JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 1988, p. 2465–2466
0095-1137/88/112465-02$02.00/0

Vol. 26, No. 11

Numerical Index of the Discriminatory Ability of Typing Systems: an Application of Simpson's Index of Diversity

PAUL R. HUNTER* AND MICHAEL A. GASTON

Division of Hospital Infection, Central Public Health Laboratory, 61 Colindale Avenue, London NW9 5HT, England

Received 25 May 1988/Accepted 10 August 1988

An index of discrimination for typing methods is described, based on the probability of two unrelated strains being characterized as the same type. This index may be used to compare typing methods and select the most discriminatory system.

Epidemiological typing of bacterial strains can be carried out by a variety of techniques, including serotyping, biotyping, and bacteriophage and bacteriocin typing. To type many of the more frequently occurring human pathogens, there is often a choice of methods available that have been developed independently by different groups. For example, there have been at least 10 phage-typing schemes described for *Pseudomonas aeruginosa* (2).

An assessment of which typing method is the most efficient has to be based on a number of factors: typability, reproducibility, and discrimination. Of these characteristics, typability and reproducibility are relatively easy to quantify and are often expressed as simple percentages. Thus, the typability of a method is the percentage of distinct bacterial strains which can be assigned a positive typing marker, and the reproducibility is the percentage of strains that give the same result on repeated testing.

The discriminatory power of a typing method is its ability to distinguish between unrelated strains. It is determined by the number of types defined by the test method and the relative frequencies of these types. These two facets of discrimination are not generally presented as a single numerical value and therefore cannot be used for a straightforward comparison of different methods. In the literature, most workers present the frequency of the most common types and the number of types and it is often left to the interested reader to make a subjective assessment of how one system compares with another.

We suggest the use of a single numerical index of discrimination (D), based on the probability that two unrelated strains sampled from the test population will be placed into different typing groups. This probability can be calculated by Simpson's index of diversity, which was developed for the description of species diversity within an ecological habitat (9). This index can be derived from elementary probability theory (1) and is given by the following equation:

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} n_j(n_j - 1)$$

where N is the total number of strains in the sample population, s is the total number of types described, and n_j is the number of strains belonging to the jth type. This equation is derived as follows. The probability that a single strain sampled at random will belong to the jth group is n_j/N. The probability that two strains sampled consecutively will belong to that group is $n_j(n_j - 1)/N(N - 1)$. These probabilities can be summed for all the described types to give the probability that any two consecutively sampled strains will be the same type. This summation can be subtracted from 1 to give the equation above. No correcting factor for small populations has been made, as typing schemes should not be validated with small samples.

The application of this index is illustrated by comparing two hypothetical typing schemes, both differentiating 20 types in a population of 100 strains. In the first, we will assume a distribution of strains of 40 in type 1, 30 in type 2, 7 in types 3 and 4, and 1 in the remaining 16 types. In the second, we assume an ideal distribution of five strains in each type. Working through the first example, we have N = 100, s = 20, n_1 = 40, n_2 = 30, n_3 and n_4 = 7, n_5 to n_{20} = 1: $D = 1 - [(40 \times 39 + 30 \times 29 + 7 \times 6 + 7 \times 6 + 1 \times 0 \ldots 1 \times 0)/(100 \times 99)] = 1 - (2514/9900) = 0.746$. Therefore, the first example has an index of 0.746. This index indicates that if two strains were sampled randomly from the population, then on 74.6% of occasions they would fall into different types. The second example is clearly more discriminating and has an index of 0.960.

This equation can be applied both to a direct comparison of the discriminating power of typing methods and to analysis of the discriminating power of combined typing schemes. Table 1 shows the discriminating indices for various *Candida albicans* typing methods. The data have been adapted from descriptions of these methods in the literature and are presented solely for the purposes of illustrating the discriminatory index. It can be seen that the discriminatory power of all the methods is fairly poor. Resistotyping is the most discriminatory method, but it is of note that DNA typing has better discrimination than either the killer system or immunoblotting, both of which are able to differentiate more types. The killer system has better discrimination than immunoblotting, despite having a higher proportion of strains in the most frequent type.

TABLE 1. Discrimination indices for some *C. albicans*-typing methods

Method (reference)	No. of types	Size (%) of largest type	Discrimination index
Resistotyping (6)	16	25	0.899
DNA typing (8)	10	35	0.868
Killer system (7)	25	52	0.724
Immunoblotting (5)	16	41	0.679
Enzyme biovars (3)	4	64	0.549
Serotyping (4)	2	68	0.438

* Corresponding author.

TABLE 2. Discrimination indices for hierarchical typing methods for *Klebsiella* spp., *E. cloacae*, and *S. marcescens*[a]

Organism(s)	Discrimination index	
	Serotyping alone	Serotyping and phage typing
Klebsiella spp.	0.962	0.997
E. cloacae	0.899	0.998
S. marcescens	0.567	0.783

[a] Strains were differentiated by serotype, and then strains of the same serotype were subdivided by phage sensitivity. Data drawn from cultures sent to the Central Public Health Laboratory, 1985 to 1987, for epidemiological investigations.

Table 2 illustrates the use of the discrimination index to validate the hierarchical typing methods for *Klebsiella* spp., *Enterobacter cloacae*, and *Serratia marcescens*. For each of the three groups, serotyping alone provides less discrimination than when serotyping is combined with phage typing to discriminate within the frequent serotypes. It also highlights the comparatively poor discrimination given by O-serotyping *S. marcescens*.

A second application for this index is in the development of nonserological typing methods, such as phage- and bacteriocin-typing schemes. In phage typing for example, there is usually a large number of candidate phages from which to select a subset for routine typing. We suggest that once the initial test (phage) is selected, the discriminatory index can be used in a simple computer program to select the subsequent phages to be included in the set. In practice, the initial phage would probably be the test with the highest discrimination (the phage that was closest to lysing 50% of test strains). The discrimination provided by each of the remaining phages used in conjunction with the first phage can be quantified with the index, and the best phage can then be selected. This process would be continued until the remaining candidate phages could not significantly improve the discrimination of the phage-typing set.

We believe that this index greatly aids comparisons between typing systems and that it would be useful if it were included in the descriptions of each new method. It must be emphasized that it is of most value for large and representative (nonlocal) collections of distinct strains. Clearly, in developing a new typing scheme, one should aim for as large a discriminatory index as possible. Of course, the acceptable level of discrimination will depend on a number of factors, but an index of greater than 0.90 would seem to be desirable if the typing results are to be interpreted with confidence.

LITERATURE CITED

1. **Armitage, P., and G. Berry.** 1987. Statistical methods in medical research, p. 49–53. Blackwell Scientific Publications, Ltd., Oxford.
2. **Bergan, T.** 1978. Bacteriophage typing of *Pseudomonas aeruginosa*, p. 169–199. *In* T. Bergan and J. R. Norris (ed.), Methods in microbiology, vol. 10. Academic Press, Inc. (London), Ltd., London.
3. **Casal, M., and M. J. Linares.** 1983. Preliminary investigation of *Candida albicans* biovars. J. Clin. Microbiol. **18**:430–431.
4. **Hasenclever, H. F., and W. O. Mitchell.** 1963. Antigenic studies of *Candida*. IV. The relationship of the antigenic groups of *Candida albicans* to their isolation from various clinical specimens. Sabouraudia **2**:201–204.
5. **Lee, W., J. Burnie, and R. Matthews.** 1986. Fingerprinting *C. albicans*. J. Immunol. Methods **93**:177–182.
6. **McCreight, M. C., and D. W. Warnock.** 1982. Enhanced differentiation of isolates of *Candida albicans* using a modified resistogram method. Mykosen **25**:589–598.
7. **Polonelli, L., C. Archibusacci, M. Sestito, and G. Morace.** 1983. Killer systems: a simple method for differentiating *Candida albicans* strains. J. Clin. Microbiol. **17**:774–780.
8. **Scherer, S., and D. A. Stevens.** 1987. Application of DNA typing methods to epidemiology and taxonomy of *Candida albicans*. J. Clin. Microbiol. **25**:675–679.
9. **Simpson, E. H.** 1949. Measurement of diversity. Nature (London) **163**:688.