Journal of General Microbiology (1977), **98**, 595–597
Printed in Great Britain

SHORT COMMUNICATIONS

Reticulation and Other Methods of Reducing the Size of Printed Diagnostic Keys

By R. W. PAYNE

Rothamsted Experimental Station, Harpenden, Hertfordshire, AL5 2JQ

(*Received* 28 *June* 1976; *revised* 27 *September* 1976)

INTRODUCTION

Long keys are now easily constructed by computer, thus compact representations have become important. Figure 1 (*a*), taken from a key for 434 species of yeast, shows a compact representation suggested by Payne, Walton & Barnett (1974). Each test (or character) description is given only once, so that both outcomes of the test (or leads) can be printed on the same line. Figure 1 (*a*) is part of a key constructed and printed by the computer program GENKEY (Payne, 1975*a*, *b*) and contains 1187 couplets. The compact representation takes 1597 lines to print, whereas the conventional forms take 2908 lines. (A full key for the 434 yeast species can be found in Barnett & Pankhurst, 1974.) Reticulation, as described in this communication, gives further savings.

Species with variable characters may be keyed out at several endpoints: see, for example, *Trichosporon cutaneum* in Fig. 1 (*a*). With many such species, duplicate subkeys may occur, for example, the subkeys in Fig. 1 (*a*) starting with 'Lactose growth' at reference numbers 37 and 42. Figure 1 (*b*) shows how to avoid printing repeated subkeys, simply by changing reference number 42 to 37 and renumbering references 43 onwards. This process, termed reticulation, has been used in keys constructed by hand (see, for example, Bonnier, 1917) but not in computer-produced keys. The example in Fig. 1 is one of 37 reticulations in the full key, which reduce the number of couplets to 1146 and the length of the printed key to 1543 lines (compact form) or 2809 lines (conventional form).

METHODS

One might detect reticulation either while constructing a key or by inspecting a stored key before printing. The former alternative has the advantage that tests can be specially selected if they produce a group of species already occurring elsewhere in the key and, if such a group occurs, the duplicate subkey need not actually be constructed. This seems to be the method used when constructing keys by hand. However unless, as when constructing a key by hand, one has clear prior knowledge of likely reticulation, the time saved by avoiding construction of duplicate subkeys would be small compared with the time taken to match lists of species with those occurring elsewhere in the key. Furthermore, the time taken to compute a key greatly exceeds that taken to print it (see Discussion), so it is important not to over-complicate test-selection procedures.

Using GENKEY, reticulation is detected by scanning the stored key before printing. The key is stored as a list of integers representing each sub-branch of the key in turn, thus subkeys are represented as sub-lists of the full list and can easily be detected.

(*a*)

		NEGATIVE	POSITIVE
33	GLYCEROL GROWTH	34	39
34	CELLOBIOSE GROWTH	35	38
35	D-XYLOSE GROWTH	36	37
36	ETHANOL GROWTH	CANDIDA SLOOFFII SACCHAROMYCES TELLURIS TORULOPSIS BOVINA TORULOPSIS PINTOLOPESII	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS SACCHAROMYCES TELLURIS TORULOPSIS BOVINA
37	LACTOSE GROWTH	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS	TRICHOSPORON CUTANEUM
38	D-GLUCOSE FERMENTATION	TRICHOSPORON CUTANEUM	HANSENIASPORA UVARUM HANSENIASPORA VALBYENSIS
39	D-GLUCOSE FERMENTATION	40	43
40	D-XYLOSE GROWTH	41	42
41	CITRATE GROWTH	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS PICHIA NONFERMENTANS TRICHOSPORON CAPITATUM	PICHIA MEMBRANAEFACIENS PICHIA TERRICOLA
42	LACTOSE GROWTH	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS	TRICHOSPORON CUTANEUM
43	D-XYLOSE GROWTH	44	CANDIDA LAMBICA CANDIDA VALIDA PICHIA MEMBRANAEFACIENS

(*b*)

		NEGATIVE	POSITIVE
33	GLYCEROL GROWTH	34	39
34	CELLOBIOSE GROWTH	35	38
35	D-XYLOSE GROWTH	36	37
36	ETHANOL GROWTH	CANDIDA SLOOFFII SACCHAROMYCES TELLURIS TORULOPSIS BOVINA TORULOPSIS PINTOLOPESII	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS SACCHAROMYCES TELLURIS TORULOPSIS BOVINA
37	LACTOSE GROWTH	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS	TRICHOSPORON CUTANEUM
38	D-GLUCOSE FERMENTATION	TRICHOSPORON CUTANEUM	HANSENIASPORA UVARUM HANSENIASPORA VALBYENSIS
39	D-GLUCOSE FERMENTATION	40	42
40	D-XYLOSE GROWTH	41	37
41	CITRATE GROWTH	CANDIDA VALIDA PICHIA MEMBRANAEFACIENS PICHIA NONFERMENTANS TRICHOSPORON CAPITATUM	PICHIA MEMBRANAEFACIENS PICHIA TERRICOLA
42	D-XYLOSE GROWTH	43	CANDIDA LAMBICA CANDIDA VALIDA PICHIA MEMBRANAEFACIENS

Fig. 1. Part of a key for 434 species of yeast: (*a*) without any reticulation; (*b*) with duplicate subkey '42 LACTOSE GROWTH – – – –' omitted by reticulation.

Although detecting reticulation during construction is inefficient, there are simple methods of influencing test selection that increase the likelihood of duplicate subkeys. These methods are applicable however reticulation is implemented.

GENKEY uses a test-selection method similar to that described by Gower & Barnett (1971). Tests are initially assessed using a criterion function that depends on the relative sizes of the groups of species that can give each possible test result. The aim is to choose a test with few variable responses and with group sizes as nearly equal as possible, since this tends to produce an efficient key. (The most efficient key is defined as that with the minimum average number of tests per identification.) Should identical groups of species occur at different points in the key, duplicate subkeys will be constructed, since the most efficient subkey for the species concerned will be the same at each point.

Subsidiary criteria distinguish between tests whose main criterion values are nearly equal. Reticulation can be treated as a subsidiary consideration since the length of the printed key is generally of secondary importance to its efficiency.

At any point in the key, a duplicate subkey can occur only if every member of the group of species at that point occurs elsewhere in the key, i.e. if, for each species, a test with respect to which it is variable has been used earlier in the key. Thus tests should be chosen that tend

to separate such 'variable' species from those that have only given fixed responses, i.e. will occur at only one endpoint. This can be performed by calculating for each test the proportion of 'variable' species in the group of species giving each possible test result, and using a criterion similar to the main criterion to select the test with group proportions as unequal as possible.

DISCUSSION

If tests with variable responses are used in a key, reticulation may reduce the length of the printed key. The saving in printing the yeast key was only 3·5 %. However this key was constructed and stored on a computer file for later printing before it was possible to influence the choice of tests to encourage reticulation, thus further savings might be possible. The computing time required to detect reticulation was about 25 % of the total time taken to read the key back from the computer file, reticulate and print it. However this was only 2 % of the time taken to construct the key.

A prospectus describing the facilities and availability of the program GENKEY can be obtained from the author.

REFERENCES

BARNETT, J. A. & PANKHURST, R. J. (1974). *A New Key to the Yeasts*. Amsterdam: North Holland Publishing Co.

BONNIER, G. E. M. (1917). *Name this Flower*. London: Dent.

GOWER, J. C. & BARNETT, J. A (1971). Selecting tests in diagnostic keys with unknown responses. *Nature, London* **232**, 491–493.

PAYNE, R. W., WALTON, E. & BARNETT, J. A. (1974). A new way of representing diagnostic keys. *Journal of General Microbiology* **83**, 413–414.

PAYNE, R. W. (1975*a*). Genkey, a program for constructing diagnostic keys. In *Biological Identification with Computers*, pp. 65–72. London: Academic Press.

PAYNE, R. W. (1975*b*). *Genkey*, Inter-University/Research Council Series, Report no. 24. Edinburgh: University of Edinburgh.