

Evo Edu Outreach (2012) 5:595–602
DOI 10.1007/s12052-012-0455-5

CURRICULUM AND EDUCATION ARTICLE

Teaching Tree Thinking to College Students: It's Not as Easy as You Think

Brenda C. Phillips · Laura R. Novick · Kefyn M. Catley · Daniel J. Funk

Published online: 18 October 2012
© Springer Science+Business Media New York 2012

Abstract The ability to understand and reason with tree-of-life diagrams (i.e., cladograms), referred to as tree thinking, is an essential skill for biology students. Yet, recent findings indicate that cladograms are cognitively opaque to many college students, leading them to misinterpret the information depicted. The current studies address the impact of prior biological background and instruction in phylogenetics on students' competence at two foundational tree-thinking skills. In study one, college students with stronger ($N=52$) and weaker ($N=60$) backgrounds in biology were asked to (a) identify all the nested clades in two cladograms and (b)

evaluate evolutionary relatedness among taxa positioned at different hierarchical levels (two questions) and included in a polytomy (two questions). Stronger-background students were more successful than weaker-background students. In study two, a subset of the stronger-background students ($N=41$) who were enrolled in an evolution class subsequently received two days of instruction on phylogenetics. As expected, these students' tree-thinking skills generally improved with instruction. However, although these students did very well at marking the nested clades, fundamental misinterpretations of relative evolutionary relatedness remained. The latter was especially, although not exclusively, the case for taxa included in a polytomy. These results highlight the importance of teaching cladistics, as well as the need to tailor such instruction to the difficulties students have learning key macroevolutionary concepts.

B. C. Phillips · L. R. Novick · D. J. Funk
Vanderbilt University,
Nashville, TN, USA

K. M. Catley
Western Carolina University,
Cullowhee, NC 28723, USA

B. C. Phillips
SDR Lab, Maxwell Dworkin 136, School of Engineering
and Applied Science, Harvard University,
Cambridge, MA 02138, USA

L. R. Novick (✉)
Department of Psychology and Human Development,
Vanderbilt University,
230 Appleton Place, Peabody #552,
Nashville, TN 37203-5721, USA
e-mail: Laura.Novick@vanderbilt.edu

K. M. Catley
Department of Biology, Western Carolina University,
132 Natural Science Building,
Cullowhee, NC 28723, USA

D. J. Funk
Department of Biological Sciences, Vanderbilt University,
Station B, Box 351634, Nashville, TN 37235-1634, USA

Keywords Evolution education · Tree of Life · Tree thinking · Cladograms · Scientific reasoning

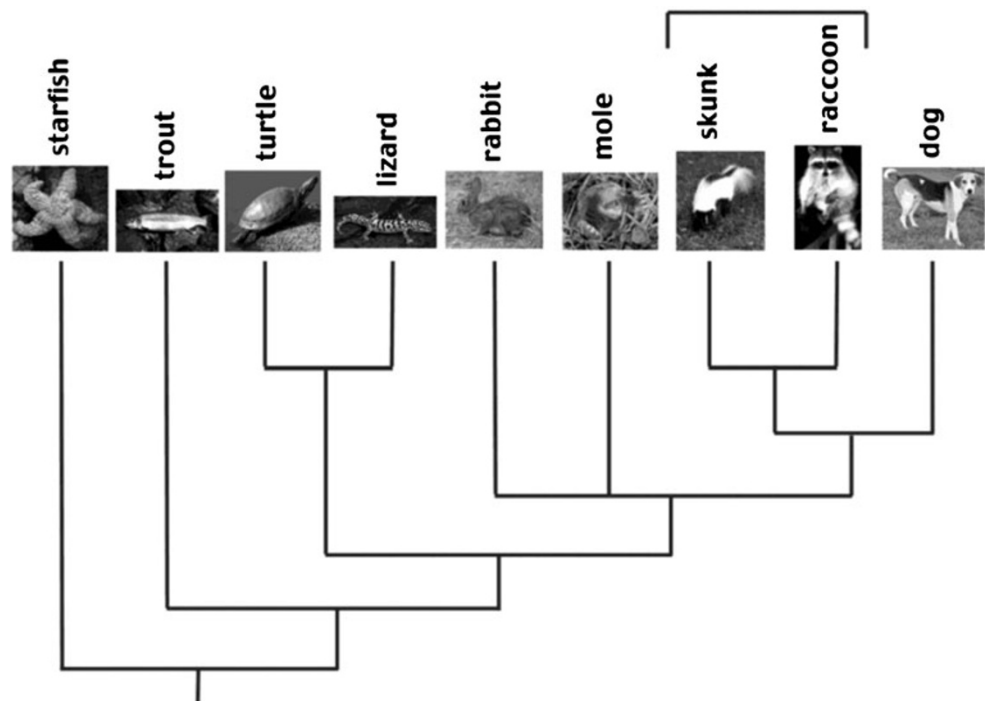
Phylogenies depicting the evolutionary relationships among extant and extinct taxa are central to the study of modern biology. Professional biologists use phylogenies to map shared characters among biological groups over historical time. Phylogenies are used by basic researchers to address fundamental questions regarding the history and diversity of life on Earth and by applied researchers to, for example, track and cure global emergent diseases such as HIV, influenza, and the West Nile virus (American Museum of Natural History 2002; Ducatez et al. 2006; Sharp and Hahn 2010; Yates et al. 2004). Given the importance of evolution in contemporary biology (National Research Council 2009), Thanukos (2010, p. 563, emphasis in the original) has argued that “to grasp modern biology, students *must* understand the basics of phylogenetics.”

Although evolutionary diagrams—specifically cladograms—depicting subsets of the Tree of Life are common in college introductory biology textbooks (Catley and Novick 2008), students are often not taught how to reason about the evolutionary relationships depicted (i.e., are not taught tree thinking), nor are they provided with sufficient information regarding the theory and processes on which phylogenies are based. It is not surprising, therefore, that prior research indicates that macroevolutionary misconceptions abound among college biology students (Gregory 2008; Meir et al. 2007; Novick and Catley 2007, 2012; Sandvik 2008; Shtulman and Schulz 2008).

Novick and Catley (2012) identified five core tree-thinking skills that are essential for understanding and reasoning with cladograms: (1) identifying characters (i.e., synapomorphies) that are inherited from a most recent common ancestor (MRCA) and shared by two or more taxa, (2) identifying a set of taxa that either do or do not share a specific character, (3) understanding the concept of a clade or monophyletic group (i.e., a group comprising an MRCA and all of its descendants), (4) evaluating relative evolutionary relatedness among a set of taxa, and (5) using evidence of most recent common ancestry to support inferences. Novick and Catley found that college students who have a rudimentary understanding of macroevolution and/or hierarchically organized diagrams may reason correctly using skills one and two despite having had little content-specific instruction in phylogenetics. However, skills three and four, which reflect core concepts concerning a cladogram's structure that are vital for tree thinking, present greater difficulty (also see Catley et al. 2012).

Accordingly, we focused on these two skills in the present research. The questions in Novick and Catley's (2012) tree-thinking assessment for skills three and four were relatively basic. For example, one type of question for skill three asked students whether (and why or why not) a marked set of two or three taxa comprise a clade in the given cladogram. Imagine for Fig. 1 that the bracket at the top of the figure enclosed skunk, raccoon, and dog, which comprise a three-taxon statement, rather than just skunk and raccoon, which comprise a sister group. Those three taxa comprise a clade because they include all the depicted descendants of their MRCA. Although students usually correctly identified a three-taxon statement as comprising a clade, especially after instruction in phylogenetics (Catley et al. 2012), Novick and Catley reported one consistent error among those who got such questions wrong: Some students wrote that the three taxa do not comprise a clade because the bracketed group includes more than just two taxa that are most closely related. Those students then tended to indicate that the way to make the group a clade is to remove the least related taxon (dog in the example based on Fig. 1). These results suggest that students may have difficulty understanding clades at a deeper level, specifically in the usual case in which they are nested such that the cladogram includes many clades. Accordingly, we evaluated students' understanding of the nesting of clades in the present studies by asking them to identify and mark all the clades in a given cladogram. We are not aware of any published research on tree thinking that has investigated students' understanding of the nesting of clades.

Fig. 1 Cladogram of deuterostome relationships with taxa oriented horizontally. Students received a version of this cladogram that included color photographs



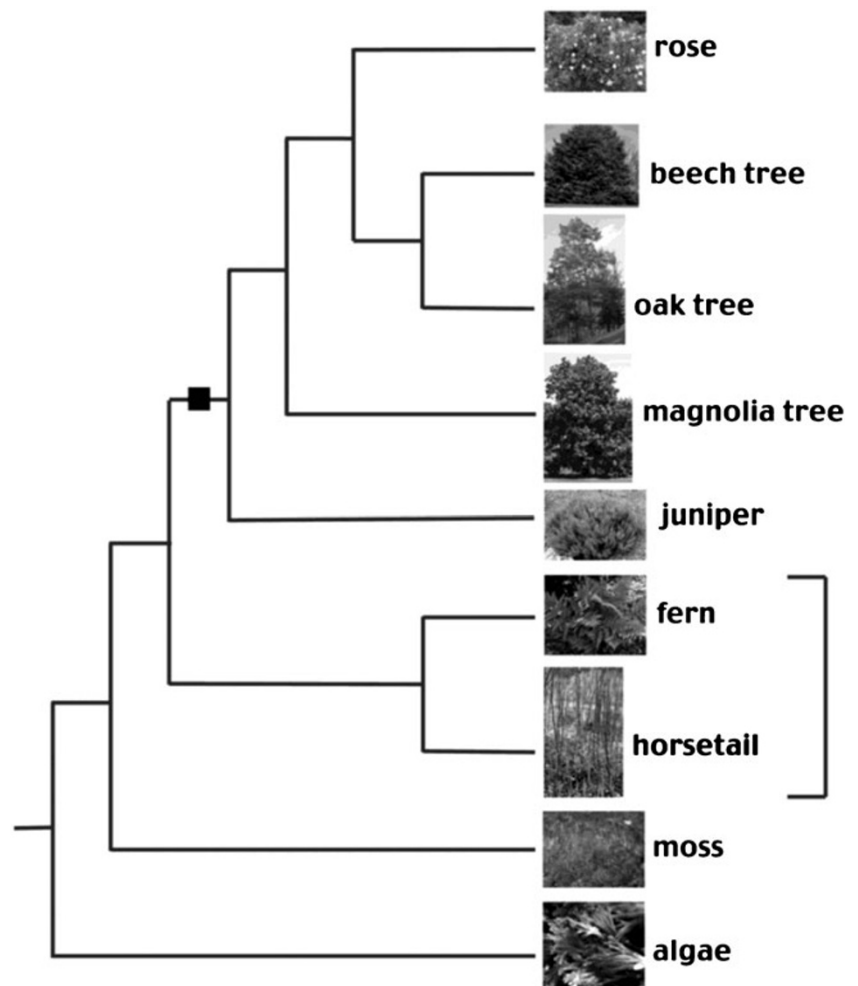
Novick and Catley (2012) also found that students have difficulty evaluating which of two taxa is the closest evolutionary relation to a third, reference taxon when the reference taxon is at an intermediate hierarchical level between the two comparison taxa. Using the taxa in Fig. 1, a comparable question is whether the trout or the skunk is the closest evolutionary relation to the lizard, and why. To appreciate the implications of the hierarchical arrangement of these taxa for understanding evolutionary relatedness, it is necessary to discern which pair of taxa has a more recent common ancestor. In this case, the lizard is more closely related to the skunk than to the trout because it shares a more recent common ancestor with the skunk than with the trout. Novick and Catley computed composite evolutionary relatedness scores to assess student understanding by averaging across accuracy (0, 1) and explanation quality (0, 0.5, 1). They found that for their sample of students who had taken the two-semester introductory biology course for majors plus at least one to five other biology courses, the mean composite score was only 0.44 for a question similar to the example given here. Catley et al. (2012) found that after instruction in phylogenetics in a zoology or evolution class,

the mean composite score increased to 0.61. The current studies provide a further evaluation of students' ability to assess evolutionary relatedness by examining their understanding of the relationships among taxa in both hierarchical and polytomous arrangements. The hierarchical arrangements enable a replication of the Catley et al. study with new cladograms and questions and serve as a comparison for the analysis of polytomous arrangements, which is new to the present research.

Overview of Studies

The present studies examined two core tree-thinking skills: identifying nested clades and evaluating evolutionary relatedness. Students answered questions about four cladograms, which were oriented either horizontally or vertically (see Figs. 1 and 2, respectively). Both cladogram orientations are found in college textbooks (Catley and Novick 2008). Although we are not aware of any arguments suggesting that one orientation might be easier for students to understand than the other, it would be important to know about

Fig. 2 Cladogram of plant relationships with taxa oriented vertically. Students received a version of this cladogram that included color photographs



such differences if they exist. Therefore, we manipulated cladogram orientation in the present studies.

A further critical aspect of the present research concerned the effectiveness of instruction in phylogenetics given in biology classes. In what ways is such instruction effective and in what ways does it need to be strengthened if biology instructors are to realize their goal of producing students who are competent at tree thinking? Study one included a sample of college students with stronger and weaker backgrounds in biology. The stronger-background students had previously completed, at minimum, a two-semester introductory biology course for biology majors. Phylogenetics is covered, briefly, in the second semester of that course. Study two included a subset of stronger-background students from study one who were recruited from an intermediate level course on evolution. Those students were tested before and after receiving two days of phylogenetics instruction in the evolution class.

Study One

Method

Subjects We tested 112 Vanderbilt University undergraduates. Sixty-nine students (34 females, 33 males, two undisclosed sex) were recruited from a paid subject pool coordinated by the psychology department. The remaining 43 students (23 females, 20 males) were enrolled in the course on evolution taught by the fourth author.

Students were divided into two groups based on their biology background: The stronger-background group had completed at least the two-semester introductory biology sequence for biology majors and premedical students, which included one to two basic lectures on phylogenies in the second-semester course; the remaining students were assigned to the weaker-background group. The 52 stronger-biology background students (28 females, 24 males) had completed an average of 3.02 semesters of biology courses that were included on a list of primarily organismal biology classes. Of these 52 students, 44 were currently enrolled in or had previously completed at least one course beyond the introductory sequence. The 60 weaker-background students (29 females, 29 males, two undisclosed sex) had completed an average of 0.28 semesters of such coursework.

Design and Procedure All students received a four-page booklet that included a cladogram (printed in color) and several questions about that cladogram on each page. Each of the four cladograms featured a different set of nine taxa, including a focal taxon (human, honeybee, dog, or rose), so named because the first question on each page asked students to explain what the diagram shows about the evolution of that taxon. For the 107 tree cladograms found in the

college introductory biology textbooks analyzed by Catley and Novick (2008), the mean number of taxa is 8.7 (Median=7; range of 2–79). Thus, our cladograms are comparable in size to those that would have been previously encountered by the students who participated in our study. Figures 1 and 2 show versions of the deuterostome and plant cladograms, respectively. The remaining two (metazoan and bilaterian) cladograms had topologies that are very similar to those shown in these figures. The four cladograms were presented in counterbalanced order across subjects. Cladogram orientation was manipulated between subjects, with students randomly assigned to receive either four horizontal or four vertical cladograms.

Students completed the booklet for this study as the first of three booklets addressing distinct conceptual questions concerning students' ability to engage in tree thinking. Students completed the booklets without using any outside resources during a single session that lasted approximately 50–75 min.

Nested Clades Questions Students were asked to mark all the clades depicted in the plant and deuterostome cladograms. For example, students received the following question for the deuterostome cladogram in Fig. 1:

A clade is a group of taxa that includes the most recent common ancestor of the group and all descendants of that ancestor. For example, the skunk and the raccoon represent a clade in the diagram. How many clades are there in this diagram (including the one already marked)? Mark each additional clade with a bracket as shown in the example.

There are seven clades in the deuterostome cladogram and eight in the plant cladogram (Fig. 2).

Evolutionary Relatedness Questions We probed students' understanding of evolutionary relatedness in two situations—when taxa are located at different hierarchical levels and when they are included in a polytomy (i.e., when three or more branches diverge from the same node). For example, for the plant cladogram in Fig. 2, students were asked “Which taxon—fern or oak tree—is the closest evolutionary relation to the juniper?” These three taxa are located at different levels in the cladogram, with the juniper occupying a hierarchically intermediate position relative to the fern and the oak tree. Although the juniper is closer to the fern than to the oak tree in this particular cladogram if one counts the number of “steps” (i.e., branching points) between them, it is more closely related to the oak tree because it shares a more recent common ancestor with that taxon. Students answered two evolutionary relatedness questions of this type (one each for the plant and bilaterian cladograms). For each question, they were also asked to provide a written explanation for their answer.

The second type of evolutionary relatedness question asked students to evaluate the relationships among three taxa that comprise a polytomy. For example, for the deuterostome cladogram in Fig. 1, students were asked:

Which of the following three statements (A, B, or C) is best supported by the scientific evidence: A. Moles are more closely related to rabbits than to raccoons; B. Moles are more closely related to raccoons than to rabbits; C. Rabbits, moles, and raccoons are all equally closely related to each other. These three taxa diverge from the same node in the tree and thus share the same MRCA. Therefore, based on this topology, the correct answer is that rabbits, moles, and raccoons are all equally closely related to each other. Students answered two evolutionary relatedness questions about polytomies (one each for the deuterostome and metazoan cladograms) and gave written explanations for their answers.

Results and Discussion

Understanding Nested Clades Students’ understanding of nested clades was assessed by calculating the average proportion of correctly marked clades across the two cladograms. The results of a two-(biology background; between)by-two (cladogram orientation; between) analysis of variance (ANOVA) indicated a main effect of biology background, $F(1, 108)=17.20, p<0.001, MSE=0.13, \eta_p^2 = 0.14$, with stronger-background students having higher accuracy scores than weaker-background students (see Table 1). Although stronger-background students successfully identified almost twice as many clades as weaker-background students, they still only managed to identify 62% of the clades on average. Clearly, the nesting of ancestry and clades is a challenging concept for students to understand. These results extend prior work showing that some college students have difficulty understanding nested clades in three-taxon statements (Novick and Catley 2012). Neither the main effect of cladogram orientation nor the biology background by orientation interaction were significant: $F(1, 108)=0.01, p>0.90, \eta_p^2 = 0.00$, and $F(1, 108)=0.27, p>0.60, \eta_p^2 = 0.00$, respectively.

Scoring Students’ Evolutionary Relatedness Responses Students received a score for accuracy (0 or 1) and for explanation quality (0, 0.5, or 1) for each evolutionary relatedness question. Composite scores (mean of accuracy and explanation quality) were then calculated across the two questions testing each aspect of understanding evolutionary relatedness (different levels and polytomy). Responses to the two question types were analyzed separately. The means are shown in Table 1.

The coding scheme used to categorize students’ written explanations was modeled on a scheme previously used to assess tree-thinking skills in college students (Novick and Catley 2012) and was verified for its appropriateness for the present research by the first and second authors based on the responses of a randomly selected subset of 20 students. The scheme included components of what would be considered a scientifically valid response as well as naïve responses that focused on erroneous factors. Eight coding categories were arranged from most to least indicative of student comprehension (i.e., in decreasing order of sophistication), and each code was assigned an explanation quality score. The first author and a research assistant independently coded the remaining written responses from both studies. The responses from the two studies were randomly intermixed, and the coders were unaware for each response whether it came from a weaker- or stronger-background student (Study 1) or from before or after instruction (Study 2). Each response received a single code (responses that met the criteria for more than one code were given the highest code in the ordering). The coders agreed on 464 responses out of a total of 532 across the two studies, for an agreement rate of 87%. Disagreements were resolved by discussion.

The best response, which received a quality score of one, was to appeal to the MRCA of the taxa in question. For example, a stronger-background student explained that rabbit, mole, and raccoon are equally closely related in Fig. 1 as follows: “As before, I believe that moles, rabbits, and raccoons are equally related as they branch from the same most recent common ancestor, ancestor X.” Another such student explained for Fig. 2 that the juniper is more closely related to the oak tree because “they share the more recent common ancestor than juniper and fern.”

Table 1 Students’ mean tree-thinking scores for the clade questions (mean proportion of clades marked correctly across two questions) and evolutionary relatedness questions (mean of accuracy and explanation quality across two questions)

Tree-thinking skill	Study 1		Study 2	
	Weaker biology background	Stronger biology background	Before instruction	After instruction
Nesting of clades	0.33	0.62	0.63	0.90
Relative evolutionary relatedness—different levels	0.19	0.42	0.45	0.59
Relative evolutionary relatedness—polytomies	0.12	0.24	0.28	0.47

A quality score of 0.5 was given for two types of explanations: (a) those indicating that certain taxa share a recent common ancestor without specifying that it is the *most recent* common ancestor and (b) those indicating that the taxa are most or more closely related evolutionarily. An example of the first type of explanation is: “They share the same recent common ancestor and are on the same branch of the diagram” (weaker background, Fig. 1). An example of the second type of explanation is: “Fern is a closer evolutionary relation with the juniper because it has had less divergences from the shared common ancestor” (stronger background, Fig. 2). All other explanations received a quality score of 0. These explanations mentioned, for example, that certain taxa share a common ancestor (all taxa share a common ancestor, so this response is uninformative), that there are fewer steps between one pair of taxa than another, that certain taxa are more closely connected, or that the pictures of certain taxa are closer to each other.

Assessing Relatedness at Different Hierarchical Levels The results of a two-by-two ANOVA on students’ composite scores revealed a main effect of biology background, again indicating that stronger-background students did better, relatively speaking, than weaker-background students, $F(1, 108)=15.18, p<0.001, MSE=0.10, \eta_p^2 = 0.12$. Yet only 15% of stronger-biology-background students received both an accuracy score of one and an explanation quality score of one across the two questions. Clearly, only a minority of these students relied on the critical macroevolutionary concept of most recent common ancestry to evaluate the relationships depicted in these cladograms. These results indicate that even students who receive college-level instruction in the biological sciences fail to perform well on questions that evaluate their understanding of evolutionary relatedness when the reference taxon occupies a hierarchically intermediate position relative to the comparison taxa.

There was no main effect of cladogram orientation, $F(1, 108)=0.00, p>0.95, \eta_p^2 = 0.00$. However, there was an interaction between biology background and cladogram orientation, $F(1, 108)=4.15, p<0.05, \eta_p^2 = 0.04$, as stronger-background students did better when the cladogram was oriented vertically as opposed to horizontally ($M=0.48$ vs. $M=0.35$, respectively), whereas weaker-background students did better when the cladogram was oriented horizontally rather than vertically ($M=0.24$ vs. $M=0.12$, respectively). It is unclear how differences in biological knowledge contribute to this pattern of responses. Nevertheless, the overall quite poor tree-thinking scores in all conditions indicate that both groups of students weighed the importance of the relative number of steps between the taxa and cladogram orientation, both phylogenetically irrelevant factors, when assessing the degree of relatedness among taxa. Meir et al. (2007) and

Novick and Catley (2012) also found that college students incorrectly think the relative number of steps separating taxa reflects the degree of evolutionary relatedness.

Assessing Relatedness in a Polytoomy A two-by-two ANOVA on students’ composite scores revealed a main effect of biology background, $F(1, 108)=10.21, p<0.01, MSE=0.04, \eta_p^2 = 0.09$, with better performance by stronger- than weaker-background students. Neither the main effect of cladogram orientation nor the biology background by orientation interaction were significant— $F(1, 108)=0.76, p>0.35, \eta_p^2 = 0.01$, and $F(1, 108)=0.05, p>0.80, \eta_p^2 = 0.00$, respectively. Thus, for the polytoomy evolutionary relatedness questions, stronger- and weaker-background students did not perform differently as a function of cladogram orientation. Students’ mean composite scores were extremely low overall (see Table 1), and only four students (two stronger and two weaker background) used the concept of most recent common ancestry to justify a correct polytoomy evolutionary relatedness response, each for only one of the two questions.

Study Two

Study two examined the extent to which deficiencies in tree thinking are easily amenable to instruction. A subset of the stronger-biology-background students from study one were tested prior to and following two in-depth lectures on phylogenetics in an evolution class.

Method

Subjects We tested 41 Vanderbilt University students (22 females, 19 males) who were enrolled in the intermediate-level Evolution course taught by the fourth author. These students comprised a subset of the stronger-biology-background students included in study one. Thus, they had all previously completed at least the two-semester introductory biology sequence. When they were first tested at mid-semester, they had completed an average of 3.23 biology courses on our list. When retested at the end of the semester, they had completed an average of 3.93 courses (assuming they passed all the biology courses in which they were currently enrolled).

Design and Procedure The design and procedure were the same as for study one. Students completed the same (horizontal or vertical orientation) cladogram booklet immediately before and four-and-a-half to five weeks after instruction in phylogenetics. The pretest was given just after the midpoint of the semester; the posttest was given at the end of the semester. The instruction included two lectures on phylogenetic theory

and terminology that covered the following concepts: characters (e.g., synapomorphies), character states (ancestral vs. derived), and character-based evidence for homology versus homoplasy; parsimony; sister groups (two taxa that share an MRCA) and monophyletic groups (i.e., clades) versus paraphyletic groups (groups that omit one or more descendants of the MRCA); polytomies and phylogenetic resolution; and structural equivalence of cladograms across rotation of branching points. These concepts were reinforced during subsequent lectures through the presentation and discussion of cladograms used to teach or illustrate other macroevolutionary concepts.

Results and Discussion

The dependent variables were computed as in study one. The means are given in Table 1. A comparison of the study two pre-instruction means to the stronger-background means in study one indicates that the study two sample is representative of the larger group from which it was drawn.

Understanding Nested Clades The results of a two-(cladogram orientation; between)by-two (time—before vs. after instruction; within) mixed ANOVA indicated that the proportion of clades marked correctly increased with instruction, $F(1, 39)=17.84$, $p<0.001$, $MSE=0.08$, $\eta_p^2 = 0.31$. Moreover, the absolute level of accuracy after instruction ($M=0.90$) indicated that students attained a good level of understanding that cladograms are composed of many nested clades. Neither the main effect of cladogram orientation nor the time by orientation interaction were significant— $F(1, 39)=0.34$, $p>0.55$, $MSE=0.13$, $\eta_p^2 = 0.01$, and $F(1, 39)=0.17$, $p>0.65$, $MSE=0.08$, $\eta_p^2 = 0.00$, respectively.

Assessing Relatedness at Different Hierarchical Levels A two-by-two mixed ANOVA on the composite scores did not reveal any significant effects— $F(1, 39)=3.39$, $p>0.05$, $MSE=0.13$, $\eta_p^2 = 0.08$, for the main effect of time; $F(1, 39)=0.80$, $p>0.35$, $MSE=0.22$, $\eta_p^2 = 0.02$, for the main effect of cladogram orientation; and $F(1, 39)=0.11$, $p>0.70$, $MSE=0.13$, $\eta_p^2 = 0.00$, for the interaction. Although the mean score was higher after instruction than before, the improvement was not large enough to be statistically significant. The mean score after instruction is comparable to what Catley et al. (2012) found in their instructional study for a similar sample of students. Even after instruction, only 41% of students received accuracy and evidence quality scores of one for both questions. At best, then, there was only a small improvement on different-levels evolutionary relatedness questions, and most students failed to demonstrate mastery after instruction.

Assessing Relatedness in a Polytoomy A two-by-two mixed ANOVA on the composite scores indicated only a main effect of instruction, $F(1, 39)=15.61$, $p<0.001$, $MSE=0.05$, $\eta_p^2 = 0.29$. Even after instruction, however, students did very poorly on these questions, with a mean score of only 0.47 on a zero-to-one scale. Moreover, the proportion of students who used the concept of most recent common ancestry (explanation score of one) to justify a correct evolutionary relatedness response for both questions increased from 0.00 before instruction to only 0.17 after instruction. Neither the main effect of cladogram orientation nor the time by orientation interaction were significant— $F(1, 39)=0.31$, $p>0.55$, $MSE=0.12$, $\eta_p^2 = 0.01$, and $F(1, 39)=0.23$, $p>0.60$, $MSE=0.05$, $\eta_p^2 = 0.01$, respectively. In sum, the results for both types of evolutionary relatedness questions highlight the difficulty biology students have using critical macroevolutionary concepts to reason about the relationships depicted in cladograms, even after instruction in this area.

General Discussion

The present research examined college students' understanding of two critical core tree-thinking concepts—most recent common ancestry and nested clades—as a function of instruction in biology generally and in phylogenetics specifically, the latter in an evolution course. We found, not surprisingly, that college students generally improved in their tree-thinking ability as a result of instruction. In study one stronger-background students, who had completed at least the two-semester introductory biology sequence for majors, did significantly better than weaker-background students on all three types of tree-thinking questions: nested clades, evolutionary relatedness at different levels, and evolutionary relatedness in a polytoomy. In study two stronger-background students enrolled in the evolution course did significantly better after instruction than before for the nested clades and polytoomy questions.

Students were most successful on the nested clades questions. After general biology instruction (study one), they had a mean proportion correct of 0.62; after phylogenetics instruction in the Evolution course (study two), accuracy increased to 0.90. It is perhaps relevant to note that these questions only required students to mark the clades; they did not also have to provide a written explanation for their responses as was required for the evolutionary relatedness questions. Nevertheless, the evolution students were highly successful at marking the nested structure of the cladograms after instruction.

However, these students were not nearly as successful at understanding the implications of this nested structure for determining evolutionary relatedness, indicating that the macroevolutionary concepts underpinning tree thinking are

difficult even for students with substantial training in the biological sciences. Success on the evolutionary relatedness questions was assessed by a composite score (on a zero-to-one scale) that included both accuracy of evaluating relative relatedness and use of the concept of most recent common ancestry to explain relatedness. After instruction in phylogenetics, there was no significant improvement in evolution students' performance on the questions concerning evolutionary relatedness at different levels, with an overall mean score across the pretest and posttest of only 0.52. Although there was significant improvement on the polytomy evolutionary relatedness questions, the average score after instruction was only 0.47. These results indicate that the concept of most recent common ancestry, which is essential for assessing relative relatedness among taxa, is difficult for college biology students to apply to evolutionary relationships represented diagrammatically in a cladogram. Instead, students attended to irrelevant presentation factors, especially the relative number of steps (i.e., branching points) between taxa, to reason about evolutionary relatedness.

Perhaps it is not surprising that the stronger-biology-background students in study one had difficulty with our tree-thinking questions, given that high school and college instruction on evolution focuses primarily on microevolution (e.g., natural selection) rather than macroevolution (Catley 2006). Yet even the stronger-background students in study two, who received two days of in-depth instruction in phylogenetics in a semester-long evolution course, in general failed to demonstrate a high level of competence at a foundational tree-thinking skill—the ability to assess relative evolutionary relatedness among taxa using the concept of most recent common ancestry—by the end of the course.

Ideally, students would acquire the core tree-thinking skills when they are first exposed to theories on microevolution and macroevolution. The results of the present research, however, as well as of previous studies by the present authors (Catley et al. 2012; Novick and Catley 2012), suggest that current instruction in college biology classes is not highly effective in this regard. We therefore hope that students' difficulties documented in these studies will be used to inform the design of instruction that may yield greater understanding of and ability to engage in tree thinking.

By representing the currently best-supported hypotheses regarding evolutionary relationships among taxa, cladograms provide critical information for biologists and other professionals in applied fields such as ecology, genomics, epidemiology, and pharmacology to support inferences that impact the welfare of our planet and of our own species (e.g., AMNH 2002; Ducatez et al. 2006; Sharp and Hahn 2010; Yates et al. 2004). Because of an increasing recognition of their utility for illuminating basic and applied issues from diverse fields of inquiry, cladograms are becoming ever more prevalently used. Clearly, then, the ability to accurately interpret these phylogenetic

diagrams and to reason correctly about the relationships depicted therein is an increasingly important aspect of scientific literacy (also see Gregory 2008). Our study demonstrates that these skills can be improved through instruction but that true mastery of them often remains elusive. We thus call for an increased focus on teaching and evaluating core tree-thinking skills in biology curricula.

Acknowledgments We thank Emily Schreiber and Marissa Mencio for their help with collecting the data and Emily Schreiber for her help in coding students' written justifications. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080621 to Vanderbilt University (Novick and Catley, co-investigators). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- American Museum of Natural History [AMNH]. 2002. Assembling the Tree of Life: harnessing life's history to benefit science and society. Brochure produced for the National Science Foundation based on three NSF Tree of Life workshops held in 1999 and 2000 at Yale University, the University of California Davis, and the University of Texas Austin (30 August 2010; www.phylo.org/sub_sections/outreach/outreach_b.php)
- Catley KM. Darwin's missing link: a new paradigm for evolution education. *Sci Educ.* 2006;90:767–83.
- Catley KM, Novick LR. Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. *BioSci.* 2008;58:976–87.
- Catley KM, Novick LR, Funk DJ. The promise and challenges of introducing tree thinking into evolution education. In: Rosengren K, Evans EM, Brem S, Sinatra G, editors. *Evolution challenges: integrating research and practice in teaching and learning about evolution.* New York: Oxford University Press; 2012. p. 93–118.
- Ducatez MF, Olinger CM, Owoade AA, De Landtsheer S, Ammerlaan W, Niesters HGM, Osterhaus ADME, Fouchier RAM, Muller CP. Avian flu: multiple introductions of H5N1 in Nigeria. *Nature.* 2006;442:37.
- Gregory TR. Understanding evolutionary trees. *Evo Educ Outreach.* 2008;1:121–37.
- Meir E, Perry J, Herron JC, Kingsolver J. College students' misconceptions about evolutionary trees. *Am Bio Teach Online.* 2007;69:e71–6.
- National Research Council. *A new biology for the 21st century.* Washington, DC: National Academies Press; 2009.
- Novick LR, Catley KM. Understanding phylogenies in biology: the influence of a Gestalt perceptual principle. *J Exp Psy Appl.* 2007;13:197–223.
- Novick LR, Catley KM (2012) Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *Am Educ Res J* (in press)
- Sandvik H. Tree thinking cannot be taken for granted: challenges for teaching phylogenetics. *Theory Biosci.* 2008;127:45–51.
- Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Phil Trans R Soc Lond B.* 2010;365:2487–94.
- Shulman A, Schulz L. The relation between essentialist beliefs and evolutionary reasoning. *Cogn Sci.* 2008;32:1049–62.
- Thanukos A. Evolutionary trees from the tabloids and beyond. *Evol Educ Outreach.* 2010;3:563–72.
- Yates TL, Salazar-Bravo J, Dragoo JW. The importance of the tree of life in society. In: Cracraft J, Donoghue MJ, editors. *Assembling the tree of life.* New York: Oxford University Press; 2004. p. 7–17.