

Rethinking Tree-Thinking: Cladograms, Ancestors & Evidence

ANDREW V. Z. BROWER

ABSTRACT

I offer comments on two recent articles in *The American Biology Teacher* by Davenport and colleagues addressing the interpretation and construction of phylogenetic trees. The “tree-thinking” literature suggests that students need to acquire a clear understanding of the meaning of phylogenetic tree diagrams. To this end, I provide clarifications of terminology and address the problematical status of “ancestors.” Cladograms are not genealogies viewed from a distance, but empirical hypotheses of relationship based on the distribution of shared derived character states. I describe an exercise employed in an introductory systematics course that emphasizes the empirical activities of character delimitation and formation of groups based on those characters.

Key Words: Tree-thinking; cladogram; phylogenetic tree; common ancestor; parsimony.

○ Introduction

Two recent “Inquiry & Investigation” articles by Davenport et al. (2015a, b) contain statements that suggest the authors have been misinformed about systematics principles. It is not my intent to chastise teachers who are doing their best to develop lessons within the theoretical framework promoted to them by “tree-thinking” authorities (Baum et al., 2005; Baum & Offner, 2008), but rather to correct what I consider to be mistakes, misconceptions, and ambiguities in these articles before they are promulgated in the curriculum and students become misguided in their understanding of the meanings of phylogenetic trees.

My own research program is in systematics, not in science education, but I will not belabor this article with a lot of citations to the systematics literature. For those desiring further insight into

phylogenetic principles, many of the points raised here are explained in greater detail by Schuh and Brower (2009).

○ What Is a Cladogram?

First of all, let us clarify the definitions of “cladogram” and associated terms. Scientific words have specific meanings, and when these are muddled by careless usage, teaching and learning are hindered. Baum and Offner (2008) rather ambiguously defined “cladogram” as “either a general term for a tree diagram, or a particular style of tree diagram in which neither the amount of change nor time is depicted.” Davenport et al. (2015a) employed a more general definition: “Evolutionary trees, known variously as cladograms, phylogenies, phylogenetic trees, and genetic trees, are representations of biologists’ working hypotheses.” While all of these should represent empirical hypotheses (they do not, always), they are not synonyms. The general term for a branching diagram depicting relationships among taxa is *dendrogram*. A dendrogram considered to represent the evolutionary history of a group is a phylogenetic tree. Phylogenetic trees (or “phylogenies”) may be representations of actual data, may be based on the judgment of unspecified evidence by a taxonomic authority (as they often were prior to the 1970s), or may be pictorial summaries of conventional wisdom regarding relationships of a group, as often found in textbooks. A cladogram is an empirical phylogenetic tree based on minimizing the number of evolutionary steps implied by the distribution of character state differences among the observed taxa. Thus, it is a graphical representation of a data analysis, not a picture of a hypothetical evolutionary chronicle. As noted by Baum and Offner (2008), a cladogram depicts branching order, not absolute amounts of evolution along the branches, or time. This is because groups on cladograms are formed on the basis of

A dendrogram considered to represent the evolutionary history of a group is a phylogenetic tree.

shared, derived character states (features unique to a single taxon provide no evidence of grouping). Further attributes of cladograms are that all observed taxa appear at the tips of the branches and that the internal nodes do not represent any sort of material “ancestors.” See more discussion of the definition of “cladogram” in Brower (2016).

○ Ancestors & Systematics: The Cart & the Horse

What is a common ancestor, and how do we know about common ancestors? The first question, from a metaphysical realist perspective, is relatively straightforward: a common ancestor is a species that lived in the past that gave rise to two or more daughter species that exist today. (Of course, this definition raises all sorts of other questions – such as “What is a species?” – that are beyond the scope of this article.) Since Darwin, the observed hierarchical pattern of the diversity of life has been explained by common ancestry, which, as he recognized, is an inductive theory that accounts for a wide range of biological observations. However, the epistemological problem of how we know about common ancestors is somewhat trickier, since they cannot be directly observed as extant taxa can.

What about fossils? Fossils are rocks (mostly), are not alive, and the only reason we recognize them as the preserved remains of once living organisms is that they share morphological features in common with living things. Despite their inferred age, there is no way to know whether a particular fossil could have been ancestral to any other taxon. In cladistics, fossils are not afforded any special prestige on account of their antiquity and are treated the same as any other taxon. On a cladogram, fossils appear as sister groups to other extant or extinct taxa, depending on the combinations of character states they exhibit.

The characteristics of hypothetical ancestors are inferred as follows. If a character state is shared by all the taxa in a clade, then parsimony strongly suggests that the ancestor of the clade also possessed that character state. If some members of a clade exhibit a character state and others do not, then it sometimes is possible to infer the ancestral state on the basis of its observed distribution, or to suggest the probability that the ancestor might have exhibited one state versus another.

The bottom line is that everything we know about common ancestors is inferred from the distributions of observed character states, either in extant taxa or in fossils that are recognized, by their own pattern of characters, to be related to some extant group. Therefore, hypotheses about ancestry represent results, or conclusions, of phylogenetic analyses.

○ A Note on Parsimony

The principle of parsimony is the centuries-old philosophical idea that, basically, simple explanations that account for all the observations are better than more complicated ones. In its simplest form, parsimony suggests that the world is a predictable place because the future resembles the past. Parsimony is a fundamental part of everyone’s daily life, and some version of it is likely involved in the abilities of other animals to evaluate risks and rewards as well. For example, Batesian mimicry works because predators are fooled

into thinking that something that looks like bad-tasting prey they have experienced before is also bad-tasting. The mimic benefits because the predator assumes that the current experience would be the same as a previous one. It is important to note that parsimony is an epistemological concept, not an ontological one: that is, parsimony does not assume that patterns are simple; it merely states that simple explanations of patterns are preferred.

In systematics, the “maximum parsimony method” selects phylogenetic hypotheses by minimizing the implied evolutionary steps. Some systematists reject parsimony in favor of more complicated model-based approaches. However, when all is said and done, the choice of one phylogenetic hypothesis over another is always based on the minimization of something (the tree with the “highest likelihood” is also the least unlikely tree, given the model and the data). Thus, the principle of parsimony is still fundamental to phylogenetic analyses, as it is to a variety of statistical methods, such as regression.

○ Pedigrees, Cladograms & Ancestors

Davenport et al. (2015a) presented their “tree-thinking” lesson in steps, progressing from a simple, familiar example to an actual phylogenetic hypothesis. They used as their first model the analogy of a small, incomplete pedigree diagram drawn in the form of a cladogram. Again, the authors are blameless for this choice, since Baum and Offner (2008) explicitly encouraged the misleading pedigree metaphor: “A helpful introduction to this material is to stress the parallels between relationships among species and among individuals within families.”

Why is a pedigree not like a cladogram? First, a pedigree of sexually reproducing biparental taxa (such as humans) is not hierarchical but is, rather, a reticulating network. Thus, there are no monophyletic groups in a pedigree. Drawing a pedigree to look like a cladogram is simply confusing, as it implies that each offspring has only one parent. Further, the meaning of “relationship” expressed by cladograms and pedigrees is quite different. Relationships in a cladogram are based, as noted above, on shared, derived character states, as determined by observation of the features of the terminal taxa. Historical narratives of ancestry and descent are inferred from these character distributions. Relationships in a pedigree are based on direct historical knowledge of who begat whom (consanguinity). Internal nodes on a cladogram are hypothetical taxa, whereas internal nodes on a pedigree represent individual people.

Davenport et al.’s (2015a) second model is a “cladogram” of languages (actually not a cladogram in the sense described above, because it has branches of differing lengths and a time scale). This is less problematical than the pedigree metaphor – cladists have noted the parallels between historical linguistics and biological evolution for a long time (Platnick & Cameron, 1977). However, there is a regrettable emphasis on “ancestors,” even though Latin, which actually is the likely “ancestor” of the Romance languages, is represented as a terminal taxon. The evolution of languages is also complicated by the cross-cultural sharing of words, such as *karate* or *quesadilla*, which in systematics is termed “horizontal transfer” and is generally considered a rare exception to the rule of descent with modification that explains the hierarchical pattern of branching of phylogenetic trees.

Davenport et al.’s (2015a) “model 3” is a phylogenetic tree of turtles from Near et al. (2005) (who described the figure as a

“chronogram”). The caption for the model says that “Molecular and radioactive dating of fossils are used to determine the phylogeny of organisms. Turtles are organisms with a long evolutionary history, making them an ideal group to study. Below is a phylogeny of some turtle genera developed using both fossil and molecular dating.” Unfortunately, this is not an accurate description of the phylogenetic analysis by Near et al. (2005), which was based on sequences of three genes (mtDNA *cyt b*, *RAG-1*, and *R35*). The topology is based on the DNA, and only the inferred ages of the nodes on the tree were based on the ages of the fossils – which were, in turn, based on radioactive dating of the rocks in which they were found. The questions associated with this model again refer to the tree as a “cladogram” (which it is not, although apparently Near et al. did generate a cladogram that was very similar to this diagram). It is not a fatal flaw to their “model 3” that Davenport et al. misunderstood the source of the tree they chose, but it is suggestive of the degree to which they are, perhaps, out of their depth.

Davenport et al.’s (2015a) “question 16” encapsulates many of the problems with the Baum “tree-thinking” worldview: “According to Model 3, which two are more closely related: *Caterrochelys*, *Apalone*, *Lyssemys*? Explain how you know, including the term ‘common ancestor’ in your explanation.” Their answer: “*Apalone* and *Lissemys*, because they share a common ancestor most recently.” Technically speaking, this is not a correct explanation: as noted several times already, there are no common ancestors on a cladogram. Thus, neither the students nor anyone else “knows” anything because of any data about a common ancestor. What we actually know (or can infer) is that the evidence (synapomorphic nucleotide substitutions, in this case) suggests that *Apalone* and *Lissemys* are more closely related to one another than either is to *Carettochelys*. We can therefore infer that they share a more recent common ancestor, but this is a result of the analysis and an explanation for the pattern of features shared in common, not a piece of evidence in itself.

○ How Cladograms Are Constructed

Davenport et al.’s second contribution to exploring phylogenetic trees (2015b) is more complex, and its difficulties are more subtle. Some explication of the lesson protocol is in order. First, the lesson

uses a fixed topology (pattern of branching) and shuffles the positions of eight artiodactyl taxa among the terminal slots. That simplifies matters considerably, as there are 135,135 possible rooted trees for eight taxa, but only some 5000 ways the taxa can be arranged on the given tree. That’s good from the perspective of getting an answer, but it is not the way most-parsimonious trees are discovered.

Second, the data provided to the students, as far as I can tell from the article, are incomplete. While the students are supposed to work out relationships of eight taxa (peccary, humpback whale, hippopotamus, pig, mouse deer, camel, elk, and dolphin), they are not given complete sets of characters for each taxon. The presence/absence of data provided is summarized in Table 1. Perhaps this is an intentional component of the “purposely confusing data” that Davenport et al. (2015b) want the students to wrestle with, but this seems to conflate the idea of data that are actually missing (such as features unobservable from a fragmentary fossil) with data that are simply not provided. All eight of the species have a skull, so why shouldn’t the students be able to look at eight skulls?

A critical component of modern phylogenetic analyses – particularly cladistic analyses – is the distinguishing of derived character states from ancestral character states; the father of cladistics, Willi Hennig (1966), referred to these as apomorphies and plesiomorphies, respectively. This is usually accomplished by including an outgroup taxon that is assumed to be outside of the group of interest. Character states may be polarized (determining which states are the plesiomorphies and which are the apomorphies) by comparing the state of a given ingroup taxon to the state of the outgroup. If they are the same, the state is probably plesiomorphic; if they are different, the ingroup character state is likely to be apomorphic. In cladistics, only shared, derived characters (synapomorphies) count as evidence of grouping. Plesiomorphies are just complementary states that do not imply common ancestry. For example, presence of feathers is a synapomorphy for birds. Absence of feathers characterizes the complementary “non-group.” Crocodiles, mammals, sea urchins, plants, bacteria, and any other organism that is not a bird all lack feathers, but these do not form a coherent group or share an inferred common ancestor that is not also common to birds.

Table 1. Inclusion of characters for the eight taxa and seven datasets provided to students in Davenport et al. (2015b).

Animal	Skull	Color	Movement	Range	Foot/Gut Morphology	DNA Insert	Hemoglobin Sequence
Camel	Present	Present	Present	Present	Present	Present	Present
Dolphin		Present	Present			Present	Present
Elk	Present	Present	Present	Present	Present	Present	
Hippo	Present	Present	Present		Present	Present	Present
Mouse Deer		Present	Present	Present		Present	
Peccary		Present	Present	Present	Present	Present	
Pig		Present	Present		Present	Present	Present
Whale	Present	Present	Present			Present	

Davenport et al. (2015b) never mention character polarity in their article, and it is not clear whether this is because (1) they are unfamiliar with the concept, (2) they feel that the concept is too difficult for the students to understand, or (3) they think it is not important to the lesson. However, they make the implicit assumption, in their discussion of the “DNA insert data,” that the insertions are apomorphic, and that therefore “the camel is an outgroup because it shares none of the inserts with any other group.” There is no evidence presented to preclude the possibility that the evolutionary events leading to sequences of differing length among the taxa must have been insertions and not deletions. The explanation offered for these sequence-length differences (“since no protein or RNA is encoded, the presence of these sequences in distinct groups can be explained on the basis of homology and not by analogy resulting from selection for any phenotype”) could apply equally well if we exchanged the word *absence* for the word *presence*, and then the hypothesis of relationships would be completely different. Thus, identifying the camel as the outgroup is an assumption, not an inference from the data.

The final dataset is beta hemoglobin amino acid sequences, again for a subset of four taxa (hippo, pig, camel, dolphin). Davenport et al. (2015b) present this from a phenetic perspective, showing a table with percent identical/similar amino acids between pairs of taxa. It is not clear to me that this conveys any phylogenetic information at all. However, if the camel is again assumed to be the outgroup, then the data shown include 15 putative synapomorphies: five uniting hippo pig and dolphin, five uniting hippo and pig, four uniting hippo and dolphin, and one uniting pig and dolphin. There are also 21 characters with one or more autapomorphic states (uninformative to cladistics). Parsimony suggests that these data support the grouping ((hippo, pig), dolphin), and that the characters supporting alternative pairs of taxa are instances of convergence.

In sum, I think Davenport et al.’s (2015b) exercise has great promise as a lesson in exploring the empirical basis of phylogenetic trees, but it is not clear to me how students are intended to integrate information from the various incomplete datasets into a comprehensive hypothesis of relationships among these animals. The implication that “modern” (genetic) data are better sources of phylogenetic evidence than skulls or mode of locomotion seems to be an artifact of the incomplete sampling more than an intrinsic attribute of the data themselves. As noted, close examination of the sequence data shows that they also contain contradictory information, and indeed, the groups they support are not the same as the putatively “correct” tree no. 4 (camel, ((peccary, pig), ((mouse deer, elk), (hippo, (whale, dolphin)))))) supported by the insertion dataset. The most parsimonious tree provides the best explanation of all the evidence, and it is important not to admit inadvertent bias toward one sort of data over another. The way that particular pieces of evidence are shown to be good or bad is solely by their congruence with the weight of other evidence.

○ A Different, Potentially Helpful Exercise

As noted, I am a systematics researcher, but I also teach systematics to undergraduate and graduate students. An exercise I use on the

first day of my class assumes little or no knowledge of systematics methods and allows students to do “hands-on” observation of character states and assessment of similarities and differences among “taxa.” The progenitor idea for this was published by Burns (1968), but I have changed the “taxa” to a group of objects that share more features in common that may be compared among one another, which I think greatly improves the heuristic value of the exercise. The materials cost a few dollars and can be obtained at the local hardware store.

I give pairs or small groups of students a collection of 8–12 screws (pointy ends) and bolts (blunt ends) of different shapes and sizes, composed of different metals and with different head types (e.g., slot head vs. Phillips head vs. hex head; flat-headed wood screws vs. machine bolts with dome-shaped heads). The students are asked to observe different characters and character states, to write these down in a simple character matrix, and to use these different features to come up with a hypothesis about which ones should be grouped together. They can draw this on a piece of paper. There is no “correct” answer, and, depending on which features students select, they may come up with very different classifications. Some will group by size or color; some will group by the “purpose” of the fasteners. Some groups will recognize more or different characters than others. I ask them to present their results by arranging the different fasteners using their scheme under a document camera that projects it on the screen for the class to see and discuss. Many students produce hierarchical, treelike arrangements, and others produce more complex diagrams of interrelationships.

The above provides a good introduction to the raw materials of phylogenetic analysis – observation of variability, and recognition and specification of different characters that can be compared among the taxa. However, obviously, different kinds of fastening hardware are not descended from a common ancestor, and the similarities and differences that are observed do not necessarily fall into a natural hierarchy, as our prior knowledge of systematics and evolution would lead us to expect to find among living taxa. Some might therefore view this as a pointless and confusing exercise that ignores common ancestry and has nothing to do with phylogenetic inference. My counter to that argument is that this sort of observational taxonomy is effectively what systematists were doing for hundreds of years prior to Darwin’s theory, and as Darwin (1859, p. 413) himself said, common descent is the explanation for “the grand fact in natural history of the subordination of group under group.” As discussed above, we still never observe common ancestry among taxa; we only infer it on the basis of shared character distributions.

If one wanted to make the exercise more “legitimate” by looking at biological entities, something similar could be done with dried beans or nuts. However, the aforementioned hardware has the advantages of being indestructible, inedible, uniform in their variation (so that different groups of students are looking at identical “taxa”), diverse in form, and relatively rich in characters that are easily observed and familiar to students.

Sometimes, as a follow-up to the initial exercise, I give my more advanced students an “outgroup” (a big galvanized deck screw that has a sort of prehistoric look to it) and tell them to use it to polarize the character states they have scored for their other bolts and screws. This allows the students to think about the problem in a phylogenetic context (albeit fanciful), to infer the direction of character state transformations, and to gain an understanding of the fact that only

the derived states represent evidence of grouping, based on inferred evolutionary transformation events.

The difference between this exercise and the pedigree and linguistic models used by Davenport et al. (2015a) is that it puts the empirical analysis of data at the forefront and moves the evolutionary interpretation to the end. There are no ancestors, and it becomes clear that there is no need to invoke common ancestry ahead of time in order to form quite detailed hypotheses of grouping. This highlights the fact that in biological systematics, genealogical relatedness is inferred from the empirical evidence, and not assumed ahead of time.

○ Conclusion

Baum's dictum (Baum & Offner, 2008) that "one can develop a solid understanding of what a phylogenetic tree represents without knowing much about how scientists actually infer the structure of those trees" is a dangerous recipe for evolutionary fundamentalism – "knowledge" based on belief without the critical capacity to trace its empirical source or understand why we believe what we believe. The fact that scientific theories can, at least in principle, be disassembled in a step-by-step manner so that their underlying evidence and assumptions are scrutable is what privileges science as a way of understanding the world that is different from other sorts of belief systems. Disregard for this sort of transparency is a widespread problem today, even among scientists themselves. Thankfully, Davenport et al. (2015b) take important steps toward remedying this deficiency, by emphasizing the empirical nature of phylogenetic hypotheses and by pointing to evidence instead of ancestors. The little hardware exercise described here is offered in a similar vein. If our aim is to instill in students an appreciation for science (and in particular for evolutionary biology) as an empirically based, rational system of knowledge – as opposed to an authority-based system of dogma – it is incumbent upon us not to conflate conclusions with evidence, and hypothetical constructs with empirically observed entities.

○ Acknowledgments

I am grateful to anonymous reviewers for comments that helped me improve the manuscript. Research in my lab is supported by a

collaborative grant, "Dimensions US-Biota-São Paulo: Assembly and evolution of the Amazon biota and its environment: an integrated approach," supported by the U.S. National Science Foundation (NSF DEB 1241056), National Aeronautics and Space Administration (NASA), and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grant 2012/50260-6).

References

- Baum, D.A., Smith, S.D. & Donovan, S.S.S. (2005). The tree-thinking challenge. *Science*, 310, 979–980.
- Baum, D.A. & Offner, S. (2008). Phylogenies & tree-thinking. *American Biology Teacher*, 70, 222–229.
- Brower, A.V.Z. (2016). What is a cladogram and what is not? *Cladistics*, 32, in press.
- Burns, J.M. (1968). A simple model illustrating problems of phylogeny and classification. *Systematic Zoology*, 17, 170–173.
- Darwin, C. (1859). *On the Origin of Species*. London: John Murray.
- Davenport, K.D., Milks, K.J. & Van Tassell, R. (2015a). Investigating tree thinking & ancestry with cladograms. *American Biology Teacher*, 77, 198–204.
- Davenport, K.D., Milks, K.J. & Van Tassell, R. (2015b). Using evolutionary data in developing phylogenetic trees: a scaffolded approach with authentic data. *American Biology Teacher*, 77, 274–283.
- Hennig, W. (1966). *Phylogenetic Systematics*. Urbana, IL: University of Illinois Press.
- Near, T.J., Meylan, P.A. & Shaffer, H.B. (2005). Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *American Naturalist*, 165, 137–146.
- Platnick, N.I. & Cameron, H.D. (1977). Cladistic methods in textual, linguistic, and phylogenetic analysis. *Systematic Zoology*, 26, 380–385.
- Schuh, R.T. & Brower, A.V.Z. (2009). *Biological Systematics: Principles and Applications*, 2nd Ed. Ithaca, NY: Cornell University Press.

ANDY BROWER teaches Principles of Systematics, Evolutionary Biology, and a variety of other courses at Middle Tennessee State University, Murfreesboro, TN 37132; abrower@mtsu.edu. He is coauthor of a textbook on systematics and an associate editor of *Cladistics*, the international journal of the Willi Hennig Society.