

How to Determine the Compatibility of Undirected Character State Trees

GEORGE F. ESTABROOK AND CHRISTOPHER A. MEACHAM

*Department of Botany, Division of Biological Sciences,
University of Michigan, Ann Arbor, Michigan 48109*

Received 4 January 1979; revised 18 April 1979

ABSTRACT

Methods for cladistic character compatibility analysis have required *a priori* estimates of the direction of evolutionary trends among the states of a character, in order to ensure consistency. Results are presented here that can enable comparative biologists, if they choose, to proceed with cladistic character analysis without *a priori* estimates of the direction of evolutionary trends. Specifying only undirected character state trees is sufficient to ensure the compatibility of a collection of characters each pair of which is compatible. The results have been implemented in a computer program, for which a FORTRAN source deck is available.

I. REVIEW OF BASIC CONCEPTS

An important aspect of modern systematics is the construction of an estimate of evolutionary relatedness among the kinds of organisms under study. Such an estimate often takes the form of a directed tree partial ordering of the collection of kinds of organisms, enlarged to include hypothetical ancestral forms when necessary. Such estimates are often based on qualitative characters, which are equivalence relations on the collection of kinds of organisms. An equivalence class of a qualitative character is called a character state. Thus, the kinds of organisms constituting a character state are all the same with respect to the basis for comparison that served to define the character. The problems and procedures associated with the choice of kinds of organisms to make up the study collection, and with the choice and construction of characters to reflect the similarities and differences among these kinds of organisms, were reviewed by Estabrook in 1972 [3].

One procedure that can help with the construction of an estimate of evolutionary relatedness is to make for each character a partial estimate of

evolutionary relatedness, and then to test each pair of these partial estimates to see if they are mutually contradictory or not. It is to this procedure that the results we present below contribute. To follow is a brief, informal review of the concepts basic to this procedure. The reader who wishes a more detailed, formal exposition of these concepts is directed to [6], [7], [8], [10], and [14]. A qualitative character is converted to a partial estimate of evolutionary relatedness by specifying a tree partial order for its character states. A qualitative character together with a tree partial ordering of its states is called a cladistic character. It is only a partial estimate because details of evolutionary relatedness among the kinds of organisms in any single state are left unresolved. A cladistic character is consistent with any complete estimate that can be derived from it by resolving the details of evolutionary relatedness within its character states. There is a homomorphism (preserving order) from a complete estimate to any cladistic character that is consistent with it (Estabrook et al., 1975 [6]). Strauch in 1978 [15] presented some examples that effectively illustrate the relationship between a complete estimate and a cladistic character that is consistent with it.

Two cladistic characters are said to be compatible if there exists a complete estimate with which each is consistent. Many writers have contributed to the evolution of this concept: Wilson (1964) [16]; Camin and Sokal (1965) [1]; Le Quesne (1969) [13]; Estabrook (1978) [4]. At most one of the possible complete estimates of evolutionary relatedness among the kinds of organisms under study is historically correct. Thus, we can conclude that at least one and possibly both of two incompatible characters are inconsistent with the truth and would be misleading if used as a basis for constructing an estimate of evolutionary relatedness. Any two characters that were consistent with the truth would, of course, be compatible.

It has been shown that for any collection of cladistic characters in which every pair is compatible there is at least one estimate of evolutionary relatedness with which every cladistic character in the collection is consistent (Estabrook et al., 1976 [8]). Such a compatible collection of cladistic characters can be used to determine a more refined partial estimate (or in some cases a complete estimate) of evolutionary relatedness. Sometimes the estimate of evolutionary relatedness that has the largest collection of cladistic characters consistent with it is considered the most credible (Estabrook and Anderson, 1979 [5]; Estabrook et al., 1977 [11]; Cichocki, 1977 [2]; Strauch, 1978 [15]).

There are three levels at which two cladistic characters can fail to be compatible: the most severe is a consequence of the character state memberships; the next most severe is a consequence of the hypothesized evolutionary proximity of the character states; and the least severe is a consequence of the hypothesized directions of evolutionary trends. These three concepts are made more precise, and a means of avoiding the least severe source of incompatibility is described for those who choose to use it.

II. NEW RESULTS

DEFINITION

Two qualitative characters are defined to be *potentially compatible* if there exist a tree partial ordering for the states of each such that as cladistic characters with these character state trees, they are compatible.

Characters that are not potentially compatible are incompatible as a consequence of the character state memberships. Estabrook and Landrum (1975) [9], and Fitch (1975) [12] independently conjectured operational criteria to test for potential compatibility. Estabrook and McMorris (1977) [10] showed that these conjectures are mathematically correct, and provided a computer algorithm to implement the test. Fitch (1975) [12], with an elegant example, has shown that a collection of pairwise potentially compatible characters is not necessarily a potentially compatible collection of characters, i.e., even though every pair of characters in a collection is potentially compatible, there still may not exist an estimate of evolutionary relatedness with which all the characters in the collection are consistent. Thus, if we wish to preserve collection wide compatibility as a consequence of pairwise compatibility of all pairs in the collection, a condition stronger than potential compatibility is needed.

DEFINITION

Two cladistic characters are defined to be *undirected tree equivalent* (u.t. equivalent) if one can be converted to the other by reversing some of the orderings in the character state tree, i.e., by changing the estimated directions of evolutionary changes, but not the proximities of character states.

Clearly u.t. equivalent is an equivalence whose classes are those cladistic characters that differ only in the hypothesized direction of evolutionary trends.

DEFINITION

Two *u.t. equivalence classes* are defined to be *compatible* if some member of one is compatible with some member of the other.

DEFINITION

In a character state tree, *one state* is defined to *cover another* if it is a next more advanced state.

DEFINITION

For every cover (pair of states in which one covers the other), the kinds of organisms in the advanced state, or in any state derived from the advanced state, taken together, comprise the kinds of organisms that are defined to be *advanced for that cover*; the rest comprise the kinds of organisms that are defined to be *primitive for that cover*.

DEFINITION

A cladistic character is defined to be "*common-equal-primitive*" if for every cover in the character state tree the number of kinds of organisms that are primitive is never less than the number that are advanced.

These concepts are illustrated in Part III, "Example."

THEOREM

Every u.t. equivalence class contains a "common equal primitive" character.

Proof. Consider any member of an equivalence class and, if it is not "common-equal-primitive," reverse the orderings of those covers for which the number of kinds that are advanced exceeds the number of kinds that are primitive. If this new ordering gives a tree, then this is a common-equal-primitive member of the class. The new ordering could fail to give a tree only if thereby some state Y was considered next after each of two different states, X and Z , i.e., $X \rightarrow Y$ or $Z \rightarrow Y$. Since the rearrangements began with a cladistic character, one of $X \rightarrow Y$ or $Z \rightarrow Y$ must have been reversed. Suppose, without loss of generality, that $X \rightarrow Y$ was reversed. Then formerly the number of kinds of organisms in state X or in states derived from state X must have exceeded half the total number of kinds. But in that case, $Z \rightarrow Y$ is not common-equal-primitive. ■

The collection of two-state cladistic characters with primitive and advanced states determined by each of the covers in a given cladistic character, can be combined to reconstruct that cladistic character, and in this sense such a collection of two-state cladistic characters is known as the binary factors of the given cladistic character. Estabrook et al. (1976) [7] have shown that two cladistic characters are compatible if and only if every binary factor of one is compatible with every binary factor of the other.

THEOREM

Two u.t. equivalence classes are compatible if and only if a common-equal-primitive member of one is compatible with a common-equal-primitive member of the other.

Proof. Only the forward implication needs proof. Observe that two cladistic characters in the same u.t. equivalence class differ in their binary factors only in that the direction of evolutionary trends in *some* binary factors may be the opposite of that in *others*. McMorris (1977) [14] has shown that if binary cladistic characters are compatible, then they are also compatible when reversed, if necessary, so that they become common-equal-primitive. McMorris's result can be applied to the binary factors of each of two compatible characters from different u.t. equivalence classes, to

provide the conclusion that some common-equal-primitive representatives of these u.t. equivalence classes must also be compatible. ■

In practice this result is applied in its logically equivalent negative form: if the common-equal-primitive members of two u.t. equivalence classes are incompatible, then so are the u.t. equivalence classes, for we know that no other pair of cladistic characters, one chosen from each class, can be compatible. In this sense the common-equal-primitive members of u.t. equivalence classes are representative of the whole class.

Those workers who wish to make no assumptions about directions of evolutionary trends prior to an analysis of cladistic character compatibility, may do so simply by using common-equal-primitive character state trees. Kent Fiala has implemented an option in the computer program CLINCH to automatically convert character state trees to common-equal-primitive if desired. A FORTRAN source deck of CLINCH is available from the authors.

III. EXAMPLES OF CONCEPTS

Designate the kinds of organisms under study with the letters *a, b, c, d, e, f, g*. A qualitative character would be



A cladistic character would be as shown in Fig. 1(a). According to this character (abc) is considered primitive; (de) evolved next from it; and (fg) then evolved from (de) .

The pair (fg) (de) is a cover in this tree; (fg) is advanced, and $(abcde)$ is primitive for this cover.

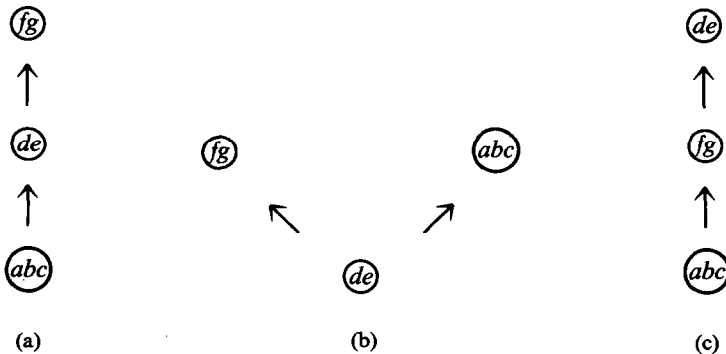


FIG. 1.

This character is not common-equal-primitive, but the character shown in Fig. 1(b) is common-equal-primitive and belongs to the same u.t. equivalence class. The pair \textcircled{abc} \textcircled{de} is a cover in this tree; its primitive class is

\textcircled{defg}

The cladistic character shown in Fig. 1(c) is *not* in the same u.t. equivalence class with the previous two characters.

REFERENCES

- 1 J. H. Camin and R. R. Sokal, A method for deducing branching sequences in phylogeny, *Evolution* 19:311–326 (1965).
- 2 F. G. Cichocki, Cladistic history of cichlid fishes and reproductive strategies of the American genera *Acarichthys*, *Biotodoma*, and *Geophagus*, Vol. I, Ph.D. Thesis, Univ. of Michigan, Ann Arbor, Mich., 1976.
- 3 G. F. Estabrook, Cladistic methodology: a discussion of the theoretical basis for the induction of evolutionary history, *Ann. Rev. Ecol. Syst.* 3:427–456 (1972).
- 4 G. F. Estabrook, Some concepts for the estimation of evolutionary relationships in systematic botany, *Systematic Botany* 3:146–158 (1978).
- 5 G. F. Estabrook and W. R. Anderson, An estimate of phylogenetic relationships within the genus *Crusea* (Rubiaceae) using character compatibility analysis, *Systematic Botany* 3:179–196 (1978).
- 6 G. F. Estabrook, C. S. Johnson, Jr. and F. R. McMorris, An idealized concept of the true cladistic character, *Math. Biosci.* 23:263–272 (1975).
- 7 G. F. Estabrook, C. S. Johnson, Jr. and F. R. McMorris, A mathematical foundation for the analysis of cladistic character compatibility, *Math. Biosci.* 29:181–187 (1976).
- 8 G. F. Estabrook, C. S. Johnson, Jr. and F. R. McMorris, An algebraic analysis of cladistic characters, *Discrete Math.* 16:141–147 (1976).
- 9 G. F. Estabrook and L. Landrum, A simple test for the possible simultaneous divergence of two amino acid positions, *Taxon.* 25:609–613 (1975).
- 10 G. F. Estabrook and F. R. McMorris, When are two qualitative taxonomic characters compatible?, *J. Math. Biology* 4:195–200 (1977).
- 11 G. F. Estabrook, J. G. Strauch, Jr. and K. L. Fiala, An application of compatibility analysis to the Blackiths' data on orthopteroid insects, *Syst. Zool.* 26:269–276 (1977).
- 12 W. M. Fitch, Relationship between prim networks and trees of maximum parsimony, in *Proceedings of the Eighth International Conference on Numerical Taxonomy*, (G. F. Estabrook, Ed.) Freeman, San Francisco, 1975.
- 13 W. J. LeQuesne, A method of selection of characters in numerical taxonomy, *Syst. Zool.* 18:201–205 (1969).
- 14 F. R. McMorris, On the compatibility of binary qualitative taxonomic characters. *Bull. Math. Biol.* 39:133–138 (1977).
- 15 J. G. Strauch, Jr., The phylogeny of the Charadriiformes (Aves): a new estimate using the method of character compatibility analysis, *Trans. Zool. Soc. Lond.* 34:263–345 (1978).
- 16 E. O. Wilson, A consistency test for phylogenies based on contemporaneous species, *Syst. Zool.* 14:214–220 (1965).