J Clin Pathol 1986;**39**:798–802

Constructing a database for low cost identification of Gram negative rods in clinical laboratories

P CLAYTON, RKA FELTHAM, CJ MITCHELL, PHA SNEATH*

*From the Public Health Laboratory, Leicester Royal Infirmary, Leicester, and the *Department of Microbiology, University of Leicester, Leicester*

SUMMARY A database was constructed for the routine identification of *Enterobacteriaceae* and Gram negative rods growing aerobically on MacConkey agar. The test methods were based primarily on multipoint inoculation technology. The final database was constructed from the laboratory results of 4989 clinical isolates and 66 reference strains and was extensively analysed and evaluated. The proposed scheme is rapid, reliable, and cheap.

Routine methods for identifying Gram negative bacilli growing on MacConkey agar vary between laboratories. Many laboratories will identify isolates from serious clinical infections, such as those from blood cultures and cerebrospinal fluids, but provide only minimal identification of organisms from specimens such as urine. One of the major factors in this differentiation between specimens, apart from the clinical or epidemiological importance of the identification, is cost.

The methods used to identify Gram negative rods growing aerobically on MacConkey agar range from conventional tests[1] to kits such as API20E.[2] A set of conventional tests is difficult to standardise, time consuming—and therefore expensive to prepare—and may take up to five days to produce a definitive result. Once results are recorded they must be subjectively compared with diagnostic tables.[1 3] A few laboratories have introduced a computer assisted form of identification.[4]

Identification kits are standardised, usually take 24 to 48 hours for interpretation of results, and provide either a full computer assisted service with microcomputer or a printout of the most common result patterns, or profiles. They are, however, expensive to purchase and for this reason tend to be used only for highly important isolates.

We report here a simple system for the identification of *Enterobacteriaceae* and other Gram negative rods that grow aerobically on MacConkey agar. The system comprises conventional tests, using substrates incorporated into agar plates and a multipoint inoculation system similar to that used for the antibiotic susceptibility testing breakpoint method.[5] This system provides an identification in 24 hours, makes use of computer assisted technology—either with an interactive terminal or a profile printout—and is simple and cheap to perform.

Accepted for publication 6 February 1986

Material and methods

MEDIA AND QUALITY CONTROL

Unless otherwise stated, all tests were performed using a 21 pin multipoint inoculator on substrate incorporated agar plates. Three of the pin positions were used for internal quality control of all batches. The organisms used were laboratory isolates BP6 *(Escherichia coli)*, BP8 *(Proteus mirabilis)*, and BP3 *(Pseudomonas aeruginosa)*. All tests were incubated at 37°C for 18 to 24 hours. Motility was also tested at room temperature.

Oxidase was tested by smearing the test organism on to filter paper moistened with a 1% (v/v) aqueous solution of tetramethyl-para-phenylenediamine dihydrochloride.

Phenylalanine deamination was tested using Mast (Mast Laboratories Ltd, Bootle, Merseyside, England) phenylalanine agar (IDM 31). A brown-black coloured colony indicated a positive result.

Gelatin hydrolysis was tested using Mast gelatin agar (IDM 24). After incubation the colonies were separated from each other using a scalpel and then the plates were flooded with acidic mercuric chloride. Clear areas around the colonies indicated positive gelatinase production.

The indole test was performed in 5 ml aliquots of peptone water[1] in bijoux bottles. A few drops of Kovacs' indole reagent were added after incubation, and a red

colour in the reagent layer indicated a positive result.

Urease production was tested using Mast urea agar base (Multipoint) (IDM 32). A positive reaction was indicated by a purple halo and colony on an orange-yellow plate.

Hydrogen sulphide production was tested using Mast hydrogen sulphide agar (IDM 25). Positive organisms showed as black colonies.

Acid from glucose, mannitol, rhamnose, sucrose, and inositol were tested using Mast carbohydrate peptone agars (IDM 3, 6, 8, 10, and 4). The indicator was bromo-thymol blue, and a positive fermentation reaction was indicated by a yellow halo round the colony on a blue plate.

Citrate utilisation was tested using Mast citrate agar (IDM 23). This was a modification of Simmons' method, and organisms able to use citrate as their sole carbon source both grew and changed the medium from green to blue.

ONPG production was tested using Mast ONPG agar base (IDM 29). After multipoint inoculation each suspension drop was separated using a sterile scalpel blade to cut thin channels between the inocula.[7] This prevented cross colouration and facilitated the interpretation of a pale yellow positive colour.

Lysine decarboxylase production was tested using Mast LDC agar (IDM 26). This was a modification of Mancani's aerobic technique,[8] and positive results were shown as a purple colour; negatives showed as yellow.

Malonate utilisation was tested using Mast malonate agar (IDM 27). A positive result was indicated by a blue colour from a green plate.

Motility was tested both at 37°C and room temperature using Mast motility test agar (IDM 28), which contained the dye triphenyltetrazolium chloride (TTC). The agar was dispensed as 2 ml quantities in each of the 25 chambers in a compartmentalised Petri dish. The test organisms were stabbed into each chamber with a straight wire, and after overnight incubation positive isolates gave a diffuse pink cloud throughout the chamber for either one of the incubation temperatures. Negative isolates gave a red inoculation track for both temperatures.

An extra test (Oxoid CN agar CM559 plus SR102 supplement) was used as a confirmatory test for *Pseudomonas* spp. This was not included in the principal database but used only as confirmation on occasions.

COMPARATIVE COSTING

A comparative costing exercise was carried out between the methods detailed here and a commercial kit (API20E, API Laboratories, Basingstoke, Hampshire).

A set of 18 organisms, chosen at random from routine clinical material, was tested with the multipoint system and through API20E kits. Table 1 shows the costs of materials, labour for media preparation, labour for setting up tests and for reading and interpretation. The costs of the media were for price lists dated September 1985. All timings were performed for each test method by two medical laboratory scientific officers. The results shown in Table 1 are the averages of these two sets of figures. The mid point of the 1984 Whitley council salary scale was used and gave a value of £0·065 per medical laboratory scientific officer's minute.

Table 1 *Comparative costing: multipoint system v API20E*

Item/18 organisms	*Multipoint system (£)*	*API20E (£)*
Medium cost	1·59	23·04
Preparation	0·845	0·00
Set up labour	1·94	3·72
Reading labour	1·745	2·92
Total for 18 organisms	6·12	29·68
Total for 1 organism	0·34*	1·65

*This value includes three quality control strains per set of 18 organisms.

DATABASE CONSTRUCTION

This was carried out in two stages, a preliminary database being used to provide a skeleton profile list from which a full list was finally prepared.

The initial database was constructed using data for 56 tests from Bascomb, Lapage, Willcox, and Curtis.[9] The computer analysis programs CHARSEP[10] and DIACHAR[11] were used to define the most useful 16 of these tests; these were extracted from the full database into a subset. The quality of this smaller database was tested using the programs OVERMAT[12] and MOSTTYP,[13] as used by Feltham and Sneath[14] for Gram positive aerobic cocci. Some of the bacterial groups were amalgamated when they could not be adequately differentiated with the chosen tests. This preliminary database was used to generate a set of octal profiles using the computer programs PROFGEN, PROFSORT, and PROFPRINT.

The program PROFGEN generated a list of the most likely six digit profiles with corresponding taxon name for each taxon within the matrix. The program

Table 2 *Sample of full profile list*

Profile	*Taxon name*	*Probability*
017404	*Citrobacter freundii*	0·961581
	Salmonella typhi	0·015363
017410	*Klebsiella rhinoscleromatis*	0·990932
017440	*Citrobacter freundii*	0·930282
	Klebsiella ozaenae	0·032819
017444	*Citrobacter freundii*	0·995975

PROFSORT used the data produced by PROFGEN to print a numerical list of profiles with the taxon name beside each profile. Specified taxon names also had asterisks printed alongside their names to denote that serology would be a useful extra diagnostic test. If any duplicate profiles were produced the output from PROFPRINT was used as reference.

The program PROFPRINT, using the output from PROFGEN, gave a computer assisted identification against the matrix for each listed profile. The output from this program printed the profile, the single highest taxon name, and probability of identification by the Willcox probability coefficient,[4] provided the highest probability was at least 0·97. If the highest value was less than 0·6 then this was regarded as unidentified; for all other values between 0·61 and 0·96 the profile, top two taxon names and their corresponding probability scores were printed (Table 2). This database was evaluated using the following procedures. A total of 5055 strains were tested against this database using the substrate incorporation methods described above. Sixty six reference strains from the Leicester clinical culture collection were used in conjunction with 4989 fresh clinical isolates. If any isolate did not identfy according to the laboratory protocol they were further identified using API20E. The multipoint methods were repeated, and if consistent, the "new" profile was incorporated into the list.

The profiles from the 5055 strains were analysed by the computer program PROFMAT to generate a new database, wholly based on the actual results obtained from the multipoint methods in our laboratory. The data for the program comprised a list of bacterial taxon names, their corresponding profiles, together with the number of times each profile for that taxon had occurred. The program carried out the analysis in three stages: sort all bacterial names together with corresponding profile data; calculate sequential

Table 3 *Improved Enterobacteriaceae plus related organisms database (B)*

	Oxidase	*PPA*	*Gelatin*	*Indole*	*Urease*	*H2S*	*Glucose*	*Mannitol*	*Rhamnose*	*Sucrose*	*Inositol*	*Citrate*	*ONPG*	*LDC*	*Malonate*	
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
Acinetobacter calcoaceticus	1	1	10	1	8	1	80	6	1	1	1	85	1	97	40	
Acinetobacter sp	1	1	7	1	1	1	10	2	1	1	1	35	1	69	25	
Aeromonas sp	99	8	99	95	1	1	99	99	10	90	3	70	99	25	1	99
Alcaligenes sp	60	1	35	1	1	1	1	1	1	1	1	99	1	25	25	80
Chromobacterium violaceum	99	1	99	1	1	1	99	5	1	20	1	50	1	1	1	99
Citrobacter diversus	1	1	1	98	50	1	99	99	99	45	10	95	98	7	90	99
Citrobacter freundii	1	1	2	2	35	70	99	99	97	59	10	90	92	1	10	90
Edwardsiella tarda	1	1	1	99	1	99	99	1	1	1	1	1	1	99	1	99
Enterobacter aerogenes	1	1	3	1	7	1	99	99	99	99	99	99	99	99	85	99
Enterobacter agglomerans	1	1	32	32	1	1	99	99	40	99	64	78	95	91	32	99
Enterobacter cloacae	1	1	7	1	20	1	99	98	88	97	5	98	90	90	85	93
Escherichia coli	1	1	1	96	1	1	99	97	85	30	1	1	79	90	1	50
Flavobacterium meningosepticum	99	1	99	1	1	1	1	1	1	1	1	1	99	99	1	
Hafnia alvei	1	1	5	1	1	1	99	91	90	10	2	4	90	99	26	50
Klebsiella oxytoca	1	1	1	99	70	1	99	99	99	99	99	95	93	93	69	
Klebsiella ozaenae	1	1	1	1	12	1	99	99	90	97	85	60	99	85	1	
Klebsiella pneumoniae	1	1	1	1	65	1	99	99	96	99	93	90	99	93	87	
Klebsiella rhinoscleromatis	1	1	1	1	1	1	99	99	77	95	55	1	22	7	93	
Morganella morganii	1	52	1	99	99	1	99	1	1	1	1	1	1	7	1	77
Moraxella sp	99	1	25	1	1	1	1	1	1	1	1	25	1	99	1	
Pasteurella sp	1	1	1	99	1	1	99	1	1	99	1	1	1	1	1	
Plesiomonas shigelloides	90	1	1	99	1	1	99	1	1	1	42	1	99	99	1	99
Proteus mirabilis	1	98	88	1	99	93	99	1	1	1	1	23	1	8	1	99
Proteus vulgaris	1	99	95	90	99	13	99	1	1	93	13	11	1	16	1	99
Providencia alcalifaciens	1	96	3	88	1	1	99	1	1	1	8	96	1	28	1	2
Providencia rettgeri	1	95	1	99	95	1	99	93	46	6	86	95	6	33	1	95
Providencia stuartii	1	96	1	99	1	1	99	1	1	5	98	99	1	69	1	76
Pseudomonas aeruginosa	99	1	98	1	1	1	19	1	1	7	1	99	3	99	93	98
Pseudomonas cepacia	1	1	45	1	1	1	99	1	1	1	1	63	1	99	63	99
Pseudomonas fluorescens	99	1	93	1	1	1	10	1	1	1	1	93	1	93	7	10
Pseudomonas maltophilia	29	1	49	1	1	1	1	1	1	1	1	23	2	99	49	71
Pseudomonas stutzeri	99	1	1	1	1	1	1	1	1	1	1	87	1	99	87	3
Salmonella sp	1	1	1	2	1	93	99	99	93	1	32	92	1	99	1	99
Salmonella typhi	1	1	1	1	1	50	99	99	1	1	1	1	1	50	1	99
Serratia sp	1	1	99	1	4	1	99	99	1	99	78	93	87	97	8	88
Shigella sp	1	1	1	7	1	1	99	66	24	6	1	1	30	18	1	
Vibrio parahaemolyticus	99	1	99	99	10	1	99	99	1	99	1	93	1	99	1	99
Vibrio sp	99	1	99	99	1	1	99	99	1	99	99	93	99	99	1	99
Yersinia enterocolitica	1	1	1	72	79	1	99	99	1	95	50	1	75	7	1	99
Yersinia sp	1	1	1	99	71	1	99	99	99	99	1	1	99	1	1	99
Yersinia pseudotuberculosis	1	1	1	1	50	1	99	99	90	1	1	1	50	1	1	99

PPA = Phenylalanine; H2S = hydrogen sulphide; LDC = lysine decarboxylase production.

summated results for each taxon; construct and print probabilities of positive results for each taxon. This database was analysed using the OVERMAT and MOSTTYP computer programs, and some species were again amalgamated to ensure a good identity (identification level of 0·95) to at least genus level.

The test order was changed after studying the conclusions obtained from the database analysis programs CHARSEP and DIACHAR. Ease of interpretation was also considered. A list of likely profiles and the bacterial taxon names was produced using PROFGEN and PROFSORT for routine use. A full list of probabilities for identification, together with likely taxa, was produced using PROFPRINT.

DATABASE EVALUATION

Two methods were used to evaluate the preliminary database A. These were repeated for the improved database B. Evaluation I was based on 100 organisms selected from a set used for evaluating the Autobac IDX.[15] Each organism had been identified using the Autobac and also API20E. Organisms were identified against the preliminary profile index for database A, either to "species" or "genus" level. For database B, the organisms were compared with the full probability profile index generated from PROFPRINT and then against a fuller list of possible profiles generated by PROFGEN or PROFSORT, with 98% or more being a positive result and 2% or less, negative.

Evaluation II was much more stringent and entailed four separate sets of 50 organisms chosen from a range of previously identified collection organisms. Each set contained single, duplicate, or triplicate cultures selected and coded at random. The same two methods for identification (Evaluation I and databases A and B) were used.

Results

Table 2 gives a sample of the full profile list. Table 3 shows the improved database, and Table 4 shows the results from the evaluations compared with those of the improved database. Using 59 random paired strains from Evaluation II, there were 22 test differences. For the 16 tests this gave a test error (p = 1·17915%). The 22 test differences comprised five for the motility test and four for the malonate test.

Discussion

Pease[16] described a rapid and economic scheme using a multipoint identification system. Only 12 tests were used to distinguish between a selection of 10 genera from the *Enterobacteriaceae*. The scheme proved reliable, flexible, and rapid in use. Problems were found with diffusion of colour through some of the media, which were solved by anaerobic incubation and an increase in agar concentration. We experienced difficulty with the ONPG and gelatin reactions until each inoculum was segregated using dividing channels. The lysine decarboxylase gave unusual, though consistent results. Mancani[8] found that inoculation with the prongs just below the agar surface gave stronger reactions. Our motility test proved extremely difficult to interpret using the standard protocol. This test was placed at the end of the list so that any reading errors would limit the likelihood of a misidentification. We found that an extra test for confirmation of *Pseudomonas* strains was useful but not essential.

Table 4 *Percentage identification rates for two evaluations*

No of strains	*Identification 98%/2% index*	*Identification full index*
I 100	96·9	94·9
II 200	98·5	91·8

Although the profile register contains many profiles where the second identity has a low probability, routine isolates seldom gave such double identities.

The importance of reproducibility has recently been reiterated, and a interlaboratory probability of errors of 6·1% for the API20E system was found.[17] The test organisms were "difficult" strains, which had been sent to the National Collection of Type Cultures computer laboratory for further identification. This value was high, however, despite extreme standardisation of production, inoculation, and interpretation using a colour chart. Other authors have shown an intralaboratory reproducibility value of 2%.[18]

Conventional tests are prone to test errors with an average 6–10%[19] between laboratories and values of 2–4% (on average) within a laboratory.[4] Our scheme has shown a high degree of intralaboratory reproducibility, but until another laboratory evaluates the method we are unable to comment on interlaboratory figures. The principal reason for test error has been subjective interpretation, which should be solved with training.

The cost and time taken to perform the tests was very attractive when compared with that of a leading commercial kit. The five fold decrease in direct costs and the almost halving of time required for setting up and reading were impressive. The inclusion in our system of three quality control strains per set of 18 organisms could be changed, so that all 21 positions were used for "unknown" strains. The use of these quality control strains permits inbuilt control of media batches and should be used if possible. Higher numbers of organisms may be inoculated per plate, for example, using 36 pins. Either of these options would make the system even more economical,

although the 36 pin option could present cross contamination problems.

The multipoint system for identification may be used in association with a multipoint antibiotic breakpoint system and becomes yet more cost effective as an integrated laboratory culture and sensitivity system.

A better degree of differentiation would be achieved for more taxa with more tests. The addition of two more tests, after due consideration, together with improvements in the present choice would necessitate reconstructing the database. This should not prove too difficult if records are kept for all the routine identifications and the procedures described in this paper are followed. We are currently investigating a similar scheme for Gram and catalase positive cocci.

We thank Michaela Foster, AM Folwell, C Geary, GT Parish, and SM Hardy for their help in producing the data for this work.

References

[1] Cowan ST, Steel KG. *Manual for the identification of medical bacteria*, Second ed. Cambridge: University Press, 1974.
[2] Holmes B, Willcox WR, Lapage SP. Identification of Enterobacteriaceae by the API20E system. *J Clin Pathol* 1978;**31**:22–30.
[3] Lennette EH, Balows A, Hausler WJ, Shadomy HJ, eds. *Manual of clinical microbiology*, 4th ed. Washington, D.C: American Society for Microbiology, 1985.
[4] Lapage SP, Bascomb S, Willcox WR, Curtis MA. Identification of bacteria by computer: general aspects and perspectives. *J Gen Microbiol* 1973;**77**:273–90.
[5] Waterworth P. Sensitivity testing by the break point method. *Antimicrob Agents Chemother* 1981;**7**:117–26.
[6] Clayton P. An evaluation of breakpoint susceptibility test methods. *Mast Matters* 1984;**5**:10–17.
[7] Buckwold FJ, Ronald AR, Harding GKM. Biotyping of Escherichia coli by a simple multiple-inoculation agar plate technique. *J Clin Microbiol* 1979;**10**:275–8.
[8] Mancani JE. Aerobically incubated medium for decarboxylase testing of Enterobacteriaceae by replica-plating methods. *J Clin Microbiol* 1979;**10**:940–2.
[9] Bascomb S, Lapage SP, Willcox WR, Curtis MA. Identification of bacteria by computer: identification of reference strains. *J Gen Microbiol* 1977;**77**:291–315.
[10] Sneath PHA. BASIC program for character separation indices from an identification matrix of percent positive characters. *Computers and Geosciences* 1979;**5**:349–57.
[11] Sneath PHA. BASIC program for the most diagnostic properties of groups from an identification matrix of percent positive characters. *Computers and Geosciences* 1980;**6**:21–6.
[12] Sneath PHA. BASIC program for determining overlap between groups in an identification matrix of percent positive characters. *Computers and Geosciences* 1980;**6**:267–78.
[13] Sneath PHA. BASIC program for determining the best identification scores possible from the most typical examples when compared with an identification matrix of percent positive characters. *Computers and Geosciences* 1980;**6**:27–34.
[14] Feltham RKA, Sneath PHA. Construction of matrices for computer-assisted identification of aerobic Gram-positive cocci. *J Gen Microbiol* 1982;**128**:713–20.
[15] Stevens M, Feltham RKA, Schneider F, Grasmick C, Schaak F, Roos P. A collaborative evaluation of a rapid, automated bacterial identification system: the Autobac IDX. *Eur J Clin Microbiol* 1984;**3**:419–23.
[16] Pease A. Biochemical identification of the Enterobacteriaceae using a multipoint inoculation system. *Med Lab Sci* 1983;**40**:349–53.
[17] Holmes B, Dawson CA. Misuse and interlaboratory test reproducibility of API 20E system. *J Clin Pathol* 1985;**38**:937–41.
[18] Butler DA, Lobregat CM, Gavan TL. Reproducibility of the Analytab (API 20E system). *J Clin Microbiol* 1975;**2**:322–6.
[19] Sneath PHA. Test reproducibility in relation to identification. *Int J System Bacteriol* 1974;**24**:508–23.

Requests for reprints to: Mr P Clayton, Acting Head MLSO, Public Health Laboratory, Leicester Royal Infirmary, Infirmary Square, Leicester LE1 5WW, England