Cladistics and Archaeological Phylogeny

Michael J. O'Brien, R. Lee Lyman, and John Darwent

ABSTRACT

Cladistics, a biological method used to create a nested series of taxa based on shared derived characters (synapomorphies), offers a means of constructing hypotheses about artifact lineages - hypotheses based on heritable continuity as opposed to simple historical continuity. Although cladistically derived trees are only hypotheses about phylogeny, they are superior both to trees created through phenetics, which employs characters without regard to whether they are analogous or homologous, and to trees created by using undifferentiated homologous characters. To date, cladistical analysis is an unused approach to examining archaeological phylogenies but one that holds considerable potential for resolving some of archaeology's historical problems.

INTRODUCTION

Archaeologists have long used changes in artifact form - specifically, changes in character states of artifacts - to measure the passage of time. If the changes in character states are ordered correctly, a historical sequence of forms is created, although independent evidence is needed to root the sequence chronologically. However, historical continuity - X is older than Y, which is older than Z - is different than heritable continuity - X produced Y, which in turn produced Z. The notion that formal similarity of cultural phenomena can be used to indicate heritable continuity originated with the use of the comparative method in linguistic studies of the late eighteenth and early nineteenth centuries (Leaf 1979; Platnick and Cameron 1977). As Franz Boas (1904:518) observed, "Owing to the rapid change of language, the historical treatment of the linguistic problem had developed long before the historic aspect of the natural sciences was understood. The genetic relationship of languages was clearly recognized

Michael J. O'Brien, R. Lee Lyman, and John Darwent, Department of Anthropology, University of Missouri, Columbia, Missouri 65211, United States. E-mail: jad065@mizzou.edu.
when the genetic relationship of species was hardly thought of (...) No other manifestation of the mental life of man can be classified so minutely and definitely as language. In none are the genetic relations more clearly established.” Boas was speaking metaphorically about the “genetic relationship of languages,” but his point was solid: linguistic similarity was the result of cultural transmission and heritable continuity.

In systematic biology in the 1940s and 1950s, formal similarity was used to create evolutionary taxonomies (e.g., Dobzhansky 1951; Mayr 1942; Simpson 1945) that were intended in some cases to reflect not only patterns of ancestry and descent but also degrees of divergence between and among taxa. Evolutionary taxonomy was based on the same axiom that underlay any approach to understanding evolutionary pathways: similarity of phylogenetically related organisms is a result of replication through genetic transmission. Overall similarity in terms of formal characters (traits) was being used to group like with like in descending order of specificity, but overlooked in the process was the notion that there were different kinds of formal similarity, each created by different processes. It wasn’t that biologists (or for that matter archaeologists) were ignorant of the differences between homologous and analogous characters - the former the result of genealogy and the latter of parallelism or convergence - and they correctly ignored analogous characters and focused only on homologous characters in their taxonomies. But they ignored the fact that there are different kinds of homologous characters and that it matters greatly how those different kinds of characters are handled during phylogenetic analysis.

In the late 1940s German entomologist Willi Hennig devised a phylogenetic method that has come to dominate other approaches - an approach that he termed phylogenetic systematics (Hennig 1950) but which in the years following publication of his treatise in English (Hennig 1966) came to be known simply as cladistics. The method is identical in purpose to evolutionary taxonomy - the creation of phylogenetic trees that show evolutionary relationships between and among taxa - but it differs from other methods of reconstructing evolutionary relationships in that it employs only a subset of homologous features - shared derived characters as opposed to shared ancestral characters - to create relationships between and among taxa.

Cladistics is a powerful tool for constructing hypotheses concerning phylogenetic histories of anything that evolves over time, including material remains found in the archaeological record (O’Brien and Lyman 1999; O’Brien et al. 2001). To date, its major use has been in the biological realm, but the basic approach is identical in logic and similar in method to tracing historical patterns of descent in languages (Platnick and Cameron 1977; Ross 1997). The logical basis for extending the use of cladistics into archaeology is the same as it is in
biology: Artifacts are complex systems, comprising any number of parts that act in concert to produce a functional unit. The kinds of changes that occur over generations of, say, projectile-point manufacture are highly constrained, meaning that new structures and functions almost always arise through modification of existing structures and functions as opposed to arising de novo. Thus, "the history of these changes is recorded in the similarities and differences in the complex characteristics of related [objects] - in the extent to which the characteristics of their common ancestors have been modified by subsequent additions, losses, and transformations" (Brown and Lomolino 1998:328). To date, most archaeological discussions of cladistics (e.g., Moore 1994a; Terrell et al. 1994) use the term indiscriminately to refer to any kind of branching phylogeny. Although such usage might be proper semantically - the Greek word *klados* means "branch" - archaeologists appear unaware (Collard and Shennan [2001], Foley [1987] and Harmon et al. [2000] are exceptions) that in the natural sciences cladistics refers to a very specific method of reconstructing phylogeny.

**THE CLADISTICS METHOD**

Figure 1, a cladogram that classifies four taxa, illustrates the conceptual underpinnings of cladistics. All the cladogram says is that based on a certain character distribution, taxa C and D are more similar to one another than either is to any other taxon. It also says that taxa B, C, and D are more similar to one another than any of the three is to taxon A. Despite the fact that a cladogram summarizes only similarity, our interests are in phylogeny. We know that taxa A through D did not appear miraculously out of nothing; rather, they evolved from one or more ancestral taxa. At this point we know little or nothing about those ancestors except for the implication, derived from our cladogram, that with respect to certain characteristics, taxa C and D look more like their immediate common ancestor than they do the one that unites them with taxon B. Likewise taxa B, C, and

![Cladogram](image)

**Figure 1.** One method of illustrating the formal relations of four taxa (A-D). Based on a certain character distribution, taxa C and D are more similar to one another than either is to any other taxon. Also, taxa B, C, and D are more similar to one another than any of the three is to taxon A.
D look more like their common ancestor than they do the one that unites them with taxon A. Convention is to place nodes at the points where branches meet and to refer to the nodes as ancestors that produced the terminal taxa (those at the branch tips) (Figure 2). In our tree, taxa C + D, together with their hypothetical common ancestor (node x), form a monophyletic group, or clade. Taxa D + C + B, together with their common ancestors (nodes x and y), form another, more inclusive clade, and taxa D + C + B + A, together with their common ancestors (nodes x, y, and z), form yet another, and the most inclusive, clade.

Cladistically built trees, like evolutionary trees in general, are constructed using only homologous characters, but the former use only a subset of those characters, termed derived characters, or apomorphies. When these derived characters are shared between taxa related by a common ancestor, such as C and D in Figure 2, they are referred to as synapomorphies. An example of synapomorphic characters is shown in Figure 3. Characters that have been inherited from an ancestor more distant than the most recent one are termed plesiomorphies, or ancestral (primitive) characters. An example of symplesiomorphic (shared ancestral) characters is shown in Figure 4. Characters that show up in lines not related directly through one common ancestor are referred to as homoplasies. An example of homoplasious characters is shown in Figure 5. One kind of homoplaspy results from character-state reversals - meaning, for example, that character A changes from state 1 to state 2 and at some point in a lineage reverts to state 1. We view this kind of homoplaspy more as an analytical problem, meaning that rarely if ever will precisely the same character state reemerge after it disappears. More
likely, the measurement scale being used makes it appear as if the new character state is a homoplasy. Another kind of homoplasy results from parallelism or convergence - organisms, perhaps because of anatomical and/or environmental constraints (the first the result of common history, the second because of similar environments), independently evolve the same character state. All but the simplest cladograms contain homoplasy, and the task of the analyst is to reduce its influence on phylogenetic reconstruction. There are numerous techniques for measuring the robustness (goodness of fit) of the arrangement of taxa, but none of them can be expected to produce more than a locally optimal solution(s). Thus it is imperative to emphasize that any phylogenetic tree is merely a hypothesis concerning genealogical relationships among the taxa under investigation.

**EXAMPLE**

In one of the few archaeological uses of cladistics of which we are aware (O'Brien et al. 2001, 2002), we examined 17 classes of projectile points (83 specimens) from the southeastern United States (Figure 6) that date roughly 11,500-10,000 radiocarbon years BP. Space does not permit us to discuss the analysis in detail, but shown in Figure 7 is the resulting phylogenetic tree. Surprisingly, the program (PAUP [version 4] [Swofford 1998]) generated but a single tree. Its consistency index, or measure of robustness, is 0.5909, which
Figure 6. Illustrations of specimens in each of 17 paradigmatic classes used in the analysis of projectile points from the southeastern United States; class names appear under each example (from O’Brien et al. 2001).
is much higher than we expected and is an indication that although homoplasious characters are present, there is still a strong phylogenetic signal created by the arrangement of apomorphic characters.

CONCLUSION

In closing, we point out that we anticipate objections to the applicability of cladistics to archaeological phenomena. Such objections might take several forms. First, it might be argued that artifacts do not breed - akin to J. O. Brew's (1946: 53) oft-quoted statement in Americanist archaeology that "phylogenetic relationships do not exist between inanimate objects." Tools certainly do not breed, but this is a red herring. Tool makers do breed, and they do pass on information to other tool makers, irrespective of whether those other tool makers are lineal descendants. Transmission creates what archaeologists have long referred to as tool traditions, or what we refer to as tool lineages. It seems naive, given what we know of the archaeological record, not to believe that tool forms are modeled on preexisting tool forms.

Second, it might be argued that although there are tool lineages, they are almost impossible to discover archaeologically because of the kind of process - cultural evolution - that produced them. And, as everyone knows, the tempo of cultural evolution is much faster than that of biological evolution, as is the mode - reticulate for cultural evolution and branching for biological evolution. The classic expression of this dichotomy is shown in Figure 8, which is from A. L. Kroeber's (1948) introductory textbook, Anthropology. Proponents of the use of reticulate models for exploring cultural patterns - models referred to by Moore (1994a, 1994b) as "rhizotic" - claim that diffusion and its attendant processes
swamp all traces of phylogenetic history, thus reducing the cultural landscape to little more than a blur of hybrid (convergent) forms (see Terrell et al. [1997] and responses therein).

We find it difficult not to agree that some cultural differentiation probably is reticulate; similarly, we do not view that as being particularly problematic to archaeological (or linguistic) analysis (see Bellwood 1996). Ward Goodenough (1997:178) makes an excellent point with respect to language: “Contact between Japan and the United States has resulted in considerable borrowing in language and culture by Japan and some reverse borrowing by the United States, but their languages and cultures retain their respectively distinct phylogenetic identities.” Further, linguists do not flip a coin to determine whether two or more languages share a phylogenetic history. Innumerable case studies have provided the basis for deciding which linguistic characters might be apomorphic - bound morphemes and vocabulary, for example - and which might be homoplastic - syntax, for example (see Nichols 1996). Thus Ross (1997:183) can state, as a biologist would, that the comparative method used in linguistics “subgroups languages not on the basis of shared similarities but according to the distribution of shared innovations relative to a reconstructed protolanguage ancestral to the whole family.” This is nothing more than separating apomorphic from plesiomorphic characters and using the former to construct a phylogeny.

Anthropologists are not the only ones to preclude cultural phenomena from phylogenetic study; several biologists and paleontologists (e.g., Gould 1987, 1996, 1997) have done the same thing. The problem is, critics fail to remember that populations within a species exchange genes all the time; it is this exchange that allows a species to remain a species. Further, there is phenotypic change that goes on within a species, and if over time enough variation is generated, cladistical analysis might indeed be able to detect the phylogenetic signal. It bears repeating that cladistics depends on identifying apomorphic character states - an endeavor the success of which is based on the kinds of analytical units used. This applies to the characters identified as well as to the taxonomic units employed. The taxa used in biological cladistics often are species, but there is no inherent reason why they must be species. All that matters is that units are built around characters that change states over time as a result of transmission. We chose paradigmatic classes for our analysis of
southeastern United States projectile points because such classes allowed us to monitor changes in character states (O'Brien and Lyman 1999). It is that ability and only that ability that allows the creation of a phylogenetic ordering of taxa - one built on heritable continuity and that has testable implications.

REFERENCES CITED

Bellwood, P.

Boas, F.

Brew, J. O.

Brown, J. H., and M. V. Lomolino

Collard, M., and S. Shennan

Dobzhansky, T.

Foley, R.

Goodenough, W. H.
Gould, S. J.
Harmon, M. J.; R. D. Leonard; C. S. VanPool, and T. L. VanPool

Hennig, W.

Kroeber, A. L.

Leaf, M. L.

Mayr, E.

Moore, J. H.

Nichols, J.
O'Brien, M. J.; J. Darwent and R. L. Lyman


O'Brien, M. J., and R. L. Lyman


O'Brien, M. J.; R. L. Lyman; Y. Saab; E. Saab; J. Darwent, and D. S. Glover


Plattnick, N. I., and D. Cameron


Ross, M.


Simpson, G. G.


Swofford, D.


Terrell, J. E.; T. L. Hunt, and C. Gosden