Development and Evaluation of a Program and Probability Matrix for the Computer-aided Identification of Non-tuberculous Mycobacteria†

Enrico Tortoli*, Vieri Boddi[1] and Valeria Penati[2]

Bacteriological and Virological Laboratory, Careggi Hospital, Florence, Italy
[1]Institute of General Pathology, Florence University, Florence, Italy
[2]Institute for Chest Diseases, TB Laboratory, Milan, Italy

Received: 8 April 1992; revised 17 June 1992

The authors developed a program for the computer-assisted identification of mycobacteria using the results of conventional laboratory tests, and founding it on an empirical probability matrix largely based on published data.

The program's performance was assessed both in a retrospective study and in a prospective one. In the former, 6184 recorded laboratory results on 419 mycobacterial isolates, which had already been identified subjectively according to routine procedures in two laboratories, were fed into the computer in order to obtain its proposed identifications. In the latter, 164 unknown isolates were characterized according to the chosen set of laboratory tests, and their results were used for automated identifications as well as identifications by the expert, who also took into account the program's suggestions.

A good agreement between the program and the expert was achieved in both studies: 86.64% in the first and 94.51% in the latter. The increased efficiency may be due to the choice of laboratory tests based on insight developed during the construction of the program, and to the fact that the microbiologist was helped by hints given during the computation.

The program therefore appears to be a valid adjunct to the microbiologist's experience for the identification of mycobacteria.

Introduction

The probabilistic approach for microbial speciation is universally accepted and utilized by all commercially available kits for bacterial identification. The final identification is usually suggested by software that computes data received from an automatic reader, referring them to a built-in probability matrix; otherwise, code books are available where numerical profiles, i.e. concise transcriptions of all the results of identification tests, are combined with corresponding species.

For mycobacteria, no miniaturized test gallery is currently available. Consequently, neither software nor code books for the identification of these microorganisms have been developed by any firm. Various taxa of bacteria, among which are mycobacteria, are, however, included in a recently described microbiological database management and analysis system (Portyrata & Krichevsky, 1992).

The aim of this paper is to assess the performance of a program, developed by two of the authors, for the computer-aided identification of mycobacteria.

The program, which may be obtained (free of charge) from the corresponding author, was written using *QuickBASIC* (Microsoft Corporation), and is designed to run on IBM-compatible microcomputers, in a DOS environment. The diagnostic matrix, unlike other parts of the program, is open, so that a user with little experience in handling ASCII files can modify it at will.

The core of the system is its probability matrix of test results from mycobacterial species (Table 1); the entries are per cent frequencies of strains of each species giving a positive result for each individual test. The current matrix includes 27 species, among which are all those most commonly isolated from clinical specimens, and 23 commonly performed identification tests.

The matrix was compiled by pooling published data from various sources (Kubica, 1973; Roberts *et al.*, 1991; Wayne *et al.*, 1984; Wayne & Kubica, 1986) according to our own empirical experience, which dictated the pooling criterion when conflicting data were found in the literature; when only qualitative information was found for a test result, we arbitrarily assigned a percentage value consistent with it. Finally, for cells with no certain data, the value was set at 50.

Upon entering each isolate's vector of test results (positive, negative, or not done) the program computes the probability of its belonging to each of the species in the matrix: if a test is scored positive the corresponding values of the matrix remain unchanged; if a test is scored negative the values are subtracted from 100; and finally, for tests not done, the figures of the matrix are ignored. In order to eliminate multiplications by zero, 100 and 0 values of the matrix were set to 99 and 1, respectively.

†Presented in part at the 12th meeting of the European Society for Mycobacteriology in Prague, 2—3 July 1991

*Address for correspondence:
Dr Enrico Tortoli
Laboratorio di Microbiologia e Virologia, Ospedale di Careggi, Viale Pieraccini 24, 50139 Firenze, Italia

BINARY Vol. 4 pp. 200-203 © 1992 Academic Press

0226-304X/92/060200+04 $03.00/0

Table 1 Probability matrix of positive test percentages.

Taxon	Tests																					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
M. africanum	50	50	1	2	1	1	1	50	50	1	99	1	1	1	1	1	99	50	50	1	1	1	50
M. asiaticum	1	1	90	90	90	1	90	10	90	1	90	90	1	50	1	1	50	1	99	70	90	50	50
M. avium/intracellulare	1	1	87	90	19	2	2	10	1	5	90	50	47	24	76	1	50	1	99	98	71	98	91
M. bovis/M. bovis-BCG	1	8	1	5	1	1	1	1	1	1	99	1	1	1	1	1	99	99	1	1	1	1	2
M. chelonae	3	8	50	90	97	1	1	10	20	59	1	96	1	90	51	90	50	90	90	90	96	50	50
M. flavescens	1	99	99	92	99	99	99	10	99	99	50	99	17	1	32	1	1	99	99	99	1	1	33
M. fortuitum	7	99	99	99	98	1	1	85	33	99	1	90	1	99	43	99	50	99	99	99	96	50	99
M. gastri	1	1	1	1	1	1	1	90	90	1	90	90	1	16	1	1	50	90	99	9	1	8	1
M. gordonae	1	1	90	88	96	90	90	10	91	1	90	90	1	1	1	1	1	8	99	90	35	87	60
M. haemophilum	1	1	1	10	1	1	1	50	1	1	90	90	1	1	1	1	90	1	90	50	50	90	50
M. kansasii	1	97	99	89	99	3	97	30	99	1	99	99	1	1	1	1	50	99	99	9	11	15	9
M. malmoense	1	1	50	90	1	1	1	10	90	1	90	86	1	50	50	1	1	71	99	90	71	90	1
M. marinum	1	1	90	67	33	1	90	50	90	17	90	90	1	1	1	6	50	90	99	67	90	83	17
M. nonchromogenicum	1	45	90	82	90	1	1	50	98	1	90	90	10	1	1	50	50	1	99	90	96	90	70
M. phlei	1	90	90	50	90	90	90	50	90	90	1	50	90	1	62	1	90	95	90	90	10	50	90
M. scrofulaceum	1	18	90	86	90	90	90	10	1	1	90	90	1	1	4	1	1	91	99	30	85	90	91
M. shimoidei	1	1	99	90	1	1	1	50	99	1	99	10	50	1	50	1	99	1	99	99	1	99	50
M. simiae	46	1	90	90	90	1	90	10	1	1	90	90	1	50	90	1	1	90	99	90	90	90	70
M. smegmatis	1	99	79	50	76	1	1	70	99	99	1	80	99	1	90	10	50	50	50	99	10	50	99
M. szulgai	1	99	99	99	99	99	99	70	80	1	99	99	1	1	50	50	50	80	99	99	1	30	50
M. terrae	1	45	90	82	90	1	1	50	90	10	90	90	1	3	1	1	50	1	99	90	96	90	59
M. thermoresistibile	1	99	50	50	50	99	99	50	50	99	1	50	99	1	1	1	50	50	50	50	1	50	50
M. triviale	4	45	90	82	92	1	1	10	90	90	90	90	10	1	1	26	90	1	99	90	96	90	50
M. tuberculosis	99	98	2	2	1	1	1	90	70	1	99	5	1	1	2	1	99	99	96	10	1	4	15
M. ulcerans	1	1	90	50	1	1	1	50	1	1	90	90	1	50	1	1	90	1	99	50	50	90	50
M. vaccae	1	90	90	50	90	90	90	70	90	63	1	90	10	1	70	1	1	50	50	90	10	50	50
M. xenopi	1	1	90	90	1	86	86	10	1	1	90	1	90	1	10	60	1	1	99	90	1	1	1

Definition of tests: (A) niacin; (B) nitrate reduction; (C) catalase after 20 min at 68°C; (D) p-nitrobenzoic acid (500 µg/ml) tolerance; (E) catalase over 45 mm of foam; (F) pigmentation in the dark; (G) pigmentation in the light; (H) ß-glucosidase; (I) Tween 80 hydrolysis (10 days); (J) NaCl 5% tolerance; (K) slow growth (>7 days); (L) growth at 25°C; (M) growth at 45°C; (N) growth in MacConkey medium; (O) tellurite reduction; (P) arylsulfatase; (Q) rough colonies; (R) urease; (S) thiophene-2-carboxylic acid hydrazide (1 µg/ml) tolerance; (T) thiacetazone (10 µg/ml) tolerance; (U) hydroxylamine hydrochloride (500 µg/ml) tolerance; (V) isoniazid (1 µg/ml) tolerance; (W) oleic acid (250 µg/ml) tolerance.

According to the Bayes' theorem of probabilities, the likelihood of a strain belonging to a given taxon corresponds to the product of the probabilities of its single independent test results. Consequently, the program multiplies the probabilities of all the tests for every species in the table and chooses the product with the highest value. Besides the identification with the best likelihood, an alternative choice is always proposed and, when a third identification exists with normalized likelihood ≥10%, this one too is suggested. The normalized probability of each identification is given by the ratio of the corresponding product to the sum of the products for all the species.

Clearly, a maximum likelihood identification will always occur, whatever the test results entered in the program; this does not imply that such an identification is correct. A measurement of the correctness of an identification is constituted by the typicalness of its test results; for its evaluation the program calculates the ratio of its computed likelihood to the product expected for the same species (the 'expected' product is the maximum likelihood possible, and is achieved when the results of all the tests performed are the most probable). The program considers the typicalness of an identification 'very good' if the above mentioned ratio is >0.01, 'good' for values ≤0.01 and >0.005, 'insufficient' for values ≤0.005 and >0.001, and 'rejectable' when the values are ≤0.001.

Additional features of the program are the emphases laid on the tests conflicting with each one of the identifications proposed, and on those tests, among the ones not done, that could result in discriminating or supporting the above mentioned identifications. A test is considered 'conflicting' when its probability for the identification proposed is less than 20%. A test not done is considered 'discriminating' between the first and the second choice identification, if the absolute value of the difference between the corresponding data in the matrix is >80. A test not done is considered 'supporting' for an identification if the corresponding value in the matrix is either >80 or <20.

Finally, the program produces a report with all the above mentioned features.

Materials and methods

In order to assess the program's performance, two separate studies where done.

Retrospective Study

Laboratory records on mycobacterial isolates which had previously been characterized and identified were used to obtain 6184 test results from 419 mycobacterial strains (mean 14.8 tests per strain); 244 of the isolates were from Milan, 175 from Florence.

BINARY Vol. 4 © 1992 Academic Press Limited

Table 2 Statistical evaluation of agreement of 419 conventional and computer-aided identifications in the retrospective study.

Taxon	N[1]	K[2]	Wrong identifications[3]
M. africanum (AF)	3	0.85	-
M. avium/intracellulare (AI)	78	0.96	-
M. bovis/M. bovis-BCG (BB)	11	0.95	1AF
M. chelonae (CH)	45	0.78	1AI,1SM,1TE,1TR
M. flavescens (FL)	10	0.33	5PH,1VA,2TH
M. fortuitum (FO)	70	0.76	2AI,16CH,4SM,1ML,1UL
M. gastri (GA)	1	1.00	-
M. gordonae (GO)	49	0.93	1PH,2VA,1XE,1AS,1SZ
M. kansasii (KA)	14	1.00	-
M. marinum (MR)	2	0.81	-
M. phlei (PH)	4	0.17	1VA,2TH
M. scrofulaceum (SC)	18	0.88	2XE,1TH
M. smegmatis (SM)	1	n.a.	1TR
M. terrae (TE)	10	0.95	-
M. triviale (TR)	8	0.85	1SM,1ML
M. tuberculosis (TB)	4	0.85	1SH
M. vaccae (VA)	1	0.33	-
M. xenopi (XE)	90	0.96	2AI,1SC

Total value of Kappa coefficient = 0.846 (standard error 0.018).
[1]Number of strains.
[2]Kappa coefficient.
[3]Abbreviations of species not included in column Taxon. AS: *M. asiaticum*; ML: *M. malmoense*; SH: *M. shimoidei*; SZ: *M. szulgai*; TH: *M. thermoresistibile*; UL: *M. ulcerans*. The figure preceding the abbreviation of each species indicates the number of identifications wrongly attributed to that species.

Table3 Sensitivity and specificity of 164 computer-aided identifications estimated in the prospective study.

Taxon	N[1]	SE[2]	SP[3]	Wrong identifications[4]
M. africanum (AF)	1	0	100	1BB
M. asiaticum (AS)	1	100	100	-
M. avium/intracellulare (AI)	22	95	100	1SI
M. bovis/M. bovis-BCG (BB)	7	86	99	1HA
M. chelonae (CH)	24	100	99	-
M. flavescens (FL)	2	100	100	-
M. fortuitum (FO)	11	82	100	2CH
M. gordonae (GO)	10	100	100	-
M. kansasii (KA)	17	94	100	1AS
M. malmoense (ML)	1	100	100	-
M. marinum (MR)	1	100	100	-
M. scrofulaceum (SC)	1	100	100	-
M. shimoidei (SM)	1	100	100	-
M. simiae (SI)	1	100	99	-
M. terrae (TE)	5	100	100	-
M. thermoresistibile (TH)	1	0	100	1PH
M. tuberculosis (TB)	5	60	100	2AF
M. vaccae (VA)	3	100	100	-
M. xenopi (XE)	50	100	100	-

[1]Number of strains.
[2]Sensitivity.
[3]Specificity.
[4]Abbreviations of species not included in column Taxon. HA: *M. haemophilum*; PH: *M. phlei*. The figure preceding the abbreviation of each species indicates the number of identifications wrongly attributed to that species.

Prospective Study

The program was introduced experimentally (in the microbiological laboratory at Florence) into the routine identification of mycobacteria, which was changed accordingly to a more systematic testing protocol; 164 consecutive strains were thus identified by the expert, taking into account the hints given by the program and its automatic algorithm. In this group of identifications, almost all the tests included in the program were performed (mean 22.0 tests per strain); both sets of final identifications were based on the overall test results.

In the first study the identifications reached by the expert and those proposed by the program were compared and their agreement was assessed using the Kappa coefficient (K) of Cohen (1960). In the prospective study the identifications reached by the expert with the program's help were taken as the gold standard on which the sensitivity and the specificity of the program were evaluated.

Results

The agreement between the computer results of the identification in the two participating laboratories did not differ significantly (p=0.42 using the chi-square test), so aggregated results are given here.

In the retrospective evaluation, the comparison between the previously achieved identification and the one proposed by the computer as the most likely, revealed an exact agreement in 86.64% of cases (363 strains); of the remainder 56 identifications, 30 (7.16%) agreed with the species proposed by the program as its second identification choice, and five (1.19%) with the third choice proposal. In 21 cases (5.01%), none of the identifications suggested by the computer corresponded to that achieved previously using conventional methods.

The use of K showed the overall agreement of the two systems to be excellent. Values of K range from 1, when the agreement is complete, to 0, when it is equal to chance-agreement. According to Landis & Koch (1977), the agreement between two raters is excellent when K is greater than 0.75 and it is good when K is greater than 0.40. In our study (Table 2) only for three species (represented by just four strains) was the agreement poor. The satisfying level of such results is emphasized by the fact that only the first choice identifications proposed by the program have been taken into account for the evaluation of the agreement.

The accuracy of the computer identifications (by the chi-square test), was strongly related to the number of tests performed (p<0.01). In fact, if the identifications are divided into various classes on the basis of the number of tests performed, it can be seen that, while the agreement of those achieved using 20 or more tests exceeds 98%, it falls to 70% for the ones based on less than 13 tests.

In our opinion, however, the most meaningful evaluation is the one based on the prospective study; in the retrospective estimation it is in fact impossible to infer which results would have been achieved if the suggestions of the program (execution of further tests, repetition of some of them, etc.) were taken into account.

In this evaluation the automatic identification was correct in 94.51% of cases (155 strains) and wrong in the remaining 5.49% (nine strains). It is unimportant that, for eight of nine misidentified strains, the second choice identification proposed by the program was correct; evidently the compliance with the suggestions was not sufficient to shift the probability to the correct taxon.

The sensitivity (number of strains correctly identified

as belonging to a single taxon/total number of strains of that taxon) and the specificity (number of strains not attributed to a given taxon/number of strains not really belonging to that taxon) of the program are reported, along with misidentifications, in Table 3.

The analysis of disagreements revealed that the only kind of computer misidentifications found most frequently, concerned, in the retrospective evaluation of the program, 16 strains of *Mycobacterium fortuitum* speciated as *Mycobacterium chelonae*. In these cases the misleading tests were often nitrate reduction, and/or growth on MacConkey agar. This kind of error was, however, almost eliminated in the prospective study thanks to the repetition of tests signalled as 'conflicting' and, above all, to the inclusion in our protocols of some discriminating tests, such as β-glucosidase.

Discussion

Mycobacterial identification is one of the fields that has been less involved in the technological revolution that has characterized modern microbiology. DNA probe technology has so far been applied to just a few species; so that conventional speciation still remains a milestone in the study of these microorganisms.

The experience of the microbiologist is a heritage which will never be replaced by software; thus, the aim of the present program is not to substitute, but to assist the microbiologist. Moreover, an inexperienced operator will reasonably misidentify less mycobacteria by resorting to the probabilistic taxonomy on which this program is based, than by perusing one of the tables reported in microbiological handbooks.

The performance of our program seems to be very good according to the reported results; the agreement values achieved (>85%) seem to be very high, considering that 100% agreement is in fact out of reach of systems involved with biological variability. The program is, of course, subject to improvement, mainly by updating the probability matrix and by adding new tests and new species as soon as further data become available, for example, by addition of data from the recently published fourth report of the International Working Group on Mycobacterial Taxonomy (Wayne *et al.*, 1991), but the occurrence of strains with unpredicted features will never be ruled out. From this point of view, what might be considered minor features of the program, such as the alternative identifications, the level of typicalness and suggestions for performing or repeating tests, qualify it as particularly useful for the microbiologist.

The identification of *Mycobacterium tuberculosis* (MT), even if it is included with other species in the matrix, is not a main aim of the current program. The characterization of this species is generally performed in microbiological laboratories on the basis of a screening limited to only few tests, and usually does not present excessive problems. The few strains of MT included in this evaluation of the program had some atypical results in one or more of the tests used in our laboratories for the screening of this species, and consequently a thorough examination was performed. Our screening for MT identification includes just six tests (niacin, nitrate reduction, 68°C catalase, pigment, growth rate, colonial morphology); needless to say, the input into the program of typical results for these tests is sufficient to evoke the correct identification.

Since important innovations to the traditional methods of identification of mycobacteria are not universally available, programs like the one assessed here still appear to be useful tools in freeing the microbiologist from the less creative part of his work.

Acknowledgement

We would like to thank P. Urbano, Institute of Microbiology of Florence University (Florence, Italy) for the critical revision of the manuscript.

References

COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educucational and Psychological Measurement* **20**, 37-46.

KUBICA, G.P. (1973). Differential identification of mycobacteria. VII. Key features for identification of clinically significant mycobacteria. *American Review of Respiratory Diseases* **107**, 9-21.

LANDIS, J.R. & KOCH, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174.

PORTYRATA, D.A. & KRICHEVSKY, M.I. (1992). MICRO-IS: A microbiological database management and analysis system. *Binary* **4**, 31-36.

ROBERTS, G.D., KONEMAN, E.W. & KIM, Y.K. (1991). Mycobacterium. In *Manual of Clinical Microbiology*, 5th edition, pp. 304-339. Edited by A. Balows, W.J. Jr. Hausler., H.D. Isenberg & H.J. Shadomy. Washington: American Society for Microbiology

WAYNE, L.G., GOOD, R.C., KRICHEVSKY, M.I., BLACKLOCK, Z., DAVID, H.L., DAWSON, D., GROSS, W., HAWKINS, J., VINCENT LEVY-FREBAULT, V., McMANUS, C., PORTAELS, F., RÜSCH-GERDES, S., SCHRÖDER, K.H., SILCOX, V.A., TSUKAMURA, M., VAN DEN BREEN, L. & YAKRUS, M.A. (1991). Fourth report of the cooperative, open-ended study of slowly growing mycobacteria by the International Working Group on Mycobacterial Taxonomy. *International Journal of Systematic Bacteriology* **41**, 463-472.

WAYNE, L.G., KRICHEVSK, Y M.I., PORTYRATA, D. & JACKSON, C.K. (1984). Diagnostic probability matrix for identification of slowly growing mycobacteria in clinical laboratories. *Journal of Clinical Microbiology* **20**, 722-729.

WAYNE, L.G. & KUBICA, G.P. (1986). Genus *Mycobacterium* Lehmann and Neumann 1986, 363[al]. In *Bergey's Manual of Systematic Bacteriology*, pp. 1435-1457. Edited by P.H.A. Sneath, N.S. Mair, M.E. Sharpe & J.H. Holt. Baltimore, MD: The Williams & Wilkins Co.