
<i>M. terrae</i> complex	27	0
<i>M. gastri</i>	4	2
<i>M. kansasii</i>	17	19 ^c
<i>M. marinum</i>	3	2
<i>M. flavescens</i>	1	0
<i>M. simiae</i>	0	2
<i>M. szulgai</i>	0	0
<i>M. malmoeense</i>	0	0
<i>M. shimoidei</i>	0	8 ^d
<i>M. asiaticum</i>	0	0
Fall in no named cluster		64

^a See footnotes b and c of Table 4.

^b Of a total of 870 strains tested, 230 (26.5%) had R scores below the rejection level, 118 (13.6%) showed best fit to the same species or complex, and 112 (12.9%) showed best fit to another species or complex.

^c Of these strains, 11 composed a tight N.T. cluster considered to be a low-catalase, environmental biovar of *M. kansasii* (18).

^d These eight strains resemble those described as *M. shimoidei* in an IWGMT study (32).

speculate that this difference is a reflection, in part at least, of a bias in the selection of strains that compose the W series. This set of cultures was accumulated during the period of 1955 to 1970, when various clinical microbiologists and research investigators were encouraged to send us problem organisms, which were not readily identified by tests then in common use. This procedure resulted in an enrichment of this collection with cultures that were aberrant in one or more of the key features that were used for identification, and these deviations from the norm could have led to their rejection in the present matrix.

In our opinion, this probability matrix should serve as a supplement to, not a substitute for, the skills of a competent microbiologist. No automated system should preempt the judgment of qualified and experienced clinical laboratory personnel. The main value of a matrix of this type is to provide a rapid analysis of available data and definitive answers for the uncomplicated majority of cases, i.e., strains that exceed the ID and R score threshold criteria. This frees the microbiologist to concentrate on strains that require additional efforts. In many instances this will simply involve repeating selected tests or identifying and performing additional tests that are included in the matrix but which may not be performed routinely on all mycobacterial cultures received in a given laboratory. In other instances, the microbiologist may be justified in overriding an R score that rejects a strain and may make an identification based on a high ID score, when his or her knowledge of the taxonomy and physiology of mycobacteria provides a strong rational basis for doing so.

One example of a decision to override a low R score is found in the category of strains identified as members of the *M. tuberculosis* complex in Tables 3 and 5. In the combined data from the IWGMT and W series strains, the decision added another 23 strains to the 93 strains specifically identified as *M. tuberculosis* or *M. bovis*, and only 1 strain disagreed with N.T. clustering behavior. The original decision to employ these special-case criteria was based on the knowledge that the distinctions between members of the *M. tuberculosis* complex have been exaggerated by a bias in the selection of tests used to identify mycobacteria, which was a consequence of the special clinical importance of members of this complex (22).

Another example is represented by 11 strains in the W series cited in footnote c of Table 6, which were recognized as members of a tight N.T. subcluster of environmental strains of *M. kansasii* but were rejected from any matrix taxon by virtue of low R scores. Originally, these strains had been distinguished from the disease-associated biovar of *M. kansasii* by virtue of their low catalase activity (feature 2 in Table 2) (18); analysis of the phenol-soluble surface antigens of four of these strains showed them to belong to the very specific *M. kansasii* serovar (20). Subsequent studies have demonstrated that mycobacteria produce two distinct classes of catalase, designated as M and T, which differ from one another in tolerance to heat and inhibitors, substrate kinetics, and serological reactivity (24). Most strains of *M. kansasii* produce both classes of catalase, but the loss of the M catalase in these environmental strains resulted in negative scores for features 2 and 16 in Table 1 (catalase, >45-mm foam and catalase resists inactivation at 68°C, respectively). When the probability matrix analysis was rerun on these 11 strains, but with their results in features 2 and 16 suppressed, all yielded R scores for *M. kansasii* in excess of 0.01, and 10 of them also had ID scores for this species above 0.99, thus permitting identification. Since cultures of

this biovar are rarely encountered, it would not be appropriate to establish a separate category in the matrix to accommodate them. Instead, whenever a strain is encountered that produces the very characteristic pattern of the photochromogenicity, Tween hydrolysis, and nitrate reduction of *M. kansasii* but falls below threshold criteria for identification, it would be appropriate to determine the effect of suppression of features 2 and 16 on the identification based on probabilistic analysis of the balance of available data.

The efficiency of the matrix can be estimated by considering only those test strains that fell into an N.T. cluster that was represented by a species or complex that was included in the matrix. The IWGMT series of strains provides a "best case" estimate of efficiency of the matrix because most of the feature frequencies recorded in the matrix were derived from analysis of this series, the data used in testing the matrix corresponded to modal results from several laboratories, and the series was assembled without special emphasis on inclusion of aberrant strains of recognized species. Of the 322 (94.2%) strains in the IWGMT series that clustered with a matrix taxon by N.T. analysis, 293 met the criteria for identification indicated in Table 3, and an additional 11 would probably have been identified if missing features were provided; thus a total of 304 strains (94.4%) reflect the probable efficiency of the matrix with this test set.

The W series, on the other hand, represents an approximation of a "worst case" test of the matrix. The feature frequencies were derived from an independent set of strains, the data represented a single performance of each test (with rare exceptions), the collection was assembled with the intention of accumulating strains that had presented difficulties in other laboratories, and results of fewer tests were available. Of the 782 (89.9%) strains that clustered into a matrix taxon by N.T., 528 met the criteria for identification (Table 5), and as many as 88 more might have been identified if missing data were provided; a total of 616 strains (83.4%) thus reflects the probable efficiency with this test set. However, a review of samples of failures suggests that carefully considered decisions to override the results of selected tests that had caused the rejection of some identifications by virtue of low R scores could have yielded correct identifications approaching 90% for this series as well.

ACKNOWLEDGMENTS

This study was supported by the Medical Research Service of the Veterans Administration.

We express our gratitude to Elaine J. Krichevsky for patient and expert data management of the data bases used in these studies and to Cynthia A. Walczak for understanding the nature of the problems and writing computer programs to cope.

LITERATURE CITED

1. Bönicke, R., E. Rohrscheidt, and E. Pascoe. 1962. Die Verbreitung der Nitratreduktase innerhalb der Gattung *Mycobacterium*. Sonderb. Naturwiss. 49:43-44.
2. David, H. L., M.-T. Jahan, A. Jumen, J. Grandy, and E. H. Lehman. 1978. Numerical taxonomy analysis of *Mycobacterium africanum*. Int. J. Syst. Bacteriol. 28:464-472.
3. Dybowski, W., and D. A. Franklin. 1968. Conditional probability and the identification of bacteria: a pilot study. J. Gen. Microbiol. 54:215-229.
4. Friedman, R. B., D. Bruce, J. MacLowry, and V. Brenner. 1973. Computer-assisted identification of bacteria. Am. J. Clin. Pathol. 60:395-403.
5. Friedman, R., and J. MacLowry. 1973. Computer identification of bacteria on the basis of their antibiotic susceptibility patterns. Appl. Microbiol. 26:314-317.

6. Johnson, R. 1979. Computer-aided identification. FDA By-Lines 9:235-250.
7. Kppler, W. 1965. Acetyl-naphthylamin-esterasen-aktivitt von Mykobakterien. Beitr. Klin. Tuberk. Spezif. Tuberk. Forsch. 130:1-4.
8. Kppler, W. 1965. Zur Differenzierung von Mykobakterien mit dem Phosphatase-Test. Beitr. Klin. Tuberk. Spezif. Tuberk. Forsch. 130:223-226.
9. Krichevsky, M. I. 1979. The microbial information system (MICRO-IS): a system overview. FDA By-Lines 9:217-222.
10. Kubica, G. P., and G. L. Pool. 1960. Studies on catalase activity of acid-fast bacilli. Am. Rev. Respir. Dis. 81:387-391.
11. Lapage, S. P., S. Bascomb, W. R. Willcox, and M. A. Curtis. 1970. Computer identification of bacteria, p. 1-22. In A. Baillie and R. J. Gilbert (ed.), Automation, mechanization and data handling in microbiology. Academic Press, Inc., London.
12. Lapage, S. P., S. Bascomb, W. R. Willcox, and M. A. Curtis. 1973. Identification of bacteria by computer: general aspects and perspectives. J. Gen. Microbiol. 77:273-290.
13. Meissner, G., K. H. Schrder, G. E. Amadio, W. Anz, S. Chaparas, H. W. B. Engel, P. A. Jenkins, W. Kppler, H. H. Kleeberg, E. Kubala, M. Kubin, D. Lauterbach, A. Lind, M. Magnusson, Z. Mikova, S. R. Pattyn, W. B. Schaefer, J. L. Stanford, M. Tsukamura, L. G. Wayne, I. Willers, and E. Wolinsky. 1974. A cooperative numerical analysis of nonscoto and nonphotochromogenic slowly growing mycobacteria. J. Gen. Microbiol. 83:207-235.
14. Murphy, D. B., and J. E. Hawkins. 1975. Use of urease test disks in the identification of mycobacteria. J. Clin. Microbiol. 1:465-468.
15. Rogosa, M., M. I. Krichevsky, and R. Colwell. 1971. Method for coding data on microbial strains for computers (edition AB). Int. J. Syst. Bacteriol. 21:1A-175A.
16. Runyon, E. H., L. G. Wayne, and G. P. Kubica. 1974. Family II. Mycobacteriaceae Chester 1897, 63, p. 681-701. In R. E. Buchanan and N. E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.
17. Tsukamura, M. 1967. Identification of mycobacteria. Tubercle 48:311-318.
18. Wayne, L. G. 1962. Two varieties of *Mycobacterium kansasii* with different clinical significance. Am. Rev. Respir. Dis. 86:651-656.
19. Wayne, L. G. 1970. On the identity of *Mycobacterium gordonae* Bojalil and the so-called tap water scotochromogens. Int. J. Syst. Bacteriol. 20:149-153.
20. Wayne, L. G. 1971. Phenol-soluble antigens from *Mycobacterium kansasii*, *Mycobacterium gastri*, and *Mycobacterium marinum*. Infect. Immun. 3:36-40.
21. Wayne, L. G. 1974. Simple pyrazinamidase and urease tests for routine identification of mycobacteria. Am. Rev. Respir. Dis. 109:147-151.
22. Wayne, L. G. 1982. Microbiology of the tubercle bacilli. Am. Rev. Respir. Dis. 125(Suppl.):31-41.
23. Wayne, L. G., L. Andrade, S. Froman, W. Kppler, E. Kubala, G. Meissner, and M. Tsukamura. 1978. A cooperative numerical analysis of *Mycobacterium gastri*, *Mycobacterium kansasii*, and *Mycobacterium marinum*. J. Gen. Microbiol. 109:319-327.
24. Wayne, L. G., and G. A. Diaz. 1982. Serological, taxonomic, and kinetic studies of the T and M classes of mycobacterial catalase. Int. J. Syst. Bacteriol. 32:296-304.
25. Wayne, L. G., T. M. Dietz, C. Gernez-Rieux, P. A. Jenkins, W. Kppler, G. P. Kubica, J. B. G. Kwapinski, G. Meissner, S. R. Pattyn, E. H. Runyon, K. H. Schrder, V. A. Silcox, A. Tacquet, M. Tsukamura, and E. Wolinsky. 1974. A cooperative numerical analysis of scotochromogenic slowly growing mycobacteria. J. Gen. Microbiol. 66:255-271.
26. Wayne, L. G., and J. R. Doubek. 1964. The role of air in the photochromogenic behavior of *Mycobacterium kansasii*. Am. J. Clin. Pathol. 42:431-433.
27. Wayne, L. G., and J. R. Doubek. 1965. Classification and identification of mycobacteria. II. Tests employing nitrate and nitrite as substrate. Am. Rev. Respir. Dis. 91:738-745.
28. Wayne, L. G., and J. R. Doubek. 1968. Diagnostic key to mycobacteria encountered in clinical laboratories. Appl. Microbiol. 16:925-931.
29. Wayne, L. G., J. R. Doubek, and R. Russell. 1964. Classification and identification of mycobacteria. I. Tests employing Tween 80 as substrate. Am. Rev. Respir. Dis. 90:588-597.
30. Wayne, L. G., H. C. Engbaek, H. W. B. Engel, S. Froman, W. Gross, J. Hawkins, W. Kppler, A. G. Karlson, H. H. Kleeberg, I. Krasnow, G. P. Kubica, C. McDermont, E. E. Nel, S. R. Pattyn, K. H. Schrder, S. Showalter, I. Tarnok, M. Tsukamura, B. Vergmann, and E. Wolinsky. 1974. Highly reproducible techniques for use in systematic bacteriology in the genus *Mycobacterium*: tests for pigment, urease, resistance to sodium chloride, hydrolysis of Tween 80, and β -galactosidase. Int. J. Syst. Bacteriol. 24:412-419.
31. Wayne, L. G., H. W. B. Engel, C. Grassi, W. Gross, J. Hawkins, P. A. Jenkins, W. Kppler, H. H. Kleeberg, I. Krasnow, E. E. Nel, S. R. Pattyn, P. A. Richards, S. Showalter, M. Slosarek, I. Szabo, I. Tarnok, M. Tsukamura, B. Vergmann, and E. Wolinsky. 1976. Highly reproducible techniques for use in systematic bacteriology in the genus *Mycobacterium*: tests for niacin and catalase and for resistance to isoniazid, thiophene 2-carboxylic acid hydrazide, hydroxylamine, and *p*-nitrobenzoate. Int. J. Syst. Bacteriol. 26:311-318.
32. Wayne, L. G., R. C. Good, M. I. Krichevsky, R. E. Beam, Z. Blacklock, S. D. Chaparas, D. Dawson, S. Froman, W. Gross, J. Hawkins, P. A. Jenkins, I. Juhlin, W. Kppler, H. H. Kleeberg, I. Krasnow, M. J. Lefford, E. Mankiewicz, C. McDermont, G. Meissner, P. Morgan, E. E. Nel, S. R. Pattyn, F. Portaels, P. A. Richards, S. Rsch, K. H. Schrder, V. A. Silcox, I. Szabo, M. Tsukamura, and B. Vergmann. 1981. First report of the cooperative, open-ended study of slowly growing mycobacteria by the International Working Group on Mycobacterial Taxonomy. Int. J. Syst. Bacteriol. 31:1-20.
33. Wayne, L. G., R. C. Good, M. I. Krichevsky, R. E. Beam, Z. Blacklock, H. L. David, D. Dawson, W. Gross, J. Hawkins, P. A. Jenkins, I. Juhlin, W. Kppler, H. H. Kleeberg, I. Krasnow, M. J. Lefford, E. Mankiewicz, C. McDermont, E. E. Nel, F. Portaels, P. A. Richards, S. Rsch, K. H. Schrder, V. A. Silcox, I. Szabo, M. Tsukamura, L. Vanden Breen, and B. Vergmann. 1983. Second report of the cooperative, open-ended study of slowly growing mycobacteria by the International Working Group on Mycobacterial Taxonomy. Int. J. Syst. Bacteriol. 33:265-274.
34. Wayne, L. G., E. J. Krichevsky, L. L. Love, R. Johnson, and M. I. Krichevsky. 1980. Taxonomic probability matrix for use with slowly growing mycobacteria. Int. J. Syst. Bacteriol. 30:528-538.
35. Willcox, W. R., and S. P. Lapage. 1975. Methods used in a program for computer-aided identification of bacteria, p. 103-119. In R. J. Pankhurst (ed.), Biological identification with computers. Academic Press, Inc., London.
36. Willcox, W. R., S. P. Lapage, S. Bascomb, and M. A. Curtis. 1973. Identification of bacteria by computer theory and programming. J. Gen. Microbiol. 77:317-330.
37. Wolinsky, E. 1979. Nontuberculous mycobacteria and associated diseases. Am. Rev. Respir. Dis. 119:107-159.