Epidemiol. Infect. (1990), **105**, 521–531
Printed in Great Britain

Evaluation of numerical typing systems for *Escherichia coli* using the API 50 CH and the PhP-EC systems as models*

I. KÜHN[1,2], A. BRAUNER[3] AND R. MÖLLBY[1]

[1] *Department of Bacteriology, Karolinska Institute, S-104 01 Stockholm, Sweden*
[2] *Department of Bacteriology, National Bacteriological Laboratory, S-105 21, Stockholm, Sweden*
[3] *Department of Clinical Bacteriology, Stockholm City Council Central Microbiological Laboratory and Karolinska Hospital, S-107 26 Stockholm, Sweden*

(*Accepted 28 July 1990*)

SUMMARY

Reproducible and discriminating typing methods are required for epidemiological investigations. Numerical typing systems analyse patterns obtained in various ways by calculating similarity coefficients between isolates. In the present study, various measures of the efficiency of a numerical typing system are quantified. These include reproducibility, accuracy, and discrimination power.

Three different numerical typing methods for *Escherichia coli* were compared using these measures: (a) Biotyping with API 50 CH system, (b) Biochemical fingerprinting with the API 50 CH system and (c) Biochemical fingerprinting with the PhP-EC system. Biotyping qualitatively measures the results of a set of biochemical reactions as + or −. Biochemical fingerprinting also uses biochemical reactions, but the tests are scored quantitatively by measuring the kinetics and intensity of each reaction.

It was found that biotyping yielded poor reproducibility. When biochemical fingerprinting analysis was used with the API 50 CHE system the reproducibility and the discrimination was good. The PhP-EC system for biochemical fingerprinting showed equal reproducibility but was superior to the API 50 CH system with regard to discrimination power.

INTRODUCTION

The identification of *Escherichia coli* rarely causes any diagnostic problems in the clinical routine laboratory. Typing of *E. coli* below the species level may be required in various epidemiological and ecological studies. Serotyping has traditionally been used as the reference method for typing *E. coli* [1, 2]. A complete serotyping requires, however, availability to many reference antiseras, and is usually only performed in certain reference laboratories. Biotyping, on the other hand, is a simple method for typing *E. coli* [3] and may be performed in any

* This paper was, in part, presented at the 2nd International Meeting on Bacterial Epidemiological Markers, Rhodos, Greece, 8–11 April, 1990

microbiological routine laboratory. In contrast to many other modern typing methods based on electrophoreograms, like characterization of whole cell proteins [4], outer membrane proteins, or plasmid content [5, 6], the interpretation of data obtained from large investigations offers no difficulty.

Biotyping is usually performed by investigation of the ability of an isolate to metabolize a number of chemically defined substrates. Identification of the strains is done by assigning a positive or negative value to each reaction (the biochemical markers of the strain). Biotyping systems are often set up specifically in the laboratory, and may involve various compounds [3, 7–9]. Some commercially available systems also exist, like the API 20E, API 50 CH and Micro ID systems. These kits are primarily designed for identification of strains belonging to the family Enterobacteriaceae, but they have also been used for biotyping of *E. coli* [10–12].

Biochemical fingerprinting may be applied on the same reactions as biotyping. When using biochemical fingerprinting, a quantitative measure of the metabolism of various substrates is obtained by measuring the speed and intensity of each reaction [13]. The 'biochemical fingerprint' of an isolate thus consist of a set of quantitative numbers.

The aims of the present study were to compare biotyping and biochemical fingerprinting, using a commercially available kit for biochemical typing of bacteria (the API 50 CH system), and also to compare biochemical fingerprinting with two different typing systems, the PhenePlate *E. coli* (PhP-EC) system and the API 50 CH.

MATERIALS AND METHODS

The PhP–EC system

The PhP-EC system (BioSys Inova, Stockholm) consists of 24 biochemical reagents which have been selected to give a high discrimination between independent *E. coli* strains [13] (Table 1). The dehydrated reagents are kept in flat-bottomed microtiter plates, with four sets of reagents for testing four isolates in each plate. Bacterial suspensions were made in a substrate containing 0·2% peptone and 0·1% bromothymole blue and were dispensed into each one of the 24 wells by the aid of a multi-channel pipette. The absorbance value for each reactions was read after 4, 7, 24 and 48 h incubation by an optical reader (Titertek Multiscan) at 620 nm [14]. The reader was connected to a microcomputer, where data were collected and stored and all calculations were automatically performed by the PhP software (BioSys Inova, Stockholm). After the final reading, the mean absorbance value from each test was calculated, yielding a quantitative result which depended on the speed and intensity of each reaction. An isolate was thus determined by a set of 24 integers, each one ranging from 0 (acidic reaction) to 25 (alkaline reaction).

The API 50 CH system

The API 50 CH system (La Balme, Les Grottes, France) consists of 49 carbohydrates and one negative control test (Table 1). The carbohydrates have been selected for identification and typing of various species. The dehydrated reagents and a pH indicator (phenol red) are kept in cups in specially designed

Table 1. *Set of reagents in the PhP-EC and API 50 CH systems. '+' indicates that the reagent is included in the system*

Reagent	PhP	S.D.	API:1	S.D.	API:3	Reagent	PhP	S.D.	API:1	S.D.	API:3
Glycerol	+	1·0	+	1·2	+	Maltose	+	1·6	+	1·5	+
Erythritol	−	·	+	0·0	−	Lactose	+	3·3	+	4·4	+
D-arabinose	+	2·2	+	2·3	+	Melbiose	+	2·8	+	3·9	+
L-arabinose	−	·	+	0·4	−	Sucrose	+	5·6	+	6·5	+
Ribose	−	·	+	1·2	+	Trehalose	−	·	+	1·7	+
D-Xylose	+	1·9	+	2·3	+	Inuline	−	·	+	0·4	−
L-Xylose	−	·	+	0·0	−	Melezitose	−	·	+	0·0	−
Adonitol	+	2·4	+	2·3	+	Raffinose	+	3·7	+	4·3	+
Beta-Methylxyloside	−	·	+	0·1	−	Amidon	−	·	+	1·2	+
Galactose	−	·	+	1·6	+	Glycogen	−	·	+	0·1	−
Glucose	−	·	+	0·1	−	Xylitol	−	·	+	0·3	−
Fructose	−	·	+	0·0	−	Gentobiose	−	·	+	1·5	+
Mannose	−	·	+	0·4	−	Turanose	−	·	+	0·0	−
Sorbose	+	4·0	+	4·3	+	Lyxose	−	·	+	0·4	−
Rhamnose	+	2·5	+	3·0	+	Tagatose	+	3·4	+	3·7	+
Dulcitol	+	2·5	+	3·5	+	D-Fucose	−	·	+	0·0	−
Inositol	−	·	+	0·5	−	L-Fucose	+	2·5	+	3·8	+
Mannitol	−	·	+	0·5	−	D-Arabitol	+	2·8	+	3·3	+
Sorbitol	+	2·5	+	2·7	+	L-Arabitol	−	·	+	0·1	−
Alpha-Methylmannoside	−	·	+	0·0	−	Gluconate	−	·	+	2·0	+
Alpha-Methylglucoside	−	·	+	0·1	−	2-Ketogluconate	−	·	+	0·0	−
N-acetylglucosamine	−	·	+	0·2	−	5-Ketogluconate	+	3·4	+	4·9	+
Amygdaline	−	·	+	0·0	−	Deoxyribose	+	4·0	−	·	·
Arbutine	+	1·4	+	2·2	+	Melbionate	+	3·9	−	·	·
Esculine	−	·	+	3·2	+	Lactulose	+	2·2	−	·	·
Salicine	−	·	+	2·7	+	Beta-methylglucoside	+	1·8	−	·	·
Cellobiose*	+	1·0	+	0·0	−	Ornithine	+	5·9	−	·	·

* Cellobiose is used as a negative control in the PhP-EC-system.

plastic trays. A bacterial suspension, in the API 50 CHE medium (La Balme, Les Grottes, France), was dispensed with a Pasteur pipette into each of the 50 wells in a plastic tray. The reactions were read visually after 3, 6, 24 and 48 h incubation at 37 °C. The indicator change of each reaction was scored from 0 (no reaction) to 4 (full positive reaction). The API 50 CH results were manually fed into the microcomputer, and were evaluated by the aid of the PhP software in four different ways:

Biochemical fingerprinting, 50 tests. The scores from all four readings were added together for each reaction. The final result from each test was thus ranging between 0 and 16 (API:1).

Biotyping, 50 tests. The results which at the reading after 24 h incubation had been assigned a value of 2–4 were coded as 1 (positive), and those which had a value of 0–1 were coded as 0 (negative) (API:2).

Biochemical fingerprinting, 26 most discriminating tests. The standard deviation of each test coded according to method API:1 was calculated for 96 distinct *E. coli* isolates, and those 26 tests which showed a standard deviation of one or more were regarded as discriminating for *E. coli* and used for further calculations (Table 1) (API:3).

Biotyping, 22 most discriminating tests. Only those 22 tests which did not yield identical results for 96 distinct *E. coli* isolates according to method API:2 (biotyping) were used (API:4).

Isolates investigated

Altogether 96 epidemiologically unrelated *E. coli* strains were tested. Fifty strains were isolated in blood-cultures from bacteremic patients and 46 strains were faecal *E. coli* isolates from healthy outpatients [15]. Twelve of these strains were assayed twice, on different occasions, for estimation of the inter-assay reproducibility.

Calculations

All calculations were performed by the PhP software (BioSys Inova, Stockholm), on an IBM compatible computer (Victor V286C).

Similarity matrix

For each method, the test results for all isolates were compared pairwise, and the similarity between each pair of isolates were expressed as the correlation coefficient (r). This resulted in a similarity matrix containing $96 \times (96-1)/2$ correlation coefficients for each method.

The correlation matrix was clustered by the UPGMA method [16] yielding a dendrogram. The cophenetical correlation between different methods was calculated as the correlation between the similarity matrixes obtained from each method. This cophenetical correlation is a measure of the similarities between the results obtained from different numerical typing methods [16].

Reproducibility and identity level

Interassay reproducibility was calculated as the mean correlation coefficient obtained from independent duplicate assays of 12 strains. The identity level for each one of the typing methods was defined as the reproducibility minus two standard deviations (SD) (95% confidence level).

Discrimination power of the typing systems

Different measures of the discrimination power of the typing systems were used:

Resolution capacity. The resolution capacity was expressed as the resolution index. This was calculated as $1\text{-}r_{mean}$, where r_{mean} indicates the mean of all correlation coefficients obtained when all tested isolates were compared to each other. The resolution index determines the suitability of a certain set of tests to discriminate among the isolates [14] and is a valuable help when selecting an optimal set of tests.

Accuracy. The accuracy is a measure of the proportion of comparisons between isolates which fall within the 'grey area', i.e. between isolates which are almost identical. A high accuracy means that this proportion is low, and that isolates which are assigned to different phenotypes by the actual typing system really are different. The accuracy index of each typing system was calculated from the similarity matrix according to the formula $1\text{-}r1/r_{tot}$, where $r1$ is the number of similarity coefficients ranging from the identity level to a value defined by the identity level minus $1 \times$ SD (the grey area), and r_{tot} is the total number of correlation coefficients.

Variety. The variety of each typing system was expressed as the number of phenotypes which could be detected at the identity level specified above.

Diversity. The diversity of each typing system was calculated using Simpson's index of diversity [17] according to the formula

$$D = 1 - \Sigma(Ni \times (Ni-1))/N \times (N-1),$$

where Ni is the number of isolates in the i:th phenotype, and N is the total number of tested isolates. This index measures the probability that two randomly selected strains will be assigned to different phenotypes, using the actual typing method. It gives a valuable measure not only on the number of phenotypes, but also on the distribution of isolates among the phenotypes.

RESULTS

Interassay reproducibility and identity level

Table 2 shows the reproducibility of the 5 typing methods, given as mean correlation between duplicate independent assays of 12 *E. coli* isolates. The identity level for each typing method was then set at a value of mean correlation minus $2 \times$ SD, thus ranging between 0·735 and 0·967 for the typing methods.

Table 2. *Interassay reproducibility calculated from duplicate assays of twelve strains*

	Typing system*				
	PhP-EC	API:1	API:2	API:3	API:4
Mean correlation	0·983	0·986	0·977	0·984	0·931
Standard deviation (S.D.)	0·008	0·012	0·029	0·014	0·098
Mean − 2 × S.D.	0·967	0·962	0·919	0·956	0·735

* PhP-EC, 24 biochemical tests. Sum of 4 readings, determined by a set of quantitative numbers ranging between 0 and 25. API:1, 50 biochemical tests. Sum of 4 readings, determined by a set of numbers ranging between 0 and 16. API:2, 50 biochemical tests, read once after 24 h, determined as positive or negative (biotyping). API:3, the 26 for *E. coli* most discriminating biochemical tests, analysed as API:1. API:4, 22 biochemical tests, analysed as API:2. For further description, see Materials and Methods.

Table 3. *Discrimination power and accuracy assayed on 96 epidemiologically unrelated* E. coli *strains*

	Typing system*				
	PhP-EC	API:1	API:2	API:3	API:4
Discrimination					
Resolution index	0·350	0·122	0·195	0·239	0·350
Variety (number of types)	67	33	30	48	21
Simpson's diversity index	0·989	0·927	0·946	0·966	0·912
Accuracy index	0·994	0·942	0·951	0·983	0·760

PhP-EC, 24 biochemical tests. Sum of 4 readings, determined by a set of quantitative numbers ranging between 0 and 25. API:1, 50 biochemical tests. Sum of 4 readings, determined by a set of numbers ranging between 0 and 16. API:2, 50 biochemical tests, read once after 24 h, determined as positive or negative (biotyping). API:3, the 26 for *E. coli* most discriminating biochemical tests, analysed as API:1. API:4–22 biochemical tests, analysed as API:2. For further description, see Materials and Methods.

The lowest reproducibility was obtained when API 50 CH was used as a biotyping system, i.e. the tests were read only once and the results were coded as + or − (Method API:2 and API:4). One reason for this is that when using this coding method, three tests, namely dulcitol, salicin and melbiose, were subject to late fermentations (between 24 and 48 h), and gave 100% variation (between positive and negative) in 2 out of the 12 isolates (data not shown). When using biochemical fingerprinting analysis, the results from these tests varied only with 35% (between 3 and 8 units out of 17 possible), and the differences in the final result did not affect the reproducibility to such an extent.

Discrimination power and accuracy of the typing systems

Ninety-six epidemiological unrelated *E. coli* strains were analysed. Table 3 shows that the PhP-EC system showed the highest discrimination, with regard to variety, resolution and diversity, and also the best accuracy. Biochemical fingerprinting with the API 50 CH system using 50 tests (API:1) had a low resolution index and diversity index, whereas if only the 26 most discriminating

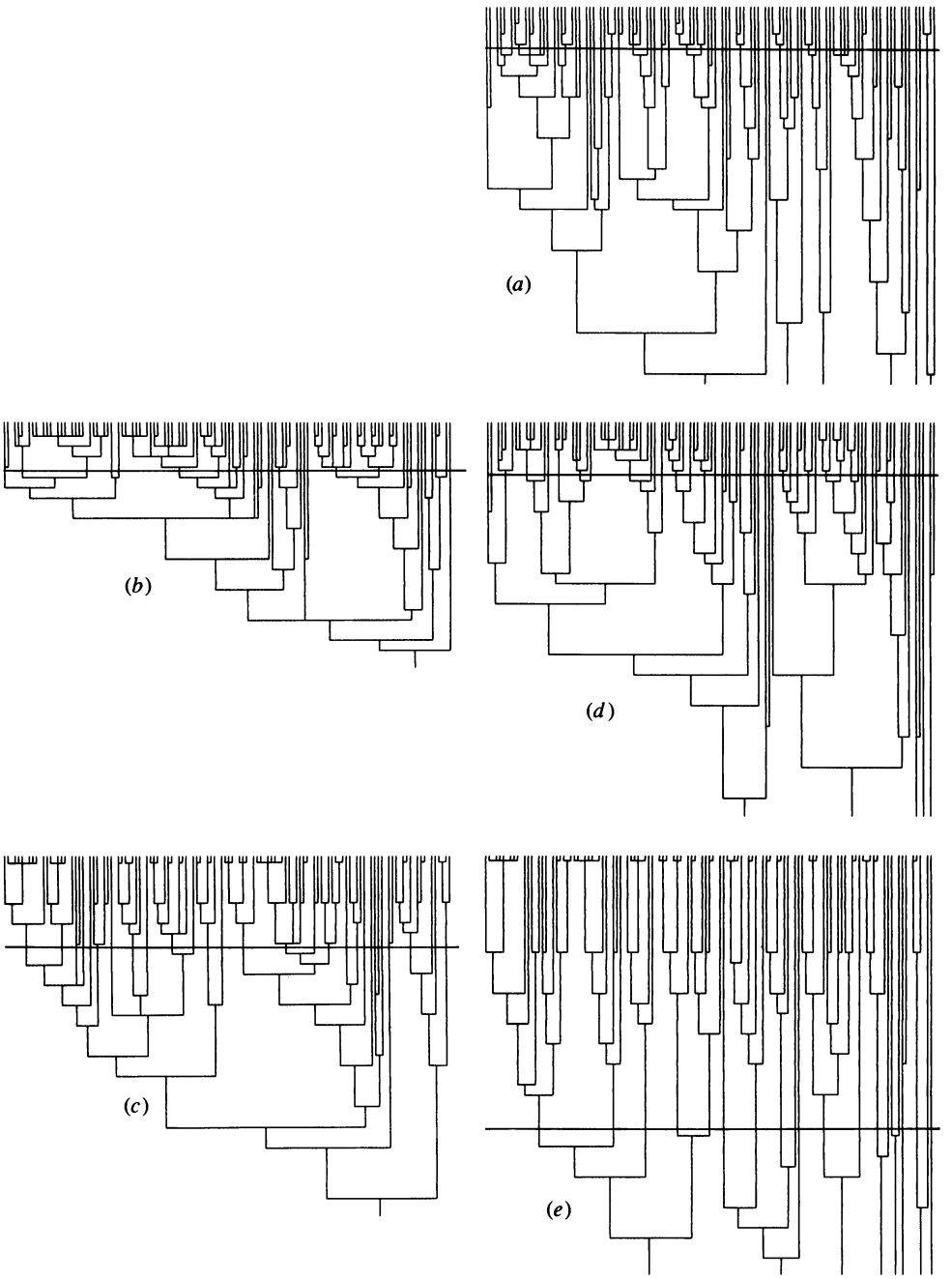

Fig. 1. Dendrograms obtained from UPGMA-clustering of 96 epidemiologically unrelated *E. coli* isolates. Five different typing methods were used. (*a*) PhP-EC (24 biochemical tests, sum of four readings); (*b*) API:1 (50 biochemical tests. Sum of four readings); (*c*) API:2 (50 biochemical tests, determined as positive or negative); (*d*) API:3 (26 biochemical tests. Sum of four readings); (*e*) API:4 (22 biochemical tests, determined as positive or negative). The identity level for each method is represented by a horizontal line in the dendrogram.

tests were included (method API:3) all measures of discrimination capacity increased. Biotyping (methods API:2 and API:4) yielded poor discrimination, due to the low reproducibility and the correspondingly low identity level.

Clustering of isolates

The similarity matrixes obtained from analysis of 96 independent isolates according to the PhP-EC system and the API 50 CH system, methods 1–4, were clustered by the UPGMA method [16] (Figs. 1*a–e*). The dendrograms thus obtained clearly visualizes the effect of the identity levels and the resolution on the discrimination power of the typing methods. Using the PhP-EC system, 16 different phenotypes containing more than one isolate (C-phenotypes) and 51 single phenotypes were found at an identity level of 0·967 (Fig. 1*a*). Using biochemical fingerprinting with the API 50 CH system and the 26 most discriminating tests, (method API:3), 15 C-phenotypes and 33 single isolates were found (Fig. 1*d*). Of the 45 isolates which fell into C-phenotypes using the PhP-EC system, 35 fell into identical groups using method API:3. The cophenetical correlation coefficient between the similarity matrix obtained from the PhP-EC system and API:3, respectively, was 0·785, indicating that both methods yield very similar correlation matrixes.

DISCUSSION

When using traditional typing methods, like phage-typing, sero-typing, colicin-typing etc., the bacteria are only assigned to a specific type, but no information is yielded about the relation between different types. By using typing systems based on several properties and numerical identification methods, it is possible to get an estimate on the overall similarity between different isolates. Modern typing methods based on electrophoreograms of cell components or chromatograms of metabolites, as well as most biochemical typing systems, give results which may be analysed by the aid of numerical methods such as those proposed here.

In the present study, we have used five variables to evaluate biochemical typing systems, namely reproducibility, accuracy, resolution, variety, and diversity. In order to measure these variables, it is necessary to use a collection of epidemiologically unrelated strains, and to make some duplicate assays.

The reproducibility determines the identity level and is therefore the basis for the other variables. Ideally, the reproducibility should be one, i.e. duplicate assays should give identical results, but unfortunately this is not the case. By applying numerical variables it is possible for each assay to calculate the reproducibility, i.e. the mean similarity between duplicate assays, and to give a statistical measure of the identity level. In the present study we have used the mean reproducibility minus $2 \times$ SD. Since the discrimination power has been calculated from results obtained from assays performed at different occasions, the reproducibility was also calculated from inter-assay analysis.

The resolution index, calculated as one minus the mean similarity coefficient among all isolates in a population (r_{mean}), is a valuable tool to design an optimal test set for a particular group of bacteria [14]. It expresses the total discriminatory potential of all reagents included in the set. If reagents giving identical values for

all isolates are included in the set, the resolution index will decrease. An optimal test set should only contain those tests which give a good discrimination, and of course a good reproducibility, for the group of microorganisms to be studied. This is clearly shown when methods API:1 and API:3 are compared (Table 3), where the resolution index, the variety, the accuracy and the diversity are increased by omitting those tests which show low variation among different isolates.

We have earlier used r_{mean} to describe the homogeniety of a studied population, and whether any epidemiological relations seem to exist in the population [18–20]. We found that populations of unrelated *E. coli* strains always showed a homogeniety of *c* 0·65 when the PhP-EC system was used. The same homogeniety was also found in the population of 96 isolates studied here.

Variety is another measure of the discriminatory power. The theoretical variety is the number of types which may be determined, using the actual typing system. For biotyping, e.g. this number is 2^T, where T is the number of tests used. For biochemical fingerprinting the theoretical variety is almost infinite. The actual variety is only relative, i.e. it expresses the number of types found in a particular study, and is only of interest for comparisons of different typing systems using a certain population.

The diversity index is a measure of the distribution of the bacterial isolates into different types. It measures the probability that two randomly selected isolates will be assigned to different groups. An advantage of the diversity index is that it can be used to compare all kinds of typing systems, like e.g. serotyping and phage-typing [17], phage-typing and biotyping [21] or serotyping and biochemical fingerprinting [14].

Another measure is the stability of the typing system. A typing system should measure properties of the bacteria which are stable during an epidemic outbreak, but which differ between different clones of bacteria. Biotyping systems normally measure properties which are stable among *E. coli* strains [22]. Some typing systems, like e.g. plasmid typing, measure characters which are more easily subject to changes during an outbreak. With the PhP-EC system we have evaluated the stability of *E. coli* phenotypes over long time periods [18], upon geographic spread [19], and during storage and subculturing [23], and found that the phenotypes in most cases were stable.

In the present study we have used the above measures to compare biotyping and biochemical fingerprinting of *E. coli* , and to compare biochemical fingerprinting of *E. coli* using the PhP-EC and API 50 CH system.

The two typing systems showed a high interassay reproducibility when speed and intensity of the reactions were evaluated together (biochemical fingerprinting). Biotyping with the API 50 CH system yielded poor reproducibility, mainly because of variations in certain test results after 24 h incubation. However, those tests results which varied between + and − after 24 h incubation for duplicate assays of an isolate, usually were due to late fermentation reactions. When using quantitative coding of these tests [0–16] they were normally assigned values between 3 and 8, yielding a maximal error in that particular test of 35%, whereas in biotyping one variation step in a test result means an error of 100% in that particular test.

With regard to discrimination, the PhP-EC system was superior to the other

methods. This can partly be explained by the fact that API 50 CH is not mainly designed for typing of *E. coli*, but is a more general system for studying the metabolism of carbohydrates by Enterobacteriaceae. The API 50 CH contains several reagents which do not discriminate *E. coli* strains. The results from those reagents increased the overall similarity between non identical isolates, and reduced the resolution capacity of the system. For the further analysis, tests with low variation among 96 independent *E. coli* isolates (standard deviation lower than 1·0), were excluded. Using the remaining 26 tests, the number of phenotypes in the material increased from 33 to 48.

PhP-EC, on the other hand, is specially designed to discriminate between different *E. coli* clones. The high reproducibility and resolution capacity of the PhP-EC system is mainly due to two things. Firstly, the evaluation of the test results is done objectively, with an optical reader, so that errors due to subjective judgements of colour nuances do not occur. Secondly, the reagents used in the PhP-EC system have been carefully selected by assaying several thousands of *E. coli* isolates with different sets of reagents, and only those reagents showing a high reproducibility and a high inter-strain variation for *E. coli* were included in the system.

Biotyping with the API 50 CH system yielded poor discrimination of the isolates, much depending on the poor reproducibility. In this case, elimination of those tests which were not discriminating for *E. coli* did not improve the discrimination power, although the resolution index was increased. This is because the identity level has to be correspondingly decreased, when tests yielding identical results for all strains, and thus have 100% reproducibility, are eliminated.

In traditional biotyping systems, the reproducibility is normally not considered when assigning an isolate to a specific biotype. The number of biotypes found in this material of 96 isolates was 63 using methods API:2 and API:4, and assuming an identity level of 1·000. However, since the reproducibility assay showed variations in certain tests, many isolates which were identical would have been assigned to different biotypes using traditional biotyping.

In conclusion, for typing purposes the performance of the API 50 CH system may be greatly improved in two ways. Firstly, by using results from several readings and coding them quantitatively, and secondly by using only those tests with a high discrimination for the group of microorganisms studied. However, it was not possible to achieve such high reproducibility and discrimination as with the PhP-EC system, which as been especially designed for biochemical fingerprinting of *E. coli*.

ACKNOWLEDGEMENTS

This study was supported by grants from the Karolinska Institute, the Swedish Board for Forestry and Agricultural Research, Magn Bergvall Foundation, Tore Nilsson Foundation, Åke Wiberg Foundation, the Swedish Diabetes Foundation, Swedish Hoechst, Nordic Insulin Foundation and the Swedish Society of Medical Science. We wish to thank Anna-Lisa Belin and Birgitta Karlsson for skilful technical assistance.

REFERENCES

1. Ørskov, F, Ørskov I, Jann B, Jann K. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol Rev 1977; **41**: 667–710.
2. Ørskov F. Genus *Escherichia*. In: Kreig NR, eds. Bergey's manual of systematic bacteriology. Baltimore: Williams and Wilkins, 9th ed, 1984 (1).
3. Gargan R, Brumfitt W, Hamilton-Miller JMT. A concise biotyping system for differentiating strains of *Escherichia coli*. J Clin Pathol 1982; **35**: 1366–1369.
4. Jackman P. Microbial identification and systematics based on whole-cell protein patterns. In: Sanna A, Morace G, eds, New horizons in microbiology. Elsevier Science Publisher B.V. 1984: 71–8.
5. Achtman M, Mercer A, Kusecek B, et al. Six widespread bacterial clones among *Escherichia coli* K1 isolates. Infect Immun 1983; **39**: 315–35.
6. Achtman M, Pluschke G. Clonal analysis of descent and virulence among selected *Escherichia coli*. Ann Rev Microbiol 1986; **40**: 185–210.
7. Crichton PB, Old DC. Biotyping of *Escherichia coli*. J Med Microbiol 1979; **12**: 473–86.
8. Crichton P, Old DC. A biotyping scheme for the subspecific discrimination of *Escherichia coli*. J Med Microbiol 1982; 223–42.
9. Okerman L, Devriese LA. Biotypes of enteropathogenic *Escherichia coli* strains from rabbits. J Clin Microbiol 1985; **22**: 955–58.
10. Godbout-Delasalle F, Higgins R. Biotyping of clinical isolates of *Escherichia coli* of animal origin, using the API 20E system. Can J Vet Res 1986; **50**: 418–21.
11. Cicmanec JF, Evans AT. Classification of urinary tract infections by biotype identification of the pathogens. J Urology 1979; **124**: 68–9.
12. Brauner AL, Boeufgras JM, Jacobson SH, Kaijser B, Källenius G, Svenson SB, Wretlind B. The use of biochemical markers, serotype and fimbriation in the detection of *Escherichia coli* clones. J Gen Microbiol 1987; **133**: 2825–34.
13. Kühn I. Biochemical fingerprinting of *Escherichia coli*. A simple method for epidemiological investigations. J Microbiol Meth 1985; **3**: 159–70.
14. Kühn I, Burman L, Ericsson L, Möllby R. Subtyping of *Klebsiella* by biochemical fingerprinting. A simple system for epidemiological investigations. J Microbiol Meth 1990; **11**: 177–85.
15. Brauner A, Hylander B, Wretlind B, Östenson CG, Kühn I. *Escherichia coli* populations from diabetic and non-diabetic patients with bacteraemia and faecal samples from healthy subjects – a comparative study. Epidemiol Infect 1990; **105**: 533–40.
16. Sneath PHA, Sokal RR. Numerical taxonomy. San Francisco: W. H. Freeman and Company, 1973.
17. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. J Clin Microbiol 1988; **26**: 2465–6.
18. Kühn I, Franklin A, Söderlind O, Möllby R. Phenotypic variations among enterotoxinogenic O-groups of *Escherichia coli* from Swedish piglets with diarrhoea. Med Microbiol Immun 1985; **174**: 119–30.
19. Kühn I, Möllby R. (1986) Phenotypic variations among enterotoxinogenic O-groups of *Escherichia coli* from various human populations. Med Microbiol Immun 1986; **175**: 15–26.
20. Kühn I, Tullus K, Möllby R. Colonization and persistence of *Escherichia coli* phenotypes in the intestines of children aged 0 to 18 months. Infection 1986; **14**: 7–16.
21. Old DC, Barker RM. Numerical index of the discriminatory ability of biotyping for strains of *Salmonella typhimurium* and *Salmonella paratyphi* B. Epidemiol Infect 1989; **103**: 435–43.
22. Crichton P, Old DC. Biotyping of *Escherichia coli*: methods and applications. In: Sussman M, ed. The virulence of *Escherichia coli*: reviews and methods. London: Academic Press, Inc., 1985.
23. Katouli M, Kühn I, Möllby R. Evaluation of the stability of biochemical phenotypes of *Escherichia coli* upon subculturing and storage. J Gen Microbiol. In Press.