THE DUAL POLYCLAVE: AN AID TO MORE EFFICIENT IDENTIFICATION

BY R. W. PAYNE

Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, U.K. and C.S.I.R.O. Division of Mathematics and Statistics, Glen Osmond, South Australia 5064;

R. R. LAMACRAFT

C.S.I.R.O. Division of Mathematics and Statistics, Glen Osmond, South Australia, 5064;

AND R. P. WHITE

Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, U.K.

(*Accepted 11 April 1980*)

SUMMARY

Polyclaves have advantages over conventional aids to identification in that the user is free to examine any convenient character. However the polyclave gives no guidance about the best character to examine next during an identification. This deficiency can be overcome by using a *dual polyclave* which indicates how each character separates the taxa not eliminated by characters already observed.

INTRODUCTION

A polyclave is a set of punched cards that describes the properties of some set of taxa. In the original form (Bianchi, 1931; Clarke, 1936, 1938) there is an edge-punched card for each taxon. Each character is allocated a set of holes, one hole for each possible state; continuous characters are represented by discrete ranges. On each card, the holes corresponding to the character-states that can occur within the taxon concerned, are notched, i.e. the card between the hole and the edge of the card is cut away.

Specimens of the taxa can be identified by observing characters one at a time: after each character-state has been determined, a rod is inserted through the appropriate hole and the notched cards are shaken out; the remainder, which correspond to the taxa that cannot exhibit the character-state concerned, are discarded. The process continues until there is only one card left.

In the alternative form of polyclave there is card for each character-state, and a position on the cards is allocated to each taxon. Each card has holes punched in the positions corresponding to the taxa that can exhibit the character-state concerned. To determine the taxa that can exhibit an observed set of character-states, the appropriate cards are superimposed and held to the light to see which positions have holes punched on every card. This form can be accommodated on standard computer cards and can be produced automatically (e.g. Morse, 1974; Pankhurst and Aitchison, 1975; Payne, 1978).

0028-646X/81/010121+06 $02.00/0

ADVANTAGES OF POLYCLAVES

Unlike most conventional methods of identification, the user of a polyclave is not forced to examine any particular character, nor to examine characters in any particular order. For example, identification keys require the user to observe a predetermined sequence of characters. A choice of character may be provided at some points in the sequence, but rarely for the initial characters (Payne and Preece, 1980, section 4).

It has been suggested that diagnostic tables (or synoptic keys) allow the characters to be chosen in any convenient order. However, Payne and Preece (1980) stress that use of a large table is feasible only if the character-states of the taxa are printed in a lexicographical order that allows the table to be used like a 'dictionary'. This is usually achieved by first ordering the taxa according to the character in the first column, then ordering taxa with identical states of the first character according to the character in the second column, and so on. It is then inconvenient to examine characters in other than column order.

DISADVANTAGES OF POLYCLAVES

Unlike the identification key, and to some extent the diagnostic table, which indicate the characters to be observed next, the polyclave gives no direct guidance. Thus time and effort may be wasted by observing characters whose states turn out to be shared by all the remaining taxa.

With the original, edge-punched, form this can be avoided by looking at the edges of the cards to find a character whose notched holes on the remaining cards are not all for the same state. However the equivalent procedure for the computer-generated form would require the user to superimpose the cards for each state of a character in turn, and check that the holes for the remaining taxa are not all on the same card. This would be tedious and prone to error.

AIDS FOR NON-SEQUENTIAL USE

Payne (1978) suggested that it might be found helpful to use a polyclave in conjunction with a list of *irredundant character sets* for the taxa concerned; these are sets of characters that are sufficient to identify all the taxa, but which contain no unnecessary characters. The user could then choose some convenient set of characters and examine them all before looking up the identification with the polyclave. However such non-sequential identification usually requires more characters than a sequential method in which characters are examined one at a time (cf. Payne and Preece, 1980). Hence these sets are useful mainly when character states take time to determine and several characters can be examined simultaneously (e.g. chemical tests to identify yeasts; Barnett, Payne and Yarrow, 1979, pp. 72, 277).

Another type of irredundant set is one that allows a particular taxon to be distinguished from all the remaining taxa (also known as a diagnostic character set; Pankhurst, 1978). These would be particularly useful to confirm or check an identification that has been made in some other way (Barnett, Payne and Yarrow, 1979, chapter 8).

Irredundant sets of either type can be produced by the computer program GENKEY (Payne, 1978).

Aids for sequential use

To help assess which characters are effective with a particular set of taxa, Payne (1975) defined the concepts of definite and partial separation.

Definite separation occurs if the set contains (at least one) pair of taxa that cannot exhibit the same state of the character. This character will eliminate at least one taxon, whatever state may be observed. For example in Table 1, which shows the possible states of taxa A to D for characters 1 to 3, the states of character 1 that can be observed with taxon A are distinct from those that can be observed with taxon B. Thus if state 1 is observed, taxon B can be eliminated; whereas if states 2 or 3 are observed, taxon A is eliminated (as well as taxa C and D).

Table 1. *States that can be exhibited by taxa A to D for characters 1 to 3*

	Character		
Taxon	1	2	3
A	1	1	2
B	2,3	1,2	2
C	1	1,3	2
D	1	1,4	2

A character gives *partial separation* if some of its states are exhibited by all the taxa in the set while others are exhibited by some, but not all, of the taxa. Some states of this character will eliminate taxa but others will not. For example, character 2 in Table 1 can eliminate taxa only if states 2, 3 or 4 are observed.

Finally a character whose states are given by either all or none of the taxa (like character 3 in Table 1) cannot eliminate any taxa, and thus gives no separation.

With the original form of polyclave, the type of separation provided by each character can be determined simply by examining the positions of notches within the sets of holes for each character. With the computer-generated form, this is not feasible and some ancillary aid must be provided.

The standard polyclave is used to determine the set of taxa that correspond to a particular set of character-states; we now wish to determine the characters that correspond to a particular set of taxa. This suggests that what is required is a 'dual' polyclave – a polyclave with a card for each taxon and positions for the character-states. In its simplest form this would have a single position for each character-state, and holes punched on each card in the positions corresponding to the character-states that can be exhibited by the taxon concerned. (This is apparently the same as the 'inverse polyclave' attributed to O. A. Chater by Pankhurst and Aitchison, 1975.)

As an example, the first four lines of Table 2 show the holes that would be punched on the cards for taxa A to D in Table 1, and the final line shows the positions where there would be unobscured holes if the cards were superimposed. Character 1 is readily seen to give definite separation as none of its positions has a hole punched on every card (so each state eliminates at least one taxon). However characters 2 and 3 both show a similar pattern and the ability of character 2 to provide partial separation can be detected only by examining the individual cards to check whether there are taxa that can exhibit additional states to state 1.

It might seem that a character that gives definite separation will always be

Table 2. *Positions where holes would be punched * in a simple dual polyclave for the taxa in Table 1*

Positions for Character	1			2				3	
State	1	2	3	1	2	3	4	1	2
Card for taxon									
A	*	.	.	*	.	.	.	.	*
B	.	*	*	*	*	.	.	.	*
C	*	.	.	*	.	*	.	.	*
D	*	.	.	*	.	.	*	.	*
Positions with holes punched on all four cards	.	.	.	*	.	.	.	.	*

Table 3. *Positions with holes punched * in the full dual polyclave for the taxa in Table 1*

Character	1						2								3			
State	1		2		3		1		2		3		4		1		1	
Position	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
Card for taxon																		
A	*	.	.	*	.	*	*	.	.	*	.	*	.	*	.	*	*	.
B	.	*	*	.	*	.	*	.	*	.	.	*	.	*	.	*	*	.
C	*	.	.	*	.	*	*	.	.	*	*	.	.	*	.	*	*	.
D	*	.	.	*	.	*	*	.	.	*	.	*	*	.	.	*	*	.
Positions punched on all four cards	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	*	*	.

superior to a character that gives only partial separation, and that it is therefore unimportant for partial separation to be easily detected. This is not necessarily true: use of a character that gives partial separation can sometimes lead to more efficient identification than using only characters that give definite separation. For example, if state 2 of character 2 – and thus taxon A – occur only rarely, character 2 may be preferable to character 1, as the former immediately identifies the three common taxa, B, C and D. Also, characters that give definite separation may be unavailable for closely-related groups of taxa.

To distinguish partial separation from no separation, each character-state requires an extra position in which a hole is punched if the taxon concerned can never exhibit that state. In Table 3, which shows the full dual polyclave for taxa A to D, these extra positions are denoted by a minus, and the positions punched if a taxon can exhibit a state are denoted by a plus. Character 3 now has position 1− punched on every card as well as position 2+. This indicates that state 2 occurs with all the taxa and that the only other state, state 1, can never occur; hence character 3 gives no separation. Conversely, although state 1 of character 2 can be exhibited by all the taxa, the fact that positions 2–, 3– and 4– do not have holes on every card shows that states 2, 3 and 4 can also occur; thus character 2 is shown to give partial separation.

The procedure for detecting the various types of separation can be defined formally as follows: select and superimpose the cards of the dual polyclave belonging to the taxa of interest, then

(1) a character gives definite separation if none of the plus positions has a hole punched on every card (although some of the minus positions may have holes on every card);
(2) a character gives only partial separation if at least one of its plus positions has a hole on every card and it also has at least one state neither of whose positions has a hole punched on every card;
(3) a character gives no separation if either the plus or the minus position of each state has a hole punched on every card.

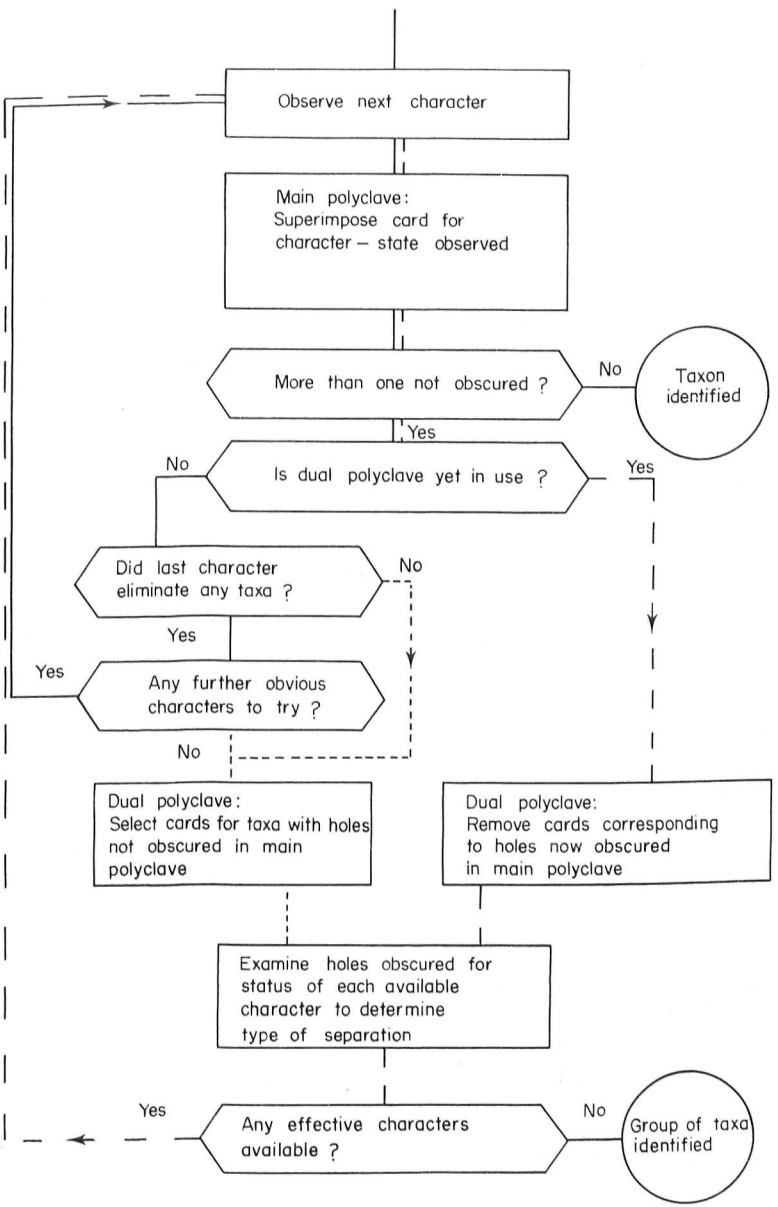

Fig. 1. Flow chart showing the use of the dual polyclave in conjunction with the main polyclave.

CONCLUSION

Deficiencies of the standard polyclave can be countered by an additional, *dual*, polyclave. This is a set of cards, one for each taxon, with a pair of positions allocated to each character-state. On the card for each taxon a hole is punched in one of each pair of positions depending on whether or not that taxon can exhibit the character-state concerned.

The procedure for using the two polyclaves is illustrated by the flow-chart in Figure 1. Most users will prefer to start with only the standard (or main) polyclave in the normal way. This corresponds to the steps, linked by continuous lines, in which obvious characters of the specimen are observed and the appropriate cards of the main polyclave are superimposed. Eventually most of the remaining characters will tend to have constant states with the taxa not eliminated by the characters so far observed (i.e. the taxa with holes punched on all the superimposed cards). It will then be difficult to find a character that can eliminate any further taxa. At this stage the steps indicated by dotted lines should be performed: select and superimpose the cards of the dual polyclave that correspond to the unobscured holes in the main polyclave. By examining the superimposed cards of the dual polyclave, as described above, the type of separation provided by each character can be determined and characters that give no further separation can be avoided. Once the dual polyclave is in use, the sequence of steps are those indicated by the dashed line: the card for each observed character-state is added to the main polyclave to obscure further holes; then the cards corresponding to these holes are removed from the dual polyclave and the next character is chosen. There are two points at which the process can terminate, depending on whether or not the specimen has sufficient characters to allow full identification.

The construction of dual polyclaves has been implemented in the computer program GENKEY (Payne, 1978). The program automatically prints the taxon name on each card of the dual polyclave and also the co-ordinates of the position of that taxon in the main polyclave. This cross-referencing is completed by printing the co-ordinates of the positions of each character-state in the dual polyclave on the appropriate card of the main polyclave. Further information about facilities and availability of GENKEY can be obtained from the authors.

REFERENCES

BARNETT, J. A., PAYNE, R. W. & YARROW, D. (1979). *A Guide to Identifying and Classifying Yeasts.* Cambridge University Press, Cambridge.

BIANCHI, A. T. J. (1931). Een nieuwe determinatie-methode. *Tectona*, **24**, 884–893. Partial English translation: *Tropical Woods*, No. 29 (1932), p. 53.

CLARKE, S. H. (1936). A use for the Paramount sorting system. *Nature*, **137**, 535.

CLARKE, S. H. (1938). A multiple entry perforated-card key with special reference to the identification of hardwoods. *New Phytologist*, **37**, 369–374.

MORSE, L. E. (1974). Computer programs for specimen identification, key construction and description printing. Publications of the Museum of Michigan State University, *Biology*, **5**, (1).

PANKHURST, R. J. (1978). *Biological Identification.* Edward Arnold, London.

PANKHURST, R. J. & AITCHISON, R. R. (1975). A computer program to construct polyclaves. In: *Biological Identification with Computers* (Ed. by R. J. Pankhurst), pp. 73–78. Academic Press, London.

PAYNE, R. W. (1975). Genkey. *Inter University/Research Council Series Report No. 24.* Program Library Unit, Edinburgh Regional Computing Centre, Edinburgh.

PAYNE, R. W. (1978). *Genkey, a Program for Constructing Identification Keys and Diagnostic Tables.* Rothamsted Experimental Station, Harpenden, Herts.

PAYNE, R. W. & PREECE, D. A. (1980). Identification keys and diagnostic tables: a review (with discussion). *Journal of the Royal Statistical Society, Series A*, **143**, 253–292.