Computer aided identification of Mycobacteria: a prototype of software to interpret the results of cultural and biochemical tests

TORTOLI E.
Laboratorio di Batteriologia e Virologia, USL 10/D, 50139 FIRENZE

RIASSUNTO

IDENTIFICAZIONE COMPUTERIZZATA DEI MICOBATTERI: PROTOTIPO DI PROGRAMMA PER L'INTERPRETAZIONE DEI RISULTATI DELLE PROVE COLTURALI E BIOCHIMICHE.
Sono stati introdotti nel programma di identificazione messo a punto dall'Autore, per valutarne l'affidabilità, i risultati dei tests identificativi relativi a 194 ceppi micobatterici isolati fra il 1975 ed il 1989 (gli stipiti tipici di M. tubercolosis sono stati esclusi dalla valutazione). Dal confronto fra l'identificazione ottenuta con l'ausilio del computer e quella "ragionata" è emersa una concordanza del 72%. Nel riesame critico delle discordanze la preferenza è andata all'identificazione computerizzata nel 12% dei casi, a quella "ragionata" nel 9% dei casi mentre nel rimanente 7% non è stato possibile effettuare una scelta.

PAROLE CHIAVE: MICOBATTERI, TASSONOMIA NUMERICA, IDENTIFICAZIONE

SUMMARY

To check an identification software developed by the Autor, 194 mycobacterial identifications (other than typical M. tubercolosis species) have been re-examined. The "computer-aided" identification and the "reasoned" one showed 72% of agreement, 7% of disagreement without possibility of giving preferences, 12% of greater plausibility of "computer-aided" report and 9% of "reasoned" identification.

KEY WORD: MYCOBACTERIA, NUMERICAL TAXONOMY, IDENTIFICATION

INTRODUCTION

The concept of numerical taxonomy was first introduced by Sneath P.H.A. in 1957. Since then the idea has reached an enormous success, so much so, that nowadays every kit for microbial identification uses it. Whereas the manual methods use coding book, the automated ones resort to a computer to convert the test pattern results into the most likely identifications.

The main exception, if not the sole one, is represented by mycobacteria; as there is no kit on sale for their identification and usually many features are considered (cultural properties, morphology, biochemical actions, tolerance to several matters or drugs). The critical point of this system is the interpretation of the tests results; it is performed by using tables that do not always agree, and in which the properties of the various species are rarely unequivocal (plus or minus); while very frequent are +/− and −/+ often placed side by side with a long series of reference marks referring to footnotes.

The aim of this study is to overcome this obstacle by developing software able to assist in determining final mycobacterial identification for every pattern of test results.

MATERIALS AND METHODS

The programme was prepared using as a basis a "preadsheet" of the type that can be found in the most widespread "integrated softwares"; i.e. a table which squares can be connected using proper formulas.

The software is made up of:

i) a matrix (two-way-table) with 25 species of Mycobacteria that can be encountered in humans, and with 22 of the most frequently tested distinctive properties (1, 5) (niacin, nitrate reduction, catalase, 68° C catalase, >45 mm catalase, pigmentation on dark and light, 5 and 10 days tween hydrolysis, NaCl tolerance, >5 days growth rate, growth at 25°C and 45°C, Mac Conkey, tellurite reduction, arylsulfatase, colony roughness, urease, susceptibility to TCH, Tb1 and NH2OH, nutrient-agar); in every square of this table (table 1) the percentages of positive results are placed for each test and each taxon (these percentages are mainly taken from the data of Wayne et al. (6) and of Kubica (2) integrated with other data found in literature (1, 3, 4, 5);

ii) a one-way-table to input the results of single tests as "positive" (1), "negative" (−1) or "not done" (0);

iii) a table, the same size as the one in (i) item, that has in every square a formula to convert into positive or negative test probabilities, the percentages of positiveness: for "positive" tests the corresponding values in the table (i), e.g. X, are divided by (X/100); for "negative" tests they are changed into (100-X)/100 and for "not done" tests the higher value between X/100 and (100−X)/100 is selected;

iv) a programme to choose the taxa with the largest absolute likelihood (Bayes theorem) of agreement with the inputted pattern of properties; at least two species are always selected, a third species is displayed if its probability percentage is greater than 0.20 of the first one. As it has not been possible to find data concerning all tests for all the considerad taxa, only the test with known positiveness percentage are weighted by the software;

v) a warning system that displays the atypical test results of the two most likely organism identifications and suggests which of the tests not done might be performed to deinitively speciate the examined isolate. If the absolute likelihood of the identification falls below 0.95 of the best possible likelihood for that species a massage for the unacceptability of the diagnosis is displayed.

To verify the reliability of this programme, all the tests concerning 194 identifications of Mycobacteria, carried out between 1975 and 1989 at the Bacteriological and Virological Laboratory of USL 10/D in Florence, have been inputted. All the strains identified as *M. tubercolosis* on the basis of the screening used in above-mentioned Laboratory (niacin +, nitrate reduction +, catalase +, 68°C catalase−) have been excluded from this appraisal.

TABLE 1
Frequency per cent of 22 features for 25 considered taxa

SPECIES	Niac	Nitr	Cat	168°C	>45CD	PigL	PigT	w5g	Tw10	NaClG	>5g
M. africanum	50	50	99	1	1	1	1	1	1	1	99
M. asiaticum	1	1	99	99	99	1	99	99	99	1	99
M. avium/M. intracell.	1	1	99	87	19	2	2	1	1	5	99
M. bovis/M. bovis BCG	1	8	99	1	1	1	1	1	1	1	99
M. chelonae	3	8	99	94	97	1	1	15	20	59	1
M. flavescens	1	99	99	99	99	83	83	99	99	99	50
M. fortuitum	7	99	99	99	98	1	1	56	56	99	1
M. gastri	1	1	99	1	1	1	1	99	99	1	99
M.gordonae	1	1	99	99	96	99	99	79	91	1	99
M. haemophilum	1	1	–	1	1	1	1	1	1	1	99
M. kansasii	1	97	99	99	99	3	94	99	99	1	99
M. malmoense	1	1	99	–	1	1	1	–	99	1	99
M. marinum	1	1	99	99	33	1	99	97	99	17	99
M. nonchromogenicum	1	–	99	99	99	1	1	99	98	1	99
M. phlei	1	99	99	99	99	99	99	99	99	99	1
M. scrofulaceum	1	18	99	99	99	99	99	1	1	1	99
M. simiae	46	1	99	99	99	1	90	1	1	1	99
M. smegmatis	1	99	99	79	76	1	1	99	99	99	1
M. szulgai	1	99	99	99	99	99	99	–	80	1	99
M. terrae	1	97	99	99	99	1	1	94	99	10	99
M. triviale	4	99	99	99	92	1	1	80	99	99	99
M. tuberculosis	99	98	99	2	1	1	1	20	70	1	99
M. ulcerans	1	1	99	99	1	1	1	1	1	1	99
M. vaccae	1	90	99	99	99	–	99	99	99	63	1
M. xenopi	1	1	99	99	1	86	86	1	1	1	99

SPECIES	25°C	45°C	CM Con	Tell	Aryl	Roug	Urea	TCH	Tb1	NH 20	Agar
M. africanum	5	1	1	1	1	99	50	50	1	–	1
M. asiaticum	99	1	–	1	1	–	1	1	–	–	–
M. avium/M. intracell.	–	47	24	76	1	–	1	99	98	99	1
M. bovis/M. bovis BCG	1	1	1	1	1	99	99	1	1	1	1
M. chelonae	75	1	99	51	99	–	99	99	99	99	–
M. flavescens	99	17	1	32	1	1	99	99	99	10	1
M. fortuitum	90	1	99	43	99	–	99	99	99	99	99
M. gastri	99	1	16	1	1	–	99	99	9	99	1
M.gordonae	99	1	1	1	1	1	8	99	99	–	1
M. haemophilum	99	0	1	1	1	99	1	–	–	–	–
M. kansasii	99	1	1	1	1	–	99	99	9	90	1
M. malmoense	86	1	–	–	1	1	71	1	99	–	–
M. marinum	99	1	1	1	6	–	99	99	67	90	1
M. nonchromogenicum	90	–	1	1	–	–	1	1	99	–	1
M. phlei	–	99	1	62	1	99	–	–	99	–	–
M. scrofulaceum	99	1	1	4	1	1	91	99	30	–	1
M. simiae	99	1	–	90	1	1	99	99	99	–	1
M. smegmatis	80	99	1	90	10	–	–	–	99	–	–
M. szulgai	99	1	1	–	–	–	80	99	99	–	1
M. terrae	90	1	3	1	1	–	1	99	99	90	1
M. triviale	–	–	1	1	26	99	1	99	–	–	–
M. tuberculosis	5	1	1	2	1	99	99	96	10	1	1
M. ulcerans	99	0	–	1	1	99	1	1	–	–	–
M. vaccae	–	–	1	70	1	1	–	–	99	–	–
M. xenopi	1	99	1	10	60	1	1	99	99	99	1

RESULTS

The data of this re-examination are expounded on table 2: 72.17% of the identifications coincide. The remaining identifications have been critically inspected, considering especially the tests able to discriminate, between the "computer-aided" and the "reasoned" taxon. the most correct one (table 3).

DISCUSSION

The results of this work can be considered encouraging, particularly if it is considered that: i) In this appraisal only the first choice "computer-aided" identification (the one with the largest absolute likelihood) has been valued. Valuing also the alternative choices, the cases in which the "computer-aided" identification seems to be incorrect fall to 4.64%; ii) the number of the tests performed in the above-mentioned Laboratory to speciate the Mycobacteria have increased with time, so the strains identified some years ago are accompained by few tests; even if the programme can theoretically work also with the imput of only one test, the reliability of the result is proportional with the number of the properties tested; iii) as tests have not been repeated but only old results have been re-examined, is has not been possible to turn the suggestions of the programme (repetition or execution of tests) to identification advantage.

All cases that show the same type of disagreement have been carefully valued looking for systematic mistakes. As it is shown in

REASONED ID. \ COMP.-AIDED ID.	M. ulcerans	M. tuberculosis	M. bovis /M. bovis BCG	M. africanum	M. marinum	M. kansasii	M. simiae	M. asiaticum	M. scrofulaceum	M. szulgai	M. gordonae	M. flavescens	M. xenopi	M. avium-intracellulare	M. gastri	M. malmoense	M. haemophilum	M. nonchromogenicum	M. terrae	M. triviale	M. fortuitum	M. chelonae	M. phlei	M. smegmatis	M. vaccae	???
M. ulcerans																										
M. tuberculosis		3		2																						
M. bovis /M. bovis BCG		1	2	1														1								
M. africanum				1																						
M. marinum																										
M. kansasii						3																				
M. simiae																										
M. asiaticum																										
M. scrofulaceum									4	1	1		3													
M. szulgai																										
M. gordonae								1		1	9		1										4			
M. flavescens											2	2										2				
M. xenopi													58													
M. avium-intracellulare				1									3	19				1								
M. gastri		1																								
M. malmoense																										
M. haemophilum																										
M. nonchromogenicum																										
M. terrae						1													4	1						
M. triviale																								1		
M. fortuitum	1															1		1	1	19	8		1			
M. chelonae														1							12		2			
M. phlei																						1		2		
M. smegmatis																			1							
M. vaccae								1																2		
???			1											1		2	1							1		

TABLE 2
Correlation between "computer-aidsad" and "reasoned" identifications

TABLE 3
Per cent results of correctness valuation of the identifications performed in computer aided-and traditional way

IDENTIFICATIONS	CORRECT		WRONG	NOT VALUABLE	
	coinciding	not coinciding		not done	done
"Computed"	72.17	11.85	8.76	—	7.22
"Reasoned"	72.17	8.76	11.85	2.06	5.16

TABLE 4
Valuation of disagreements found more than once

PAIRS OF SPECIES "REASONED"/"COMPUTER-AIDED"	IDENTIFICATIONS CORRECT "REASONED"	"COMP.-AIDED"	NOT VALUABLE
M. fortuitum/M. chelonae	7	1	—
M. gordonae/M. vaccae	—	2	2
M. terrae/M. triviale	—	2	2
M. scrofulaceum/M. xenopi	—	1	2
M. avium-intracell./M. xenopi	—	3	—
M. tubercolosis/M. africanum	1	1	—
M. flavescens/M. gordonae	—	2	—
M. flavescens/M. phlei	—	2	—
M. chelonae/M. smegmatis	—	1	1
M. phlei/M. vaccae	1	1	—

table 4 the software, in only one case, made more than once the same mistake: 7 patterns of tests, that have probably to be attributed to species *M. chelonae*, have been interpreted as *M. fortuitum*. On 6 of these occasions nevertheless the second identification was exact. To correct this anomaly the insertion in the matrix of a test (probably an antimicrobial susceptibility test) able to distinguish between these two taxa is considered.

The identification programme described here, even if does not intend to be used as an alternative to the rational interpretation, can give a valid starting-point for a consistent appraisal of the pattern of properties of the mycobacterium examined and can provide useful suggestions about tests to repeat or to be done to check an identification.

References

1. **Kleeberg H.H., Koornhof H.J., Palmhert H.:** Laboratory manual of tuberculosis methods. Tb Resp. Inst., Pretoria 1980, 211.
2. **Kubica G.P.:** Differential identification of Mycobacteria. VII key features for identification of clinically significant Mycobacteria. Am. Rev. Respir. Dis., 1973, 107, 9-21.
3. **Mandler F., Peona V.:** Micobatteriosi e micobatteri non tubercolari. Masson Italia Editori, Milano 1987, 126.
4. **Meyer L., David H.L.:** Mycobactériologie en santé publique - Publication du centre national de réference pour la tuberculose et les mycobactéries. Istitut Pasteur, Paris 1979.
5. **Sommers H.M., Good R.C.:** *Mycobacterium.*In: Lennette E.H., Ballows A., Hausler Jr. W.J., Shadomy H.J. (Eds): Manual of clinical microbiology. American Society for Microbiology, 4th ed., Washington 1985, 216-48.
6. **Wayne L.G., Krichevsky M.I., Portyrata D., Jackson C.K.:** Diagnostic probability matrix for identification of slowly growing Mycobacteria in clinical laboratories. J. Clin. Microbiol., 1984, 20, 722-9.