

An Idealized Concept of the True Cladistic Character

G. F. ESTABROOK

Department of Botany, University of Michigan, Ann Arbor, Michigan

AND

C. S. JOHNSON, JR.

Department of Mathematics, Bowling Green State University, Bowling Green, Ohio

AND

F. R. MC MORRIS

Department of Mathematics, Bowling Green State University, Bowling Green, Ohio

Communicated by Stan Ulam

ABSTRACT

Estimating evolutionary relationships is basic to the objectives of systematics. Comparative data, structured as taxonomic characters, are usually the essential considerations on which such estimates are founded. Some taxonomic characters are more useful than others for structuring plausible estimates of evolutionary relationship. Thus, one of the primary challenges to the systematist is the construction of taxonomic characters most useful for this purpose. Since taxonomic characters are the result of action on the part of the systematist, they must be defined operationally. It is our hope, however, that these operationally defined characters will conform to an ideal that, itself, cannot be operationally defined insofar as the concept depends on history that is, usually, inherently unknowable. It is, nonetheless, essential to our conceptual methods that this ideal concept be well defined. Here we present a series of definitions leading to a clear ideal concept of *true cladistic character*. This series includes definitions of the concepts: evolutionary unit, qualitative taxonomic character, monophyletic group, divergent character, true cladistic character, operational cladistic character, and the post-factum ideal relation between an operational cladistic character and an estimate of cladistic history. A concise characterization of true cladistic character is presented and proved.

I. INTRODUCTION

Curiosity about the past has stimulated the speculations of critical thinkers in a wide variety of disciplines. Curiosity about the history of the evolution of the earth, and in particular the organisms that inhabit the

earth, continues to be one of the basic driving forces in biological research. The field of systematics includes many biological researchers whose primary objective is to make a plausible speculation about the evolutionary history of some particular group of organisms. How one can induce, from whatever might be known about a group of organisms, a plausible speculation about the evolutionary history of that group comprises an important part of the theoretical methods of systematics (Estabrook [3]). It is intended that the results reported here will strengthen these theoretical methods.

II. IDEALIZED CONCEPTS AND OPERATIONISM

It is important to recognize and maintain the distinction between an idealized concept and its operational interpretation (Sokal and Camin [16], Sneath and Sokal [14, pp. 17–18]). We practice empirical science by doing. Thus, when reporting empirical results it is important to describe what was done, as well as the conclusions that were drawn (Estabrook [4]; but also see Hull [9] for a more extensive discussion). Many workers strongly advocate operational descriptions of what was done. These describe the actual operations that were performed by the investigator to carry out his empiricism. Such descriptions facilitate constructive criticism of the results by other workers, and make the data more valuable for testing hypotheses other than the one for which they were gathered. However, it is important to realize in the context of our interest in history, that we rarely, if ever, observe or measure history directly. All our measurements, even those of fossils, are made, relatively speaking, in the present. Thus we are presented with the problem of defining concepts whose existence is predicated on the past, speculations on the precise form of which will be the objective of our empiricism (Hull [8]). These concepts do not admit of operational definitions: there are no operations we can perform other than in the present. Such definitions will be made presupposing a knowledge of the very thing we are trying to estimate, and thus the concepts so defined will be called *idealized concepts*.

III. IDEALIZED EVOLUTIONARY UNITS

We will employ this approach in defining an evolutionary unit (EU). The precedented candidate for evolutionary unit is the biological species, or a monophyletic group of biological species. How best to define a biological species is a point of some disagreement among systematists, but fundamental to virtually all versions of the concept is the idea that two members of the same species can exchange “evolutionarily significant” amounts of genetic information, while members of different species cannot (Simpson [13], Mayr [11]). Thus the relation “breeds with” corresponds in essence

with the biological species concept. Many workers realize (see e.g. Sokal and Crovello [17]) that this does not operationally define the concept biological species. But even if we admit the concept of the biological species as a useful idealized concept there are other difficulties with it. A natural step would be to define a biological species as an equivalence class of the relation "breeds with". This would work if the relation "breeds with" were an equivalence relation. The difficulties come with transitivity (Simpson [14]). For many groups of organisms that we have recognized and now recognize as species insofar as they bear Latin binomials, it is necessary to ignore the fact that some members of the same species cannot breed, or may in fact breed with members of other species. Furthermore, there seems to be no way to define the relation "breeds with" with sufficient care that its general application will determine for us, in all cases, groups we wish to recognize as species. Application of a biological species concept becomes especially difficult in the spatial and temporal vicinity of a cladistic event, the splitting of one phyletic line into two or more lines.

It is not essential to construe every organism that ever lived as a member of an EU. Members alive during the phyletic evolution of an EU, i.e., the continuous change in its genetic construction through time while the integrity of the EU is maintained by the absence of cladistic events, are easily construed as belonging to a group with a single evolutionary fate. Organisms alive in the vicinity of a cladistic event may not be so easily construed. We will simply exclude them from membership in any evolutionary unit, thus giving us discrete EU's.

Many biologists feel that, at least for a large number of taxa, the branching pattern of the phyletic lines can be adequately described as a tree, i.e., each EU was derived from a unique immediately ancestral EU (Hennig [7]). Such feelings minimize the importance of hybridization and other mechanisms for mixing genes from distinct phyletic lines into a new EU during the evolution of such groups (Wagner [19]). Is it true that a collection of evolutionary units is a collection of disjoint groups of organisms for which the cladistic history, i.e., a description of the relationships among the cladistic events that derived the collection from some common ancestor, can be represented with a tree diagram? An answer to this question is important before we endeavor to make estimates of cladistic history, in order to be aware of the basic form our estimates are to take. Clearly the answer to this question depends, in part, on how we choose to define a collection of EU's. In order to ensure an affirmative answer, we present a definition of the idealized concept.

First we need to recall the following definitions.

DEFINITION. A *tree* (lower) *semilattice* is a set T together with a partial order \leq such that $a \wedge b = \text{g.l.b. } \{a, b\}$ exists for all $a, b \in T$ and satisfies the

condition that $a \leq c$ and $b \leq c$ imply $b \leq a$ or $a \leq b$.

DEFINITION. A collection of evolutionary units is a collection S' of disjoint groups of organisms alive at any time for which the cladistic history is a tree lower semilattice on S' with \wedge (meet) representing "most recent common ancestor".

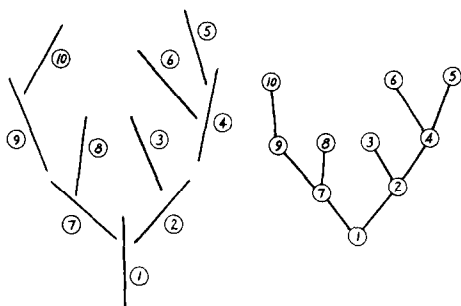
DEFINITION. An evolutionary unit is a member of the collection of evolutionary units under study.

DEFINITION. The "ancestor" relation \mathbf{A} on S' is defined in the usual way with $a\mathbf{A}b$ if and only if $a = a \wedge b$.

The Hasse diagram of \mathbf{A} is a directed tree graph. The relationship between the branching pattern of the phyletic lines and the corresponding Hasse diagram is shown in Fig. 1. This definition of evolutionary unit is an idealized concept because it is predicated on the existence of an evolutionary history of a certain form, which existence can never be operationally determined. However, in order to conduct a study for the purpose of making an estimate of cladistic history, organisms assumed to represent observable evolutionary units are measured and compared. We must have *done* something to determine what the source of data is to be. An operational interpretation in action of the idealized concept is made in order to proceed with the study. An operational description of this interpretive action is an important inclusion in the published report. How operational interpretations are made is an important component of the artistry inherent in successful empiricism. It is not our intent to discuss this aspect of systematic method here. We will proceed with the confidence that a firmer grasp of essential idealized concepts contributes to more effective operational procedures, whatever other considerations a worker might also choose to invoke in devising those judged appropriate to his problem. Hereafter, we will assume that an operational collection of observable evolutionary units has been established by the systematist.

IV. QUALITATIVE TAXONOMIC CHARACTERS

The practice of taxonomy generally proceeds through the recognition of bases for comparing observable operational evolutionary units. Let us denote by S this collection of operational evolutionary units. We assume that S is contained in some S' , so that $S' - S$ will be the ancestral evolutionary units derived from $\wedge S = \wedge S'$ but having no observable representatives in S . Thus the study collection S is an operational concept, while S' (including unobservable ancestors) is an idealized concept. To serve us effectively as a qualitative taxonomic character, a basis for comparison must consistently determine whether any pair of EU's from S is the same



with respect to that basis (Estabrook and Rogers [5]). With respect to an effective basis, an EU should be considered

- (i) the same as itself (reflexive);
- (ii) the same as a second EU if that second EU is the same as it (symmetric);
- (iii) the same as a second EU if there is a third EU the same as each (transitive).

Thus we make the following definition.

DEFINITION. A qualitative taxonomic character for a study collection S of observable evolutionary units is an equivalence relation K on S . We will call an equivalence class of K a *character state* of the character K , and for $a \in S$ we will denote by $K(a)$ the equivalence class of K to which a belongs.

Note that the concept of a qualitative taxonomic character is operational: presented with an arbitrary descriptive construct for S , it can be determined whether that construct is a qualitative taxonomic character or not. Examples of a qualitative taxonomic character are easily made, and the following two examples are from Takhtajan [17]. Suppose S is a collection of putative plant species. K_1 might be a type of wood parenchyma (storage cells); we might produce three equivalence classes from considerations of similar and different types of wood parenchyma: class A with diffuse parenchyma cells, class B with bands of parenchyma cells, and class C with parenchyma cells surrounding the vascular tissue. K_2 might be the number of gaps through which pass strands of vascular tissue at the leaf nodes; we might recognize three classes of plant species equivalent with respect to this basis for comparison: class A of plants with a single gap, class B of plants with three gaps, and class C of plants with many gaps. Suppose S contained only EU's maximal in the lower semilattice, (S', \wedge) , that represents the true

cladistic history. This is in accord with Hennig's 1966 preference for classifying only contemporaneous taxa. What, then, should be the ideal relationship between a qualitative taxonomic character K , and (S', \wedge) , the true cladistic history for S , if K is to be most useful as a basis for estimating true cladistic history? To discuss this question further we need the following definition.

DEFINITION. A set $M \subseteq S$ is *monophyletic* if and only if $(\wedge M)\mathbf{A}a$ for $a \in S$ implies $a \in M$.

This definition is an idealized concept, as it depends on knowledge of "recent common ancestor", \wedge . If all the states of K were monophyletic then K would be most useful for estimating cladistic history; we would know that two members of a given character state of K have a more recent common ancestry than either has with a nonmember. It is unnecessarily restrictive, however, to require that all the states of K be monophyletic, for there is a much more general class of characters that are equally useful for estimating true cladistic history. Moreover, there is some evidence to suggest that evolution proceeds in such a way that virtually none of the observable results of the processes could be expected to admit, in a natural way, of characters all of whose states are monophyletic. Here, we define our concept of paraphyly (divergence).

DEFINITION. The states of a qualitative taxonomic character K for a collection S of operational EU's are *paraphyletic* if $(\wedge K(a))\mathbf{A}(\wedge K(b))\mathbf{A}a$ implies $K(a) = K(b)$. If K has paraphyletic states, we say that K is *divergent*.

This definition is an idealized concept. As defined, the concept of paraphyletic does not apply to an arbitrary subset of S , but is a property enjoyed collectively by all the states of a character. It might seem natural to suggest that an arbitrary subset $X \subseteq S$ is paraphyletic if there exists a paraphyletic character K for S such that X is a state of K . However, a moment's reflection yields that this suggested definition is degenerate in that any arbitrary subset of the maximals in (S', \wedge) would be paraphyletic. It is therefore not a useful concept.* Concepts of paraphyly have been advanced by Hennig and Simpson, but there is much controversy concerning the utility of these concepts, and their implications are not well understood by many workers. It is not our intent to review here the history of these concepts. Notice that at least one state of any divergent character is monophyletic. By definition, S itself is monophyletic, and furthermore, S

*However we can define an idealized concept of paraphyly for an arbitrary subset $M \subseteq S'$ as follows. $M \subseteq S'$ is *paraphyletic* if and only if $b \in M$ and $(\wedge M)\mathbf{A}a\mathbf{A}b$ imply $a \in M$. The states of a divergent qualitative character for S can be extended to disjoint paraphyletic subsets of S' (cf. the discussion of cladistic characters in V).

will perforce contain a monophyletic group M . Now $S \setminus M$ will either be monophyletic or conform to the concept of paraphyletic in the sense of Hennig. The character states M and $S \setminus M$ will be paraphyletic in our sense. To this extent, our concept includes Hennig's as a special case.

To be most useful for estimating true cladistic history, a character K should be divergent. Divergence for a character K assumes that all the EU's belonging to a given state, $K(a)$, say, could have evolved from $\wedge K(a)$ with no change in the property characteristic of $K(a)$, and that changes in the properties characteristic of the states of K could have occurred only on phyletic lines leading directly to the recent common ancestors of the states of K themselves (LeQuesne [10]). Thus, for a divergent character, it is possible that the evolutionary changes producing the observable differences upon which K is structured, correspond uniquely to the states of K as they can be operationally observed.

V. CLADISTIC CHARACTERS

As we pointed out above, the concept of a qualitative taxonomic character is operational in that it depends only on the observable properties of the members of a collection S of operational evolutionary units. Such a character will be ideally related to the true cladistic history (S', \wedge) , if certain desirable relationships among the historically true evolutionary changes and the states of K are logical possibilities. If we include in the concept of a qualitative taxonomic character, K , estimates of the relationships among the evolutionary changes to which the states of K would ideally correspond, we will call the result a cladistic character (Camin and Sokal [1], Farris, et al. [6]). To make this concept clear, we will make the following definitions.

DEFINITION. A cladistic character on S is a map $K: S \rightarrow P$ where P is a tree semilattice.

DEFINITION. A cladistic character on S' is an onto map $K: S' \rightarrow P$, where P is a tree semilattice.

This concept is idealized. P is called the *character state tree* of K . Notice that nonvoid sets of the form $S \cap K^{-1}(K(a))$ for $a \in S$ reconstruct the states of K construed as a qualitative taxonomic character. In addition, a cladistic character may hypothesize the existence of character states (subsets of S') that have no representatives in S but that are comprised entirely of unobservable ancestral EU's. Furthermore, states with representatives in S may also contain unobservable ancestral EU's.

Thus, a cladistic character is itself a partial estimate of cladistic history, asserting, in the spirit of estimates, the possible *existence* of ancestral

conditions, i.e. character states (although the forms of these conditions, where representations of them are available for us to see, need not be specified), together with possible evolutionary relationships among those character states, expressed in the form of a tree diagram P (Estabrook [2]). What should be the ideal relationship between a cladistic character K and the true cladistic history (S', \wedge) of S , towards whose plausible estimate K is to be structured? We would suggest that there are three natural properties that should be met for K to be ideally related to (S', \wedge) .

(i) A character state should contain its own most recent common ancestor. Clearly this is desirable to eliminate the need for any evolutionary change in a character during the evolution of the members of a character state from the most recent common ancestor for that state.

(ii) If a first EU is an ancestor of a second EU, then the state of which the first is a member should be equal, or should be ancestral in the character state tree P , to the state of which the second is a member. This constraint is desirable to ensure that P recognizes enough ancestor-descendant relationships to account for the historically true course of history represented by (S', \wedge) .

(iii) Lastly, if one character state is ancestral to another in the character state tree P , then the most recent common ancestor for the one state should be ancestral in (S', \wedge) to the most recent common ancestor for the other.

This last constraint is similar to (ii) but prevents P from recognizing more ancestor-descendant relationships than true history warrants. Constraints (ii) and (iii) together insure that the *direction* of evolutionary trends suggested by P is consistent with true history.

Since these three constraints characterize a cladistic character that makes a true partial estimate of cladistic history, we will define a true cladistic character as follows.

DEFINITION. A cladistic character $K: S' \rightarrow P$ is a *true cladistic character* if, for $a, b \in S'$, the following conditions are satisfied:

- (i) $\bar{a} \in K^{-1}(K(a))$, where $\bar{a} = \wedge K^{-1}(K(a))$;
- (ii) $a \mathbf{A} b$ implies $K(a) \leq K(b)$, where \leq is the ordering in P ;
- (iii) $K(a) \leq K(b)$ implies $\bar{a} \mathbf{A} \bar{b}$.

This definition is of an idealized concept.*

VI. MATHEMATICAL CHARACTERIZATION OF TRUE CLADISTIC CHARACTER.

*Notice that a true cladistic character has states that are paraphyletic subsets of S' .

A remarkably simple mathematical characterization of the ideal concept of true cladistic character is given in the following theorem.

THEOREM

The cladistic character $K: S' \rightarrow P$ is true if and only if K is a lower semilattice homomorphism (i.e., $K(a \wedge b) = K(a) \wedge K(b)$ for all $a, b \in S'$).

Proof. Assume K is true. From (ii) we have $K(a \wedge b) \leq K(a)$ and $K(a \wedge b) \leq K(b)$. Let $K(z)$ be an arbitrary element of P below both $K(a)$ and $K(b)$. We must show $K(z) \leq K(a \wedge b)$. Now $K(z) \leq K(a)$ and $K(z) \leq K(b)$ imply $\bar{z} \mathbf{A} \bar{a}$ and $\bar{z} \mathbf{A} \bar{b}$ by (iii). Using (i) and (ii) we obtain $K(z) = K(\bar{z}) \leq K(\bar{a} \wedge \bar{b}) \leq K(a \wedge b)$. Therefore

$$K(a \wedge b) = \text{g.l.b. } \{K(a), K(b)\} = K(a) \wedge K(b).$$

For the converse, assume K is a homomorphism. Property (ii) is immediate. For (i), assume $\{z_1, \dots, z_n\} = K^{-1}(K(a))$. Then $\bar{a} = z_1 \wedge z_2 \wedge \dots \wedge z_n$ and $K(\bar{a}) = K(z_1 \wedge \dots \wedge z_n) = K(z_1) \wedge \dots \wedge K(z_n) = K(a) \wedge \dots \wedge K(a) = K(a)$. To prove (iii), let $K(a) \leq K(b)$. From (i) we have $K(\bar{a}) = K(a)$ and $K(b) = K(\bar{b})$. Thus $K(a) = K(a) \wedge K(b) = K(\bar{a}) \wedge K(\bar{b}) = K(\bar{a} \wedge \bar{b})$, which implies that $\bar{a} \wedge \bar{b} \in K^{-1}(K(a))$. Hence $\bar{a} \mathbf{A} \bar{a} \wedge \bar{b}$, since $\bar{a} = \bigwedge K^{-1}(K(a))$. Since $\bar{a} \wedge \bar{b} \mathbf{A} \bar{b}$, the proof is complete.

We remark that nowhere did we use the fact that S' and P are trees. Hence the above theorem is true for arbitrary (finite) semilattices.

One of the ultimate objectives of a systematic study of a collection S is to erect a plausible estimate (S^*, \wedge^*) of the true cladistic history (S', \wedge) of S . Of course, for this estimate, $S \subseteq S^*$. Having made this estimate we can determine in retrospect which of our cladistic characters would be true if our estimate were correct. In this spirit, we will offer two final definitions.

DEFINITION. An *operational cladistic character* is an onto map $K: S^* \rightarrow P$ where P is as before and S^* is an estimate of S' .

DEFINITION. An operational cladistic character K is *ideally related* to an estimate (S^*, \wedge^*) of the true cladistic history of S , if K is a lower semilattice homomorphism from S^* to P .

Note that both these definitions are operational.

REFERENCES

- 1 J. H. Camin and R. R. Sokal, A method for deducing branching sequences in phylogeny, *Evolution* **19**, 311–326 (1965).
- 2 G. F. Estabrook, A general solution in partial orders for the Camin-Sokal model in phylogeny, *J. Theor. Biol.* **21**, 421–438 (1968).

- 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, Theoretical methods in systematic and evolutionary studies, *Progress in Theor. Biol.* **2**, 23–86 (1972).
- 5 G. F. Estabrook and D. J. Rogers, A general method of taxonomic description for a computed similarity measure, *BioScience* **16**, 789–793 (1966).
- 6 J. S. Farris, A. G. Kluge, and M. J. Eckardt, A numerical approach to phylogenetic systematics, *Syst. Zool.* **18**, 83–92 (1970).
- 7 W. Hennig, *Phylogenetic Systematics*, University of Illinois Press, Urbana, 1966, 263 pp.
- 8 D. L. Hull, Certainty and circularity in evolutionary taxonomy, *Evolution* **21**, 174–189 (1967).
- 9 D. L. Hull, The operational imperative: sense and nonsense in operationism, *Syst. Zool.* **17**, 438–457 (1968).
- 10 W. J. LeQuesne, A method of selection of characters in numerical taxonomy, *Syst. Zool.* **18**, 201–205 (1969).
- 11 E. Mayr, The biological meaning of species, *Biol. J. Linn. Soc.* **1**, 311–320 (1969).
- 12 G. Szasz, *Introduction to Lattice Theory*, Academic Press, New York, 1963, 229 pp.
- 13 G. G. Simpson, *The Major Features of Evolution*, Columbia Univ. Press, New York, 1953, 434 pp.
- 14 G. G. Simpson, *Principles of Animal Taxonomy*, Columbia Univ. Press, New York, 1961, 247 pp.
- 15 P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman Press, San Francisco, 1973, 573 pp.
- 16 R. R. Sokal and J. H. Camin, The two taxonomic areas of agreement and conflict, *Syst. Zool.* **14**, 176–195 (1969).
- 17 R. R. Sokal and T. J. Crovello, The biological species concept: A critical evaluation, *Amer. Natur.* **104**, 127–153 (1970).
- 18 A. Takhtajan, *Foundations of the Evolutionary Morphology of Angiosperms*, Moscow and Leningrad, 1964 (in Russian).
- 19 W. H. Wagner, Biosystematics and evolutionary noise, *Taxon* **19**, 146–151 (1970).