Forum

Primary Homology Assessment, Characters and Character States

Julie A. Hawkins, Colin E. Hughes and Robert W. Scotland[1]

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, U.K.

We discuss contrasting approaches to cladistic character definition and thus to cladistic data matrix compilation. The conventional approach considers character states as alternate forms of the "same thing" (the character). A review of the challenges to this convention is presented, and their implications evaluated. We argue that the recognition of structures which are alternate forms is a vital stage of primary homology assessment and is equivalent to the conceptualization of a transformational homology. Such a view complies with the demand that characters are independent and that character states are hierarchically related. We identify one justifiable solution to the inapplicable data coding problem (coding for organisms which have red tails, blue tails or no tails), and show that alternative approaches to character definition support spurious solutions which deny the relation of structures which are "the same but different". We propose that the term character can be defined, in a cladistic context, as the descriptive label referring to a transformational homology evidenced by the similarity criterion.

© 1997 The Willi Hennig Society

[1]Author for correspondence. E-mail: Robert.Scotland@ Plant-Sciences.ox.ac.uk

INTRODUCTION

Homology is "the relation that systematists and comparative anatomists use in generating hypotheses of relationship" (Patterson, 1982: 29). Recognition that every homology statement involves both the generation of a proposition of homology and the legitimation of that hypothesis through congruence (see Rieppel, 1988 for discussion), led de Pinna (1991) to propose the terms *primary* and *secondary* homology in reference to these two stages. Two stages are also involved in formulation of a hypothesis of primary homology (Brower and Schawaroch, 1996). First, comparative morphological or molecular study of organismal variation is used to define characters which in turn define columns in the data matrix. Second, characters are partitioned and coded as character states which are assigned to terminal taxa as one column in the data matrix. The data matrix may thus be viewed as a set of primary homology statements.

The data matrix, and therefore the procedure of primary homology assessment, is the prime determinant of the outcome of cladistic analysis (Brady, 1983; Bryant, 1989). In contrast to the conceptual and

methodological explicitness and rigour of parsimony analysis, assessment of primary homology remains a contentious, subjective, and yet ultimately crucial and influential, step in any cladistic analysis (Pimentel and Riggins, 1987: 201; Bryant, 1989: 221; Pogue and Mickevich, 1990: 319, 359; de Pinna, 1991: 380; Stevens, 1991: 573, 506; Smith, 1994: 34; Pleijel, 1995: 309). Quite simply, "different workers will perceive and define characters in different ways" (Smith, 1994: 34).

We believe that two points have contributed to growing confusion surrounding primary homology assessment. First is the notion that characters and character states need not be distinguished. Second is the denial of the role of transformational homology in homology assessment. These views have resulted in the increasingly common use of methods incompatible with the conventional view of cladistic characters and character states.

The aim of this paper is to explore the implications of some different approaches to cladistic character definition. We do not provide complete answers, but seek to explain the fundamental importance of transformational homology, independence and an additive relation of character states in primary homology assessment. We begin by introducing what we consider to be the conventional approach, presented explicitly by Platnick (1979), and arising directly from the work of Hennig (1966).

Hennig (1966) described "character conditions" as plesiomorphic or apomorphic; he considered character conditions to be related in that they are actual evolutionary transformations from an original condition. Platnick (1979) also considered character states as transformations. He argued that "a character is a theory, a theory that two attributes which appear different in some way are nonetheless the same" and presented a clear definition of the term character in a cladistic context: "a character consists of two or more different attributes (character states) found in two or more specimens that, despite their differences can be considered alternate forms of the same thing (the character)" (Platnick, 1979: 542).

Although Platnick (1979) accepted the notion of character states as modifications, he also argued that since all characters can be seen as modifications of other characters, "the grouping of character states within a character can be seen as just arbitrarily delimiting clusters of separate characters that are increasingly more restricted in generality (i.e. that form nested sets of increasingly modified versions of other characters)" (Platnick, 1979: 543). Patterson (1988: 604) also down-played the distinction between characters and character states:

"Many systematists find it necessary or useful to distinguish these (characters and character states) two concepts in discussing homologous features; for example, in mammals the character "cochlea" may have the states "curved" (monotremes) or "spiral" (therians), or in angiosperms the character "flower" may have the states "red" or "blue". In the same way, in comparative molecular biology the character "position 86 in myoglobin" may have the states "Ile", "Leu" etc. In agreement with many other systematists (e.g. see Wiley, 1981: 9; Schoch, 1986: 75; Ax, 1987: 108), I find this distinction neither necessary nor useful. The essence of systematics is hierarchy, and in a hierarchic framework homologous "characters" and their "states" each represent characters—but at more and less inclusive levels, just as "spiral cochlea" delimits a subset of organisms having "cochlea" and "myoglobin 86 Ile" delimits a subset of those having "myoglobin 86". So in what follows, no distinction is necessary or intended between characters or features and their states."

We believe that the distinction between characters and character states is fundamental to the operational task of constructing a data matrix. The views of Platnick (1979) and Patterson (1988) add to the confusion surrounding character and character state concepts, and could be interpreted as justifying extreme approaches to character coding, contrary to the conventional views of cladistic character coding.

Patterson (1982) contrasted taxic and transformational approaches to the study of homology. He explained that taxic homologies are concerned with the monophyly of groups whereas the transformational approach is concerned with imagined or observed transformation of one structure into another. The taxic approach advocated by Patterson (1982) was viewed as falling within the realm of systematics and hierarchy, whereas transformational homologies were viewed as "empty transformations that lead to no new hypotheses of grouping". Such a strong stance might be misinterpreted as an implicit criticism of the conventional approach to character delimitation, which depends on the recognition of structures which are the same but different. We agree with de Pinna (1991: 376), who reviewed the ways in which the term transformational homology has been used and suggested that proposals that a structure in one organism or taxon is a transformation of one in another are "identical to the conjecture of primary homologies". Brower and Schawaroch (1996) show that the recognition of structural correspondence is the first of the two steps of

primary homology assessment; their interpretation of characters and character states is compatible with what we refer to as the conventional approach.

It can be seen that the views expressed by Patterson (1982, 1988) might be misinterpreted and taken to support the position of coding each anatomical variable as a separate character, an approach we refer to as nominal variable coding. Wilkinson (1995a) compared "composite coding" and "reductive coding", two approaches which coincide with conventional coding and nominal variable coding; he argued that neither approach has a monopoly of advantages nor dangers. The logical extreme of abandoning the conventional approach to character delimitation was advocated by Pleijel (1995) who defended a nominal variables approach on the grounds that it "avoids statements regarding homology of character states within a character" and circumvents the problem that "relationships between the different states within a character are never questioned" (Pleijel, 1995: 312, 313). Similar schemes have been defended and implemented by others building cladistic data matrices, e.g. Bateman et al. (1992) and Grimes (1995) who opted for a uniformly two-state data matrix. The former included a number of nominal variables; the latter suggested that all characters should be coded, at least initially, as nominal variables and then modified in the light of a preliminary parsimony analysis whereby when "distribution on the cladogram indicates that the binary characters are better scored as a multistate character, the characters have been rescored as states of one character" (Grimes, 1995: 102). One obvious outcome of such approaches is that weighting problems arise when primary homology assessment is inconsistent, i.e. when a mixture of nominal variables and conventional characters are employed, since a conventionally coded binary character (in one column) has 50% of the weight of a nominal variable coding of the same variation (in two columns). However, we believe that there are more fundamental problems with this approach. Given the wide spectrum of different approaches evident amongst cladistic practitioners, and that, as stated earlier, primary homology assessment is the prime determinant of the outcome of cladistic analysis, there is a need for clarification. Here we present what we consider to be justification for the conventional approach to character definition.

AN EXAMPLE—RELATIONSHIP OF RED AND BLUE TAILS

Maddison (1993) discussed the consequences of inapplicable data in cladistic analyses with reference to an example of birds without tails or with red or blue tails. Table 1 shows a hypothetical data set comprising six taxa (A,B,C,D,E,F), which relates to Maddison's (1993) tail colour problem. Two of the taxa (A,B) have no tails, whereas four taxa (C,D,E,F) have tails. Given that existing data show that (C,D,E,F) form a group which have tails, what are the possible relationships of red-tailed and blue-tailed birds? The raw data in Table 1 can be coded in a number of ways. Three approaches are explored here, though others might be proposed. The different codings represent different views of the variation and therefore different primary homology assessments. Columns 1 and 2 represent two hypotheses of primary homology (one for tail presence and one for tail colour), column 3 represents one hypothesis of primary homology which includes both tail presence and colour, and columns 4 and 5 represent tail colour treated as two separate primary homology assessments. It is not hard to find examples of each of these approaches used, often arbitrarily in published cladistic analyses. On theoretical grounds, Pleijel (1995) attempts to reduce primary homology assessment to scoring all variation as simply present or absent, i.e. as columns 1, 4 and 5. Maddison (1993) contrasted two of the alternatives (i.e. columns 1 and 2 vs. column 3), while Smith (1994) suggested that a solution to the

TABLE 1

Data and Character Coding Based on the Example of Maddison (1993)

Taxa	Raw data	Column coding				
		1	2	3	4	5
A	No tail	0	?	0	0	0
B	No tail	0	?	0	0	0
C	Red tail	1	0	1	0	1
D	Red tail	1	0	1	0	1
E	Blue tail	1	1	2	1	0
F	Blue tail	1	1	2	1	0

Column 1: 0=absence of tail; 1=presence of tail.
Column 2: ?=inapplicable; 0=red; 1=blue.
Column 3: 0=absence of tail; 1=red tail; 2=blue tail
Column 4: 0=absence of blue; 1=presence of blue.
Column 5: 0=absence of red; 1=presence of red.

problem of coding taxa with missing entries due to character inapplicability is to "recast character definitions to avoid this problem…by amalgamating two or more binary characters into a single multistate character" (Smith, 1994: 43), i.e. as column 3.

The analyses (Table 2) were undertaken using these data; the full set of solutions is presented in Fig. 1A–D. Analyses and results are summarized in Table 2. Analysis 1 resulted in three cladograms (Fig. 1A–C), Analysis 2 resulted in four cladograms (Fig. 1A–D) and Analysis 3 resulted in one cladogram (Fig. 1C). The observation that different codings discover different cladograms illustrates that primary homology assessment and character coding are critically important. Maddison (1993) was concerned primarily with the difficulties of coding relative to inapplicable data and the treatment of missing data by cladistic parsimony programs. We expand that view and argue that his example also illustrates the necessity of taking a theoretically and operationally justifiable approach to primary homology assessment. Primary homology statements and solutions for the red tail–blue tail problems are here re-evaluated in the light of three theoretical issues: the nature of anatomical evidence, the importance of character independence, and hierarchical relations of character states.

CHARACTERS AND CHARACTER STATES

Anatomical Relationship and the Similarity Criterion

Comparative anatomy is central to cladistics (Rieppel, 1988). For the purposes of this discussion we accept the view of others (Patterson, 1982; Rieppel,

1988; de Pinna, 1991) that topographic correspondence (position within the whole) and ontogenetic and compositional similarity are the criteria used for the determination of anatomical relationship and the establishment of primary homology propositions.

Approaches to character definition which simply equate primary homology assessment with the identification of discrete variables as separate characters (e.g. Pleijel, 1995), or unite as one character variables which cannot be correctly considered as related (e.g. Smith,
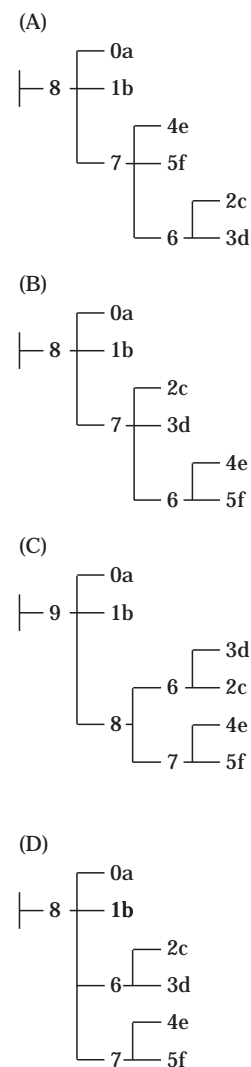

FIG. 1. Most parsimonious cladograms generated by the three analyses of the data given in Table 1. (A–C) are the three solutions for Analysis 1; (A–D) are the four solutions for Analysis 2; (C) is the single solution for Analysis 3.

TABLE 2

Summary of the Three Analyses Performed on the Data in Table 1, Showing Coding Options and Resultant Cladograms

Analysis	Columns used	Coding	Cladograms
1	1 and 2	Conventional	A, B and C
2	3	Multistate	A, B, C and D
3	1, 4 and 5	Nominal variables	C

1994) fail to recognize the role of anatomical relationship. Rather than being applied in order to distinguish similar structures which might be considered "the same but different", anatomical observations are used to support the observation of the "sameness" of one character state. Thus it is not redness and blueness which are considered as homologues and therefore states of the character "tail colour". Rather, redness is considered a primary homology in that organisms with red tails may be grouped by the synapomorphy "redness", which then becomes a character defining the group of red-tailed organisms. Individual observations become characters because there is an implied transformation to an unspecified state, "not red". The first of the two processes of primary homology assessment, the determination of topographic identity, has been subsumed into the second, that of determining character state identity.

Anatomical relationship, recognized through topographic correspondence, is what allows us to specify a "conditional phrase". Patterson (1982) drew attention to this important aspect of homology statements, citing Bock (1969), who explained that all homology propositions are meaningless without a conditional phrase, i.e. homologous as what? The example given by Patterson, taken from Bock (1974: 387), was that "the wing of birds and the wing of bats are homologous as the forelimb of tetrapods". In the case of red and blue tails, the conditional phrase in our interpretation is tail colour which can either be blue or red. When a nominal variable approach is taken, the conditional phrases are "presence of a blue tail" and "presence of a red tail". If the first character is coded as "absence of a blue tail (0); presence of blue tail (1)", what does the absence of a blue tail actually mean? In our view, the observations are that tails are either blue or red in colour, such that if you don't have a red tail you have a blue tail and vice versa. In other words, to code those taxa with red tails as having absence of a blue tail makes as little sense as coding absence of a purple tail or green tail. Patterson (1982) described the complement relation as a binary character which refers to the absence of a structure relative to the presence of the structure, i.e. absence of backbone/presence of backbone when the observation is that some taxa have a backbone and others do not. In the tail example, presence and absence of blue do not have a complement relationship; because of the observation of red, tail colour is present in two distinct ways.

If we consider the example of aligned DNA sequences, one nominal variable conditional phrase is "presence of adenine at a specified site", and character states describe the presence or absence of adenine at that site. In our view it is nonsensical to consider absence of adenine as a character state, or presence/absence of adenine as a character, since the equivalence of absence of adenine is the presence of guanine, cytosine or thymine. In the red tails and blue tails example, our preference for coding as two characters (columns 1 and 2) is justified in that there are two distinct conditional phrases appropriate to these data, one relating to whether a tail is present or absent and another relating to tail colour.

de Pinna (1991: 377) wrote that "conjectures of primary homology which do not conform to the criterion of similarity simply do not exist". Recent confusion in the partitioning of morphological variation into characters and character states suggests that the criterion of similarity is widely misunderstood, misapplied, or disregarded. The similarity criterion demands that any structures of different terminal taxa which appear to be "the same but different" should be subject to topographic, and perhaps ontogenetic and compositional, evaluation. Any structures which are topographically (and ontogenetically and compositionally) similar should be considered as a single primary homology statement and coded as one column in a data matrix.

Character Independence

One assumption required for parsimony analysis is that characters should be logically independent of one another (Felsenstein, 1982; Farris, 1983; Riggins and Farris, 1983). This is because if two characters are logically or functionally related, homoplasy in one would imply homoplasy in the other.

> "The fact that they [independent characters] suggest a similar phylogenetic conclusion, when they do, is significant. To the extent that there is no independence among characters, there is no particular meaning in their agreement—that agreement is just an expression of mutual dependence" (Goloboff, 1995: 103).

Lack of objective methods for detecting dependent characters and removing their effects from the analysis was perceived by Felsenstein (1982) to be the greatest weakness of existing methods of phylogenetic inference. As stated by de Pinna (1991: 380)

"the decision whether two or more attributes comprise a single transformation series or two or more independent series is one of the most basic, albeit confusing issues in systematics. It is a decision that is made very early in any character analysis, and rarely questioned subsequently. From that perspective, the distinction between "character" and "character state", frequently downplayed in the literature, becomes surprisingly relevant."

The term independence is often confused with notions of functional complexes of characters that always evolve in concert, with issues of physical, functional or genetic dependence and with attempts to remove dependent characters from the analysis. Independence of this type is unlikely to be discovered as part of a systematic study. However, there is another type of independence that relates to the logical sorting of characters and character states. In our example, to treat red and blue as separate, and therefore putatively independent presence/absence characters, is to ignore the fact that they are logically related ("dependent") as attributes of the same thing, i.e. tail colour. We conclude, that to code variables (character states) in a single column is to pursue the conjecture based on anatomical observation that the two variables are not independent.

Hierarchy

Primary homology conjectures imply notions about the relations (homology) between character states that are independent of character congruence. This is the nub of the issue discussed here: primary homology assessment as relation. Platnick (1979) distinguished the additive relation from the non-additive relation. He was concerned that it was problematic to use the terms character and character state because "the concept of a character state is potentially misleading. To view some character X as being composed of three states implies that the character states are alternatives, when they are actually additions" (Platnick, 1979: 543). In this context additivity is concerned with character state hierarchy, and should not be confused with ordering of character states. We believe that the cladistic practice of coding a character as states in a single column adequately deals with Platnick's concern for treating variables as having hierarchical or additive relationship. The conceptualization of a binary primary homology statement is equivalent to the imposition of an hierarchical relationship.

Returning to the data on red and blue tails, we have shown above that there is only one theoretically valid way to code these data, i.e. as columns 1 and 2 (Analysis 1). This coding implies that redness and blueness are hierarchically related. Here we consider the sets of trees discovered in all analyses in terms of the character state relationships inherent within them. Two schemes of relationship are compatible with our argument. When blue is plesiomorphic with the red homologue forming a subgroup (E,F(C,D)), or red is plesiomorphic with the blue homologue forming a subgroup (C,D(E,F)) redness and blueness are hierarchical, i.e. the character states of tail colour (red and blue) are not alternatives, but modifications one of another. A third arrangement is apparent amongst the set of trees discovered by the analyses, such that ((C,D)(E,F)), with the character states of tail colour treated not as modifications but as alternatives.

One of the cladograms (Fig. 1C), which appears under all three sets of codings, is not compatible with our primary homology assessment. In the case of Analysis 1, the solution is possible as Hennig86 treats node 8 as equivocal for character 2 and nodes 6 and 7 as unequivocal. In the case of Analysis 2, once again node 8 has equivocal optimization but this time, as it is a multistate character, node 8 can be optimized as 0, 1 or 2. If the node is optimized as a 0 then this interpretation is that the group consisting of (C,D,E,F) has neither tails nor tail colour, when in fact those taxa have both. The alternative optimizations as being 1 or 2 are equally problematic because they result in either node 6 or 7 being a zero-length branch. It is not surprising that Analysis 3 discovers the solution (Fig. 1C) which shows a non-additive relation, since treating variables as separate and independent characters is misleading as it presumes that any one character may be transformed into any other (Pimental and Riggins, 1987). A second spurious solution (Fig. 1D) is discovered only by Analysis 2. In this case, multistate character coding results in more solutions, and unnecessary loss of resolution. This solution (Fig. 1D) not only shows that red and blue are independent of each other, but also tails (as red or blue) appear twice. This is problematic. The usual view in cladistics is that the evidence for multiple origin (homoplasy) stems from incongruence with other data. However, in the case of Analysis 2 it is inappropriate primary homology assessment alone that allows tails to appear twice. There are no other data;

the presence or absence of a tail is an observed fact not reflected in the cladogram shown in Fig. 1D.

In the example worked in this paper the data matrix was chosen deliberately as it contains missing values and results in some cladograms with zero-length branches. Both these issues (missing values and zero-length branches) have been discussed recently in the literature (Platnick et al., 1991; Scotland and Williams, 1993; Maddison, 1993; Coddington and Scharff, 1995; Wilkinson, 1995b). We chose to explore Maddison's (1993) example here because we believe that primary homology assessment is directly relevant; those pursuing solutions to the inapplicable data problem need to consider the theoretical basis of primary homology assessment. Only one of the two approaches to coding inapplicable data outlined by Maddison (1993) is justified by primary homology. Smith's (1994) suggestion of recasting character definition is not a valid option.

The most theoretically robust approach to the inapplicable data problem uses two characters. However, this approach entails the use of question marks. Equivocal optimization of missing data allows the primary homology assessment to be overturned such that blue and red tails can be equally parsimoniously treated as alternatives. This means that cladistic solutions which have equivocal optimizations cannot reflect the view that character states are modifications (with hierarchical relationship) and not alternatives. There can be no real solution to the treatment of inapplicable data until new algorithms are available which take account of the problem.

DISCUSSION AND CONCLUSIONS

The conclusions we present here complement those of Brower and Schawaroch (1996), who distinguished character definition and character state delimitation as two distinct processes which together comprise primary homology assessment. We have reiterated the distinction between characters as comparable categories and character states as hypotheses of grouping, and emphasized the importance of transformational homology, evidenced by topographic correspondence, in the formulation of primary homology statements. Finally we seek to dispel the confusion surrounding the use of the terms "character" and "character state", and to clarify the meaning of transformational homology in the context of primary homology assessment.

Much of the confusion surrounding the correct use of the terms "character" and "character state" can be attributed to the use of the terms simultaneously in two different contexts. On one hand, the terminology is concerned with the data matrix, and the terms used to describe the column (character) and individual entries within a column (character state). On the other hand, it is concerned with the cladogram, and the term used to describe shared attributes which characterize groups. Hennig (1966: 89) alluded to this conflict when he wrote

> "They [character conditions=states] are "characters" in the sense that they distinguish their bearers from one another, but we must always be aware of the fact that "characters" that can be compared are basically only character conditions…produced by transformation."

Patterson (1988) argued that both characters and their states represent characters. The key to Patterson's (1988) argument is that a character, e.g. flower, can delimit a group which includes the group delimited by a state of that character, e.g. red flower. Although we agree that red-flowered organisms may comprise a subset of organisms having flowers (just as organisms with spiral cochlea may comprise a subset of organisms having cochlea) we consider "red" and "blue" to be states of the character "flower colour" and view a second character "flower" as comprising states "present" or "absent". The group of organisms with flowers is the same group as organisms with the most plesiomorphic of flower colour states (Fig. 2), but this is a special case because one of the characters is a presence-absence character, and the second describes states of presence. There is no "great chain of characters" (Platnick, 1979: 543), and the notion of characters and character states as fundamentally the same is wrong. Patterson's (1988) argument is misleading if taken to mean that characters (column descriptors) and character states (entries in the data matrix) both become characters (shared attributes which distinguish groups). We suggest that it is more useful to consider characters as column descriptors than as shared attributes which define groups for two reasons. Firstly, once the data matrix is constructed it is character states and not characters that, through the application of parsimony, provide resolution of the problem at hand. Secondly, the term synapomorphy adequately

describes "characters" that "distinguish their bearers from one another" (Hennig, 1966: 89), without causing conceptual problems. We conclude that the distinction between character and character state is of critical importance for the construction of a cladistic data matrix.

In widely accepted terminology, characters are columns and character states are the variation within columns such that primary homology assessment constitutes a whole (character) with parts (character states). It is not obvious why this terminology should not continue to be used. However, it is important to clarify that a character is operationally nothing more than the descriptive label which unites a set of character states, in line with Jardine's (1969) original view of characters as *nouns* and character states as *adjectives.* We propose that the term character thus be defined, in a cladistic context, as the descriptive label referring to a transformational homology evidenced by the similarity criterion.

The recognition that identification of empirical patterns constitutes a study prior to and independent of theories of process is central to pattern cladistics (Brady, 1985). Brady (1994) discussed the meaning of "transformation" in the context of pattern cladistics, and distinguished between perception and explanation. Our interpretation of Brady's reasoning is that it is crucial to distinguish historical explanations of transformation, such as ancestral red tails transformed into derived blue tails, from perceptions based upon the *principe des connexions* (Geoffroy Saint-Hilaire, 1830) which serve to establish relations of similarity on the basis of topological criteria. Nelson (1994: 127) argued:

> "Characters seen as part of the same transformation series were claimed by early cladists to be homologous, and it is now evident that the claim is defective, for its construes the transformation series as "fins"–"arms"–wings, that is as including symplesiomorphy. If the series is construed as fins–arms–wings then the meaning of the claim is trivial. The series is merely a more complicated way of saying that wings are arms and that arms are fins (wings=arms=fins)—in other words, characters considered homologous (part of the same transformation series) are homologous."

We agree with Nelson (1994) that wings=arms=fins, united by the conditional phrase paired appendages, and note that the primary homology of fins, arms and wings is a perception of comparative anatomy based upon topographic relations, compositional and ontogenetic similarity, independent of any claim of historical transformation involving symplesiomorphy. Thus, although transformational homology is central
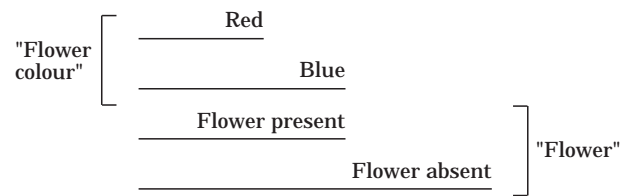

FIG. 2. The character "flower colour" has states "red" and "blue"; the character "flower" has states "flower present" and "flower absent". It is wrong to equate the character "flower" with the character state "flower present" although the presence of flowers is at the same level of generality as the most plesiomorphic of flower colour states.

to the conventional definition of a character, the definition is not incompatible with the precepts of pattern cladistics. Our view of a primary homology statement is that it is equivalent to a statement of transformational homology supported by the similarity criterion.

Similar to Nelson's (1994) claim that fins=arms =wings as paired appendages, it follows that red=blue as tail colour and that adenine=cytosine=guanine =thymine as bases of nucleic acid. Attempts to reduce all hypotheses of primary homology to being directly equivalent to hypotheses of synapomorphy by coding and treating the presence and absence of each homologue as a column in a matrix (Pleijel, 1995) reduces cladistics to congruence alone. To allow primary homology decisions to emerge as part of these results is to retreat from the vital task of primary homology assessment, and tends towards operationalism. Any retreat from the fundamentally important task of homology assessment is a retreat from theory.

ACKNOWLEDGEMENTS

REFERENCES

Ax, P. (1987). "The Phylogenetic System". John Wiley and Sons, Chichester.

Bateman, R. M., DiMichele, W. A., and Willard, D. A. (1992). Experimental cladistic analysis of anatomically preserved arborescent Lycopsids from the Carboniferous of Eurasia, an essay on palaeobotanical phylogenetics. *Ann. Missouri Bot. Garden* **79**, 500–559.

Bock, W. J. (1969). Comparative morphology in systematics. *In* "Systematic Biology, Proceedings of an International Conference Conducted at the University of Michigan, Ann Arbour, Michigan, June 14–16, 1967", pp. 411–448. National Academy of Sciences, Washington.

Bock, W. J. (1974). Philosophical foundations of classical evolutionary classification. *Syst. Zool.* **22**, 375–392.

Brady, R. H. (1983). Parsimony, hierarchy and biological implications. *In* "Proceedings of the Second Meeting of the Willi Hennig Society. Advances in Cladistics 2" (N. Platnick, and V. Funk, Eds), pp. 49–59. Columbia University Press, New York.

Brady, R. H. (1985). On the independence of systematics. *Cladistics* **1**, 113–126.

Brady, R. H. (1994). Explanation, description, and the meaning of "transformation" in taxonomic evidence. *In* "Models in Phylogeny Construction. Systematics Association Special Volume 52" (R. W. Scotland, D. J. Siebert, and D. M. Williams, Eds), pp. 11–30. Clarendon Press, Oxford.

Brower, A. V. Z., and Schawaroch, V. (1996). Three steps of homology assessment. *Cladistics* **12**, 265–272.

Bryant, H. N. (1989). An evaluation of cladistic and character analyses as hypothetico-deductive procedures, and the consequences for character weighting. *Syst. Zool.* **38**, 214–227.

Coddington, J., and Scharf, N. (1994). Problems with zero-length branches. *Cladistics* **10**, 415–423.

de Pinna, M. C. C. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**, 367–394.

Farris, J. S. (1983). The logical basis of phylogenetic analysis. *In* "Proceedings of the Second Meeting of the Willi Hennig Society. Advances in Cladistics 2" (N. Platnick, and V. Funk, Eds), pp. 7–36. Columbia University Press, New York.

Felsenstein, J. (1982). Numerical methods for inferring evolutionary trees. *Quart. Rev. Biol.* **57**, 379–404.

Geoffroy Saint-Hilaire, E. (1830). "Principes de Philosophie Zoologique, discutés en Mars 1830, au Sein de l'Académie Royale des Sciences". Pichon et Dider, Paris.

Goloboff, P. A. (1995). Parsimony and weighting: A reply to Turner and Zandee. *Cladistics* **11**, 91–104.

Grimes, J. (1995). Generic relationships of Mimosoideae tribe Ingeae, with emphasis on the New World *Pithecellobium* complex. *In* "Advances in Legume Systematics. Part 7. Phylogeny" (M. D. Crisp, and J. J. Doyle, Eds), pp. 101–121. Royal Botanic Gardens, Kew.

Hennig, W. (1966). "Phylogenetic Systematics". University of Illinois Press, Urbana, Illinois.

Jardine, N. (1969). The observational and theoretical components of homology: A study on the morphology of the dermal skull-roofs of rhipidistian fishes. *Biol. J. Linnean Soc.* **1**, 327–361.

Maddison, W. P. (1993). Missing data versus missing characters in phylogenetic analysis. *Syst. Biol.* **42**, 576–581.

Nelson, G. J. (1994). Homology and systematics. *In* "Homology, the Hierarchical Basis of Comparative Biology" (B. K. Hall, Ed.), pp. 101–149. Academic Press, San Diego.

Patterson, C. (1982). Morphological characters and homology. *In* "Problems of Phylogenetic Reconstruction" (K. A. Joysey, and A. E. Friday, Eds), pp. 21–74. Academic Press, London and New York.

Patterson, C. (1988). Homology in classical and molecular biology. *Mol. Biol. Evol.* **5**, 603–625.

Pimentel, R. A., and Riggins, R. (1987). The nature of cladistic data. *Cladistics* **3**, 201–209.

Platnick, N. J. (1979). Philosophy and the transformation of cladistics. *Syst. Zool.* **28**, 537–546.

Platnick, N. I., Griswold, C. E., and Coddington, J. A. (1991). On missing entries in cladistic analysis. *Cladistics* **7**, 337–343.

Pleijel, F. (1995). On character coding for phylogeny reconstruction. *Cladistics* **11**, 309–315.

Pogue, M. G., and Mickevich, M. F. (1990). Character definitions and character state delineation: The bête-noire of phylogenetic inference. *Cladistics* **6**, 319–361.

Rieppel, O. (1988). "Fundamentals of Comparative Biology". Birkhauser Verlag, Berlin.

Riggins, R., and Farris, J. S. (1983). Cladistics and the roots of angiosperms. *Syst. Bot.* **8**, 96–101.

Schoch, R. M. (1986). "Phylogeny Reconstruction in Paleontology". Van Norstrand-Reinhold, New York.

Scotland, R. W., and Williams, D. M. (1993). Multistate characters and cladograms: When are two stamens more similar to three stamens than to four? A reply to Lipscomb. *Cladistics* **9**, 343–350.

Smith, A. (1994). "Systematics and the Fossil Record: Documenting Evolutionary Pathways". Blackwell Scientific, Oxford.

Stevens, P. F. (1991). Character states, morphological variation and phylogenetic analysis: A review. *Syst. Bot.* **16**, 553–583.

Wiley, E. O. (1981). "Phylogenetics. The Theory and Practice of Phylogenetic Systematics". John Wiley and Sons, New York.

Wilkinson, M. (1995a). A comparison of two methods of character construction. *Cladistics* **11**, 297–308.

Wilkinson, M. (1995b). Arbitrary resolutions, missing entries, and the problem of zero-length branches in parsimony analysis. *Syst. Biol.* **44**, 108–111.