

Article

Assessment of Student Learning Associated with Tree Thinking in an Undergraduate Introductory Organismal Biology Course

James J. Smith,^{*†} Kendra Spence Cheruvilil,^{*‡} and Stacie Auvenshine[‡]

^{*}Lyman Briggs College, Michigan State University, East Lansing, MI 48825-1107; [†]Department of Entomology and [‡]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824

Submitted August 5, 2011; Revised April 22, 2013; Accepted April 23, 2013
Monitoring Editor: Marshall David Sundberg

Phylogenetic trees provide visual representations of ancestor–descendant relationships, a core concept of evolutionary theory. We introduced “tree thinking” into our introductory organismal biology course (freshman/sophomore majors) to help teach organismal diversity within an evolutionary framework. Our instructional strategy consisted of designing and implementing a set of experiences to help students learn to read, interpret, and manipulate phylogenetic trees, with a particular emphasis on using data to evaluate alternative phylogenetic hypotheses (trees). To assess the outcomes of these learning experiences, we designed and implemented a Phylogeny Assessment Tool (PhAT), an open-ended response instrument that asked students to: 1) map characters on phylogenetic trees; 2) apply an objective criterion to decide which of two trees (alternative hypotheses) is “better”; and 3) demonstrate understanding of phylogenetic trees as depictions of ancestor–descendant relationships. A pre–post test design was used with the PhAT to collect data from students in two consecutive Fall semesters. Students in both semesters made significant gains in their abilities to map characters onto phylogenetic trees and to choose between two alternative hypotheses of relationship (trees) by applying the principle of parsimony (Occam’s razor). However, learning gains were much lower in the area of student interpretation of phylogenetic trees as representations of ancestor–descendant relationships.

INTRODUCTION

Many introductory organismal biology students seem to have a difficult time understanding the evolutionary connections among organismal groups. The way introductory biology courses are taught is partly to blame. Brewer (1996) noted that historical and comparative approaches, which are important for really understanding the significance of evolutionary theory, are given short shrift in our students’ evolutionary

biology education. In contrast, natural selection and the functional perspective, with its emphasis on what a structure does today as opposed to its historical evolutionary roots, is often the focus of evolutionary modules within Introductory Biology courses (Brewer, 1996). Padian (2008) pointed out that most high school and college biology textbooks cover the principles and data of microevolution (genetic and population changes) and speciation fairly well. What they do not do, however, is cover what is known about the major evolutionary changes over time or above the species or population level (macroevolution), and how these changes are understood (Padian, 2008).

Phylogenies and tree-thinking instruction can provide tools to bridge the gap between classic historical approaches to teaching evolution and the more traditional emphasis on natural selection and microevolutionary change. After all, Darwin used a phylogenetic tree as the sole graphic illustration to explain evolution by natural selection in *On the Origin of Species* (Darwin, 1859). Baum *et al.*’s (2005) call to arms argued that “tree thinking,” or having a mindset that considers

DOI: 10.1187/cbe.11-08-0066

Address correspondence to: James J. Smith (jimsmith@msu.edu).

© 2013 J. J. Smith *et al.* CBE—Life Sciences Education © 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

phylogenetic trees as testable hypotheses of evolutionary relationships within and between groups of organisms, should be a major theme in our students' evolution training. These authors pointed out that phylogenetic analysis (inference of phylogenetic trees to interpret ancestor–descendant relationships) is rarely used outside the realm of professional evolutionary biologists. In addition, the recently published Advanced Placement Biology curriculum (College Board, 2011) considers tree reasoning to be essential knowledge (“Essential knowledge 1.B.2: Phylogenetic trees and cladograms are graphical representations (models) of evolutionary history that can be tested”).

However, having students learn about and use phylogenies is not trivial (for review, see Meisel, 2011). Students hold several misconceptions that prevent them from using phylogenies effectively and that present “fundamental barriers to understanding how evolution operates” (Meir *et al.*, 2007, p. 76). Gregory's (2008) review of student understanding of phylogenetic trees listed 10 prominent student misconceptions; among these are: branches appearing later in a tree represent “higher” taxa; similarity indicates relatedness; and relatedness of taxa can be determined by counting the number of nodes separating them. Recent papers by Novick *et al.* (2010) and Halverson (2011) have explored cognitive aspects of student interpretation of phylogenetic trees as representations. For example, Novick *et al.* (2010) found that simply representing trees with synapomorphies labeled on the branches improves student comprehension of the representation. These latter types of studies link educational initiatives that seek to incorporate tree-thinking approaches into the classroom to fundamental research questions about how students learn.

Previously, we described the restructuring of an introductory organismal biology course for majors to reincorporate some of the study of biodiversity that had been lost upon moving to an inquiry-based laboratory framework (Smith and Cheruvilil, 2009). Our curricular interventions consisted of a series of instructional activities in the classroom and laboratory that focused on tree thinking to teach organismal diversity in an evolutionary context (Smith and Cheruvilil, 2009). We used phylogenies to provide a problem-based experience, following the advice of Perry *et al.* (2008), who pointed out the need to engage college students in active, hands-on curricula relevant to macroevolution. The laboratory exercises were based on those proposed by Singer *et al.* (2001) and Giese (2005), and were similar in spirit to exercises developed by Julius and Schoenfuss (2006), who used a set of vertebrate skulls for character matrix construction and subsequent phylogenetic analysis to increase students' scientific literacy. These exercises were implemented in an introductory organismal biology course in the Lyman Briggs College (LBC) at Michigan State University (MSU), a residential college for students majoring in the natural sciences.

In this paper, we report our assessment of the student learning gains associated with these curricular interventions over the course of two semesters of instruction and the effects that these interventions had on student understanding of phylogenetic trees and evolution. To carry out our work, we designed, validated, and implemented a Phylogeny Assessment Tool (PhAT), in which students were given two phylogenetic trees, or alternative hypotheses, representing possible evolutionary relationships of five taxa (Figure 1). Before and after instruction, we used the PhAT to assess how well our stu-

dents had learned how to map characters onto the two phylogenetic trees, to apply the principle of parsimony to decide which of the two given hypotheses (trees) was more likely to be the correct hypothesis, and to interpret phylogenetic trees as representations of a set of ancestor–descendant relationships. This last part provided insight into our students' understanding of “recency of common ancestry,” a core concept of tree thinking that is the basis of the misconception that “similarity indicates relatedness” (Gregory, 2008).

METHODS

Study Context

The Course. The study was carried out in the LBC introductory organismal biology course (LB144) at MSU during the Fall semesters of 2008 and 2009. This four-credit course consists mainly of suburban and rural Michigan students, with ~10% of them coming from groups traditionally underrepresented in the sciences. About half of the students take LB144 as freshmen, and half as sophomores. The students enter the class with a range of biology backgrounds and math preparedness. LB144 is taken by science majors (mainly human biology and physiology), and includes many preprofessional students.

During both semesters, LB144 students attended two 80-min class sessions per week and a 3-h combined recitation and lab. The 80-min classes (80–110 students) used a “bookends” instructional model (Johnson *et al.*, 1998), with mini-lectures interspersed with small group exercises, individual writing, personal response pad (i.e., clicker) questions, and other active/collaborative learning activities. The combined recitations/labs met in sections of 16–24 students each, which were staffed by an LBC professor or a graduate teaching assistant, along with two undergraduate learning assistants. Students in the recitation/labs worked in teams of three to four students, yielding four to six research teams per lab section.

During Fall 2008, 216 LB144 students were enrolled in two lecture sections and 10 recitation/lab sections. During Fall 2009, 168 students were enrolled in two lecture sections and eight recitation/lab sections. During both Fall 2008 and Fall 2009, LB144 was team-taught by two LBC biology professors, with one of us (K.S.C.) serving as one of the two instructors in each semester. Within each semester, all class and recitation/lab sections were taught by equivalent teaching teams and were provided the same instructional materials and activities.

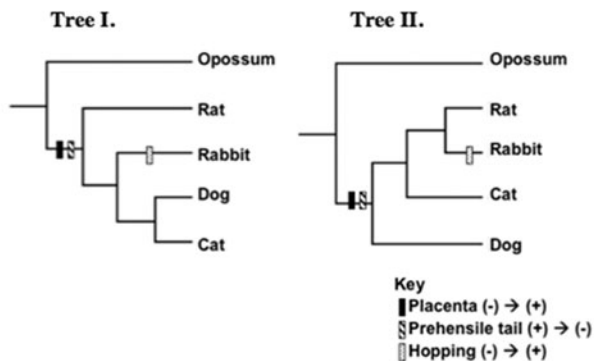
Learning Goals. We implemented a set of curricular interventions for the classroom and recitation/laboratory designed to help our students understand evolution. These interventions were centered on a tree-thinking approach that used phylogenetic trees as testable hypotheses of evolutionary relationships within and between groups of organisms (Smith and Cheruvilil, 2009). Our premise was that, by developing and using “tree-thinking” skills, our students would be able to: 1) use data to discern between two alternative hypotheses (phylogenies); 2) use objective criteria (e.g., the principle of parsimony) to argue which of two alternative hypotheses was better supported by data; and 3) demonstrate an understanding of phylogenetic trees as depictions of

The Phylogeny Assessment Tool (PhAT)

The table below lists six characteristics of the opossum, dog, rat, cat and rabbit. For each animal, (+) indicates that the character is present and (-) indicates that the character is absent.

Character	Opossum	Dog	Rat	Cat	Rabbit
1. Placenta	-	+	+	+	+
2. Prehensile (grasping) tail	+	-	-	-	-
3. Hopping locomotion	-	-	-	-	+
4. Large canine teeth	+	+	-	+	-
5. Expanded metatarsals (foot bones)	-	+	-	+	+
6. Large incisor teeth	-	-	+	-	+

A) For the two phylogenetic trees (diagrams that depict evolutionary relationships between organisms) below (Tree I and Tree II), we have mapped characters 1, 2 & 3 onto each tree. Now evaluate these two trees by mapping characters 4, 5 & 6 onto each tree. Be sure to use clear labels and provide a key to your proposed changes as shown for characters 1, 2 & 3.



- B) Which of the two phylogenetic trees (Tree I or Tree II) do you think is a better hypothesis of the evolutionary relationships of opossums, dogs, cats, and rabbits? Explain your reasoning.
- C) In the tree that you chose in part (B) above, which of the two animals are closest relatives? Explain your reasoning.

Figure 1. The Phylogeny Assessment Tool (PhAT). Each of the three parts of the PhAT (A, B, and C) was administered to students as a pretest and a posttest during Fall 2008 and Fall 2009. Depicted are the questions as asked during Fall 2008. During Fall 2009, parts B and C were each split into two parts that asked the main question and, separately, asked the open-ended rationale for that answer (“Explain your reasoning”).

ancestor–descendant relationships based on recency of common ancestry. Our overarching objective based on these three goals was to help our students understand evolution, not as a set of “just so” stories but as a rigorous framework for constructing knowledge that is amenable to testing using generally accepted scientific methods.

Instructional Activities. Much of the classroom and recitation/lab content and activities during Fall 2008 and 2009 that specifically related to tree thinking followed those described in detail in Smith and Cheruvilil (2009). Briefly, students were assigned the book chapter “Phylogeny and Tree of Life” (Campbell *et al.*, 2008) to read prior to class. This was followed by classroom content and exercises about phylogenetics and tree thinking that engaged students in practice interpreting phylogenetic trees that were built with morphological, developmental, and molecular biology data. Students were given instruction on mapping characters on alternative trees and choosing the most likely hypothesis of evolutionary relationships using the principle of parsimony (Smith and Cheruvilil, 2009).

Introductory exercises in recitation/lab followed these classroom instructional activities. These exercises engaged students in making comparisons among organisms by having them collect morphological data, build data tables, and then apply these collected data to a comparison of different

hypotheses of evolutionary relationships among five mammalian skeletons. These introductory exercises were followed by 2 wk of observations and dissections of exemplars representing nine animal phyla, and a capstone phylogeny experience. During the capstone experience, students completed a team assignment that brought together the information from their weeks of observations and dissections with the tree-thinking techniques practiced during the introductory exercises to evaluate two competing hypotheses of how the nine animal phyla are related (Smith and Cheruvilil, 2009). Thus, 4 wk of recitation/lab time and 1 wk of class were devoted to thinking about and practicing tree-thinking skills and macroevolutionary patterns and processes.

Assessment of Student Learning

Design of the PhAT. We developed and used a three-part PhAT (Figure 1) to assess student learning of tree-thinking skills. We chose to develop our own assessment tool for the purposes of this study, primarily because existing assessment tools in the area of phylogenetics, such as the Tree-Thinking Quiz (Baum *et al.*, 2005) or the Conceptual Inventory of Natural Selection (Anderson *et al.*, 2002), did not explicitly assess students’ abilities to map traits on trees, apply the principle of parsimony, and explain common ancestry as the basis of phylogenetic reasoning.

Part A of the PhAT was designed to assess student understanding of how to interpret and manipulate (i.e., map characters onto) phylogenetic trees. Part B asked students to use “tree thinking” to solve a biological problem (compare two hypotheses), and part C asked students to relate these activities to more abstract evolutionary concepts (e.g., common ancestry). Parts B and C were each split into two parts (B' and B'' and C' and C'', respectively), such that we first asked the main question and separately asked for the open-ended rationale for a student's answer (i.e., explain your reasoning). Splitting these questions allowed us to assess to what extent the learning gains resulted from identifying the correct tree (in part B) or the correct pair of taxa (in part C) compared with the extent to which gains resulted from applying correct logic.

Development and Validation of the PhAT. The PhAT was developed in consultation with experts both in tree thinking and the scholarship of teaching and learning (SoTL) to produce an assessment tool for selected aspects of undergraduate tree thinking. The PhAT was conceived during a meeting of the Tree Reasoning in Evolution Education (TREE) Working Group at the National Evolutionary Synthesis Center (NES-Cent) in May 2008, which included leading experts in all aspects of phylogenetic trees and extensive discussions of how students interact with and interpret phylogenetic trees. The PhAT itself was developed during a research residency (J.J.S.) in July 2008 in the American Society of Microbiology's Biology Scholars Program, which focused on the development of SoTL projects. The initial PhAT was developed over the course of this 3-d residency in an iterative process that involved trials of prototypes of the PhAT with groups of residents and staff involved in the research residency (20–25 college biology instructors who teach across the range of the biology curriculum). The SoTL experts and biology instructors at the research residency provided feedback at each intermediate stage of PhAT development, and the iterative presentation and feedback process led to the production of the finalized PhAT (Figure 1). As a measure of content validity, posttests from year 1 were examined and analyzed by two independent verifiers (J.J.S. and Dr. Louise Mead, Education Director for the BEACON Center for the Study of Evolution in Action at Michigan State University) in addition to S.A., who scored all the pre- and posttests for both Fall 2008 and Fall 2009. This process indicated that the PhAT responses obtained from the students were indicative of expected responses.

Institutional Review Board. Before data collection for the research study was begun, the project (including the PhAT) was submitted to the MSU Institutional Review Board (IRB X08-1057). The project was designated as “exempt.”

Administration and Scoring of the PhAT. During both semesters, the PhAT was administered as a pretest during the first week of class as an ungraded, in-class exercise. During Fall 2008, posttests were given during December as a part of the final exam for the course (total exam worth 150 points out of 900 class points total), whereas during Fall 2009, posttests were part of the second of three lab exams that occurred during a November class session (total exam worth 24 points out of 1000 class points total). All pre- and posttests for both years were scored by one of the authors (S.A.), and

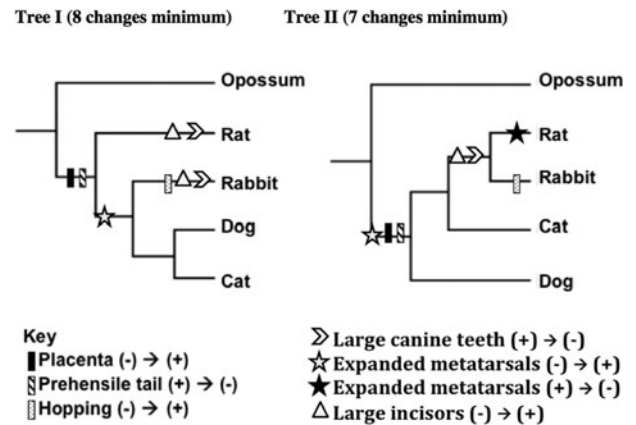


Figure 2. Answer Key for the PhAT. The key for part A is shown, with symbols mapped onto both tree I and tree II. Note that the figure shows only one of the multiple correct ways to map the minimum of eight evolutionary changes onto tree I; all correct mappings were given full credit. For example, for both “Large incisors” and “Large canine teeth,” independent losses could have occurred on the branches leading to rat and rabbit, or a loss could have occurred in the common ancestor of dog and cat, followed by a secondary gain in the common ancestor to dog and cat. On the other hand, there is only one way to map characters minimally onto tree II. For the first part of part B (part B'), the correct answer is tree II, which requires seven evolutionary changes, as opposed to tree I, which requires a minimum of eight changes. Refer to Table 2 for correct responses to part B''. For part C, the closest relatives in tree I are the dog and cat, while in tree II, the closest relatives are the rat and rabbit. In both cases, the reason is they share a common ancestor more recently than any other pair of taxa. Refer to Table 2 for correct responses to part C''.

scoring of the PhAT for this study was independent of any grading that occurred for the course itself.

Scoring the Three Parts of the PhAT. Figure 2 shows the scoring rubric that was used for part A of the PhAT, which connects to the idea that having synapomorphies displayed on branches allows students to understand trees better (Novick *et al.*, 2010). In part A, students were asked to map the evolution of large canine teeth, expanded metatarsals, and large incisors onto two different phylogenetic trees (alternative hypotheses) that were chosen out of the 15 possible phylogenetic trees for the relationships of rats, rabbits, dogs, and cats, with opossums as the outgroup (see Smith and Cheruvilil, 2009). Students were provided with three characters (placenta, prehensile tail, hopping locomotion) already mapped onto both trees, plus a key illustrating how to indicate mapped evolutionary changes (Figure 1). During both semesters, 0.5 point was awarded for each character correctly mapped on both tree I and tree II (Figure 2), for a total of three points (no partial credit). The multiple equally parsimonious mappings of characters onto tree I (see Figure 2 legend) were all given full credit.

In part B, students were first asked to choose one of the two trees shown in part A as the better of the two hypotheses. Part B was scored separately for the choice of the correct tree (part B') and for the reasoning for choosing a tree (part B''). During both semesters, students earned one point for an answer of tree II in part B', with no partial credit awarded. Students were then asked to provide the reasoning that

they used to choose one tree over the other. On the basis of our instructional activities, we anticipated that students would apply the principle of parsimony to choose a preferred hypothesis. Students earned full credit (two points) in part B' for answering "principle of parsimony," "fewer number of [evolutionary] events or character changes or traits reappearing/disappearing," or "Occam's razor." Partial credit was awarded to students who wrote "fewer number of ticks/marks," but did not explain in biological terms what their answer meant (one point).

In part C, students were asked to identify which two taxa were the closest relatives in a phylogenetic tree, thereby demonstrating an understanding of phylogenetic trees as representations of ancestor-descendant relationships. Students were asked to indicate which animals they thought were the two closest relatives in either tree I or tree II (whichever tree the student put down as the answer to part B) and to explain their reasoning for picking a particular pair of taxa (i.e., more recent shared ancestry). Students were awarded one point for the answer "rat and rabbit" for part C' if they chose tree II in part B, and one point for the answer "dog and cat" if they had incorrectly chosen tree I in part B. In response to "Explain your reasoning" (part C''), we expected students to argue that the two most closely related organisms are the ones that share the most recent common ancestry (two points). Students were awarded partial credit for the answer "they are sister taxa" (one point).

Learning Gains. Absolute learning gains were used to assess student learning on each separate part of the PhAT, as well as the combined total learning gain. Absolute learning gains were calculated as (posttest score – pretest score), both

as absolute gain and percentage gain. Student's *t* test (one-tailed; paired) was used to test for significant gains (differences between mean pretest and mean posttest scores) within semesters. Student's *t* test (two-tailed; unequal variance) was used to test for significant differences in learning gains between semesters on each separate part of the PhAT, as well as the combined total. Normalized learning gains (Meltzer, 2002; Weber, 2009), calculated as (posttest percent – pretest percent)/(100% – pretest percent), were also used in *t* tests to assess differences in the combined total gains on the PhAT.

Tests for Association between Different Aspects of Tree Thinking. Tests of association were used to examine connections between a student's abilities to map characters onto trees (part A), apply the principle of parsimony (parts B' and B'' combined), and view trees as representations of the recency of common ancestry (parts C' and C'' combined). Chi-squared tests of independence, as implemented in the interactive calculation tool developed by Preacher (2001), were used to test for associations between students' responses to parts A, B, and C of the PhAT in all three combinations (A and B, A and C, and B and C) for both semesters. Before the tests were conducted, data were pooled into two categories for the PhAT: a student either had a learning gain (positive value) or no gain (zero or negative value) for parts A, B, or C.

RESULTS

Learning Gains

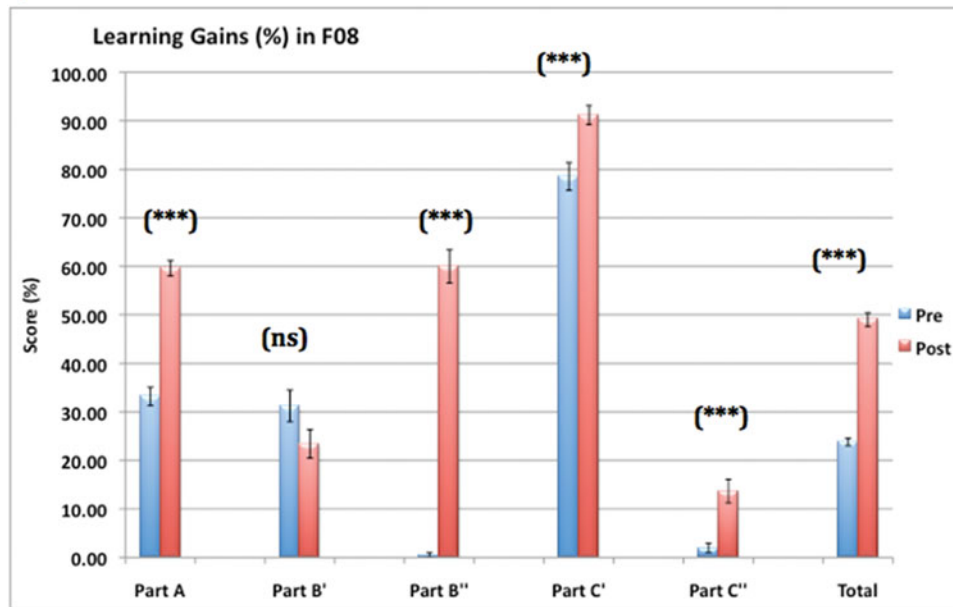
Significant learning gains were observed for the PhAT as a whole and each part of the PhAT, except part B' during Fall 2008 (Table 1 and Figure 3). Similarly, significant learning

Table 1. Mean pre- and posttest scores and learning gains on the PhAT during Fall 2008 (*n* = 205) and Fall 2009 (*n* = 163)

		Absolute scores					
	Fall 2008	Part A (3 points) Score (± SE)	Part B' (1 point) Score (± SE)	Part B'' (2 points) Score (± SE)	Part C' (1 point) Score (± SE)	Part C'' (2 points) Score (± SE)	Total (9 points) Score (± SE)
Pre		1.00 (± 0.06)	0.31 (± 0.03)	0.01 (± 0.01)	0.79 (± 0.03)	0.04 (± 0.03)	2.14 (± 0.07)
Post		1.79 (± 0.05)	0.23 (± 0.03)	1.20 (± 0.07)	0.91 (± 0.02)	0.27 (± 0.05)	4.41 (± 0.12)
Gain		0.79 (± 0.07)	-0.08 (± 0.04)	1.19 (± 0.07)	0.13 (± 0.03)	0.23 (± 0.05)	2.27 (± 0.14)
Fall 2009							
Pre		0.85 (± 0.06)	0.43 (± 0.04)	0.19 (± 0.04)	0.76 (± 0.03)	0.20 (± 0.45)	2.43 (± 0.11)
Post		1.90 (± 0.07)	0.49 (± 0.04)	1.71 (± 0.05)	0.86 (± 0.03)	0.31 (± 0.05)	5.27 (± 0.13)
Gain		1.06 (± 0.08)	0.06 (± 0.05)	1.52 (± 0.06)	0.10 (± 0.04)	0.10 (± 0.06)	2.84 (± 0.15)
		Percent scores					
	Fall 2008	Part A (3 points) Percent (± SE)	Part B' (1 point) Percent (± SE)	Part B'' (2 points) Percent (± SE)	Part C' (1 point) Percent (± SE)	Part C'' (2 points) Percent (± SE)	Total (9 points) Percent (± SE)
Pre		33.17 (± 1.88)	31.22 (± 3.24)	0.49 (± 0.49)	78.54 (± 2.86)	1.95 (± 0.97)	23.79 (± 0.81)
Post		59.59 (± 1.58)	23.41 (± 2.97)	60.00 (± 3.43)	91.22 (± 1.98)	13.66 (± 2.40)	48.97 (± 1.37)
Gain		26.42 (± 2.16)	-7.80 (± 4.23)	59.51 (± 3.51)	12.68 (± 3.41)	11.71 (± 2.55)	25.18 (± 1.50)
Normalized gain ^a							30.40 (± 2.78)
Fall 2009							
Pre		28.22 (± 1.98)	42.94 (± 3.89)	9.51 (± 2.02)	76.07 (± 3.35)	10.12 (± 2.27)	26.99 (± 1.26)
Post		63.50 (± 2.25)	49.08 (± 3.93)	85.28 (± 2.61)	85.89 (± 2.74)	15.34 (± 2.55)	58.52 (± 1.42)
Gain		35.28 (± 2.55)	6.13 (± 5.41)	75.77 (± 3.15)	9.82 (± 4.19)	5.21 (± 2.97)	31.53 (± 1.68)
Normalized gain ^a							40.32 (± 2.55)

^aCalculated as (posttest percent – pretest percent)/(100% – pretest percent).

a) Fall 2008



b) Fall 2009

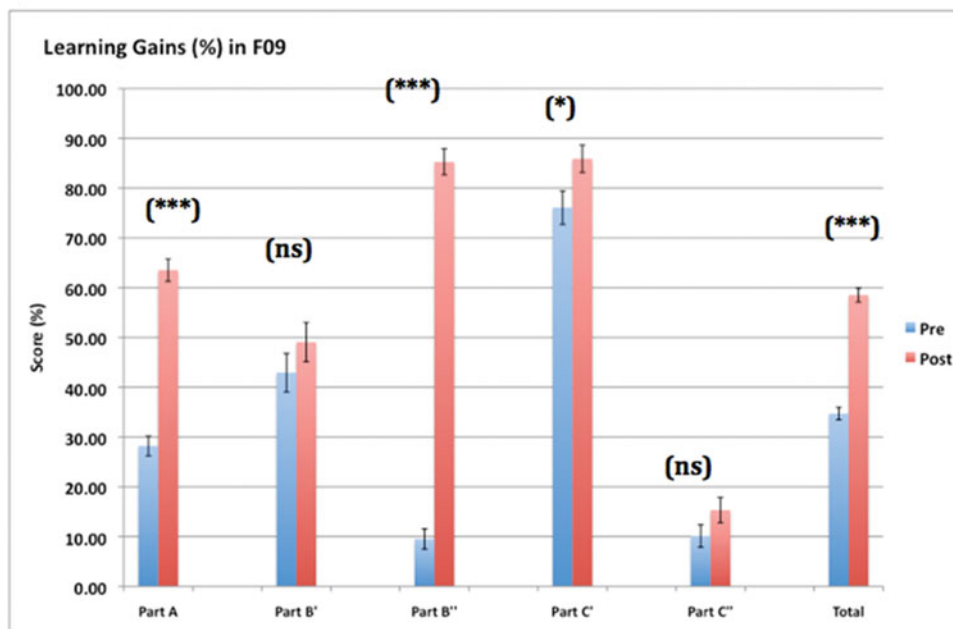


Figure 3. Learning gains on the PhAT during (a) Fall 2008 ($n = 205$) and (b) Fall 2009 ($n = 163$). Pre- and posttest scores (mean values [%] \pm SE) are indicated for each part of the PhAT, as well as the total score on the PhAT. Differences between pre- and posttest scores represent learning gains. *, significant at $\alpha = 0.05$; **, significant at $\alpha = 0.01$; ***, significant at $\alpha = 0.001$; ns, not significant.

gains were observed for the PhAT as a whole and each part of the PhAT, except part B' and part C'' during Fall 2009 (Table 1 and Figure 3). In both semesters, we found the largest learning gains for part B'', which asked students to explain their reasoning for choosing one tree over the other as a preferred hypothesis (59.51 and 75.77% in 2008 and 2009, respectively). There were no significant gains in either semester for part B', which asked students to choose the preferred tree. Smaller learning gains were observed for part C'', which asked the

students to explain their rationale for choosing the two closest relatives in tree. In Fall 2008, the gain was 11.71%, while in Fall 2009, the gain in part C'' was not significant. Total scores on the PhAT increased 25.18 and 31.53% in Fall 2008 and Fall 2009, respectively (Table 1). When total learning gains were normalized, these numbers increased to 30.40 and 40.32% in Fall 2008 and Fall 2009, respectively (Table 1).

As mentioned above, most of the learning gain for part B was in part B'' (the second subpart of the question), or

Table 2. Student responses to the open-ended questions in PhAT parts B' and C'

	Fall 2008		Fall 2009	
	Pretest	Posttest	Pretest	Posttest
Why is the tree you chose better?				
Principle of parsimony ^a	0	32	0	78
There are less events ^a	1	31	7	33
Fewer is better/simpler/Occam's razor ^a	0	34	2	24
Fewer marks ^b	0	34	12	10
Organisms with more similar traits grouped together	81	30	43	10
Arranged better/less complicated	32	32	28	4
Answer does not fit within these 6 categories	97	23	73	5
Why are the two taxa you chose closest relatives?				
They share the most recent common ancestor ^a	0	17	6	9
They evolved from/share a close or most common ancestor ^a	3	10	8	9
Evolved from same ancestor ^a	0	0	4	10
They are sister species (or taxa)	0	51	0	18
They share same/most characteristics/traits	153	127	93	88
They come off the same branch	22	4	16	25
Answer does not fit with the above 6 categories	33	7	38	5

^a Answer received full credit.

^b Answer received partial credit.

the explanation given by students for choosing one tree over the other as a preferred hypothesis. On the other hand, no significant learning gains were observed for part B', which asked students to identify which of the two trees was a better hypothesis based on the data. In both semesters, a large number of students (75% in Fall 2008 and 67% in Fall 2009) chose the incorrect tree in part B' because they had mapped characters incorrectly in part A; these students subsequently applied correct tree-reasoning logic to decide upon the best hypothesis (i.e., parsimony).

For part C of the PhAT, learning gains were more or less equivalent across the two subparts of the question, and across years (12.68% for part C' and 11.71% for part C'' in Fall 2008, and 9.82% for part C' and 5.21% for part C'' in Fall 2009; Table 1). However, the pretest scores for these two parts of part C were highly dissimilar. In both semesters, the pretest scores for part C' (which two taxa are closest relatives) was greater than 70%, while the pretest scores for part C'' was 10% or less in both semesters.

For the extended response portions of part B (part B', Why is the tree you chose better?) and part C (part C', Why are the taxa you chose closest relatives?), we established key phrases/categories that allow us to score the written responses from the students (Table 2). In both semesters, for part B', we observed a striking increase from pretest to posttest in the number of students providing what were considered to be correct (or partially correct) responses. During Fall 2008, 60.6% of the students (131/216) indicated, "principle of parsimony," "there are less events," or "fewer is better/simpler/Occam's razor" (full credit), or "fewer marks" (partial credit) on the posttest as compared with only 0.5% (1/211 students) on the pretest. The increase was also remarkable during Fall 2009, when 88.4% of the students (145/164) gave full- or partial-credit responses on the posttest as compared with 12.7% (21/165) on the pretest.

One potential concern was that in their responses to part B', students might be exhibiting "procedural display" (Bloomer

et al., 1989), in which a correct answer that receives full credit is given, but without any evidence of genuine understanding. Nearly all of the learning gains in part B' occurred by virtue of using the word "parsimony" (or some related phrase) in the explanation for why a particular tree was chosen (75.77% gain). Thus, we examined the student responses to determine what proportion of these responses might be characterized as "minimalist." No students responded simply with the word "parsimony" or phrase "principle of parsimony" without an explanation of what that word or phrase meant. A set of representative responses is shown in Table 3.

For part C'', we observed a small but significant increase in fully or partially correct responses from the pre- to the posttests in Fall 2008, while there was no significant difference before and after instruction in Fall 2009. During Fall 2008, 12.5% of the students (27/216) indicated the full-credit responses "They share the most recent common ancestor" or "They evolved from/share a close or most common ancestor" on the posttest, as compared with 1.4% (3/211) on the pretest (Table 2). A full 87.5% of the students gave incorrect responses on the posttest during Fall 2008 (189/216), with the most common incorrect responses being, "They share the same/most characteristics/traits" (58.8%; 127/216) and "They are sister species (or taxa)" (23.6%; 51/216). Results for part C'' during Fall 2009 were similar to those obtained during Fall 2008, although the number of students who indicated on the posttest, "They are sister species (or taxa)" fell to 11.0% (18/164; Table 2).

Associations between Different Aspects of Tree Thinking: Comparisons across Semesters

We also tested for associations between students' responses to the different parts of the PhAT. For these comparisons, the scores for parts B' and B'' were pooled, as were the scores for parts C' and C''. During Fall 2008, we observed a significant

Table 3. Sample student written responses for part B'' of the PhAT

Part B''. Explain the reasoning that you used in your answer for part B'. In other words, how did you decide that one tree was better than the other?

"Evolution of any kind is unlikely, so by Occam's Razor the tree with the fewest events is most likely. Occam's Razor states that the solution with the fewest assumptions is usually right."

"The total change in traits in phylogenetic tree I is 8 changes, in tree II it is 7, the principal [*sic*] of parsimony says the simplest explanation is most likely the correct one, seven changes is simpler than 8 so tree II was chosen."

"The reasoning is that [in] Tree I the animals would go through 8 evolutionary events and in tree II 9 events. Because animals tend to evolve in the least complex manor [*sic*], or with as few changes as possible, tree I makes more sense for these phyla because of the less events."

"Based on the rule of parsimony and Ockham's razor, which states a phenomena should 'make the least # of assumptions possible,' I based my decision by counting the evolutionary changes in each figure. Tree I has 8 while Tree II has 7. Since 7 has less, it is simpler and [therefore] better."

"Tree II has fewer evolutionary events than Tree I and thus follows the rule of parsimony more closely (which calls for fewest assumptions to be made and simplest explanations to be explored first)."

association between gains on parts B and C of the PhAT (χ^2 test: $p < 0.05$), while in Fall 2009, there was a significant association between gains on parts A and C of the PhAT (χ^2 test: $p < 0.05$; Table 4). None of the other tests for association was significant.

Comparisons across semesters indicated that observed learning gains on the PhAT as a whole, and for parts A, B', and B'', were significantly higher during Fall 2009 than they were during Fall 2008. During Fall 2009, we observed a 35.28% gain in part A as compared with a gain of 26.42% during Fall 2008; for part B', we observed a gain of 6.13% during Fall 2009 as compared with a 7.80% loss during Fall 2008; and for part B'', there was a gain of 75.77% in Fall 2009 compared with 59.51% in Fall 2008 (Figure 4). All of these gains/losses were significant (t test: $p < 0.001$). On the other hand, the observed learning gains on parts C' and C'' were not significantly different between Fall 2008 and Fall 2009 (t test: $p > 0.1$).

DISCUSSION

What Students Learned about Tree Thinking and Evolution

Students during both semesters made significant gains in their abilities to "map" characters parsimoniously onto two competing phylogenies (alternative hypotheses) in part A of the PhAT. Gains were modest, however, and even during Fall 2009, when the gain in part A was highest, mean posttest score was only 63.5%, indicating that many students still lacked mastery in mapping characters onto trees after instruction. By far the most common mistake made in part A was for students to indicate multiple parallel gains on terminal branches rather than indicating a single trait gain more basally in the tree, followed by a subsequent loss, which would have resulted in fewer evolutionary changes overall.

Students needed to have three pieces of information when thinking about how to map each individual character onto

Table 4. Correlations between learning gains on different parts of the PhAT

LB144 Fall 2008 ($n = 205$)			LB144 Fall 2009 ($n = 163$)		
A vs. B ($p = 0.405$)			A vs. B ($p = 0.823$)		
Part B			Part B		
Part A	No gain	Gain	Part A	No gain	Gain
Gain	44	89	Gain	21	108
No gain ^a	28	44	No gain ^a	5	29
A vs. C ($p = 0.104$)			A vs. C ($p = 0.043$)		
Part C			Part C		
Part A	No gain	Gain	Part A	No gain	Gain
Gain	87	46	Gain	66	63
No gain ^a	55	17	No gain ^a	24	10
B vs. C ($p = 0.030$)			B vs. C ($p = 0.255$)		
Part C			Part C		
Part B	No gain	Gain	Part B	No gain	Gain
Gain	86	48	Gain	73	64
No gain ^a	56	15	No gain ^a	17	9

^aIncludes zero and losses.

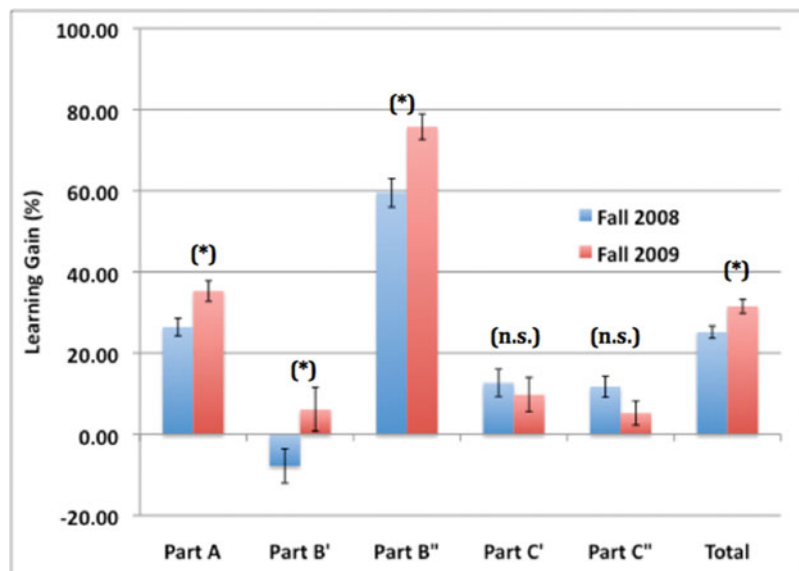


Figure 4. Comparison of learning gains (%) on the PhAT during Fall 2008 ($n = 205$) and Fall 2009 ($n = 163$). Significant increases in learning gains were observed between Fall 2008 and Fall 2009 for parts A, B', and B'' of the PhAT, as well as for the total score on the PhAT. Significance of differences across years was determined using t tests (two-tailed). *, significant at $\alpha = 0.05$; **, significant at $\alpha = 0.01$; ***, significant at $\alpha = 0.001$; n.s., not significant.

the tree: 1) for any single character the best mapping provides the fewest number of changes; 2) in general, gains are more likely than losses, but losses are possible; 3) the outgroup indicates the ancestral condition and provides a polarity for the gain and loss of traits. This final point seems particularly important for the mapping of the large canine teeth and perhaps provides an explanation of why students gave more steps to tree II. Our sense is that the instructional activities used in the courses did not make these three points explicit. Our data indicate that students did not learn these concepts as well as we might have liked. This appears to be a strong finding within our study, and our results can inform future instruction, pointing toward an emphasis on explicit instruction and practice with respect to gains and losses and as seeking the most parsimonious character reconstruction possible.

The strongest learning gains observed during both Fall 2008 and Fall 2009 occurred for part B of the PhAT, with 85.28% of the students indicating the principle of parsimony in part B' as the reason to "prefer" one tree (hypothesis) over the other during Fall 2009. Thus, the students appear to have learned that different tree topologies can be interpreted as alternative hypotheses, and that parsimony (Occam's razor) can be used to decide which of two alternative hypotheses is better supported by evidence (observational data). Parsing out the two halves of part B indicated that almost none of the observed gain occurred with respect to being able to choose the most parsimonious tree (6.13% gain; Table 2). This was caused, in part, by virtue of students incorrectly indicating tree I as the preferred tree in part B of the PhAT, because they mapped traits incorrectly in part A. Many of these students subsequently applied appropriate tree-thinking logic in part B'' and received full credit for this part of the PhAT. In addition, many students chose tree I as the better tree (an incorrect response), presumably because "dog" and "cat," which are shown as sisters in tree I, have all of the same characteristics in the data table. It would be interesting to test the effect of this aspect of the data as a distracter for students.

Despite strong gains in both parts A and B of the PhAT during both Fall 2008 and Fall 2009, students had a difficult time learning/understanding the concept that taxa are grouped together on a phylogenetic tree by virtue of how recently they shared a common ancestor. Although students were able to indicate which taxa were closest relatives on a tree (part C'), they were not able to indicate why they should be considered closest relatives, and the concept of (recency of) common ancestry was rarely mentioned in posttest responses (Table 2). Most students instead indicated that organisms are more closely related, because they are most similar to each other.

Inference of relatedness by virtue of similarity is common ("Tree-thinking Misconception #4" in Gregory, 2008) and may hinder students' ability to see the representations of ancestor-descendant relationships depicted by trees. Catley *et al.* (2010) also found that few of their students used the terms "ancestor" or "descendent" to describe aspects of phylogenies, and even fewer used the term "common ancestor." This problem may be compounded when students learn to carry out phylogenetic analysis using molecular sequence data. In many introductory molecular phylogenetics lessons, students build similarity/distance matrices and construct genetic distance-based dendrograms (e.g., Maier, 2001; Flammer, 2007), rather than taking a cladistic approach with its emphasis on shared ancestry. In lessons such as these, it needs to be made clear to students that the similarity used to build trees has arisen due to shared ancestry (the organisms are similar, because they are related), an important point that is easy to omit.

Comparing Student Evolutionary Understanding across Semesters

We found significantly larger learning gains on both parts A, B', and B'' of the PhAT during Fall 2009 as compared with Fall 2008. Thus, the teaching team may have done a better job of teaching students how to map characters onto trees and apply the principle of parsimony during Fall 2009 than

during Fall 2008, perhaps through the application of scientific teaching approaches (Handelsman *et al.*, 2007). Partly in response to documentation of the small positive impacts of course activities on student evolutionary understanding during Fall 2008, instructors made changes to the course for Fall 2009. In fact, the entire semester was “infused with evolution,” with the goal of linking each class session to some aspect of evolutionary theory. Although not all of these linkages were *phylogenetic*, many of them were. Additionally, the Fall 2009 students individually read a scientific journal article about phylogenetics (Li *et al.*, 2008), prepared answers to questions about the article, and then had a group discussion of the paper and its implications in class. Although not tested directly, these instructional interventions implemented during Fall 2009 may have contributed to the observed differences in student learning gains on parts A and B of the PhAT between Fall 2008 and Fall 2009. This instructional feedback loop provides an example of how assessing the impact of instructional activities on student learning through the use of a pre/posttest design can positively impact teaching and learning, and how assessment with the PhAT can be used to guide instruction.

The PhAT as an Assessment Tool

While the PhAT may provide a useful tool for assessing aspects of tree thinking, it is appropriate to ask exactly which aspects of tree thinking are measured by the PhAT, and how. For example, we know that students improved in their abilities to map traits onto phylogenetic trees (a tree-thinking skill). What is less clear is what it means when students are unable to map traits onto a phylogenetic tree. What do they fail to grasp? In these cases, it is unclear whether the error is in transferring and/or translating the elements of the character matrix to the tree, students do not understand the concept of a synapomorphy, or students simply do not know they are being asked to map traits *parsimoniously*. Future work with the PhAT will have to take into consideration the fact that the ability to map characters (parsimoniously or not) is a skill in and of itself and could be assessed separately from the ability to map characters parsimoniously.

One weakness with the PhAT as implemented was that part B', which was scored as correct if the students chose tree II and incorrect if they chose tree I, did not allow us to give full credit to students who chose tree I based upon a parsimony criterion. Many students (close to 30% in the Fall 2008 posttests) mapped characters onto tree II nonparsimoniously; chose tree I as the best tree, because it required fewer evolutionary steps; and then received partial or full credit for their rationale for choosing their tree. This was the reason we originally split out parts B' and B'' as separate items and may explain the lack of observed learning gains for part B' in both Fall 2008 and Fall 2009. It was also possible for students to select the right tree in part B', but for the wrong reason. This was a less serious issue, and it only occurred in ~5% of the Fall 2008 posttests.

Finally, as noted by other authors (e.g., Catley *et al.*, 2010), the concept of common ancestry as the basis of evolutionary relatedness appears to be a difficult one for students to grasp. Using the PhAT, we found that students continuously indicated that relatedness was due to similarity (not common ancestry), a misconception that persisted even after instruc-

tion. The problem may have been exacerbated in our case due to our choice of phylogenetic trees for comparison. Tree II, which is the more likely tree than tree I (via parsimony criteria), does not show “dog” and “cat” as sister taxa, despite the fact that the character matrix clearly shows these two taxa sharing all characteristics. The grouping in tree II may thus simply be counterintuitive for students, given the character matrix. Future use of the PhAT could include showing two trees that are consistent both intuitively and with respect to parsimony.

The PhAT also does not appear to be an isomorphic measure of a single entity. While significant positive associations were observed between student responses to parts B (B' and B'' pooled) and C (C' and C'' pooled) of the PhAT during Fall 2008 and between parts A and C in Fall 2009, these associations were neither strong nor consistent. Further work needs to be undertaken to determine whether there is an association between a student's ability to map characters correctly onto a phylogenetic tree, an understanding of parsimony as a criterion for evaluating the strength of hypotheses, and/or a student's understanding of a tree as a depiction of ancestor-descendant relationships.

CONCLUSIONS

Evolutionary (or phylogenetic) trees provide an excellent framework for students to organize information about groups of organisms within an evolutionary context. Phylogenies provide explicit hypotheses about ancestor-descendant relationships. In addition, although not explored in the current study, phylogenies can be used to incorporate a temporal component to the evolutionary process (assuming equal evolutionary rates), and they provide a way to help students see how biological traits are shared across taxonomic groups. Phylogenies can also move evolutionary biology lessons from the level of microevolution to the level of macroevolutionary change, which is often missing in standard curricula (Padian, 2008). Finally, by using parsimony as an optimality criterion, phylogenies provide an excellent avenue for students to conduct hypothesis testing in introductory organismal biology courses.

Using the PhAT, we found that the tree-thinking curriculum within the introductory organismal biology course at the LBC at MSU (Smith and Cheruvilil, 2009) provided students with the tools necessary to improve their tree-thinking skills. The PhAT provides a short, open-ended tool that can be used to assess student ability to map characters on phylogenetic trees; apply an objective criterion to decide which of two trees (alternative hypotheses) is “better”; and demonstrate an understanding of phylogenetic trees as depictions of ancestor-descendant relationships. In the two semesters studied here, we observed significant learning gains on each aspect of the PhAT, with the strongest learning gains observed for students' abilities to map characters onto phylogenetic trees and apply the principle of parsimony to evaluate competing hypotheses. The PhAT can be readily adaptable to other courses that include phylogenetic trees and tree thinking and can form the basis of future research to provide deeper insight into student learning and understanding of phylogenetic trees.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Wei Wang of the MSU College of Agriculture and Natural Resources Statistical Consulting Service for his assistance; Drs. Cheryl Murphy and Gerald Urquhart for allowing us to administer the pre/posttests in their courses; and the students of LB144 Fall 2008 and Fall 2009 for taking part in this study. We thank Dr. Louise Mead for her comments on the manuscript and for her assistance with exploring the nature of the signal present in student responses to the PhAT. We thank four anonymous reviewers for their constructive comments on earlier versions of the manuscript, which greatly improved the final product. We also thank our colleagues in the TREE Working Group at NESCent, and the ASM Biology Scholars Program for guidance and feedback during the development of the PhAT.

REFERENCES

- Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach* 39, 952–978.
- Baum DA, Smith SD, Donovan SSS (2005). The tree-thinking challenge. *Science* 310, 979–980.
- Bloome D, Puro P, Theodorou E (1989). Procedural display and classroom lessons. *Curric Inq* 19, 265–291.
- Brewer S (1996). A problem-solving approach to the teaching of evolution. *Bioscene* 22, 11–17.
- Campbell NA, Reece JB, Urry LA, Cain ML, Minorsky PV, Wasserman SA, Jackson RB (2008). *Biology*, 8th ed., Saddle River, NJ: Benjamin Cummings.
- Catley KM, Novick LR, Shade CK (2010). Interpreting evolutionary diagrams: when topology and process conflict. *J Res Sci Teach* 47, 861–882.
- College Board (2011). *AP Biology Curriculum Framework 2012–2013*, Princeton, NJ: College Board.
- Darwin C (1859). *On the Origin of Species by Means of Natural Selection*, London: Charles Murray.
- Flammer L (2007). Whale Ankles and DNA: Whale Connections ... Discovered. www.indiana.edu/~ensiweb/lessons/wh.a%26d.les.html (accessed 1 July 2011).
- Giese AR (2005). Using inquiry and phylogeny to teach comparative morphology. *Am Biol Teach* 67, 412–417.
- Gregory TR (2008). Understanding evolutionary trees. *Evol Educ Outreach* 1, 121–137.
- Halverson KL (2011). Improving tree-thinking one learnable skill at a time. *Evol Educ Outreach* 4, 95–106.
- Handelsman J, Miller S, Pfund C (2007). *Scientific Teaching*, New York: Freeman.
- Johnson DW, Johnson R, Holubec E (1998) *Advanced Cooperative Learning*, 3rd ed., Edina, MN: Interaction Book Company.
- Julius ML, Schoenfuss HL (2006). Phylogenetic reconstruction as a broadly applicable teaching tool in the biology classroom: the value of data in estimating likely answers. *J Coll Sci Teach* 35, 40–45.
- Li C, Wu X, Rieppel O, Wang L, Zhao L (2008). An ancestral turtle from the Late Triassic of southwestern China. *Nature* 456, 497–501.
- Maier CA (2001). Building phylogenetic trees from DNA sequence data: investigating polar bear and giant panda ancestry. *Am Biol Teach* 63, 642–646.
- Meir E, Perry J, Herron J, Kingsolver J (2007). College students' misconceptions about evolutionary trees. *Am Biol Teach* 69, 71–76.
- Meisel RP (2011). Teaching tree-thinking to undergraduate biology students. *Evol Educ Outreach* 3, 621–628.
- Meltzer DE (2002). The relationship between mathematics preparation and conceptual learning gains in physics: a possible "hidden variable" in diagnostic pretest scores. *Am J Phys* 70, 1259–1268.
- Novick LR, Catley KM, Funk DJ (2010). Characters are key: the effect of synapomorphies on cladogram comprehension. *Evol Educ Outreach* 3, 539–547.
- Padian K (2008). Trickle-down evolution: an approach to getting major evolutionary adaptive changes into textbooks and curricula. *Integr Comp Biol* 48, 175–188.
- Perry J, Meir E, Herron J, Stal D, Maruca S (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE Life Sci Educ* 7, 193–201.
- Preacher KJ 2001. Calculation for the chi-square test: an interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Available from <http://quantpsy.org/chisq/chisq.htm>.
- Singer F, Hagen JB, Sheehy RR (2001). The comparative method, hypothesis testing and phylogenetic analysis—an introductory laboratory. *Am Biol Teach* 63, 518–523.
- Smith JJ, Cheruvilil KS (2009). Using inquiry and tree-thinking to "March through the animal phyla." *Evol Educ Outreach* 2, 429–444.
- Weber E (2009). Quantifying student learning: how to analyze assessment data. *Bull Ecol Soc Am* 90, 501–511.