J. Math. Biology 10, 367–373 (1980)

Journal of
Mathematical
Biology

When is One Estimate of Evolutionary Relationships a Refinement of Another?

G. F. Estabrook[1] and F. R. McMorris[2]

[1] Department of Botany, University of Michigan, Ann Arbor, MI 48109, USA
[2] Department of Mathematics, Bowling Green State University, Bowling Green, OH 43403, USA

Abstract. A new way to view a certain type of taxonomic character is presented and several fundamental results are rederived using this approach.

Key words: Taxonomic character – Compatibility – Tree of subsets.

Introduction

In the 1960's, several authors proposed and discussed concepts of evolutionary consistency for pairs of taxonomic characters considered potentially useful for estimating the branching pattern of the evolutionary tree for a collection of kinds of organisms (Wilson, 1965; Camin and Sokal, 1965; LeQuesne, 1969). Estabrook et al. (1975, 1976a, 1976b) explicated many of these concepts formally and established some of their mathematical properties sufficient to construct computer programs to analyze a collection of characters to reveal patterns of internal consistency.

In the present paper, we give a new mathematical characterization of the concept of cladistic character without changing its basic meaning; we define a semilattice structure on the collection of cladistic characters; we define evolutionary consistency (compatibility) in terms of this semilattice; and finally, we re-establish, with the very simple proofs made possible by this approach, the major theorems of Estabrook et al. (1975, 1976a, 1976b).

Characterization of Cladistic Characters

A cladistic character may be considered to be an estimate of evolutionary relationships among the members of a collection of organisms. Often such an estimate is based on one basis for comparing the kinds of organisms in the collection. For example, if only the shape of the flower petal is used as a basis for comparing members in a collection of plants, then such an estimate could be a cladistic character which might be called "the shape of the flower petal". Most commonly a cladistic character, especially one based on a single basis for comparison, will not serve to distinguish all pairs of kinds of the organisms in the collection under study. Also, usually some kinds of organisms that are ancestors of some but not all of those in the study are not available for examination. For these

0303 – 6812/80/0010/0367/$01.40

reasons, a cladistic character is usually only a partial estimate of evolutionary relationships, which leaves some questions about relationships between some pairs of the kinds of organisms unresolved. Although it is intended for a cladistic character to have a real observational basis among the kinds of organisms under study, its formal definition is concerned only with mathematical structure. We now redefine "cladistic character" and related concepts after Estabrook et al. (1975). All sets in this note are finite.

Definition 1. A *lower semilattice* is a set P together with a binary relation $\leqslant$ on P such that for $p, q, r \in P$ the following conditions are satisfied:

1. $p \leqslant p$.
2. $p \leqslant q$ and $q \leqslant r$ imply $p \leqslant r$.
3. $p \leqslant q$ and $q \leqslant p$ imply $p = q$.
4. For $p, q \in P$, there is an $s \in P$ such that $s \leqslant p, q$ and if $t \in P$ is such that $t \leqslant p, q$ then $t \leqslant s$.

Conditions 1, 2 and 3 mean that $\leqslant$ is a partial order on P and P is called a partially ordered set. Condition 4 says that each pair of elements in P, p and q, has a greatest lower bound s, denoted by $p \wedge q$. Whenever we discuss a partially ordered set, we will use $\leqslant$ to denote the partial order leaving it to the reader to determine from the context what partial order is being considered.

Definition 2. A *tree lower semilattice* is a lower semilattice P such that for $p, q, r \in P$, if $p \leqslant r$ and $q \leqslant r$ then $p \leqslant q$ or $q \leqslant p$.

Definition 3. Let S be a set. A *cladistic character* on S is a function $K: S \to P$ where

1. P is a tree lower semilattice

and

2. for each $p \in P$, there exists a subset T of S such that $p = \wedge \{K(a): a \in T\}$ where $\wedge$ denotes the greatest lower bound in P.

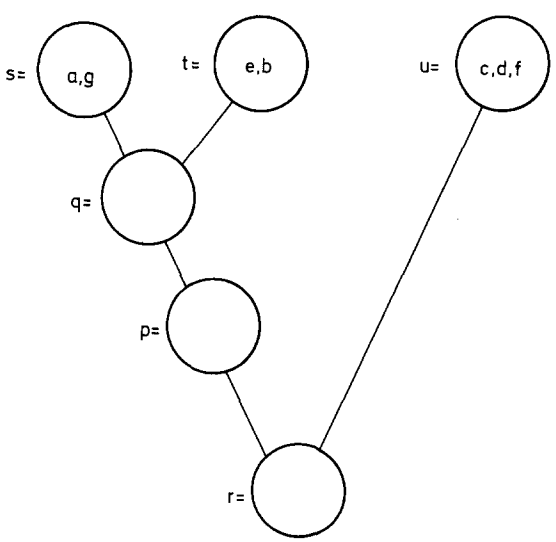

Fig. 1. Diagram of $P = \{p, q, r, s, t, u\}$, $S = \{a, b, c, d, e, f, g\}$, $K(a) = s$, $K(b) = t$, $K(c) = u$, etc. State p fails to meet Condition 2 of Definition 3

One may think of the elements of P as descriptions of the various states of the basis for comparison that are observed among the members of the study collection S of kinds of organisms, plus states that are thought to have possibly existed for ancestors of some of the members in S. The map K associates with $a \in S$ the state $K(a) \in P$ that describes a. The ordering on P expresses an estimate of how the various states might have evolved from one another. P is called the *character state tree* for K.

We point out that the definition of cladistic character presented here differs from that which has previously appeared in that every element of P must be the greatest lower bound of the image of some subset of S. In Fig. 1, removal of p from P would result in a cladistic character in the present sense. This in no way effects the descriptions of observable kinds of organisms. Inclusion in P of character states that do not meet this weak requirement contribute only to *lengths* of branches in the estimated evolutionary tree, and do not affect branching pattern. This simplification allows us to give our new characterization.

Definition 4. A *tree of subsets* of S is a collection $\mathscr{K}$ of nonempty subsets of S such that

1. $S \in \mathscr{K}$

and

2. if $A, B \in \mathscr{K}$ and $A \cap B \neq \emptyset$, then $A \subseteq B$ or $B \subseteq A$.

If $\mathscr{K}$ is a tree of subsets of S and $A, B \in \mathscr{K}$ then define $A \leqslant B$ if and only if $B \subseteq A$. It is easy to check that $\mathscr{K}$ is a tree lower semilattice with respect to this order with $A \wedge B = \cap\{X \in \mathscr{K} : A \cup B \subseteq X\}$.

Theorem 1. *The trees of subsets of S are in one-to-one correspondence with the cladistic characters on S.*

Proof. We first describe how to construct a cladistic character from a tree of subsets $\mathscr{K}$. Define the map $K: S \to \mathscr{K}$ by $K(a) = A$ where A is the set with the smallest number of elements containing $a \in S$. Then K is a mapping into the tree lower semilattice $\mathscr{K}$. We need to now show that every element of $\mathscr{K}$ is the greatest lower bound of the K-image of some subset of S. Let $B \in \mathscr{K}$. If there is an $x \in S$ such that $K(x) = B$ we are done. If not, there exist $y, z \in B$ and $X, Y \in \mathscr{K}$ satisfying the following: $X \cap Y = \emptyset$, $z \in X$, $y \in Y$, $B \subsetneq X$, $B \subsetneq Y$, X and Y cover B (i.e., there is not element of $\mathscr{K}$ between B and X or Y, with respect to set inclusion). Now, X and Y both contain a set that is a K-image of an element in B. Label these sets X_1 and Y_1. Then $B = X_1 \wedge Y_1$. Note that $B = \wedge\{K(x) : x \in B\}$.

Let $K: S \to P$ be a cladistic character on S. To construct a tree of subsets from this cladistic character, set $A(p) = \{a \in S : p \leqslant K(a)\}$ and let $\mathscr{K} = \{A(p) : p \in P\}$. We assert that $\mathscr{K}$ is a tree of subsets of S. Each $A(p)$ is nonempty by Condition 2 in Definition 3 and $S \in \mathscr{K}$ since if m is the minimal element of P, then $A(m) = S$. Now suppose $A(p) \cap A(q) \neq \emptyset$ with $a \in A(p) \cap A(q)$. Then $p, q \leqslant K(a)$ so from Definition 2 either $p \leqslant q$ or $q \leqslant p$. If $p \leqslant q$ and $x \in A(q)$, then $p \leqslant q \leqslant K(x)$ implies $x \in A(p)$ so $A(q) \subseteq A(p)$. If $q \leqslant p$ then $A(p) \subseteq A(q)$ similarly. Note that $\mathscr{K}$ is isomorphic to P.

It is easy to see that these two constructions are inverses of each other so that the proof is complete.

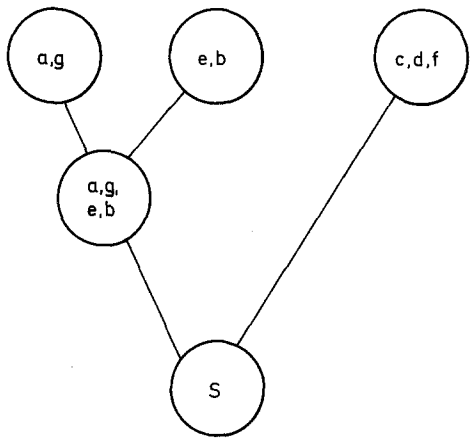

Fig. 2. Diagram of the tree of subsets of S that is equivalent to the cladistic character resulting from the removal of state p from P in Fig. 1

We may now envision cladistic characters as trees of subsets of S. With this approach we are able to rederive properties that previously were established by longer and more complicated arguments. For the remainder of the paper we will use K, K_1, L, etc. for cladistic characters and $\mathscr{K}, \mathscr{K}_1, \mathscr{L}$, etc. for the associated trees of subsets.

Cladistic Character Compatibility

As earlier remarked, a cladistic character on S is usually only a partial estimate of evolutionary relationships. When a second cladistic character asserts all that a first cladistic character does but goes on to resolve relationships among additional members in S as well, then the second is said to be a refinement of the first.

Definition 5. Let K and L be cladistic characters with corresponding trees of subsets $\mathscr{K}$ and $\mathscr{L}$. K is a *refinement* of L if and only if $\mathscr{L} \subseteq \mathscr{K}$.

Thus we see that when K is a refinement of L and $K \neq L$, then K has more character states than L to distinguish more organisms in S or to make more detailed statements about possible evolutionary relationships.

Earlier work has been concerned with analyzing collections of cladistic characters in order to choose internally consistent subcollections for use in estimating evolutionary relationships. Towards this end the concept of "compatibility" was introduced. It has been defined in various ways (Wilson, 1965; Camin and Sokal, 1965; LeQuesene, 1969; Estabrook, 1972) and some applications can be found in Estabrook et al. (1977), Strauch (1978), and Estabrook and Anderson (1978).

Definition 6. Two cladistic characters are *compatible* if and only if there is a cladistic character that is a refinement of both of them. A collection of cladistic characters is a *compatible collection* if and only if there is a cladistic character that is a refinement of every member in the collection.

Definition 6 is equivalent to that which we proposed earlier as Definition 2.4 of Estabrook et al. (1976b). We will now prove this fact. Rather than take the space to restate our earlier definitions and theorems, we simply indicate the results and leave it to the interested reader to check the details.

Assume K_1 and K_2 are compatible cladistic characters in the sense of Definition 2.4. Suppose K_i: $S \to P_i$, $i = 1, 2$. By Lemma 2.2 and Theorem 2.3 of Estabrook et al. (1976b), the map K: $S \to T \subseteq P_1 \times P_2$ defined by $K(x) = (K_1(x), K_2(x))$ for all $x \in S$ is a cladistic character on S and one can choose the tree lower semilattice T such that ρ_i: $T \to P_i$ are onto homomorphisms. Let $A(p) \in \mathscr{K}_1$, so we have $p \in P_1$ and $A(p) = \{x \in S: p \leqslant K_1(x)\}$. Set $I = \{r \in P_1: p \leqslant r\}$ and let $q = \wedge \rho_1^{-1}(I)$. We assert that $A(p) = A(q) \in \mathscr{K}$ so that $\mathscr{K}_1 \subseteq \mathscr{K}$. The case $\mathscr{K}_2 \subseteq \mathscr{K}$ is similar. If $x \in A(p)$ then $p \leqslant K_1(x) = \rho_1(K(x))$ so $q \leqslant K(x)$ which implies $x \in A(q)$. If $x \in A(q)$ then $q \leqslant K(x)$ so $\rho_1(q) \leqslant \rho_1(K(x)) = K_1(x)$. But $\rho_1(q) = \rho_1(\rho_1^{-1}(I)) = \wedge \rho_1 \rho_1^{-1}(I) = \wedge I = p$. Hence $p \leqslant K_1(x)$ and $x \in A(p)$.

Conversely, if $\mathscr{K}_1, \mathscr{K}_2 \subseteq \mathscr{K}$ we must show that there exists a tree semilattice S^* extending S and extensions K_1^* and K_2^* of K_1 and K_2, which are homomorphisms into P_1 and P_2, to homomorphisms onto P_1 and P_2 respectively. Recall that $\mathscr{K}$ is a tree lower semilattice. Define $\mathscr{K}^*$, an extension of $\mathscr{K}$, as follows: Consider $x \in S$ and let $A(x)$ be as before. If $A(x) = \{x\}$ do nothing. If not, make $A(x) < \{x\}$ with no elements between $A(x)$ and $\{x\}$. Do this for all $x \in S$ and then we have that $\mathscr{K}^*$ is a tree lower semilattice whose maximal elements are $\{x\}$ for $x \in S$. If $\{x\}$ is identified with x then $\mathscr{K}^*$ is a tree lower semilattice extending S. Now define K_1^*: $\mathscr{K}^* \to \mathscr{K}_1 \cong P_1$ by $K_1^*(A) = B$ where B is the set of $\mathscr{K}_1$ containing A with the smallest number of elements. Then it can be shown that K_1^* is a homomorphism onto P, extending K_1. K_2^* is constructed similarly.

Let $\mathscr{I}$ denote the set of all trees of subsets of S. Then $\mathscr{I}$ is a semilattice ordered by set inclusion and we see from Definition 6 that the cladistic characters K_1 and K_2 are compatible if and only if $\mathscr{K}_1$ and $\mathscr{K}_2$ have an upper bound in $\mathscr{I}$.

Theorem 2. *The cladistic characters K_1 and K_2 are compatible if and only if $\mathscr{K}_1 \cup \mathscr{K}_2$ is a tree of subsets of S.*

Proof. If $\mathscr{K}_1 \cup \mathscr{K}_2$ is a tree of subsets then clearly it is an upper bound of $\mathscr{K}_1$ and $\mathscr{K}_2$ in $\mathscr{I}$ so K_1 and K_2 are compatible.

Suppose K_1 and K_2 are compatible. Then there exists a tree of subsets $\mathscr{L}$ such that $\mathscr{K}_1, \mathscr{K}_2 \subseteq \mathscr{L}$ and so $\mathscr{K}_1 \cup \mathscr{K}_2 \subseteq \mathscr{L}$. Since $S \in \mathscr{K}_1$ and $S \in \mathscr{K}_2$ then $S \in \mathscr{K}_1 \cup \mathscr{K}_2$. If $A, B \in \mathscr{K}_1 \cup \mathscr{K}_2$, then $A, B \in \mathscr{L}$ and since $\mathscr{L}$ satisfies part 2 of Definition 4 so does $\mathscr{K}_1 \cup \mathscr{K}_2$. Hence $\mathscr{K}_1 \cup \mathscr{K}_2$ is a tree of subsets.

Definition 7. Let K be a cladistic character. A *binary factor* of K is a cladistic character with its tree of subsets of the form $\{S, A\}$ where $A \in \mathscr{K}$ and $A \neq S$.

The following result was proved in Estabrook et al. (1976a) by a very laborious algebraic argument.

Theorem 3. *The cladistic characters K_1 and K_2 are compatible if and only if all their binary factors are pairwise compatible.*

Proof. From Theorem 2 we have that K_1 and K_2 are compatible if and only if for every $A \in \mathscr{K}_1$ and $B \in \mathscr{K}_2$ where $A \cap B \neq \emptyset$, either $A \subseteq B$ or $B \subseteq A$. The result then follows by Definition 7.

One of the purposes of a cladistic character compatibility analysis is to discover compatible collections of cladistic characters for use in estimating evolutionary relationships based on many characters at once. Typically a collection of fifty or more cladistic characters is structured as tentative hypotheses of evolutionary relationships for some collection S of kinds of organisms. This collection of characters is then searched for compatible subcollections. It is thus valuable to determine that the compatibility of every pair of cladistic characters in a collection is sufficient for a compatible collection. This is the main result of Estabrook et al. (1976b). We reprove it here easily with the aid of trees of subsets.

Theorem 4. *Let $K_i: S \to P_i$, $i = 1, \ldots, n$ be a collection of cladistic characters. Then this is a compatible collection if and only if K_i is compatible with K_j for all $1 \leqslant i < j \leqslant n$.*

Proof. Suppose the collection is compatible. Then there exists a tree of subsets $\mathscr{M}$ such that $\bigcup_{i=1}^{n} \mathscr{K}_i \subseteq \mathscr{M}$. Thus $\mathscr{K}_i \cup \mathscr{K}_j \subseteq \mathscr{M}$ for all i, j so K_i and K_j are compatible.

Conversely suppose K_i and K_j are compatible for all i, j. Then $\mathscr{K}_i \cup \mathscr{K}_j$ is a tree of subsets. We assert that $\mathscr{N} = \bigcup_{i=1}^{n} \mathscr{K}_i$ is a tree of subsets. Clearly $S \in \mathscr{N}$. Let $A, B \in \mathscr{N}$ where $A \cap B \neq \emptyset$. Suppose $A \in \mathscr{K}_i$ and $B \in \mathscr{K}_j$. Since $\mathscr{K}_i \cup \mathscr{K}_j$ is a tree of subsets, $A \subseteq B$ or $B \subseteq A$ and the proof is complete.

Conclusion

Several of the definitions and theorems stated above have been stated by us in earlier publications. However, Definition 4, the concept of a tree of subsets; Theorem 1, the characterization of cladistic characters as trees of subsets; Definition 5, the concept of refinement are new here. But to us, the most significant aspect of this exposition is that, through the concept that trees of subsets are semilattice ordered with the relation "is a refinement of", we gain a more direct access to the mathematical properties of cladistic characters.

References

Camin, J. H., Sokal, R. R.: A method for deducing branching sequences in phylogeny. Evolution **19**, 311–326 (1965)

Estabrook, G. F.: Cladistic methodology: A discussion of the theoretical basis for the induction of evolutionary history. Ann. Rev. Ecol. Syst. **3**, 427–456 (1972)

Estabrook, G. F., Anderson, W. R.: An estimate of phylogenetic relationships within the genus *Crusea* (*Rubiaceae*) using character compatibility analysis. Syst. Bot. **3**, 179–196 (1978)

Estabrook, G. F., Johnson, C. S., Jr., McMorris, F. R.: An idealized concept of the true cladistic character. Math. Biosci. **23**, 263–272 (1975)

Estabrook, G. F., Johnson, C. S., Jr., McMorris, F. R.: A mathematical foundation for the analysis of cladistic character compatibility. Math. Biosci. **29**, 181–187 (1976a)

Estabrook, G. F., Johnson, C. S., Jr., McMorris, F. R.: An algebraic analysis of cladistic characters. Discrete Math. **16**, 141–147 (1976b)

Estabrook, G. F., Strauch, J. G., Jr., Fiala, K. L.: An application of compatibility analysis to the Blackith's data on Orthopteroid insects. Syst. Zool. **26**, 269–276 (1977)

LeQuesne, W. J.: A method of selection of characters in numerical taxonomy. Syst. Zool. **18**, 201–205 (1969)

Strauch, J. G., 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)
Wilson, E. O.: A consistency test for phylogenies based on contemporaneous species. Syst. Zool. **14**, 214–220 (1965)

Received July 16, 1979/Revised April 15, 1980