

*Research Article***Deconstructing Evolution Education:  
The Relationship Between Micro- and Macroevolution**Laura R. Novick,<sup>1</sup> Emily G. Schreiber,<sup>1</sup> and Kefyn M. Catley<sup>2</sup><sup>1</sup>*Department of Psychology and Human Development, Peabody College #552,  
Vanderbilt University, 230 Appleton Place, Nashville, TN 37203-5721*<sup>2</sup>*Biology Department, Western Carolina University, Cullowhee, NC**Received 8 January 2014; Accepted 9 June 2014*

**Abstract:** With applications of Tree of Life data becoming ever more prevalent in everyday contexts, tree thinking has emerged as a vital component of scientific literacy. This article reports a study testing the hypothesis that instruction in natural selection, which is the primary focus of US evolution education at the high school and introductory college levels, does little to promote tree-thinking skills and that explicit instruction in understanding evolutionary tree diagrams is required. Testing this contention required the creation of a novel tree-thinking assessment and an instructional intervention, both guided by deep knowledge of evolutionary biology and of science education research. College students ( $N = 124$ ) with weaker versus stronger backgrounds in biology were randomly assigned to control versus instructional conditions and were also assessed for their knowledge of natural selection. Although knowledge of natural selection and ability to engage in tree thinking were correlated, a short instructional booklet that provided a basic introduction to evolutionary trees predicted tree-thinking success more strongly than did either knowledge of natural selection or previous college coursework in biology. Clearly, tree thinking and natural selection are dissociable constructs that must both be taught for students to grasp the full gamut of evolutionary patterns and processes. © 2014 Wiley Periodicals, Inc. *J Res Sci Teach* 51: 759–788, 2014

**Keywords:** Tree of Life; tree thinking; cladograms; evolution education; science literacy

Just as beginning students in geography need to be taught how to read maps, so beginning students in biology should be taught how to read trees and to understand what trees communicate (O'Hara, 1998, p. 327).

Humans are most closely related, evolutionarily, to chimpanzees. Humans, chimps, and the other primates, in turn, share an ancestor with all other mammals. Mammals share a most recent common ancestor (MRCA) with reptiles. Mammals and reptiles share an ancestor with arthropods. As a group, animals share an ancestor with fungi. Animals and fungi, together, share an ancestor with plants. And so on. If we go back far enough in time, we find that all living things have a single ancestor in common. This history of “descent with modification” (Darwin, 1859) can be represented in the form of a very large, branching Tree of Life (ToL). The set of skills required to understand and reason with such information depicted in diagrammatic representations of the ToL is referred to as tree thinking. Tree thinking is an important aspect of 21st century science literacy.

---

Contract grant sponsor: Institute of Education Sciences, U.S. Department of Education; Contract grant number: R305A080621.

Correspondence to: L. R. Novick; E-mail: laura.novick@vanderbilt.edu

DOI 10.1002/tea.21161

Published online 7 July 2014 in Wiley Online Library (wileyonlinelibrary.com).

Evolutionary biologists distinguish macroevolution, which includes tree thinking, from microevolution, which includes natural selection. Although knowledge of natural selection and the ability to engage in tree thinking are likely correlated, no research has considered the relationship between students' understanding of these two facets of evolution. Basic instruction in evolution in the United States, particularly in high school and in introductory biology classes in college, focuses primarily on natural selection. It is our contention that such instruction does little to promote tree-thinking skills and that explicit instruction in understanding evolutionary trees is required. The present article reports the results of a study that addresses this research question.

### The Tree of Life

Assembling the diagrammatic ToL, which shows the patterns of evolutionary relationships among all extant and extinct living things, is one of modern biology's grandest goals (see <http://www.phylo.org/atol/>). Progress in this endeavor has been accelerating since the latter part of the last century. The number of papers in the *Science Citation Index* containing *phylogeny* or *phylogenetics* in the title, abstract, or keywords, rose from 186 in 1982 to nearly 5,000 in 2001 (Hillis, 2004). Incredibly, in 2001 1 paper out of every 200 across all fields of science was on phylogeny (Hillis, 2004). Against this backdrop, it is not surprising that an understanding of phylogenetic relationships has resulted in critical insights for basic biology as well as related translational applications. To take just a few current basic research examples from many in the burgeoning field of evolutionary developmental (Evo-Devo) biology, whether they concern the distribution of Hox gene complexes across animal taxa (Carrol, Grenier, & Weatherbee, 2005) or the evolution of flowering pattern genes in angiosperms (Chanderbali et al., 2010), novel data only make sense in light of the phylogenetic history of the taxa under study.

Translational applications have yielded important benefits in diverse areas such as human health, agriculture, biotechnology, ecology, and forensics (e.g., American Museum of Natural History [AMNH], 2002; Freeman, 2011; Futuyma, 2004; Yates, Salazar-Bravo, & Dragoo, 2004). The (confirmed and potential) benefits for human health are extensive: Predicting the epidemiology of emergent and established diseases (e.g., West Nile virus, influenza, hantaviruses, malaria) based on phylogeny has saved countless human lives (e.g., AMNH, 2002; Yates et al., 2004). For example, a recent phylogenetic analysis found that *Plasmodium falciparum*, which is confined to *Homo sapiens* and results in the deadliest form of malaria, originated in gorillas and not chimpanzees, our sister taxon, as previously thought (Liu et al., 2010). This is critically important information for the 207 million people infected with and the 3.4 billion people at risk from malaria, as well as for those engaged in fighting the disease. From the field of ecology, it is clear that understanding the life support systems of our planet, which are produced by Earth's biodiversity, depends on the ToL: Ecological relationships in space are the product of evolution over time and cannot be understood or managed without a phylogenetic perspective.

Given contemporary issues that involve evolutionary relationships, a basic understanding of phylogenetic trees is important for lay people as well as bioscience majors. Indeed, a recent analysis suggests that such diagrams are increasingly common in informal science institutions such as natural history museums (MacDonald & Wiley, 2012). Thus, the ability to engage successfully in tree thinking, far from being an arcane skill, is required for a scientifically literate citizenry in the 21st century (Novick & Catley, 2013; Thanukos, 2009).

For example, as discussed by Novick, Catley, and Funk (2011), knowledge of where snakes and birds fall in the ToL enabled the first author and her son to infer that his pet snake was potentially in harm's way sitting on the kitchen counter during operation of the oven's self-cleaning cycle. Consider, as well, the 2012 public health crisis in the United States in which 750 people developed difficult-to-treat fungal infections (including meningitis) from injections of a

tainted steroid pain medication. This crisis clarifies why it is essential to understand that fungi are more closely related to animals than to any other taxa (Novick & Catley, 2014b): Because fungi are the sister group to animals, most antifungal drugs that target the pathogen also adversely affect the patient (Marcos, Gandia, Harries, Camona, & Munoz, 2012). Both of these examples highlight the insufficiency of understanding microevolutionary processes such as the origin and maintenance of genetic variation, genetic drift, gene flow, and natural selection and the necessity of being able to read evolutionary trees to learn about, and reason from, previously unknown evolutionary relationships. Futuyma (2013) argues that an understanding of common ancestry, as depicted in phylogenetic trees, is critical for appreciating why the US National Institutes of Health funds basic research on a wide range of taxa, including *E. coli*, fruit flies, zebrafish, rats, and chimpanzees, in service of their mission to improve human health. Cracraft and Donoghue (2004b, p. 554) further argue that understanding the ToL can help people move away from an anthropocentric view of the world, which, as they note, “is bound to influence the ethical picture people develop about the importance of life forms other than our own and how these have been inextricably linked to our own well-being over time.”

### Microevolution Versus Macroevolution

#### *The Perspective of Evolutionary Biology*

As an organizing principle, explanatory framework, and predictive model for understanding the history and mechanisms of life, knowledge of evolution, broadly construed, is a critical component of general science literacy. Evolutionary biologists commonly distinguish microevolution and macroevolution (e.g., Freeman & Herron, 2007; Futuyma, 2013). Microevolutionary processes include, for example, mutation, genetic drift, natural selection, and adaptation. Macroevolution, with its multiple levels of emergent properties (from genes to individuals to populations to higher taxa) interacting over time and space to yield the formation, radiation, and extinction of species and higher-order natural groups, only makes sense in terms of the patterns of relationships it generates (Catley, 2006). The job of the evolutionary biologist is to discover the processes that produced these patterns. Phylogenetic trees are macroevolutionary models that summarize scientists' current hypotheses concerning patterns of evolutionary relationships and the historical evidence that supports those patterns. Tree thinking, therefore, falls into the realm of macroevolution.

Of course, microevolution and macroevolution are intimately related, with the latter processes being largely emergent patterns of the former. Heritable qualitative and quantitative changes in gene frequencies are the result of microevolutionary processes such as mutation, natural selection, genetic drift, and adaptation. These changes lead to speciation, which is the process that links micro- and macroevolution. When a population is divided, selection continues separately in each group. The resultant traits (gene frequencies prior to fixation), when fixed in each population, become the characters that are the markers of the speciation event itself. Groups of species arising from a MRCA become clades, which either persist or become extinct over time. Thus, the microevolutionary mechanisms that operate on the individual and the population are inextricably linked and melded over time (see Figure 5 in Catley, 2006) to become the historical patterns of extinction and cladogenesis among species we observe at the macro level and that are so succinctly illustrated in phylogenetic trees.

Tree thinking, using phylogenetic representations, is the only way biologists have to understand the complex and emergent interactions among these various hierarchical levels (Catley, 2006). Given that tree thinking is a requirement for a complete understanding of evolution and is a key component of 21st century science literacy, science educators must find ways to make

this mode of thinking accessible to students (e.g., Baum, Smith, & Donovan, 2005; Catley, 2006; Catley, Lehrer, & Reiser, 2005; Goldsmith, 2003; Gregory, 2008; O'Hara, 1998).

### *The Perspective of Science Education*

A glance through any high school biology textbook, however, will attest to the fact that macroevolutionary topics, including tree thinking, are poorly represented. Indeed, microevolution in the form of natural selection is typically taught as a surrogate for evolution in US schools (Catley, 2006). At the undergraduate level, students typically complete an introductory biology course having gained a rudimentary knowledge of the molecular and genetic mechanisms that underpin heritable qualitative and quantitative changes in gene frequencies (see Dauer, Momsen, Speth, Makohon-Moore, & Long, 2013, for a recent study of students' understanding of these links). The instruction might also have touched on some of the relevant microevolutionary mechanisms. Mechanisms that operate above the level of the population and in the domain of macroevolution, however, are usually given short shrift.

It is our contention that the microevolutionary, natural-selection-based understanding of evolution that comprises the majority of evolution instruction at the high school and undergraduate levels in the United States is insufficient to provide students with either an adequate understanding of phylogenetic trees or an ability to engage effectively in tree thinking. More specifically, although knowledge of natural selection and skill at tree thinking are likely correlated, given the interrelatedness of microevolutionary and macroevolutionary phenomena, we believe that explicit instruction in understanding evolutionary trees is required for a high level of tree-thinking skill.

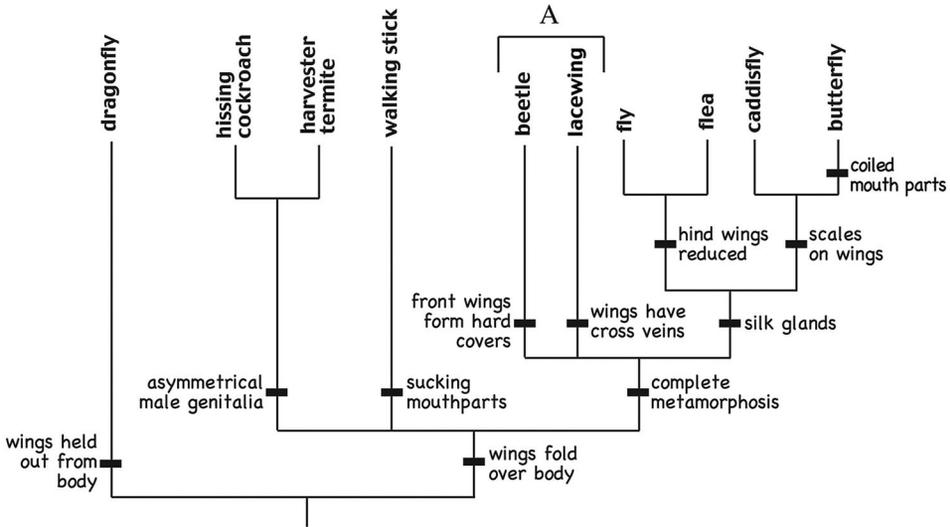
Several observations motivate these hypotheses. First, although microevolution and macroevolution are linked, this link is evidently not transparent. Studying, say, the rapid radiation of species within a clade requires knowledge of the adaptive advantage that drove it (e.g., a morphological character, a particular hox gene sequence, or a novel biochemical pathway). This in turn requires knowledge of how such novel attributes arise and become fixed in populations (e.g., point mutations, polyploidy, genetic drift, natural selection). Second, schematic diagrams used to depict scientific concepts and processes do not transparently convey meaning (e.g., Hegarty & Stull, 2012; Yeh & McTigue, 2009). Thus, if educators expect students to view such diagrams as more than abstract designs composed of assorted graphic elements, they must provide instructional support for how the diagrams convey conceptual meaning (Novick, 2006a). This requirement is well appreciated in chemistry (e.g., Chang & Linn, 2013; Cooper, Corley, & Underwood, 2013; Cooper, Underwood, Hilley, & Klymkowsky, 2012; Stull, Hegarty, Dixon, & Stieff, 2012). Third, a growing literature on students' understanding of evolutionary trees, discussed in the next section, suggests that there are significant deficiencies in students' ability to engage in tree thinking.

## Scientific Background

### *Tree Thinking in Evolutionary Biology*

The standard contemporary depiction of (subsets of) the ToL is a cladogram—a branching tree that depicts hypothesized evolutionary relationships in terms of nested sets of taxa (Baum & Smith, 2013; Hennig, 1966; Thanukos, 2009). Taxa (singular, taxon) are biological categories ranging from species (e.g., *Canis lupus*, *Panthera tigris*) to higher-order groups (e.g., Carnivora, Chordata). Consider Figure 1, which depicts a miniscule portion of the ToL involving 10 insect taxa. This cladogram shows that flies and fleas are more closely related to each other than they are to any other taxon on that cladogram because they share an MRCA that evolved the novel character

The students in a basic biology class are learning about evolutionary relationships among taxa. According to biologists, the following cladogram provides this information about the indicated taxa, which are various insects. Use this cladogram to answer the questions on this page.



1. Do the bracketed taxa labeled “A” constitute a clade? **yes** **no** (circle one)

Explain your answer:

If you answered no, which taxa need to be removed and/or added to the group to make it a clade?  
 (Make sure you indicate whether the taxa you list should be added or removed.)

2. Given that termites digest cellulose, which taxon is most likely to share this character?

Explain your answer:

Figure 1. A cladogram that appeared on the tree-thinking assessment, with associated questions testing students’ understanding of what constitutes a valid biological group (clade) and ability to make MRCA-based inferences. Adapted with permission of Sage Publications from Figure 4 in Novick and Catley (2013). (Modifications were to change the synapomorphy markers from circles to horizontal bars and to include only a subset of the original questions, whose wording was altered slightly.)

of having reduced hind wings. Thus, both flies and fleas have reduced size hind wings, but the other insects on the cladogram do not.

A group of taxa comprising a MRCA and all its descendants is called a clade. For example, flies and fleas comprise a clade on the cladogram in Figure 1, as do flies, fleas, caddisflies, and butterflies, as well as those four taxa plus beetle and lacewing. A clade consisting of only two taxa is called a sister group. Thus, cladograms are composed of nested clades, which are the only meaningful biological groups (Baum & Smith, 2013; Hennig, 1966; Thanukos, 2009).

The hind wings character shown in Figure 1 is called a synapomorphy. Synapomorphies are derived (i.e., newly evolved in the ancestor in question), invariant (therefore informative) physical, molecular, or behavioral characters shared by all members of a clade. In this article, we use synapomorphy and the less precise but more familiar term character interchangeably.

Another critical concept for understanding cladograms is that of a three-taxon statement, which is a set of three taxa in which two of the three (e.g., mammals and reptiles) share a more recent common ancestor with each other than they do with the third taxon (e.g., amphibians). The three-taxon statement is the essential building block of cladograms. That is, cladograms, no matter how large or complex, are composed of nested three-taxon statements.

Novick and Catley (2013) described five core tree-thinking skills, some of which included multiple subskills, used by biologists when reasoning about the ToL: (a) Identifying a character shared by two or more taxa due to inheritance from their MRCA; (b) identifying a set of taxa that either did or did not evolve from an ancestor with a particular character; (c) recognizing or identifying groups of taxa that do or do not comprise a clade; (d) evaluating relative evolutionary relatedness; and (e) making inferences based on shared ancestry (i.e., evolutionary history).

### *Alternative Cladogram Formats*

Cladograms are typically drawn in one of two diagrammatic formats—rectangular and diagonal. Although the rectangular format (e.g., see Figure 1) is much more common in the primary literature in biology (Novick & Catley, 2007), the diagonal format is slightly more prevalent in high school and college textbooks (Catley & Novick, 2008). Biologists presumably are successful in applying the aforementioned tree-thinking skills regardless of the format in which the cladogram is drawn. Students, however, find the diagonal format to be considerably more difficult to understand for a variety of perceptual reasons (Catley, Novick, & Funk, 2012; Novick & Catley, 2007, 2013; Novick, Catley, & Funk, 2010; Novick, Stull, & Catley, 2012). Therefore, following Novick and Catley's (2013) recommendation that the diagonal format not be used in introductory instructional materials, the instructional booklet we created for the present study used the rectangular format.

### *Research on Students' Tree-Thinking Skills*

Students' conceptual problems understanding natural selection have been well studied (e.g., Anderson, Fisher, & Norman, 2002; Bishop & Anderson, 1990; Demastes, Good, & Peebles, 1996; Jensen & Finley, 1996; Opfer, Nehm, & Ha, 2012; Sandoval & Reiser, 2003). Only recently, however, has research been conducted to document issues of understanding topics in macroevolution. In particular, starting around 2007, there has been a growing empirical literature on students' ability to engage in tree thinking (Catley et al., 2012; Catley, Phillips, & Novick, 2013; Halverson, Pires, & Abell, 2011; Meir, Perry, Herron, & Kingsolver, 2007; Morabito, Catley, & Novick, 2010; Novick & Catley, 2007, 2013, 2014b; Novick, Catley, & Funk, 2011; Novick et al., 2010, 2012; Novick, Shade, & Catley, 2011; Perry, Meir, Herron, Maruca, & Stal, 2008; Phillips, Novick, Catley, & Funk, 2012; Sandvik, 2008).

Although some empirical reports of students' ability to engage in tree thinking suffer from problems such as small samples or insufficient information about materials, data analyses, and results, the general consensus from the body of work is clear: (a) college students who have taken little or no college-level biology often do fairly well at the low-level tree-thinking skills of identifying characters and taxa. (b) However, even students who have taken introductory biology for biology majors and beyond show weak understanding of critical, higher-level skills such as evaluating relative evolutionary relatedness and identifying and recognizing clades. (c) Although phylogenetics instruction in a college biology class generally leads to improved tree thinking, significant deficiencies remain. None of this research, however, has examined the relationship between students' ability to engage in tree thinking and their understanding of natural selection, which is the focus of the present study.

### Overview of the Present Study

We noted earlier that tree thinking is an important aspect of science literacy, just as map reading is critical for geographic literacy (O'Hara, 1998). The present study primarily served the research goal of testing the relationship between (a) students' predominantly natural-selection-based understanding of evolution provided by extant instruction in evolution in high school and college biology classes and (b) their ability to engage in tree thinking. To measure understanding of natural selection, students completed the Conceptual Inventory of Natural Selection (CINS; Anderson et al., 2002), a widely used and easy to administer and score assessment of this construct. Although there is a published macroevolution assessment instrument (Nadelson & Southerland, 2010), it has serious validity issues (Novick & Catley, 2012). Therefore, we developed a new assessment to evaluate tree-thinking skills. Many of the items on this assessment were taken or adapted from our previous experimental materials. Information about students' performance on the items tested in our earlier research is provided in the method section, where our outcome measure is described in detail, to provide validity evidence for that measure.

In addition to examining the relationship between understanding of natural selection and success at tree thinking, the present study was part of an important broader effort, which included two subsequent studies conducted in biology classrooms (Novick & Catley, 2014a), to develop and evaluate curriculum materials for understanding cladograms and engaging in tree thinking that are suitable for college students with varying levels of knowledge of biology. Consistent with the conclusions of the National Research Council [NRC] (2012) report on discipline-based science education research generally, the body of research summarized in the previous section strongly argues that a new approach to teaching tree thinking is needed, one that (a) is informed by a deep understanding of both the biological science behind tree thinking and the difficulties students encounter in trying to comprehend and reason from cladograms and (b) leverages knowledge concerning effective instructional practices.

To evaluate the extent to which extant instruction in evolution in high school and college biology classes provides students with an ability to engage effectively in tree thinking, one group of undergraduates completed a tree-thinking assessment without any explicit instruction from us and therefore had to rely on their previous instruction in evolution to answer the questions. A second group of students completed the assessment after reading a short instructional booklet, described in the next section, that we created to provide a basic understanding of cladograms. This condition enabled a test of our hypothesis that extant instruction in evolution provides an insufficient basis for engaging effectively in tree thinking and that explicit instruction in understanding the relevant diagrammatic representation (i.e., cladograms) is required.

The students in our study represented a cross-section of the undergraduate population and were recruited to fit into two biology background groups: Stronger background students had taken the two-semester introductory biology sequence for majors (and possibly other biology courses) but had not taken an evolution course; weaker background students had taken little or no biology coursework in college. This sampling method ensured that the students in our study varied widely in their knowledge of biology in general and natural selection in particular, which was critical for testing our hypotheses. Students in both groups were randomly assigned to conditions.

### The Instructional Booklet

Because the students in this study represented a cross-section of the undergraduate population at the university from which they were recruited, the instruction had to be suitable for students without extensive knowledge of biology. At the same time, it had to be scientifically robust enough to be engaging for biology majors, for whom we planned to use it in our subsequent studies

(Novick & Catley, 2014a). With these considerations in mind, we developed a short, self-paced instructional booklet to provide a basic understanding of cladograms.

There are four reasons why we expected our short booklet to have a large impact on students' ability to engage in tree thinking: (a) As noted earlier, we do not believe existing instruction, with its primary focus on microevolution, conveys a good understanding of macroevolutionary topics, so additional instruction is needed. (b) We began our development work with deep knowledge from evolutionary biology of the critical concepts concerning cladograms that need to be taught. The content of the instructional booklet is described in the first subsection below. (c) Explanation of these concepts was informed by our knowledge, from two sources, of the difficulties students encounter while trying to conceptualize this material: (1) an extensive empirical data base from our laboratory documenting the difficulty undergraduates have engaging in tree thinking, including both accuracy in answering such questions and explanations of the reasoning behind their answers, and (2) the third author's successes and challenges in teaching these concepts in introductory and advanced college biology classes. (d) Pedagogical features of the instruction were informed by research from educational and cognitive psychology. These features are described in the second subsection, where supporting evidence is also provided.

### *Tree-Thinking Concepts Covered*

In this section, we describe the instructional booklet created for the present study. The current version of this booklet<sup>1</sup> (used in Study 3 in this series; see Novick & Catley, 2014a) is available on the first author's web site: [www.vanderbilt.edu/peabody/novick/evol\\_diagrams.html](http://www.vanderbilt.edu/peabody/novick/evol_diagrams.html). The content of the booklet covered key concepts included in other instructional sources on evolution generally and phylogenetics specifically (e.g., Baum & Smith, 2013; Futuyma, 2013; Gregory, 2008; Hennig, 1966; Wiley, Siegel-Causey, Brooks, & Funk, 1991). The first of two sections was labeled *A First Course in Understanding Evolutionary Trees*. It began with foundational terminology and concepts, including taxon, synapomorphy, cladogram, MRCA, and three-taxon statement. Students were shown how time (past and present) is mapped onto a cladogram and were introduced to the ideas that most recent common ancestry determines evolutionary relatedness and that common ancestry per se is uninformative because all taxa share a common ancestor if you go far enough back in time. The structure of cladograms as being compiled from a series of three-taxon statements was highlighted. The first section ended by teaching the concept of a clade (including the special case of a sister group) and showed how cladograms are composed of multiple, nested clades.

The second section was labeled *A Deeper Understanding of Evolutionary Trees*. It began with an in-depth discussion of how to determine relative evolutionary relatedness among taxa by determining which taxa have a more recent common ancestor and explained why two common alternative bases for determining evolutionary relatedness (horizontal distance between taxa and number of vertical steps between taxa) are incorrect strategies. The booklet then explained the concept of a polytomy—a set of three or more taxa that descended from the same ancestor for which there is insufficient evidence to resolve the structure into a three-taxon statement. In the cladogram in Figure 1, beetle, lacewing, and the clade supported by the character silk glands form a polytomy. This concept seems to be one of the more difficult ones for students to learn (Phillips et al., 2012). Finally, students were introduced to the fact that any particular cladogram they might see is a very small subset of the complete ToL. In service of understanding this important concept, students were taught how to “prune” taxa off a cladogram, how to collapse a set of taxa to a single higher-order group, and how to merge separate cladograms with overlapping sets of taxa to a single cladogram that preserves the relationships depicted in the separate cladograms. The

instructional booklet concluded with a brief discussion of why it is important to understand evolutionary trees. This instructional content spanned 12 pages.

*Pedagogical Features of the Instruction*

To create the best instruction we could within the constraints of a short, self-paced, self-administered booklet, we incorporated five pedagogical features into the design of our instruction. All of these features, which are described in the following paragraphs, are supported by research evidence concerning their effects on attention, cognition, and/or learning. Because all features were incorporated into the booklet, it is not possible to determine which are primarily responsible for the beneficial effects of instruction we find. The research question under investigation here, however, does not require us to localize the source of instructional benefits.

- (1) In addition to being explained verbally, each major concept (e.g., three-taxon statement, polytomy) was illustrated in a figure. Given the large literature in cognitive and educational psychology documenting the benefit of combining verbal and visual presentations of information for learning, memory, and problem solving (e.g., Brunyé, Taylor, Rapp, & Spiro, 2006; Hegarty & Just, 1993; Mayer, 1989, 2009; Paivio, 1986), this is one of the recommended practices in the Institute of Education Sciences (IES) guide for improving student learning (Pashler et al., 2007). For example, Figure 2, taken from the instructional booklet, illustrates a three-taxon statement, indicates how a

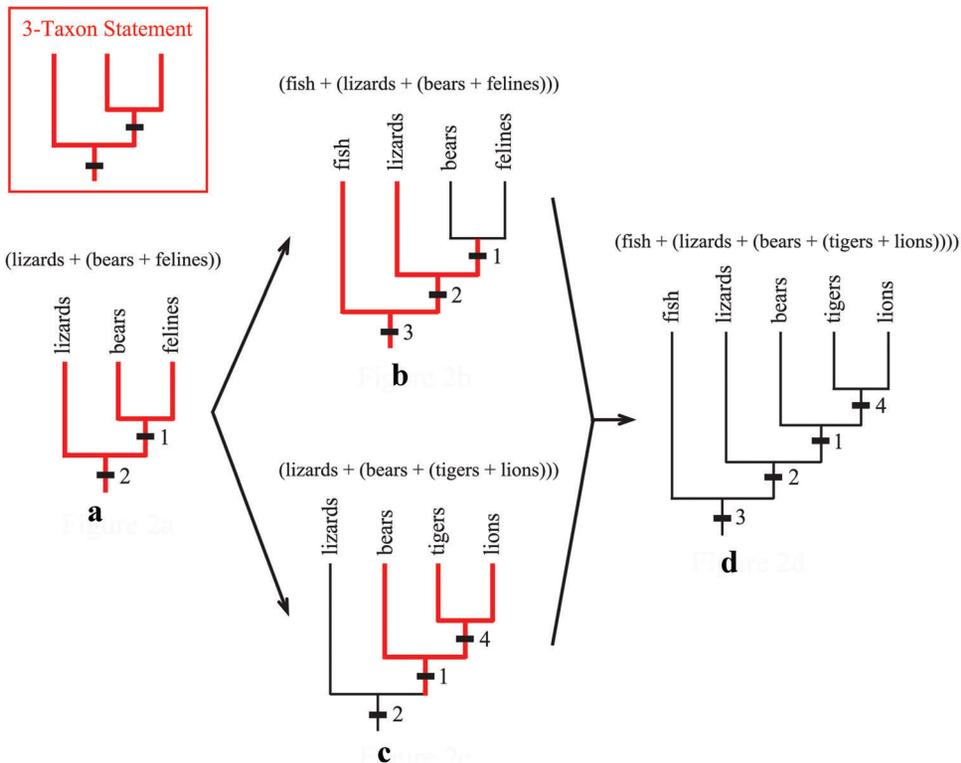


Figure 2. Figure 2 from the instructional booklet. The version students received was in color. The thicker gray lines in the figure here were red in the instructional booklet to highlight and call attention to three-taxon statements.

cladogram is composed of nested three-taxon statements, and shows how a cladogram's nested structure can be written in parenthetical notation.

- (2) The relevant part of each cladogram in a figure was color-coded to draw students' attention to the concept under consideration. Color is one of only a few stimulus features that is known to automatically capture attention (e.g., Treisman & Gelade, 1980; Wolfe, 1994). Without attention, there can be no learning. For example, in the version of Figure 2 printed in the instructional booklet, each new three-taxon statement was printed in red ink, whereas the rest of the cladogram was printed in black ink. A different color was used for each new concept.

In a second application of this principle, each new term (e.g., taxon, synapomorphy, polytomy) was printed in green ink, italicized, and underlined in the sentence in which it was defined. Such highlighting of new terminology is common in textbooks.

- (3) Each cladogram in the instructional booklet, except for the one illustrating a polytomy, was based on those shown in Figure 2. This was done to demonstrate that all of the concepts being taught are interrelated. A voluminous literature on memory in college students has shown that interrelated concepts are easier to remember than ones that are viewed as isolated or unrelated (e.g., Hulse, Egeth, & Deese, 1980; Kahana, 2012). This is important because, to mimic standard practice in classroom assessment, the students in the instructional condition had to complete the tree-thinking assessment as a closed-book test. The taxon relationships depicted throughout the instructional booklet are consistent with current research in evolutionary biology (e.g., Cracraft & Donoghue, 2004a; [tolweb.org](http://tolweb.org)). The polytomy example used different taxa because we could not find a polytomy in the portion of the ToL used for the remainder of the examples. Because our prior research has found that college (and high school) students have more difficulty attending to cladogram structure when the depicted relationships conflict with their (incorrect) prior knowledge (Novick & Catley, 2014b), we selected taxa for which the accepted evolutionary relationships are consistent (or at least do not conflict) with students' existing folkbiological taxonomy.
- (4) Two common misconceptions concerning how to evaluate evolutionary relatedness between taxa that have been identified in the empirical literature on tree thinking (Meir et al., 2007; Novick & Catley, 2013)—referring to the horizontal distance between the taxa at the top of the cladogram or to the number of vertical steps between the taxa—were explicitly noted in the text as incorrect and an explanation for why those methods are flawed was provided. This part of the instruction constitutes refutational text, which is text that “acknowledges students' alternative conceptions about a topic, directly refutes them, and introduces scientific conceptions as viable alternatives” (Mason, Gava, & Boldrin, 2008, p. 291). Refutational texts have been found to be superior to traditional ones in science education, especially when the text explains why the misconception is incorrect (e.g., Hynd, McWhorter, Phares, & Suttles, 1994; Kendeou, Smith, & O'Brien, 2013; Kowalski & Taylor, 2009; Mason et al., 2008).
- (5) The key concepts introduced in the first section of the booklet were reinforced by giving students an opportunity to test their comprehension of the material up to that point. Two practice tests were inserted at appropriate places in this section and labeled *Practice What You've Learned*. These questions were intended to be straightforward, near-transfer items that would reinforce the basic concepts being taught and alert students to gaps in their understanding. Each set had a cladogram at the top of the page with the taxa labeled with letters and the synapomorphies denoted by single-digit numbers. The first set of practice problems included four true-false questions about evolutionary relatedness (e.g., “Taxon A is more closely related to Taxon C than to Taxon B”), two yes-no questions about the parenthetical notation for three-taxon statements (see Figure 2), and three fill-in-the-blank questions about which character was possessed by the MRCA of certain taxa. The second practice set included four yes-no questions

asking whether a set of 2–6 named taxa comprise a clade, two yes–no questions about whether two named taxa form a sister group, and two fill-in-the-blank questions asking which taxa belong to the clade defined by the MRCA with a certain numbered character. Immediately following each practice set was a feedback page that gave the correct answers accompanied by an explanation. Students were invited to turn back in the booklet to review any concepts that remained unclear after reading the answer page. Given the many studies documenting the effectiveness of this kind of testing for promoting student learning (for reviews see Carpenter, 2012; Jaffe, 2008), this is another of the practices recommended in the IES guide for improving student learning (Pashler et al., 2007).

## Method

### *Students*

One hundred twenty-seven undergraduates at a highly selective, private, Research I university in the southeastern United States were recruited from a paid subject pool and paid \$25 for their participation. Students were excluded from participating if they were currently enrolled in the second-semester introductory biology class for majors and/or were enrolled in or had completed the evolution class. The data from three students were excluded from the analyses for the following reasons: One student had taken the evolution class, one student was unable to complete the study within the allotted 2 hours, and one student informed us after the study that she had reported her academic background inaccurately.

At the end of the study, students completed a background questionnaire on which they were asked whether they were enrolled in or had taken any of 10 (primarily organismal) biology and 3 historical geology classes. These courses were chosen because they included the general introductory courses and upper-level classes that may cover topics related to macroevolution. Students who had completed the two-semester introductory biology sequence for majors were assigned to the stronger biology background group ( $n = 63$ ). The remaining students were assigned to the weaker biology background group ( $n = 61$ ). The stronger background students had taken an average of 2.19 ( $s = 0.43$ , range of 2–4) of the listed classes, compared with 0.41 ( $s = 0.62$ ; range of 0–2) classes for the weaker background students. The stronger background students were sophomores (year 2) through seniors (year 4), with a mean year in school of 3.13. The weaker background group included first-year students through seniors, with a mean year in school of 2.26.

Students' majors were categorized as biology-related (e.g., biology, neuroscience, biomedical engineering), other STEM (science, technology, engineering, mathematics), and non-STEM (everything else, including undecided). Among the stronger background students, 75% had a biology-related major, 6% had another STEM major, and 19% had a non-STEM major; 68% hoped to attend medical school. The comparable percentages for the weaker background group were 11%, 21%, and 67% for the three types of majors and 13% for medical school attendance. Clearly, the two groups are quite different samples of college students.

A key reason for selecting these two samples of students was to ensure sufficient variability in knowledge of natural selection to enable us to evaluate the potential relationship between understanding of natural selection and success at tree thinking. Examination of students' CINS (Anderson et al., 2002) scores shows that we were successful. The mean CINS score was 13.85 ( $s = 3.95$ ). Individual scores ranged from a low of 3 to the maximum possible score of 20.

### *Design and Procedure*

Within each of the two biology background groups, approximately half of the students were randomly assigned to the instruction ( $n = 31$  weaker,  $n = 31$  stronger) and no-instruction (control) conditions ( $n = 30$  weaker,  $n = 32$  stronger). Students in the instructional condition completed the self-paced instructional booklet and associated practice problem booklet at the beginning of the experimental session. This took approximately 30 minutes. To save paper, the practice problem pages were placed in a separate two-page booklet (with a cover page). The instructional booklet told students when to turn to the practice problem booklet, and the end of each page in the practice booklet directed students to return to the instructional booklet. This allowed us to re-use the instructional booklets.

Because best practice in experimental design requires that students in the two conditions be equally fatigued when they complete the tree-thinking assessment (e.g., Stangor, 1998), students in the control condition spent approximately the same amount of time completing three cognitively engaging individual differences tests (a subset of the science reasoning section of the ACT college entrance exam and two spatial reasoning tests). Then all students completed the tree-thinking assessment, a second set of inference problems, the CINS (Anderson et al., 2002), and the background questionnaire. These tasks were self-paced. The inference booklet was a pilot task for a separate project examining how superficial similarity between taxa affects students' ability to draw inferences based on evolutionary relatedness. Those data will not be reported here. The entire procedure was completed within 2 hours.

Because several technical terms appeared in the test booklet that likely were unfamiliar to students in the control condition, students in both conditions received a half sheet of paper to refer to while completing the test booklet that defined three terms: (a) "A **taxon** is any taxonomic category ranging from a species (e.g., blue jay) to a higher-order group (e.g., birds, amniotes, vertebrates). The plural of taxon is **taxa**." (b) "A **cladogram** is a type of diagram that biologists use to depict evolutionary relationships among a set of taxa." (c) "A **clade** is a group of taxa that includes the most recent common ancestor of the group and all descendants of that ancestor."

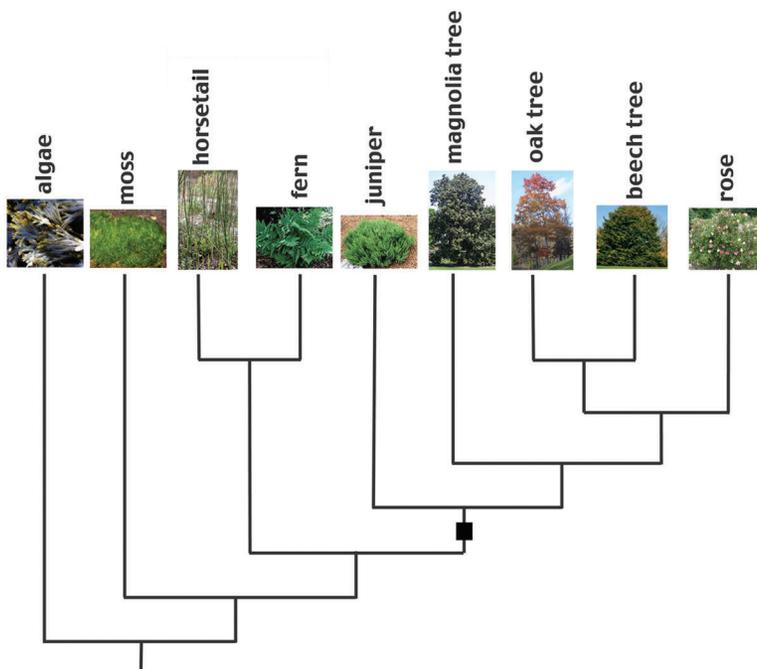
### *Tree-Thinking Assessment*

*Overview.* As noted earlier, Novick and Catley (2013) described, and provided pretest data for, five core tree-thinking skills. The present assessment did not include questions testing the low-level skills of identifying characters and taxa because Novick and Catley found that even college students with weaker backgrounds in biology did fairly well at answering such questions about relationships that were depicted in the rectangular cladogram format. Instead, we focused on questions that probed students' ability to use the critical higher-level skills of determining relative evolutionary relatedness (in both resolved and polytomous topologies), identifying individual and nested clades, and making inferences based on shared ancestry (i.e., evolutionary history).

Of the 37 questions on the assessment, 25 evaluated students' ability to engage in these core aspects of tree thinking. These questions were taken from, modified from, or modeled after those tested in our earlier research for which we found that students had difficulty (Novick & Catley, 2013, 2014b; Phillips et al., 2012), even after instruction in phylogenetics in a zoology or evolution class (Catley et al., 2012; Phillips et al., 2012). An additional 10 questions required students to interpret a given cladogram's status as a subset of the ToL. Although evolutionary biologists regularly need to reason with the idea that specific cladograms are subsets of the complete ToL, we are not aware of any prior research testing students' understanding of this concept. Thus, those questions are used for the first time in the present study.

Finally, two questions evaluated students’ ability to reason about relationships in a cladogram that conflicted with their prior knowledge. These items were included to take advantage of a captive sample to ask questions required for another study (Novick & Catley, 2014b). Thus, we will not discuss them here, and they are not included in the outcome measures for this study.

For all of the skills except subsets of the ToL, students received a cladogram at the top of the page and had to answer one or more questions about the information depicted in that cladogram. Immediately above the cladogram, we printed the following (e.g., see Figure 1): “The students in a basic biology class are learning about evolutionary relationships among taxa. According to biologists, the following cladogram provides this information about the indicated taxa, which are various insects. Use this cladogram to answer the questions on this page.” These sentences were the same for all cladograms except for the phrase indicating the taxonomic category to which the taxa belong (e.g., insects, plants, and animals, respectively, for the cladograms shown in Figures 1, 3, and 4).

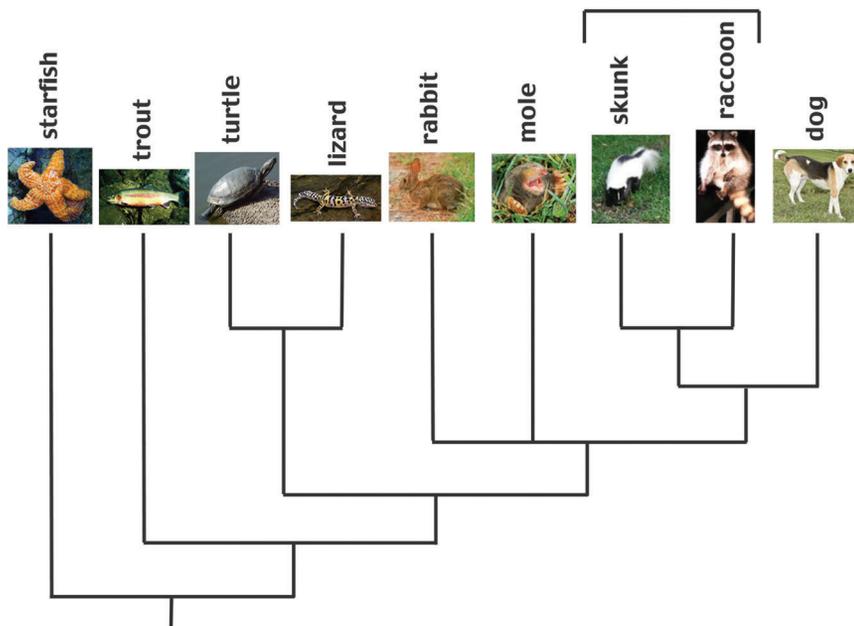


1. Which taxon—fern or oak tree—is the closest evolutionary relation to the juniper? \_\_\_\_\_  
 Explain your answer:

2. Suppose that, given new discoveries, biologists now realize that the black square on the diagram should be interpreted as representing 5 novel characters rather than just 1 as was thought when the diagram on this page was constructed. Given this new information, would your response to part (a) of this question change:  
**yes**      **no**      (circle one)

Explain your answer:

Figure 3. A cladogram that appeared on the tree-thinking assessment, with associated questions testing students’ understanding of levels of evolutionary relatedness in a resolved structure. Students received a version of this cladogram that included color photographs.



- Which of the following three statements (A, B, or C) is best supported by the scientific evidence:
  - Moles are more closely related to rabbits than to raccoons.
  - Moles are more closely related to raccoons than to rabbits.
  - Rabbits, moles, and raccoons are all equally closely related to each other.

Explain your answer:

- Two taxa that constitute a clade are enclosed in brackets in the diagram. Mark each additional clade with a bracket as shown in the example.

*Figure 4.* A cladogram that appeared on the tree-thinking assessment, with associated questions testing students' understanding of evolutionary relatedness in a polytomy and the nesting of clades. Students received a version of this cladogram that included color photographs. Adapted with permission of Springer Science + Business Media from Figure 1 in Phillips et al. (2012). (Modifications were to add text below the cladogram and to include the original color photographs in the online publication.)

The test items included requests both to use a cladogram to answer a question and to explain one's answers to such questions (e.g., see Figure 1). An initial question and associated request for an explanation were counted as two questions because each part was scored separately.

*Validity and Reliability.* The validity of a test for the intended interpretation of the scores, in this case as a measure of students' ability to engage in tree thinking, can be supported by a variety of types of evidence (AERA, APA, & NCME, 1999; Kaplan & Saccuzzo, 2009). Content-related evidence for the validity of our assessment comes from several sources: The tree-thinking skills were identified by the third author based on his doctoral training, expertise (e.g., Catley, 1994), and considerable experience teaching this area of biology. The specific questions call for the kinds

of reasoning in which professional biologists engage when examining phylogenetic trees. Criterion-related evidence for validity documents that scores on the test are related to performance on an external criterion measure. As we describe in the following subsections, for the questions that were taken from our earlier research on tree thinking in college students, there is strong evidence that tree-thinking skill is related to biology background: Students who had taken more biology classes in which topics related to macroevolution were likely to have been covered received higher scores than did students who had taken fewer such classes. Evidence concerning the reliability of the individual tree-thinking measures is presented in the Results Section.

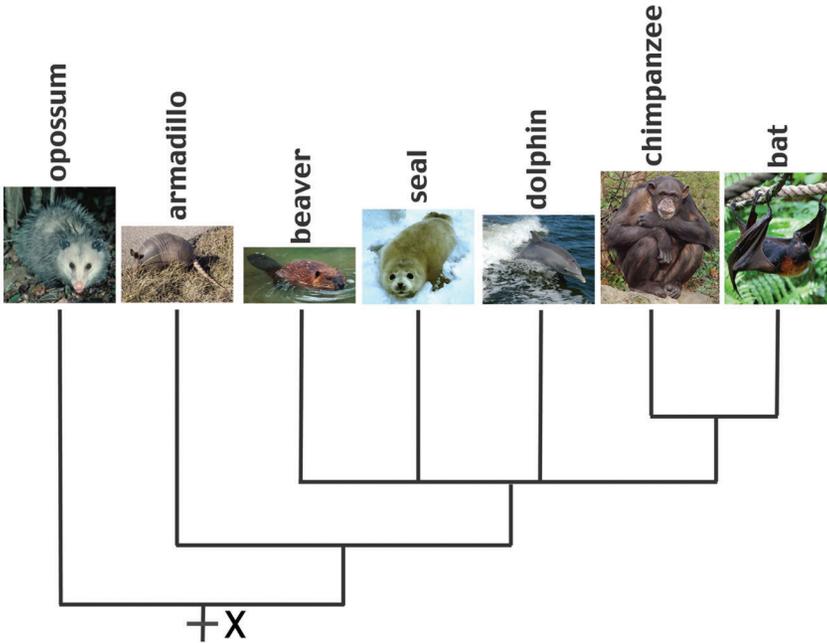
*Evolutionary Relatedness: Resolved.* Eight questions tested students' understanding that a reference taxon located at an intermediate hierarchical level between two comparison taxa is more closely related to the taxon with which it shares a more recent common ancestor. These questions were grouped into two sets of four, each for a different cladogram. Figure 3 shows one cladogram and its four questions. As can be seen in Figure 3, the relationships among the queried taxa are fully resolved (i.e., the three taxa form a three-taxon statement).

Novick and Catley (2013) found that both weaker and stronger background students did poorly on a composite measure incorporating accuracy and explanation quality for questions like the first two in Figure 3. On a 0–1 scale, weaker and stronger background students had means of 0.23 and 0.44, respectively. Even after 2 days of phylogenetics instruction in a zoology or evolution class, biology students were only moderately successful on such questions ( $M = 0.60$ ; Catley et al., 2012; Phillips et al., 2012). Students tend to get these questions wrong because they count the number of steps, or levels, between the reference taxon and each of the comparison taxa, and the questions were designed such that the reference taxon was closer by this measure to the taxon to which it is less closely related (see Figure 3).

The follow-up, black box question (see Figure 3), which was new to this assessment, probed whether students understood that the number of character differences between two taxa is irrelevant to determining evolutionary relationship; that is, that the only relevant factor is most recent common ancestry. Students also provided a written explanation for their answer to this question.

*Evolutionary Relatedness: Polytoamy.* Four evolutionary relatedness questions probed students' understanding of polytomies. These items comprised two relatedness questions and two follow-up explanation questions. Figure 4 illustrates one such pair of questions. Phillips et al. (2012) found that weaker and stronger background students had means of 0.12 and 0.24, respectively, on a composite measure (on a 0–1 scale) incorporating accuracy and explanation quality for these questions. Biology students did better after 2 days of phylogenetics instruction in an evolution class, but their scores were still very low on these questions ( $M = 0.47$ ).

*Identifying and Evaluating Clades.* Seven questions required students to recognize a clade or to evaluate whether indicated taxa comprise a clade. Two such questions are shown in Figure 5. For the first question, students were asked which of three subsets of taxa from the cladogram comprise a valid biological group. The second question asked students to explain their answer. This pair of questions and a similar pair involving a cladogram showing relationships among types of yeast came from Novick and Catley (2014b). For these two cladograms, the introductory text printed above the cladogram also included the following as the next-to-last sentence: "The students understand that all of the taxa shown in this diagram share a common ancestor marked by the X." In Novick and Catley's study, unselected college students (mostly weaker background students) had a mean score (0–1 scale) across these four questions of 0.42.



1. The following students disagree about what subsets of these taxa are valid biological groups. Which student's subset is best supported by evolutionary evidence? \_\_\_\_\_

Saul: *dolphin + chimpanzee*

Aaron: *beaver + seal + dolphin + chimpanzee + bat*

Reuven: *beaver + seal + dolphin*

Explain your answer:

Figure 5. A cladogram that appeared on the tree-thinking assessment, with associated questions testing students' understanding of what constitutes a valid biological group (clade). Students received a version of this cladogram that included color photographs. Adapted with permission of Wiley from Figure 2 in Novick and Catley (2014b). (Modifications were to remove text above and to the left of the cladogram, add text below the cladogram, and include the original color photographs in the online publication.)

In a third set of three questions (see Figure 1), taken from Novick and Catley (2013), two taxa were marked by a bracket at the top of the cladogram and students were asked whether those two taxa comprise a clade (they do not). Follow-up questions asked students to provide a written explanation for their response and, if they had answered that the bracketed group is not a clade, to indicate which taxa need to be removed and/or added to the group to make a clade. In Novick and Catley's study, weaker and stronger background students had mean composite scores across the three questions (0–1 scale) of 0.46 and 0.72, respectively. Catley et al. (2012) found that students enrolled in a zoology or evolution class had mean scores on these questions before and after 2 days of phylogenetics instruction of 0.67 and 0.79, respectively.

*Nested Clades.* Two questions tested students' knowledge that cladograms are composed of multiple, nested clades. One question is shown in Figure 4. Students saw a cladogram with one clade already marked and had to similarly mark the remaining clades. The other question was the same but involved a different cladogram. Phillips et al. (2012) found that weaker and stronger background students had mean proportions of correctly marked clades of only 0.33 and 0.62,

respectively.<sup>2</sup> After 2 days of instruction on phylogenetics in an evolution class, stronger background students did very well on these questions ( $M = 0.90$ ).

*Inference.* For two cladograms, students were asked to make an inference and then to provide a supporting explanation. One pair of questions, shown in Figure 1, came from Novick and Catley (2013). The other pair of questions was new. In each case, answering the first question required making an inference based on the evolutionary relationships shown because the mentioned character (e.g., ability to digest cellulose) is not on the cladogram. Novick and Catley found that weaker and stronger background students had mean composite scores (0–1 scale) across the two inference questions on the cladogram in Figure 1 of 0.46 and 0.69, respectively.

*Subsets of the Tree of Life.* These questions attempted to assess students' understanding that any given cladogram is a subset of the larger ToL and can be manipulated to study or highlight different parts of the larger tree. Two questions were adapted from Baum et al.'s (2005) tree-thinking challenge; the remaining eight questions were new for the present assessment.

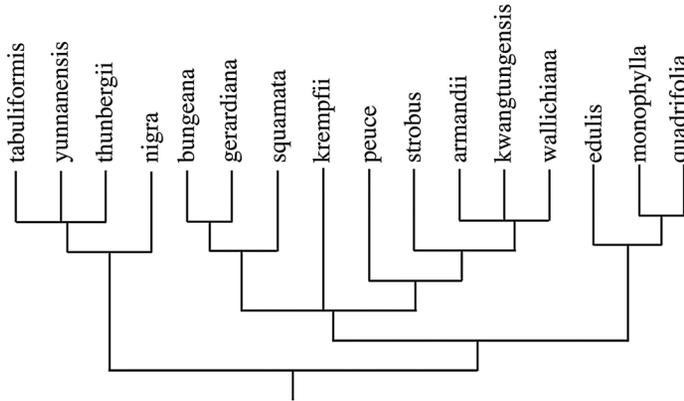
Three questions assessed students' ability to prune taxa off a tree, and two questions assessed their ability to collapse several taxa to a single higher-order group. The two collapsing questions and two of the pruning questions involved giving students a cladogram and asking them to redraw it with certain taxa pruned off or with certain taxa collapsed to a named higher-order group (e.g., see Figure 6). These four questions involved two cladograms showing relationships among pine trees, each of which was printed on two separate pages, once for a pruning question and once for a collapsing question. One cladogram included 10 taxa; the other included 16 taxa. The remaining pruning question was adapted from Baum et al.'s (2005) question 6 on quiz I by translating all the cladograms from the diagonal to the rectangular format. This question was multiple-choice rather than free response.

A sixth question tested students' understanding of pruning and collapsing by asking them to identify errors in a cladogram showing relationships among 10 fish taxa: After being told a set of taxa to prune and a set of taxa to collapse, a hypothetical student's incorrectly drawn tree was given, and students were asked to circle each error on this drawing and to explain how each error should be corrected. There were three errors (two involving pruning and one involving collapsing), although students were not told that. Whitley, Novick, and Fisher (2006) found that asking students to identify errors is a good way to measure their understanding of a concept.

Finally, four questions assessed students' ability to merge two or three smaller trees to a single tree that preserves the relationships depicted in all the smaller trees. For three questions, students were told that two (or three) researchers studied different taxa, some of which overlapped, and that a third (fourth) researcher was interested in studying all of the presented taxa. Students were then asked to draw a single cladogram that depicted the evolutionary relationships among all of the given taxa. The individual trees to be merged contained 4–6 taxa each. The final merging question was adapted from Baum et al.'s (2005) tree-thinking challenge (question 5 on quiz II) by printing the tree with the root at the bottom and placing the table showing the taxa above rather than below the tree. Like the other question from Baum et al., this one was multiple-choice.

*Scoring the Assessment.* Students' responses to most questions received an accuracy score of 0 (incorrect) or 1 (correct). The nested clades questions, however, received proportion correct scores to allow for partial credit in marking the clades. Also, students' responses to the questions requiring a written explanation received an explanation quality score of 0, 0.5, or 1. Explanations received a score of 1 if they referred to most recent common ancestry in the way that is appropriate to the question. They received a score of 0.5 if they referred to evolutionary relatedness, recent common ancestry, or the presence of some of the descendants, which suggested partial

Consider the following cladogram, which shows evolutionary relationships among pine trees.



**Pruning question:**

1. Redraw the cladogram above without the following four taxa:

- (a) *nigra*
- (b) *bungeana*
- (c) *gerardiana*
- (d) *kwangtungensi*

Make sure the cladogram you draw preserves all of the relationships among the remaining taxa.

**Collapsing question:**

1. Redraw the cladogram shown above incorporating the following two revisions:

- (a) put together the three taxa *balfouriana*, *longaeva*, and *aristata* into a single group labeled *Balfourianae*.
- (b) put together the four taxa *heldreichii*, *halepensis*, *brutia*, *pineta* into a single group labeled *Pinaster*.

Make sure the cladogram you draw preserves all of the relationships among the remaining taxa.

Figure 6. A cladogram that appeared on the tree-thinking assessment, with associated questions (asked on separate pages) testing students’ understanding of pruning and collapsing.

understanding of the relevant tree-thinking concepts. All other explanations received a score of 0 because they referred to factors that are irrelevant for understanding cladograms (e.g., horizontal or vertical distance between taxa). This coding scheme was adapted from those used by the authors in their earlier research (Novick & Catley, 2013; Phillips et al., 2012). A composite score was created for each tree-thinking skill by averaging across all questions testing that skill.

Two composite variables were computed for the ToL subsets skill. One composite, which we will refer to as change tree, included scores for the nine pruning, collapsing, and merging questions, all of which required students to produce or recognize a tree that has been changed correctly according to the instructions. Each question received an accuracy score of 0 or 1, and these scores were averaged to yield the composite. Examination of the means for the pruning, collapsing, and merging questions verified that collapsing across these questions does not obscure any findings related to condition, which is our main interest. The second composite was computed from the identifying errors question. Students received a score of 0 or 1 for correctly identifying each of the three errors and a score of 0, 0.5, or 1 for the quality of their explanation for how to fix the error. Responses that identified as an error a part of the cladogram that was correct were ignored. The six scores were averaged to yield the composite. We kept the identifying errors

question separate from the other ToL subsets questions because it required a deeper level of cognitive analysis: Students had to figure out what would be the correct answer after both pruning and collapsing taxa on the same cladogram (a skill they had not practiced), compare the given answer to the one they generated, identify the discrepancies, and then explain how to fix the errors. Whitley et al. (2006) found in their instructional study concerning computer programming that this type of debugging question was a particularly good measure of student understanding.

Students' explanations were scored independently by the first and second authors, who agreed on 90% of the scores. Discrepancies were resolved by discussion.

## Results

The primary goal of this study was to test two interrelated hypotheses concerning the relationship between knowledge of natural selection, a component of microevolution, and ability to engage in tree thinking, a component of macroevolution: (a) The natural-selection-based understanding of evolution that comprises the majority of evolution instruction at the high school and undergraduate levels in the United States is insufficient to provide students with an ability to engage effectively in tree thinking; and (b) although knowledge of natural selection and tree thinking skills are likely correlated, given the interrelatedness of micro- and macroevolutionary phenomena, explicit instruction in understanding cladograms is required to promote practical competence at tree thinking. Students' understanding of natural selection was provided by their scores on the CINS (Anderson et al., 2002). Ability to engage in tree thinking was provided by scores on the novel assessment we created for this study.

Although many of the questions on the tree-thinking assessment were tested in our prior research, some were newly written for this study. Thus, it is important to investigate performance on the individual tree-thinking skills to determine how best to compute a composite measure of tree thinking. Accordingly, the results section is divided into three parts. In part one, we present the results for Novick and Catley's (2013) core tree-thinking skills and the ToL subsets questions. Each of these measures was analyzed with a 2 (condition: instructional vs. control)  $\times$  2 (biology background: weaker vs. stronger) between-subjects ANOVA. The results of these ANOVAs are given in Table 1. Effect size is reported as  $\eta_p^2$ . Following Cohen's (1988) guidelines for proportion of variance accounted for, 0.01, 0.09, and 0.25 are the minimum values taken to indicate, respectively, a small, medium, and large effect. The relevant means for each skill are given in Table 2, along with an estimate of Cronbach's alpha (internal consistency reliability).

In part two, we compare knowledge of natural selection, completion of our instructional booklet, and biology background as predictors of success at tree thinking. In part three, we present data from the instructional condition on how performance on the Practice What You've Learned sections relates to success on the tree-thinking assessment.

### *Performance on the Individual Tree-Thinking Skills*

If prior, natural-selection-based instruction in evolution is insufficient for a high degree of skill at tree thinking and direct instruction in understanding evolutionary trees is required, then students in the instructional condition should receive significantly higher scores on our tree-thinking measures than students in the control condition. As shown in Table 1, the statistical analyses revealed a significant main effect of condition for all five core tree-thinking skills: Evolutionary relatedness with both resolved and polytomous structures, evaluating clades, nested clades, and inference. The means shown in Table 2 indicate that tree-thinking scores were higher in the instructional condition than in the control condition. For the two clade skills and inference there was also a main effect of biology background, with stronger background students having higher scores than weaker background students. The absence of an interaction between the two

Table 1  
Results of the 2 × 2 between subject ANOVAs

Measure	Condition	Biology Background	Interaction
Evol. Rel.: resolved (MSE = 0.08)	$F = 80.81,$ $p < 0.001, \eta_p^2 = 0.40$	$F = 1.28,$ $p > 0.25, \eta_p^2 = .01$	$F = 0.03,$ $p > 0.85, \eta_p^2 = 0.00$
Evol. Rel.: polytomy (MSE = 0.09)	$F = 61.46,$ $p < 0.001, \eta_p^2 = 0.34$	$F = 2.04,$ $p > 0.15, \eta_p^2 = 0.02$	$F = 1.28,$ $p > 0.25, \eta_p^2 = 0.01$
Evaluating clades (MSE = 0.06)	$F = 23.18,$ $p < 0.001, \eta_p^2 = 0.16$	$F = 17.49,$ $p < 0.001, \eta_p^2 = 0.13$	$F = 7.65,$ $p < 0.01, \eta_p^2 = 0.06$
Nested clades (MSE = 0.10)	$F = 28.55,$ $p < 0.001, \eta_p^2 = 0.19$	$F = 13.85,$ $p < 0.001, \eta_p^2 = 0.10$	$F = 2.75,$ $p = 0.10, \eta_p^2 = 0.02$
Inference (MSE = 0.06)	$F = 10.87,$ $p < 0.01, \eta_p^2 = 0.08$	$F = 6.45,$ $p < 0.05, \eta_p^2 = 0.05$	$F = 0.14,$ $p > 0.70, \eta_p^2 = 0.00$
Identifying errors (MSE = 0.06)	$F = 5.62,$ $p < 0.05, \eta_p^2 = 0.05$	$F = 0.49,$ $p > 0.45, \eta_p^2 = 0.00$	$F = 0.71,$ $p > 0.40, \eta_p^2 = 0.01$
Change tree (MSE = 0.03)	$F = 0.42,$ $p > 0.50, \eta_p^2 = 0.00$	$F = 9.39,$ $p < 0.01, \eta_p^2 = 0.07$	$F = 0.02,$ $p > 0.90, \eta_p^2 = 0.00$
CINS (MSE = 14.57)	$F = 0.18,$ $p > 0.65, \eta_p^2 = 0.00$	$F = 9.51,$ $p < 0.01, \eta_p^2 = 0.07$	$F = 1.93,$ $p > 0.15, \eta_p^2 = 0.02$

Note: Evol. Rel. is short for evolutionary relatedness.

All  $F(1, 120)$ .

independent variables for all measures except evaluating clades indicates that the instructional booklet was equally effective for both groups of students. For evaluating clades, weaker background students had means of 0.36 and 0.71 in the control and instructional conditions, respectively. The comparable means for the stronger background students are 0.68 and 0.77, respectively. Thus, there was a large difference between the two student groups in the control condition, but with instruction the weaker background students improved nearly to the level achieved by the stronger background students, who also improved with instruction.

The two ToL subsets measures yielded discrepant results. For change tree, the ANOVA indicates that there was only a main effect of biology background (see Table 1), with higher scores

Table 2  
Mean scores on the tree-thinking assessment

Tree-Thinking Skill	Coeff. Alpha	Overall Mean	Control Condition	Instructional Condition	Weaker Biology Background	Stronger Biology Background
Core skills (0–1)						
ER: resolved	0.89	0.45	0.22	0.68		
ER: polytomy	0.79	0.42	0.20	0.63		
Clade: evaluate	0.79	0.63	0.53	0.74	0.54	0.72
Clade: nested	0.96	0.73	0.59	0.88	0.63	0.83
Inference	0.56	0.68	0.61	0.75	0.63	0.74
Subsets of ToL (0–1)						
Identify errors	0.61	0.72	0.67	0.77		
Change tree	0.53	0.83			0.79	0.88
Composite (0–6) <sup>a</sup>		3.63	2.81	4.45	3.30	3.95

ER is short for evolutionary relatedness.

<sup>a</sup>Sum of the core tree-thinking skills and identifying errors (i.e., change tree is excluded).

found for stronger than weaker background students (see Table 2). The results for the identifying errors measure, however, are similar to those found for the core tree-thinking skills: There was a main effect of condition only, with higher scores for students in the instructional condition.

### *Predicting Success at Tree Thinking*

To compare the relative effectiveness of knowledge of natural selection and completion of our instructional booklet for success at tree thinking, we need composite measures of both natural selection understanding and tree-thinking ability. Students' understanding of natural selection was measured by total number correct on the CINS (Anderson et al., 2002). An ANOVA on these data indicated that stronger background students had significantly higher scores than weaker background students (see Table 1), with means of 14.89 and 12.77, respectively. This is only a small effect, as stronger background students, on average, answered only two additional questions correctly. Consistent with the fact that students were randomly assigned to the instructional and control conditions, there was not a significant difference in CINS scores as a function of condition.

We computed a composite tree-thinking score by summing across the scores on six individual skills. We included all the core tree-thinking skills because they were identified as critical skills by Novick and Catley (2013) and because each showed significant improvement with instruction. We also included the identifying errors variable because it, too, was sensitive to instruction. We excluded the change tree variable, however, as those test items, which were new to this study, turned out to be poorly conceived. Not only was there no difference between the control ( $M = 0.82$ ) and instructional ( $M = 0.84$ ) conditions (based on a dependent measure that was a composite of nine items), but accuracy was quite high in both conditions. Evidently, students were able to use their general knowledge of hierarchical diagrams, which is very good (Novick, 2006b), to answer those questions. Put another way, the results suggest that those items tested general knowledge of hierarchies rather than concepts specific to understanding cladograms.

Because each component skill variable was scored on a 0–1 scale, the overall composite is on a 0–6 scale. The means are shown in Table 2. Students in the control and instructional conditions had means of 2.81 ( $s = 1.03$ ) and 4.45 ( $s = 1.20$ ), respectively. With Cohen's  $d = 1.47$ , this is a very large overall effect of our short instructional booklet.

To investigate the contribution of understanding natural selection to success at tree thinking, we conducted a stepwise multiple regression analysis with the tree-thinking composite as the dependent variable and CINS scores, condition (control =  $-1$ , instructional =  $1$ ), and biology background (weaker =  $-1$ , stronger =  $1$ ) as independent variables. As the first variable to enter the equation, condition was the largest individual predictor of tree-thinking competence, yielding  $R = 0.59$  and accounting for 35% of the variability in tree-thinking scores. CINS score entered second, yielding  $R = 0.78$ . Condition plus understanding of natural selection together accounted for 61% of the variability in tree-thinking scores. Biology background entered last, yielding  $R = 0.79$ . Overall, the three predictors accounted for 63% of the variability in tree-thinking scores. The final regression equation is:

$$\hat{T} = (0.85)(\text{condition}) + (0.17)(\text{CINS}) + (0.16)(\text{bio}) + 1.29$$

Figure 7 shows predicted tree-thinking composite scores based on the regression equation as a function of condition, biology background, and whether students' CINS scores were one standard deviation below, at, or one standard deviation above the mean. The lighter and darker bars show, respectively, predicted scores for the control and instructional conditions and clearly demonstrate the large effect of our short, research-based introduction to evolutionary trees. In contrast, comparing the predictions for students with weaker (solid bars) versus stronger (striped

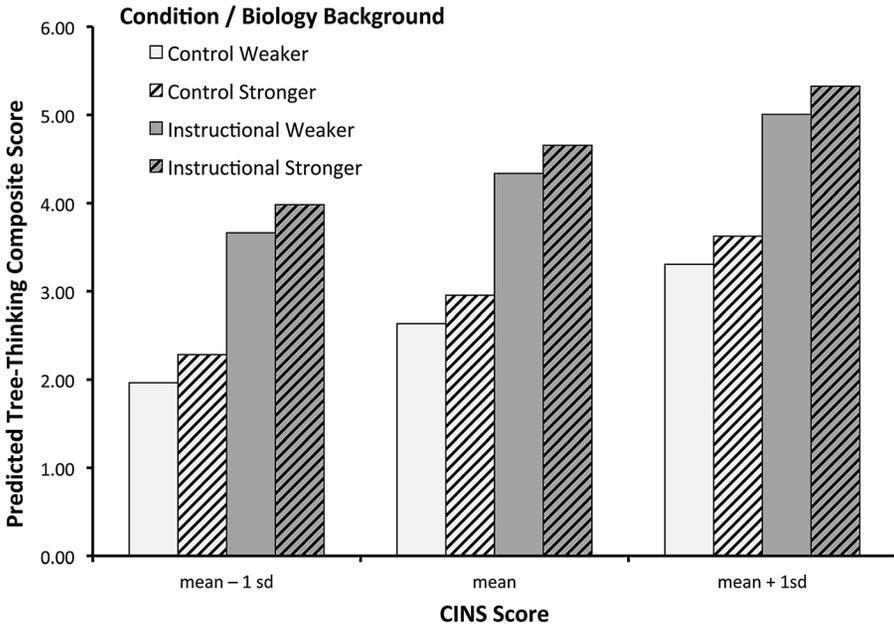


Figure 7. Predicted composite tree-thinking scores as a function of condition, biology background, and understanding of natural selection (CINS score).

bars) backgrounds in biology shows how little prior college coursework in biology mattered for success at tree thinking.

Turning now to the contribution of knowledge of natural selection, we see that control condition students with a mean CINS score can be expected to earn an average tree-thinking composite score of only 47% correct (average computed across the predictions for the two biology background groups). Even more telling, such students with a CINS score one standard deviation above the mean can be expected to earn an average tree-thinking score of just 58%. In contrast, a CINS score one standard deviation below the mean combined with about 30 minutes of self-directed instruction on evolutionary trees yields an average predicted tree-thinking score of 64%. An average CINS score plus our instructional booklet raises the predicted tree-thinking score to 75%. The additive effects of our instructional booklet and prior knowledge of natural selection to predicting tree-thinking ability suggested by Figure 7 were confirmed by a follow-up regression analysis that indicated that the interaction of instruction and CINS scores did not make a significant contribution to predicting tree-thinking ability.

#### *The Relationship Between Practice and Learning*

Students in the instructional condition were given a single proportion correct score for the two sets of practice problems combined. They did very well on these problems, with an overall mean of 0.93 ( $s = 0.11$ ; range of 0.53–1.00). There was no difference between weaker and stronger background students,  $F(1, 60) = 1.14$ ,  $p > 0.25$ ,  $MSE = 0.01$ ,  $\eta_p^2 = 0.02$ . The high degree of accuracy suggests we were successful in our endeavor to write straightforward, near-transfer items that would serve to reinforce the basic concepts being taught. Nevertheless, accuracy on the practice problems was positively correlated with the composite tree-thinking measure,  $r = 0.47$ ,  $p < 0.001$ .

To determine the contribution of practice problem performance and biology background to overall success at tree thinking, we conducted a stepwise multiple regression analysis on the instructional condition data ( $N=62$ ) to predict scores on the composite tree-thinking measure from these two predictors. Performance on the practice problems entered at the first step and was the only variable to make a significant individual contribution to the regression. This contribution is summarized by the previously reported correlation.

### Discussion

Biologists are currently actively engaged in the Herculean task of assembling the complete ToL, a hierarchical diagram depicting evolutionary relationships among all extant and extinct taxa, as evidenced by the explosion of published research on phylogenies with the turn of the new century (Hillis, 2004). This Tree in progress, despite its tentative and fragmentary nature, is yielding many important benefits to humankind in areas of interest and relevance both to scientists and the general public, including human health, agriculture, and forensics (e.g., AMNH, 2002; Freeman, 2011; Futuyma, 2004, 2013; Liu et al., 2010; Yates et al., 2004). Thus, “a scientifically-literate citizenry prepared to understand and tackle 21st century issues will need at least a basic understanding of the science behind the ToL” (Novick & Catley, 2013, p. 140; also see Thanukos, 2009).

#### *On the Relationship Between Tree Thinking and Understanding Natural Selection*

Our results clearly refute the implicit notion in biology education that to know natural selection is to know evolution. Instruction in how to interpret diagrammatic representations of the ToL, which are central to the domain of macroevolution (provided by a short, research-based, self-administered, instructional booklet we developed), and understanding of the microevolutionary process of natural selection (assessed by the CINS; Anderson et al., 2002) made independent and additive contributions to predicting accuracy on our composite tree-thinking measure. Moreover, our instructional booklet was the more important predictor. We believe the success of our booklet derives from our combined science-based and science-education-based approach to designing instruction, discussed in the next section.

As we noted earlier, tree thinking, using phylogenetic representations, is the only way biologists have to understand the complex and emergent interactions among the various hierarchical levels that comprise the broad sweep of processes we refer to as evolution (Catley, 2006). As biology educators, we need to move beyond a myopic focus on natural selection and find ways to make this mode of thinking accessible to students. This means we must teach students specifically how to interpret evolutionary relationships depicted in cladograms, as those diagrams, like scientific diagrams more generally (Hegarty & Stull, 2012; Yeh & McTigue, 2009), do not transparently convey meaning. The instructional booklet we created for this study provides a starting point for accomplishing this goal. We are not aware of other, research-based attempts to improve students’ ability to engage in tree thinking. We therefore call on other researchers to engage in this important work.

More generally, it appears that there may be some movement in the direction of focusing greater attention on macroevolutionary patterns and processes as the starting point for understanding evolution, at least at the undergraduate level. The new edition of Futuyma’s (2013) well-respected and widely used undergraduate evolution textbook now starts from a macroevolutionary rather than a microevolutionary perspective. *The Tree of Life: Classification and Phylogeny, Patterns of Evolution, and A History of Life on Earth* are three of the first five chapters that focus largely on macroevolutionary themes. *Natural Selection and Adaptation* appear as chapter 11, in the second half of the textbook.

### *A Science-Based and Science-Education-Based Approach to Designing Instruction*

As a critical initial step in a larger research program to provide college students with a basic understanding of cladograms and a corresponding ability to engage in tree thinking, we drew upon two sources to develop a short, self-paced instructional booklet. First, we drew upon expert knowledge from the field of evolutionary biology concerning how cladograms are structured and what it means to engage in tree thinking (Baum & Smith, 2013; Futuyma, 2013; Hennig, 1966; Wiley et al., 1991). Such knowledge also informed Novick and Catley's (2013) description of tree-thinking skills used by biologists when reasoning about the ToL. The items on our assessment instrument were designed to mimic the actual ways that evolutionary biologists use these skills to interpret the information expressed in cladograms. This strong contribution of disciplinary knowledge to the framing of questions and the design of research is characteristic of what a recent NRC (2012) report refers to as discipline-based education research.

The second source of information that informed the development of our instructional and test booklets was previous science education research on the nature of college students' understanding, and lack thereof, of tree thinking (Catley et al., 2012, 2013; Halverson et al., 2011; Meir et al., 2007; Morabito et al., 2010; Novick & Catley, 2007, 2013, 2014b; Novick et al., 2010; Novick, Catley, & Funk, 2011; Perry et al., 2008; Phillips et al., 2012; Sandvik, 2008). In particular, knowing the difficulties that college students encounter when trying to engage in tree thinking provided critical information regarding both what to teach and how to teach it. For example, such knowledge enabled us to provide refutational text to counteract students' common misconceptions about how cladograms depict relative evolutionary relatedness.

### *The Positive Impact of Our Research-Based Instructional Booklet*

Given that our instruction in understanding evolutionary trees consisted solely of a self-paced instructional booklet that took students roughly 30 minutes to complete, the benefit of being in the instructional condition was quite notable. As indicated in Table 1, there was a significant increase in tree-thinking accuracy as a result of completing the instructional booklet for six of the seven dependent measures: The core tree-thinking skills identified by Novick and Catley (2013) plus the identifying errors item. A Cohen's  $d$  value of 1.47 for the tree-thinking composite based on these six measures (i.e., everything except pruning and collapsing taxa and merging trees—change tree in Table 2) indicated quite a large overall effect of instruction. Although there was no delay between completing the instructional booklet and taking the test, students were not allowed to refer back to the instructional booklet to answer the test questions.

Another way to evaluate the effectiveness of our instructional booklet is to compare the results for this condition to those found in two earlier studies involving stronger biology background students in an evolution or zoology class who received 2 days of phylogenetics instruction in that class, after which cladograms were used throughout the remaining lectures to illustrate the macroevolutionary concepts being conveyed (Catley et al., 2012; Phillips et al., 2012). To ensure comparability of the student samples, we restrict this comparison to the results for just the stronger background students in the present study. As reported earlier, students in the previous studies had means of 0.60, 0.47, 0.79, and 0.90, respectively, for questions assessing understanding of evolutionary relatedness among taxa in a resolved topology, evolutionary relatedness among taxa in a polytomy, whether a set of taxa comprise a clade, and cladograms as comprising nested clades. Stronger background students in the instructional condition in the present study had means of 0.70, 0.64, 0.77, and 0.94, respectively, for these question types. Thus, for the two types of clade questions, our self-paced instructional booklet led to comparable performance as the much longer classroom instruction. For the two types of

evolutionary relatedness questions, our instructional booklet led to better performance. This was especially the case for the very difficult polytomy questions. In comparing these results, it is important to keep in mind that although the students in the biology classes received considerably more instruction than the students in the present study, which was reinforced over a longer period of time, the information on tree thinking was embedded within material not directly related to that topic, and the students' ability to engage in tree thinking was assessed after a longer delay.

### *Next Steps*

The results of this study demonstrated that (a) even a good understanding of natural selection is insufficient to enable students to be successful at tree thinking (see Figure 7) and (b) our short instructional booklet is highly effective in promoting practical competence at tree thinking in students with varying backgrounds in biology in a controlled environment when an immediate posttest is given. It is an open question, however, whether the instruction would be similarly effective in the less well-controlled classroom environment, especially if the posttest were delayed, as is typically the case in college classes.

Accordingly, we have conducted two curriculum studies in college biology classes (Novick & Catley, 2014a). The first of these studies combined an experimental versus business-as-usual instruction group comparison with a pretest/posttest design. Students enrolled in the second semester of an introductory biology class for majors participated as a normal part of this course. Students in the third author's two sections received an enhanced version of our instruction; students in the other two sections received business-as-usual instruction from another instructor. All students completed a pretest before instruction and a posttest at the end of the semester. The second classroom study used a pretest/posttest design. Advanced biology majors and masters students participated as a normal part of their coursework in the biology of arthropods class taught by the third author. This study provided further enhanced instruction appropriate for the greater biological knowledge of the students. All students received this instruction and completed both a pretest and posttest to assess learning. The results of both studies provided evidence that our tree-thinking instruction is effective in diverse classroom settings (Novick & Catley, 2014a).

### *Final Thoughts*

Given that natural-selection-based evolutionary theory is the default in biology education in the United States, high school and undergraduate students in this country are at a distinct disadvantage when it comes to understanding the full sweep of evolutionary biology. The present study tested, and found support for, the hypothesis that knowledge of microevolution alone is not a strong predictor of success in tree thinking and that direct instruction in evolutionary tree diagrams is required. Testing this contention required the creation of a novel tree-thinking assessment and a research-based instructional intervention centered around cladograms, the standard diagrammatic depiction of macroevolutionary relationships. Our results clearly showed that both of these artifacts were effective. In particular, we demonstrated significant and relatively painless gains in understanding of and ability to reason with evolutionary trees among college students with a wide variety of backgrounds in biology after just 30 minutes of self-directed instruction. We call on biology educators to broaden their instruction to include tree thinking and on researchers to explore ways to effectively merge both micro- and macroevolutionary perspectives into a pedagogical framework that provides students access to the full power and grandeur of evolutionary thinking.

## Notes

<sup>1</sup> The current version of our instructional booklet differs from that used in the present study in several ways: (a) It includes a title, authors, copyright information, and scientific references; (b) the explanations of several of the concepts were modified in an attempt to clarify the discussion and improve students' success at tree thinking (see Novick & Catley, 2014a); and (c) two new practice problem sections were added in the second half of the booklet, after the two practice sections included for the present study.

<sup>2</sup> Students were also asked how many clades there are in the diagram including the one already marked. Because Phillips et al. (2012) found that many students in their study gave responses to this question that were inconsistent with their bracketing of the taxa, they elected not to analyze the numerical responses. We encountered the same inconsistency in the present data. To facilitate comparison of performance in our study to what Phillips et al. found, we made the same analysis decision.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080621 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. A preliminary report of this study was submitted to Vanderbilt University by Emily Schreiber for her undergraduate honors thesis. She is now a graduate student at Xavier University.

## References

- AERA, APA, & NCME. (1999). Standards for educational and psychological testing. Washington, DC: AERA.
- American Museum of Natural History. (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. Retrieved August 30, 2010, from [http://www.phylo.org/sub\\_sections/outreach/outreach\\_b.php](http://www.phylo.org/sub_sections/outreach/outreach_b.php)
- Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching*, 39, 952–978.
- Baum, D. A., & Smith, S. D. (2013). *Tree thinking: An introduction to phylogenetic biology*. Greenwood Village, CO: Roberts and Company.
- Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree thinking challenge. *Science*, 310, 979–980.
- Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, 27, 415–427.
- Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. B. (2006). Learning procedures: The role of working memory in multimedia learning experiences. *Applied Cognitive Psychology*, 20, 917–940.
- Carpenter, S. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, 21, 279–283.
- Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2005). *From DNA to diversity: Molecular genetics and the evolution of animal design* (2nd ed.). Malden, MA: Wiley-Blackwell.
- Catley, K. M. (1994). Descriptions of new *Hypochilus* species from New Mexico and California with a cladistic analysis of the Hypochilidae (Araneae). *American Museum Novitates*, 3088, 1–27. Retrieved from: <http://digitallibrary.amnh.org/dspace/handle/2246/4985>, June 8, 2010.
- Catley, K. M. (2006). Darwin's Missing Link: A novel paradigm for evolution education. *Science Education* 90, 767–783.

Catley, K. M., Lehrer, R., & Reiser, B., (2005) Tracing a prospective learning progression for developing understanding of evolution. Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievement. Manuscript may be requested from Catley.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience*, 58, 976–987.

Catley, K. M., Novick, L. R., & Funk, D. J. (2012). The promise and challenges of introducing tree thinking into evolution education. In K. Rosengren, E. M. Evans, S. Brem, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 93–118). New York, NY: Oxford University Press.

Catley, K. M., Phillips, B. C., & Novick, L. R. (2013). Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. *Research in Science Education*, 43, 2327–2348.

Chanderbali, A. S., Yoo, M. J., Zahn, L. M., Brockington, S. F., Wall, P. K., Gitzendanner, M. A., . . . , & Soltis, P. S. (2010). Conservation and canalization of gene expression during angiosperm diversification accompany the origin and evolution of the flower. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 22570–22575.

Chang, H.-Y., & Linn, M. C. (2013). Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching*, 50, 858–886.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure-property relationships. *Journal of Research in Science Teaching*, 50, 699–721.

Cooper, M. M., Underwood, S. M., Hilley, C. Z., & Klymkowsky, M. W. (2012). Development and assessment of a molecular structure and properties learning progression. *Journal of Chemical Education*, 89, 1351–1357.

Cracraft, J., & Donoghue, M. J. (Eds.). (2004a). *Assembling the tree of life*. New York: Oxford University Press.

Cracraft, J., & Donoghue, M. J. (2004b). Assembling the tree of life: Where we stand at the beginning of the 21st century. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 553–561). New York: Oxford University Press.

Darwin, C. (1859). *On the origin of species by means of natural selection* (1st ed.). London: John Murray.

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50, 639–659.

Demastes, S. S., Good, R. G., & Peebles, P. (1996). Patterns of conceptual change in evolution. *Journal of Research in Science Teaching*, 33, 407–431.

Freeman, S. (2011). *Biological science: Evolution, diversity, & ecology* (Vol. 2). Boston, MA: Benjamin Cummings.

Freeman, S., & Herron, J. C. (2007). *Evolutionary analysis* (4th ed.) Upper Saddle River, NJ: Prentice Hall.

Futuyma, D. J. (2004). The fruit of the tree of life. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 25–39). New York, NY: Oxford University Press.

Futuyma, D. J. (2013). *Evolution* (3rd ed.). Sunderland, MA: Sinauer.

Goldsmith, D. W. (2003). The great clade race: Presenting cladistic thinking to biology majors & general science students. *The American Biology Teacher*, 65, 679–682.

Gregory, T. R. (2008). Understanding evolutionary trees. *Evolution: Education and Outreach*, 1, 121–137.

Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Science Education*, 95, 794–823.

- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717–742.
- Hegarty, M., & Stull, A. T. (2012). Visuospatial thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning*. New York, NY: Oxford University Press.
- Hennig, W. (1966). *Phylogenetic systematics*. Urbana, IL: University of Illinois Press.
- Hillis, D. M. (2004). The tree of life and the grand synthesis of biology. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 545–547). New York, NY: Oxford University Press.
- Hulse, S. H., Egeth, H., & Deese, J. (1980). *The psychology of learning* (5th ed.). New York, NY: McGraw-Hill.
- Hynd, C. R., McWhorter, J. Y., Phares, V. L., & Suttles, C. W. (1994). The role of instructional variables in conceptual change in high school physics topics. *Journal of Research in Science Teaching*, 31, 933–946.
- Jaffe, E. (2008). Will that be on the test? *Observer*, 21(10), 18–21.
- Jensen, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching*, 33, 879–900.
- Kahana, M. J. (2012). *Foundations of human memory*. New York, NY: Oxford University Press.
- Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing: Principles, applications, and issues* (7th ed.). Belmont, CA: Wadsworth.
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 854–865.
- Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology*, 36, 153–159.
- Liu, W., Li, Y., Learn, G. H., Rudicell, R. S., Robertson, J. D., Keele, B. F., . . . & Hahn, B. H. (2010). Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature*, 467, 420–425.
- MacDonald, T., & Wiley, E. O. (2012). Communicating phylogeny: Evolutionary tree diagrams in museums. *Evolution: Education and Outreach*, 5, 14–28.
- Marcos, J. F., Gandia, M., Harries, E., Carmona, L., & Munoz, A. (2012). Antifungal peptides: Exploiting non-lytic mechanisms and cell penetration properties. In K. Rajasekaran, J. W. Cary, J. M. Jaynes, & E. Montesinos (Eds.), *Small wonders: Peptides for disease control* (ACS Symposium Series, 1095. pp. 337–357). Washington, DC: American Chemical Society.
- Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology*, 100, 291–309.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81, 240–246.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
- Meir, E., Perry, J., Herron, J. C., & Kingsolver, J. (2007, September). College students' misconceptions about evolutionary trees. *The American Biology Teacher Online*, 69(7), 71–76.
- Morabito, N., Catley, K. M., & Novick, L. R. (2010). Reasoning about evolutionary history: The effects of biology background on post-secondary students' knowledge of most recent common ancestry and homoplasy. *Journal of Biological Education*, 44, 166–174.
- Nadelson, L. S., & Southerland, S. A. (2010). Development and preliminary evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. *The Journal of Experimental Education*, 78, 151–190.
- National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. S. R. Singer, N. R. Nielsen, & H. A. Schweingruber (Eds.). Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.
- Novick, L. R. (2006a). The importance of both diagrammatic conventions and domain-specific knowledge for diagram literacy in science: The hierarchy as an illustrative case. In D. Barker-Plummer, R. Cox, & N. Swoboda (Eds.), *Diagrammatic representation and inference* (Diagrams 2006), LNAI 4045 (pp. 1–11). Berlin: Springer-Verlag.

- Novick, L. R. (2006b). Understanding spatial diagram structure: An analysis of hierarchies, matrices, and networks. *The Quarterly Journal of Experimental Psychology*, 59, 1826–1856.
- Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, 13, 197–223.
- Novick, L. R., & Catley, K. M. (2012). Assessing students' understanding of macroevolution: Concerns regarding the validity of the MUM. *International Journal of Science Education*, 34, 2679–2703.
- Novick, L. R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *American Educational Research Journal*, 50, 138–177.
- Novick, L. R., & Catley, K. M. (2014a). Fostering 21st century evolutionary reasoning in undergraduates: A research-based approach to teaching tree thinking. Manuscript in preparation.
- Novick, L. R., & Catley, K. M. (2014b). When relationships depicted diagrammatically conflict with prior knowledge: An investigation of students' interpretations of evolutionary trees. *Science Education*, 98, 269–304.
- Novick, L. R., Catley, K. M., & Funk, D. J. (2010). Characters are key: The effect of synapomorphies on cladogram comprehension. *Evolution: Education and Outreach*, 3, 539–547.
- Novick, L. R., Catley, K. M., & Funk, D. J. (2011). Inference is bliss: Using evolutionary relationship to guide categorical inferences. *Cognitive Science*, 35, 712–743.
- Novick, L. R., Shade, C. K., & Catley, K. M. (2011). Linear versus branching depictions of evolutionary history: Implications for diagram design. *Topics in Cognitive Science*, 3, 536–559.
- Novick, L. R., Stull, A. T., & Catley, K. M. (2012). Reading phylogenetic trees: Effects of tree orientation and text processing on comprehension. *BioScience*, 62, 757–764.
- O'Hara, R. J. (1998). Population thinking and tree thinking in systematics. *Zoologica Scripta*, 26, 323–329.
- Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, 49, 744–777.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning (NCER 2007–2004)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ncer.ed.gov>
- Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D. (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE Life Science Education*, 7, 193–201.
- Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2012). Teaching tree thinking to college students: It's not as easy as you think. *Evolution: Education and Outreach*, 5, 595–602.
- Sandoval, W. A., & Reiser, B. J. (2003). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345–372.
- Sandvik, H. (2008). Tree thinking cannot taken for granted: Challenges for teaching phylogenetics. *Theory in Biosciences*, 127, 45–51.
- Stangor, C. (1998). *Research methods for the behavioral sciences*. Boston, MA: Houghton Mifflin.
- Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition and Instruction*, 30, 404–434.
- Thanukos, A. (2009). A name by any other tree. *Evolution: Education & Outreach*, 2, 303–309.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Whitley, K. N., Novick, L. R., & Fisher, D. (2006). Evidence in favor of visual representation for the dataflow paradigm: An Experiment Testing LabVIEW's Comprehensibility. *International Journal of Human-Computer Studies*, 64, 281–303.
- Wiley, E. O., Siegel-Causey, D., Brooks, D. R., & Funk, V. A. (1991). *The compleat cladist (Special Publication No. 19)*. Lawrence, KS: The University of Kansas Museum of Natural History.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.

Yates, T. L., Salazar-Bravo, J., & Dragoo, J. W. (2004). The importance of the tree of life to society. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 7–17). New York: Oxford University Press.

Yeh, Y.-F. Y., & McTigue, E. M. (2009). The frequency, variation, and function of graphical representations within standardized state science tests. *School Science and Mathematics*, 109, 435–449.