

# Snakes and Eels and Dogs! Oh, My! Evaluating High School Students' Tree-Thinking Skills: An Entry Point to Understanding Evolution

Kefyn M. Catley · Brenda C. Phillips · Laura R. Novick

Published online: 15 May 2013

© Springer Science+Business Media Dordrecht 2013

**Abstract** The biological community is currently undertaking one its greatest scientific endeavours, that of constructing the Tree of Life, a phylogeny intended to be an evidenced-based, predictive road map of evolutionary relationships among Earth's biota. Unfortunately, we know very little about how such diagrams are understood, interpreted, or used as inferential tools by students—collectively referred to as tree thinking. The present study provides the first in-depth look at US high school students' competence at tree thinking and reports how they engage cognitively with tree representations as a precursor to developing curricula that will provide an entry point into macroevolution. Sixty tenth graders completed a 12-question instrument that assessed five basic tree-thinking skills. We present data that show patterns of misunderstandings are largely congruent between tenth graders and undergraduates and identify competences that are pivotal to address during instruction. Two general principles that emerge from this study are: (a) Students need to be taught that cladograms are an authoritative source of evidence that should be weighted more than other superficial or ecological similarities; (b) students need to understand the vital importance and critical difference between most recent common ancestry and common ancestry. Further, we show how the objectives of this study are closely aligned with US and International Standards and argue that scientifically-literate citizens need at least a basic understanding of the science behind the Tree of Life to understand and engage in twenty-first century societal issues such as human health, agriculture, and biotechnology.

**Keywords** Evolution education · Tree thinking · Cladograms · Diagrammatic reasoning

---

K. M. Catley (✉)  
Western Carolina University, Cullowhee, NC, USA  
e-mail: kcatley@wcu.edu

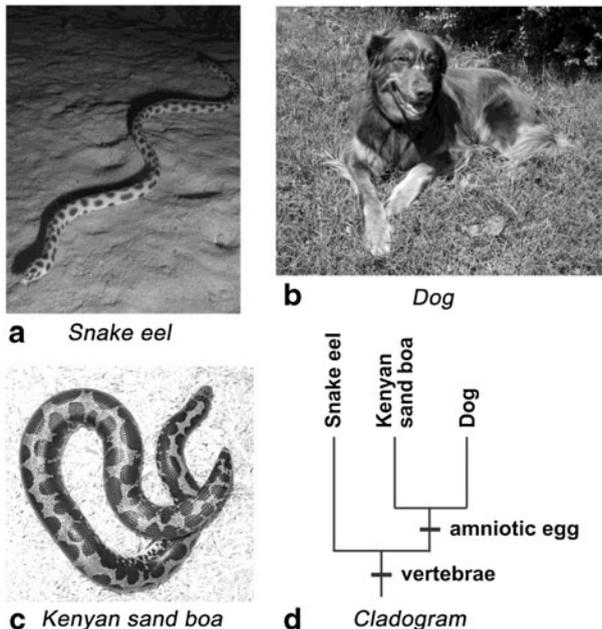
B. C. Phillips  
Harvard University, Cambridge, MA, USA

L. R. Novick  
Vanderbilt University, Nashville, TN, USA

## On the Importance of Understanding the Tree of Life

Consider the following classification question, which might arise in a high school biology class: Does the animal in Fig. 1c belong with the animal in Fig. 1a or Fig. 1b? One student might respond that the sand boa and snake eel go together because both move by slithering, neither has legs nor fur, and both have dark blotches on their skin. Another student, echoing Darwin (1859), might respond that the sand boa and dog are more alike (i.e., are more closely related) because they both have amniotic eggs as a result of sharing a more recent common ancestor (see Fig. 1d). Because of this critical event in the history of tetrapod life on Earth, both the sand boa and the dog are adapted to life on land. The second student, relying on the evolutionary relationships depicted in the *cladogram* in Fig. 1d, engaged constructively with this diagram in an activity referred to as *tree thinking*.

Tree thinking is the type of reasoning biologists employ when considering questions of evolutionary relatedness. It is based on the paradigm of cladistics (Hennig 1966; Thanukos 2009), a method of grouping taxa based on shared, derived characteristics, with preference given to the *most recent common ancestor* (MRCA) of groups and not simply the common ancestor. Tree thinking has become the dominant model for organizing, retrieving, and predicting information within evolutionary biology. Natural groupings—known as *clades* or monophyletic groups—that comprise the MRCA and all its descendants are proposed based on the testable evidence of the presence of a shared, *derived* character (synapomorphy) that is



**Fig. 1** An example of biological classification: **a** a snake eel [from Tree of Life Web Project. 2006. Anguilliformes. true eels. Version 18 August 2006 (temporary). <http://tolweb.org/Anguilliformes/68835/2006.08.18> in The Tree of Life Web Project, <http://tolweb.org/>] (photograph copyright 2004 Felix Esteban; reproduced in black-and-white under the Creative Commons Attribution-NonCommercial-NoDerivs License—Version 2.0: <http://creativecommons.org/licenses/by-nc-nd/2.0/>]; **b** a dog (photograph copyright 2006 Novick); **c** a Kenyan sand boa (photograph copyright 2004 Novick); **d** a cladogram indicating evolutionary relationships among these three taxa

restricted to members of the proposed clade. The resulting phylogeny (or cladogram; see Fig. 1d) is a “road map” of the evolutionary relatedness among taxa and a powerful tool for scientists and students alike, allowing them to explore and test hypotheses regarding the history of life on the planet.

The global biological community is currently undertaking one its greatest scientific endeavors, that of constructing the Tree of Life [ToL; American Museum of Natural History (AMNH) 2002]. This phylogeny will be an evidenced-based, predictive road map of evolutionary relationships among all extant and extinct living things. The under construction ToL has already had important societal impacts on health, agriculture, and biotechnology [American Museum of Natural History (AMNH) 2002; Futuyma 2004; Hillis 2004; Yates et al. 2004]. As Novick and Catley (2013) noted, scientifically literate citizens need at least a basic understanding of the science behind the ToL to understand and engage in twenty-first century societal issues. This means students need to be able to engage in tree thinking, as scientists and science educators have been arguing for a long time (e.g., Catley 2006; Catley et al. 2005; Baum et al. 2005 Goldsmith 2003; Gregory 2008; Halverson et al. 2011 Nickels and Nelson 2005; O’Hara 1988).

The existing research on tree thinking has focused on undergraduates (often but not exclusively those enrolled in biology courses beyond the introductory level) (Catley et al. 2012; Catley et al. 2010; Halverson et al. 2011; Meir et al. 2007; Morabito et al. 2010; Novick and Catley 2007; Novick et al. 2010a, 2011a, b; Phillips et al. 2012; Sandvik 2008). We would argue, however, that high school, or even middle school, is not too early to teach students how cladograms are constructed, interpreted, and used as inferential tools, as this topic is part of the standards-based curriculum, which covers the unity, diversity, and classification of life of many countries. However, an examination of the UK, New Zealand, and Australian curriculum standards reveals that, while they provide ample amounts of appropriate content to be mastered (e.g., classification, evolutionary processes, adaptation and speciation), and while teaching these topics could be greatly enhanced by using a tree thinking approach, such a perspective does not seem to be advocated. Moreover, the opportunity to leverage tree thinking as a high value pedagogical and scientific tool is largely ignored in current US textbooks (Catley and Novick 2008) and the preeminent Australian and New Zealand biology text *Biozone* (see this paper).

### **International Science Curriculum Standards Related to Tree Thinking**

The use of trees to study evolution and classification is required at certain grade levels by the USA, UK, Australia, and New Zealand science standards. However, their potential to be used as examples in nature of science (NOS) studies with regard to hypotheses testing and inference as required by almost all curricula standards has yet to be realized.

*USA* Tree thinking has general relevance to science literacy and is critical specifically for interpreting ToL data. Moreover, understanding how to interpret cladograms is critical to several curriculum standards in the North Carolina Standard Course of Study (NCSCS), the US state in which the present study of high school students took place, as well as to the US National Science Education Standards (National Research Council 1996), on which it and other state standards are currently based. Moreover, the framework for the US’ Next Generation Science Standards highlights evolution as a cross-cutting concept and a core disciplinary idea of the life sciences (National Research Council 2012).

Consider the five unifying concepts of the biology curriculum, as stated in the NCSCS but that are also more widely applicable: Systems, order, and organization; Evidence, models, and explanation; constancy, change, and measurement; evolution and equilibrium; and form and function. Competence as a tree thinker requires domain knowledge and understanding that embraces all five unifying concepts. In particular, it includes understanding that the evidential basis of phylogenetic trees comes from derived characters (form and function=adaptations), that patterns of relationship (constancy, change, and measurement) require process explanations (evidence, models, and explanation), and that these processes involve micro and macroevolutionary mechanisms such as natural selection and speciation (evolution and equilibrium), which result in natural predictive classifications (systems, order, and organization). More specifically, in the new “Essential Standards” (North Carolina Standard Course of Study 2011) effective in all North Carolina schools in the 2012–2013 school year, we find the following: “Bio.3.4.1 Generalize what shared anatomical structures (homologies) tell us about evolution... Bio.3.5 Analyze how classification systems are developed upon speciation; Generalize the changing nature of classification based on new knowledge generated by research on evolutionary relationships... Bio.3.5.2 Analyze the classification of organisms according to their evolutionary relationships (including dichotomous keys and phylogenetic trees); Compare organisms on a phylogenetic tree in terms of relatedness and time of appearance in geologic history.”

*UK* Evolution as a stand-alone topic is strangely missing from the current UK Standards (UK Department of Education, 2012: <http://www.education.gov.uk/schools/teachingandlearning/curriculum/secondary/b00198831/science/ks3/attainment/organisms>), although they are currently being revised. Even so, opportunities exist to utilize a phylogenetic perspective in studying taxa and their place in time and space under the current Attainment Target 2: “Organisms, their behaviour and the environment.” While only mandated at levels 4 and 5 of target 2 (and missing in 6–8), level 4 states “They recognise that evidence can support or refute scientific ideas, such as in the identification and grouping of living things,” and at level 5, “They recognise that both evidence and creative thinking contribute to the development of scientific ideas, such as the classification of living things.” However, the general principals of NOS, including hypothesis building, testing, and inference inherent in tree thinking, also lends itself very well to use in the attainment target 1: “How science works.”

*Australia* As part of their Foundation Year 2 (5–8 years of age) Australian students are required to “...use the senses to observe and gather information, describing, making comparisons, sorting and classifying to create an order that is meaningful” (Australian Curriculum, Assessment and Reporting Authority 2008 downloaded 10/31/2012 from <http://www.australiancurriculum.edu.au/Science/Curriculum/F-10>). Later, in year 3 (8–12 years of age), students are required to understand that “Living things can be grouped on the basis of observable features and can be distinguished from non-living things.” Moreover, under Nature and Development of Science, we find “Science involves making predictions and describing patterns and relationships.” Years 4 and 6 have more of an ecological focus and classification is not highlighted. However, in year 5, students are required to know that “Living things have structural features and adaptations that help them to survive in their environment.” By year 7 (12–15 years of age), the standards are very explicit: “There are differences within and between groups of organisms; classification helps organise this diversity.” Some of the elaborations provided for this standard are: considering the reasons for classifying such as identification and communication, grouping a variety of organisms on the basis of similarities and differences in particular features, considering how

biological classifications have changed over time, classifying using hierarchical systems, such as kingdom, phylum, class, order, family, genus, and species.

Years 8 and 9 are centered on organismal systems and ecology respectively, while in year 10 (15–18 years of age) students are introduced to evolution: “The theory of evolution by natural selection explains the diversity of living things and is supported by a range of scientific evidence.”

Some of the elaborations provided for this standard are the following: describing biodiversity as a function of evolution; evaluating and interpreting evidence for evolution, including the fossil record, chemical, and anatomical similarities; and geographical distribution of species. It is difficult to see how students could understand and investigate these broad and powerful concepts without the benefit of a tree thinking habit of mind.

*New Zealand* New Zealand published their new science standards in 2007 (Ministry of Education 2007), and they were fully implemented in 2010. The new contextual living world (biology) standards have capitalized on the opportunity to elevate the central role and importance of evolution in the life sciences. The overview of the biology strand states that: “Students [will] develop an understanding of the diversity of life and life processes, of where and how life has evolved, of evolution as the link between life processes and ecology, and of the impact of humans on all forms of life” (Ministry of Education 2007). However, “while the 2007 New Zealand science curriculum has implemented some important steps to strengthen the presence of evolution and the nature of science, the document lacks meaningful examples that help teachers to develop their classroom teaching in these areas” (Campbell and Otrell-Cass 2011, p. 447). They go on to call for significantly increased efforts in teacher professional development as they see many teachers lacking in the resources and experience needed to give NOS and particularly evolution the attention demanded by the new curriculum. Campbell and Otrell-Cass (2011, p. 450) further state, “In emphasising the significance of biological evolution, the 2007 New Zealand curriculum has sent a strong signal to science teachers and educators about the need to take action in developing their students’ understanding of the importance and implications of biological evolution.”

### Example Curriculum Materials

As an exemplar of a respected and widely used biology curriculum, *Biozone* course workbooks (AP biology Australian edition) and the modular *Evolution* text were examined in respect to their use of and reference to phylogenetic trees. These workbooks are available in USA/International, UK, New Zealand, and Australian editions. In general, while students are exposed to phylogenetic diagrams/cladograms in these textbooks, they are far and few between. Even so, students are presumably expected to be able to interpret them as no instruction or annotation of the diagrams is provided.

The chapter on classification in *Year 12 Biology* (Australian edition) disappointingly presents but a single tree in the diagonal ladder format—a format shown to be significantly more difficult to interpret than the rectangular tree format (Novick and Catley 2007). Moreover, the chapter does not use a phylogenetic, character-based perspective at all, but is instead, a traditional march through the phyla. Despite the fact that one of the key concepts of this chapter is “Organisms are put into taxonomic categories based on shared, derived characters,” this critical and fundamental concept of modern classification is not addressed further in the chapter. In the chapter

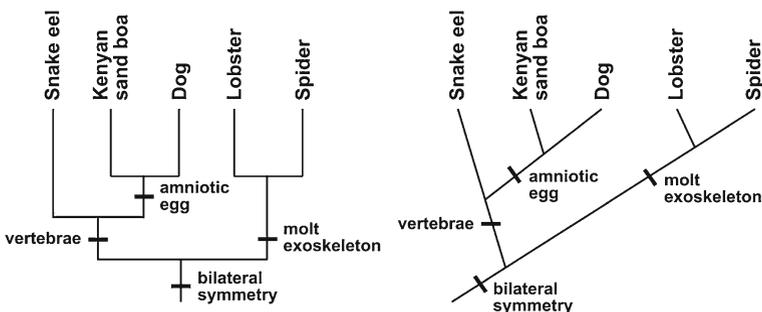
providing evidence of evolution, we find but three trees and one ladder, plus the much reproduced diagram of horse evolution based on Simpson's work from the 1960s which has been shown empirically to encourage anagenetic as opposed to cladogenetic thinking (Catley et al. 2010; Novick et al. 2011b).

In the chapter in *Year 12 Biology* on evolution, we find one valid cladogram and two simplified phylogenetic diagrams that follow no standard format. Also included is another nontree diagram, purporting to depict adaptive radiation in mammals. In spite of the fact that a stated key concept of this chapter is "Large scale patterns of evolution involve the diversification and extinction of species," the tools that biologists use to understand and explore these patterns, phylogenetic trees, are poorly utilized. The same general pattern is followed in the *Biozone* modular Evolution text.

### An Imperative to Investigate Tree Thinking in Secondary Students

The importance of students understanding evolution through the lens of tree thinking is finally being recognized within the education community (e.g., Baum et al. 2005; Catley 2006; Catley et al. 2005; Goldsmith 2003; Gregory 2008; Halverson, et al. 2011; Nickels and Nelson 2005; O'Hara 1988). The implementation of the new North Carolina "Essential Standards," the focus on evolution in the US' Next Generation Science Standards and the 2007 NZ standards, the revision of the UK standards, and the well established Australian standards that cover evolutionary processes and classification at a number of grade levels, all lend a sense of urgency to research on secondary students' ability to engage in tree thinking. Designing effective tree-thinking curricula requires understanding how students currently approach such tasks. Because previous research has focused almost exclusively on college students, the present study investigated tree-thinking skills in high school students.

Cladograms are typically drawn in one of two formats: a rectangular tree format and a diagonal ladder format (see Fig. 2). Early studies showed that even college students enrolled in biology classes beyond the introductory level have difficulty understanding the ladder format. Novick and Catley (2013) manipulated how the cladograms were drawn and found that students had much greater difficulty understanding the ladder than the tree format. They argued, therefore, that the ladder format should not be used in high school or college introductory biology classes. Following this suggestion, the present study of high school students used the tree format exclusively.



**Fig. 2** Tree and ladder cladograms depicting the same set of relationships among five taxa

## Five Core Tree-Thinking Skills

Novick and Catley (2013) described five core tree thinking skills (some of which involved multiple subskills) that underpin evolutionary biologists' ability to understand and reason with cladograms. We tested high school students' competence at a reduced set of the subskills, as outlined in Table 1. Because our prior research (Novick and Catley 2013; Catley et al. 2012; Phillips et al. 2012) indicates that college students do not understand unresolved structures (i.e., situations in which more than two lineages emerge from a branching point), it made little sense to probe high school students' understanding of such structures. In addition, we excluded questions presented in nested parenthetical notation that we expected would be unfamiliar to high school students.

The five core tree-thinking skills involve basic comprehension of cladograms (skills I and II), understanding the relationships depicted in these diagrams (skills III and IV), and using the relationships to support inferences (skill V). These cognitive skills are ordered from lower to higher (I and II, then III and IV, then V) such that they approximate Bloom's Taxonomy (Bloom 1956). Novick and Catley (2013) considered mastery of these five skills to be the minimum required to be literate in tree thinking. Novick and Catley assessed these skills in two groups of college students, who varied in their college coursework in biology. We focus here on the performance of the weaker background students, who had taken 0–2 college biology classes ( $M=0.59$ ), with none of those classes above the introductory level because they provide the more relevant comparison group for high school students. Throughout this article, when we refer to the performance of *college students* in Novick and Catley's study, we mean this subset of students when answering questions about cladograms in the tree format. In addition, we restrict the questions to those that are the same as or isomorphic to the ones used in the present study to provide the most appropriate comparison for the present results.

Skill I involves understanding that the terminal taxa in a cladogram inherited the characters possessed by their ancestors. Questions assessing this skill probe students' understanding that the character possessed by the MRCA of two taxa is the one they

**Table 1** Tree-thinking skills assessed, along with the associated question types

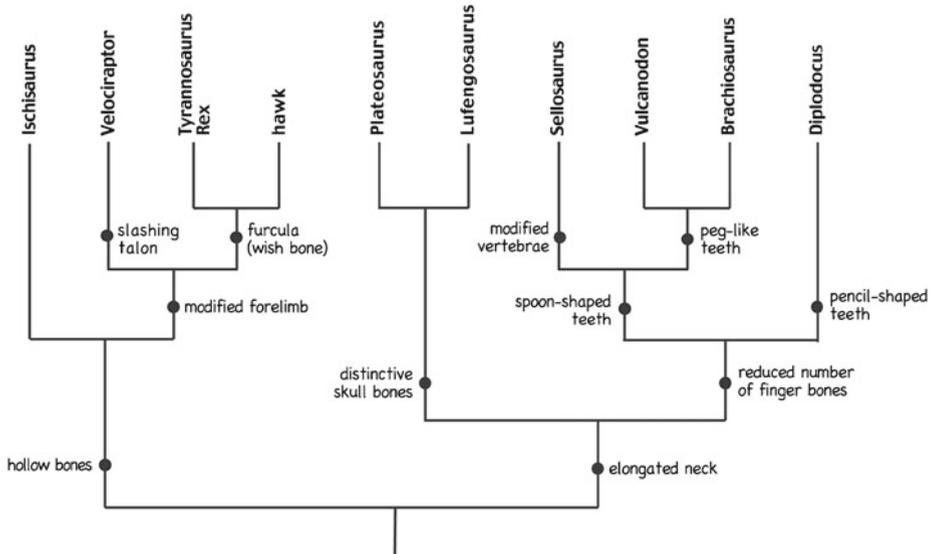
Tree-thinking skill assessed
I. Identify the character shared by two or more taxa due to inheritance from their MRCA
A. Identify the character that is the marker for the MRCA of two taxa that comprise a clade
B. Identify the character that is the marker for the MRCA of more distantly related taxa
II Identify a set of taxa based on character information provided
A. Identify taxa that evolved from a taxon with a specified character
B. Identify taxa that did not evolve from a taxon with a specified character
III. Understand the concept of a clade
A. Recognize that bracketed taxa do not comprise a clade
1. Taxa located at different levels
B. Recognize that bracketed taxa comprise a clade
IV. Evaluate relative evolutionary relatedness
A. Reference taxon is the most derived of the three taxa
B. Reference taxon is at an intermediate level between the remaining two taxa
V. Use evidence of most recent common ancestry to support an inference regarding a shared character

This table is adapted from Table 2 in Novick and Catley (2013)

encounter first when tracing back from each taxon to their branching point. For example, for the cladogram in Fig. 3, students could be asked to name the character possessed by the MRCA of *Sellosaurus* and *Brachiosaurus* (answer=spoon-shaped teeth). College students did quite well at answering such questions ( $M=0.84$ ; Novick and Catley 2013).

Skill II involves the ability to divide a set of taxa into two groups at a point of divergence specified by a particular character—i.e., to identify the taxa that did (skill II.A) or did not (skill II.B) evolve from an ancestor with a particular character (e.g., see question 1 in Fig. 3). College students did quite well on these questions overall ( $M=0.86$ ; (Novick and Catley 2013), although they were

**Use this diagram showing evolutionary relationships among dinosaurs to answer the following questions. Note that the dots on the diagram indicate characters.**



1. List all taxa (might be one or more) that did not evolve from an ancestor that had a reduced number of finger bones. [*Ischisaurus*, *Velociraptor*, *Tyrannosaurus rex*, *hawk*, *Plateosaurus*, *Lufengosaurus*.]
2. Which taxon—*Plateosaurus* or *Diplodocus*—is the closest evolutionary relation to *Sellosaurus*? [*Diplodocus*.] Explain your answer: [*They share a most recent common ancestor*.]
3. Given that birds are warm-blooded, which taxon is most likely to share this character? [*Tyrannosaurus rex*.] Explain your answer: [*They share a most recent common ancestor*.]

**Fig. 3** The dinosaur cladogram and associated questions. The answers are provided in italics and brackets following each question

significantly less likely to identify the correct taxa when the question was worded negatively (as in Fig. 3) rather than positively. We expected the same pattern of results for high school students.

Skill III involves understanding the concept of a clade, as defined earlier. We assessed two aspects of this skill—understanding that a certain set of taxa do not (skill III.A) versus do (skill III.B) comprise a clade. For example, given the cladogram in Fig. 3, we could ask whether *Sellosaurus* and *Vulcanodon* comprise a clade. The answer is no because there is another descendant of their MRCA (located at the character *spoon-shaped teeth*), namely, *Brachiosaurus*. If we instead asked whether *Sellosaurus*, *Vulcanodon*, and *Brachiosaurus* comprise a clade, the answer would be yes. College students did less well on these questions that require understanding relationships, with a mean accuracy of 0.70 (Novick and Catley 2013).

Skill IV involves understanding that relative evolutionary relatedness depends on how recently taxa shared a common ancestor. Taxa that share a more recent common ancestor are more closely related than taxa whose MRCA is more ancient. Novick and Catley (2013) distinguished two aspects of this skill, which they predicted to differ in difficulty. The easier subskill (skill IV.A) is illustrated by question 2 in Fig. 3. In this case, the reference taxon, *Sellosaurus*, is the mostly highly derived of the three taxa, which means that the correct answer can be found by going down levels in the hierarchy until finding the branching point with one of the two comparison taxa. The comparison taxon whose branching point with the reference taxon is encountered first (*Diplodocus* in the example) is the correct answer because that taxon shares a more recent common ancestor with the reference taxon. Nearly all college students answered this type of question correctly ( $M=0.94$ ). However, when the comparison taxon (e.g., *Velociraptor*) was at an intermediate hierarchical level between the two comparison taxa (e.g., *Ischisaurus* and hawk; skill IV.B), accuracy dropped to 0.44, which is not different from chance (Novick and Catley 2013). Because of how the three taxa were situated in their cladograms, these results indicate that students often (incorrectly) responded based on the spatial proximity of the taxa. We expected the same pattern of results for high school students.

Finally, skill V involves using most recent common ancestry to make inferences about shared characters, such as for question 3 in Fig. 3. The correct answer for this question is *Tyrannosaurus rex* because that taxon shares a more recent common ancestor with birds (represented by hawk on the cladogram) than does any other taxon on the cladogram. Answering this question requires making an inference because information about warm-bloodedness per se is not provided on the cladogram. The only information available concerns evolutionary relationships among a set of dinosaur taxa, including one that belongs to the bird class (Aves).

Novick and Catley (2013) found that college students had more difficulty answering this dinosaur inference question than a comparable question about insect taxa (means of 0.27 and 0.62, respectively). They hypothesized that this was because the correct answer to the dinosaur question is counter to students' folkbiological belief that birds are a distinct kind of living thing, rather than an extant dinosaur. Because the dinosaur and insect cladograms were drawn in different formats, however, it was impossible to know whether the cause of the accuracy difference was students' prior knowledge or the format of the cladogram, which was found in other cases to affect students' responses. The present study provides a clear test of the prior knowledge hypothesis because both cladograms were drawn in the tree format.

## Overview of the Present Study

This study provides the first in-depth look at high school students' baseline competence at tree thinking. Students completed a 12-question instrument that assessed the five basic tree-thinking skills outlined in Table 1. These questions were taken from the instrument used by

Novick and Catley (2013), although in a few cases, the questions were asked about a different set of taxa (to equalize the number of questions asked about each cladogram) or about a cladogram drawn in a different format (see comment above concerning the dinosaur cladogram). As noted earlier, we excluded questions that were expected to be too difficult for high school students to understand. Students' success, or lack thereof, at answering these questions will provide both an indication of their current level of competence at tree thinking and guidance for creating a curriculum to teach tree thinking to high school students enrolled in a basic biology class.

We also evaluated students' general understanding of how cladograms depict ancestry. In a cladogram, no taxon that labels a terminal branch is the ancestor or descendant of any other taxon. Instead, common ancestors of sets of nested taxa are indicated by characters placed at the points of divergence. For example, in Fig. 3, the MRCA of *Tyrannosaurus rex* and hawk is denoted by the character furcula (wishbone), which that ancestor passed on to these descendants. Nevertheless, students may misinterpret how ancestry is represented on a cladogram, which would be evident in statements about one taxon being the ancestor or descendant of another taxon.

## Method

### Subjects

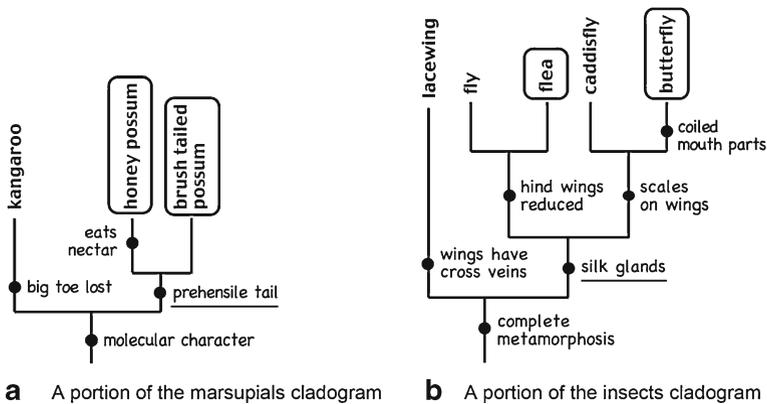
The subjects were 60 students (31 male and 29 female) enrolled in a biology class at a public high school in a rural area of western North Carolina. The life science curriculum at this school follows the North Carolina Standard Course of Study and the students are considered typical of a public high school in this region. All students were tenth graders except for one ninth grader and one 11th grader ( $M=15.65$  years, range of 15–17 years). The sample included 56 white, 2 ethnic minorities, and 2 students of unreported ethnic identity.

### Materials

*Overview* Subjects received a four-page cladogram reasoning booklet that included four cladograms with three questions about each (e.g., see Fig. 3). Each cladogram showed 7–10 taxa and 10–12 characters. The taxonomic groups included placental mammals (Fig. 1 in Novick and Catley 2013), marsupial mammals (Fig. 2 in Novick and Catley 2013), insects (Fig. 4 in Novick and Catley 2013), and dinosaurs (Fig. 3 here). The first three cladograms were identical to those used by Novick and Catley. The dinosaur cladogram was used by Novick and Catley in the ladder format; it was redrawn in the tree format for this study. The four cladograms were collated in two different orders. (Cladogram order did not have a significant effect on students' performance. Thus, this factor was not included in the analyses reported here.)

*Questions* Each of the five tree-thinking skills (see Table 1) was represented by two to four specific questions (12 questions total), with some containing multiple parts. In most cases, these questions were identical to those Novick and Catley (2013) used with college students. As noted earlier, in a few cases, we wrote isomorphic questions for different cladograms. A complete list of the questions can be found in the [Appendix](#).

Two character identification questions required students to identify the character that supports the MRCA of (a) two taxa that comprise a clade (skill I.A; e.g., prehensile tail for honey possum and brush tailed possum in Fig. 4a) and (b) two



**Fig. 4** A portion of the insect cladogram (a) and marsupial mammals cladogram (b) used to assess students' understanding that the terminal taxa inherited the characters possessed by their MRCA. These figures were adapted from Figs. 2 and 4, respectively, in Novick and Catley (2013). The two taxa mentioned in the question are circled, and the correct response is underlined

more distantly-related taxa (skill I.B; e.g., silk glands for fleas and butterflies in Fig. 4b). These questions were worded as follows: “What character was possessed by the most recent common ancestor of fleas and butterflies?” Two taxa identification questions tested students' ability to identify the taxa that share (skill II.A) or do not share (skill II.B; see question 1 in Fig. 3) a particular character.

Two questions assessed students' understanding of clades (skill III). Because *clade* is a technical term with which students were expected to be unfamiliar, they were given a definition: “A *clade* is a group that includes the most recent common ancestor of the group and all descendants of that ancestor.” They were asked whether the bracketed taxa “labeled A” form a clade (a bracket was placed over 2–3 adjacent taxon names above the cladogram, and the letter *A* was placed above the bracket). Students had to circle *yes* or *no* and provide a written explanation supporting that response. If students responded *no*, they were further asked to indicate how to make the group a clade (i.e., to identify which taxa need to be added or removed). One question required students to recognize that a marked set of taxa do not comprise a clade (skill III.A); a second question required them to recognize that the taxa do comprise a clade (skill III.B).

Because evolutionary relatedness is the quintessential tree-thinking concept, we devised four questions to test students' understanding that evolutionary relatedness is determined by most recent common ancestry. Students were asked to identify which of the two comparison taxa is most closely related to a third, target taxon. For two questions, the target taxon was the most highly derived of the three taxa (skill IV.A). For example, for question 2 in Fig. 3, *Sellosaurus* is the most derived (evolved most recently) and *Plateosaurus* the least derived (evolved least recently) of the three taxa. For the other two questions, the target taxon was located at an intermediate hierarchical position between the two comparison taxa (skill IV.B). For both question types, students were asked to explain their answer.

Finally, two questions required students to make an inference about a shared character, not depicted on the cladogram, based on most recent common ancestry (skill V). One question referenced taxa in the dinosaur cladogram (question 3 in Fig. 3); the other question concerned taxa in the insect cladogram.

## Procedure

Students completed two problem booklets, which constituted data collection for different studies. The cladogram reasoning booklet was completed second. The first booklet contained 24 problems. For each problem, students were told about the enzymes that two pictured taxa use to help regulate cell function and were asked which of these enzymes they thought a third pictured taxon uses for this function: e.g., alligators use enzyme X and groundhogs use enzyme Y; which enzyme do you think ostriches use? The results of this study are presented elsewhere (Novick et al. 2011a). After the second booklet, students completed a questionnaire that asked for background information such as sex and year in school. Students participated in one of several group sessions conducted in a classroom at their high school but outside the normal school day. Students worked individually, without access to outside resources, to complete the tasks and received \$10 for their participation.

## Results

### Overview

The close-ended questions received accuracy scores of 1 (correct) or 0 (incorrect). For each skill, we conducted a one-factor repeated-measures analysis of variance (ANOVA) on these scores, with question type as the within-subjects factor. An alpha level of .05 was the criterion for statistical significance.

Students' open-ended explanations were content coded using criteria modeled after those used by Novick and Catley (2013) to assess college students' responses to such questions. Some codes were identical to those used in the earlier study; others were determined based on common responses provided by the high school students. The codes were ordered from most to least sophisticated. The second and third authors independently coded the written responses, giving each response a single code. In the rare cases that multiple codes applied, the response was given the highest quality code in the ordering. The two coders also evaluated students' understanding of how ancestry is mapped onto cladograms by coding whether students ever gave an explanation stating (incorrectly) that one terminal taxon is the descendant or ancestor of another terminal taxon: e.g., "they are the ancestors of every other organism in the study."

Overall, there were 500 agreements and 40 disagreements, for an agreement rate of 92.6 %. The coders resolved the discrepancies by discussion. The codes for each question are reported in the sections where they are relevant. The results for the ancestry mapping measure are discussed after those for the five tree-thinking skills.

### Identifying Shared Characters

The two questions differed in whether students were asked to identify the shared character for a clade consisting of two most closely related taxa (Fig. 4a) or two more distantly related taxa (Fig. 4b). Replicating Novick and Catley's (2013) results for college students, there was not a significant effect of question type,  $F(1, 59)=2.32$ ,  $MSE=0.09$ , partial  $\eta^2=.04$ . With an overall mean accuracy of 0.68, however, the high school students did much worse than the college students, who had a mean of 0.84 for similar questions,  $t(59)=-3.04$ ,  $SEM=0.05$ .

We coded students' errors into three categories. Students who identified a character possessed by just one of the two mentioned taxa (e.g., *eats nectar* in Fig. 4a; *hind wings*

*reduced* or *scales on wings* in Fig. 3b) or the character possessed by the *least* recent common ancestor of the two mentioned taxa (located at or near the bottom of the cladogram) received the character code. This accounted for 44 % of students' errors. Another consistent error was to identify the taxon situated one level below the MRCA of the two mentioned taxa and immediately to the left of the clade that includes those taxa (33 %)—either *kangaroo* (Fig. 4a) or *lacewing* (Fig. 4b). The remaining errors (23 %) included references to other taxa or characters (e.g., *butterfly*, *complete metamorphosis*) and noninterpretable responses (e.g., “I don't know”).

### Identifying Taxa Based on Common Descent

The two questions differed in whether students were asked to identify the taxa that did (a positively worded statement) or did not (a negatively worded statement) evolve from an ancestor that had a particular character. The ANOVA indicated a significant effect of question type,  $F(1, 59)=15.63$ ,  $MSE=0.15$ , partial  $\eta^2=.21$ , with higher accuracy for the positively than the negatively worded question ( $M=0.75$  vs.  $M=0.47$ , respectively). The college students in Novick and Catley's (2013) study showed the same effect ( $M=0.91$  vs.  $M=0.81$ , respectively). Again, the high school students ( $M=0.61$ ) did much worse than the college students ( $M=0.86$ ),  $t(59)=-5.12$ ,  $SEM=0.05$ .

### Comprehending Clades

*Accuracy* For one question, the bracketed taxa comprised a clade. For the other question, the grouping did not include *all* descendants of the MRCA and therefore did not comprise a clade. The ANOVA indicated no effect of question type,  $F(1,59)=0.00$ ,  $MSE=0.20$ , partial  $\eta^2=0.00$ , with an overall mean accuracy of 0.48, which is not different from chance,  $t(59)=-0.33$ ,  $SEM=0.05$ . Again, this is much worse than the accuracy rate for the college students in Novick and Catley's (2013) study ( $M=0.70$ ),  $t(59)=-4.30$ ,  $SEM=0.05$ . Another difference between the high school and college results is that college students did better on the not-a-clade ( $M=0.78$ ) than the is-a-clade ( $M=0.62$ ) question.

*Evidence cited* Students' explanations for their clade responses received one of six evidence codes, which are presented in order from most to least sophisticated in Table 2. The most scientifically valid explanations mentioned (a) that the bracketed taxa did (is-a-clade question) or did not (not-a-clade question) include the MRCA and *all* its descendants (e.g., “They do not share all descendants of the most recent common ancestor...”) and (b) the presence (is-a-clade question) or absence (not-a-clade question) of *all* the descendants (e.g., “because the character

**Table 2** The percentage of students' explanations for the clade questions that received each evidence code as a function of question type

Code	Is-a-clade question	Not-a-clade question	Mean across questions
ANCEST	8	12	10
SC	5	0	3
MANC	15	7	11
CONN	8	7	7
SIM	20	25	23
OTH	43	50	47

Percentages sum to 100 (within rounding error) in each column

that they both possess contains more taxa than is shown”, marsupial mammals), or (c) the common ancestor and its associated character (e.g., “all three come from ancestors who walked on palm of foot”, placental mammals). Because these three types of explanations were very rare, we assigned all such responses to a composite ancestry category (ANCEST).

The next best response was to refer to the derived character (synapomorphy) shared by the taxa (SC). For example, for the placental mammal cladogram, one student wrote: “They are a clade because they all have the ‘walks on palm of foot’ trait.” As can be seen in Table 2, these responses also were very rare.

The remaining four codes were as follows: (a) missing ancestor (MANC), references to a missing ancestor or an ancestor not shown in the cladogram; (b) connection (CONN), any reference to the taxa being closely linked or connected by the lines in the cladogram; (c) similarity (SIM), statements regarding the taxa having or not having similar characteristics; and (d) other (OTH), any other explanation. OTH responses were idiosyncratic or non-interpretable. Examples of responses that received these four codes are given in the paragraphs that follow.

Perhaps the strongest evidence from the written responses that students had difficulty understanding the concept of a clade is that nearly half of their explanations received the OTH code. Examples of such responses include (a) “has lots underneath it,” (b) “they are the highest up on the taxa,” (c) “they are not related in any way,” and (d) “I cannot find a conclusion.”

Nearly another quarter of students’ explanations received the SIM code. Examples of such explanations include: “They all have the same characteristics” and “they have the most in common.” It is difficult to know the extent to which such responses were based on the evidence provided in the cladograms or on students’ prior knowledge of the taxa.

Looking at the responses from the top end of the coding scheme, only 10 % of students made use of key evolutionary concepts regarding shared ancestry and/or descendants (ANCEST). All such explanations were given by students who interpreted the clade status of the bracketed taxa correctly. Only a further 3 % of students referred to the relevant character evidence.

Similarly, small percentages of the justifications received the MANC and CONN codes. MANC responses indicate that students did not understand that the characters in the cladograms are markers for the ancestors: e.g., “This group only contains descendants, there is no ancestor shown” and “no single ancestor.” CONN responses mentioned the connections in the cladogram: e.g., “they broke off closest to each other” and “because they all have traits that are connected.”

*Clade Modifications* Students who indicated that the bracketed taxa in either question did not comprise a clade were asked to propose a modified grouping that does constitute a clade. For the not-a-clade question, for which the grouping did not include all descendants of the MRCA, students had to add taxa to create a clade. Of those who correctly indicated that the bracketed taxa did not comprise a clade, only 33 % listed the correct taxa to be added.

For the is-a-clade question, the marked grouping included all descendants of the MRCA, so no proposed changes were needed. Novick and Catley (2013) found that 50 % of college students who said the bracketed taxa for this question do not comprise a clade proposed to create a clade by removing the least related taxon from the group, suggesting that they thought a clade can only include two taxa that are most highly derived. Similarly, in the current study, 42 % of high school students who responded incorrectly that the marked grouping does not comprise a clade provided such a response.

### Evaluating Evolutionary Relatedness

*Accuracy* To assess students’ understanding of evolutionary relatedness, we computed a mean accuracy score across the two questions of each type. The ANOVA indicated a

significant effect of question type,  $F(1, 59)=15.76, p<.001, \text{MSE}=0.42, \text{partial } \eta^2=.21$ , with students doing better when the reference taxon was the most highly derived of the three taxa than when it was located in an intermediate position ( $M=0.84$  vs.  $M=0.60$ , respectively). These results are very similar to what Novick and Catley (2013) found for college students ( $M=0.94$  and  $M=0.44$ , respectively). High school students' accuracy on the questions regarding an intermediately located reference taxon did not differ significantly from chance,  $t(59)=1.85$ .

*Evidence cited* Students' written justifications for their evolutionary relatedness responses were coded into eight categories. Table 3 shows the mean percentage of justifications receiving each code (collapsed across the four questions) as a function of accuracy on the relationship question. We report students' explanations for these questions as a function of accuracy because there were few differences as a function of question type (the exceptions are described below).

As for the clade questions, referencing the MRCA of the two taxa was the most scientifically valid explanation. However, no students gave such an explanation. The next best response was to write that the target taxon and the selected taxon shared an ancestor with the relevant character (i.e., the one possessed by the MRCA). For example, one student wrote that *Diplodocus* and *Sellosaurus* (see Fig. 3) are closest evolutionarily "because they both come from an ancestor with reduced numbers of finger bones." Other students mentioned the relative degree of relatedness between taxa: e.g., "because it's on the same branching thing as *Sellosaurus*, which means it's more closely related." Because these two types of explanations were rarely given, we assigned them to a composite ancestry-relatedness (ANC-REL) category. ANC-REL explanations were more common for students who provided correct than incorrect answers.

Of the remaining codes, shared character (SC), described earlier, was the most sophisticated. Students who provided an SC explanation were reasoning about the structure of the diagram (i.e., branching events) when asked to evaluate the relative relatedness of taxa. It is not surprising, therefore, that such explanations were associated with correct responses. For example, a student who correctly answered the dinosaur question wrote "Diplodocus shares the reduced number of finger bones trait." Another student correctly identified the koala as the closest evolutionary relation to the brush tailed possum in the marsupial cladogram and explained "the Koala and the brush tailed possums evolved from one pair of lower teeth." Students were much more likely to provide SC responses for correct answers when the target taxon was located at an intermediate level rather than the most derived position in the cladogram (62 vs. 24 %).

**Table 3** Evaluating evolutionary relatedness: the mean percentage of explanations receiving each evidence code as a function of accuracy

Evidence cited	Correct response	Incorrect response
MRCA	0	0
ANC-REL	8	1
SC	43	5
STEPS	8	10
CONN	18	33
KNOW	1	11
SIM	17	23
OTH	6	16

Percentages sum to 100 (within rounding error) in each column

A small percentage of students referenced the relative number of evolutionary steps and/or derived characteristics between taxa (STEPS) for both correct and incorrect responses. For example, for the dinosaur question, one student wrote: “Diplodocus is 2 links away where as [*sic*] plateosaurus [*sic*] is 3 away.” Many more students relied on evidence regarding how closely connected (CONN) the taxa were to one another, especially to explain incorrect judgments. For example, one student wrote that *Plateosaurus* was the closest evolutionary relation to *Sellosaurus* because “They can both trace down to the same place.” Another student, who also answered incorrectly, wrote that the giraffe is the closest evolutionary relation to the horse because “they are connected earlier [*sic*] on the diagram than giraffes.”

An equally small percentage of students relied on their prior folkbiological knowledge (KNOW) to reason about the taxa in question, particularly to support incorrect judgments. These justifications often referred to the physical characteristics of taxa. For example, referring to the giraffe and horse, one student wrote: “I think so because they share characteristics such as body hair (coat, mane, tail, etc.) and hooves.” Another student wrote: “Because they are kind of similar to horses. Minus the giraffe has a much longer neck, but they eat like each specie [*sic*].” To receive the KNOW code, a student’s explanation had to include a reference to specific information that was not depicted on the cladogram. General statements about shared physical similarities were coded as SIM (described earlier): e.g., “has similar characteristics” or “they have more in common.” SIM responses were quite common. The remaining highly idiosyncratic responses were coded as OTH.

#### Inferences about Characters and Shared Ancestry

*Accuracy* One of the most important tree-thinking skills is the ability to make inferences regarding taxa that may share characters derived from a MRCA. Two questions assessed students’ inferences about a shared character that was not depicted on the cladogram. Novick and Catley (2013) found that college students were more likely to get the insect than the dinosaur inference question correct (means of 0.62 and 0.27, respectively). Although the cladograms depicting those taxa were drawn in different formats, Novick and Catley (2013) argued that the difference was likely due to the different sets of taxa depicted. They hypothesized that students without strong backgrounds in biology had difficulty thinking of birds as dinosaurs (see Fig. 3), which impaired their reasoning on that question. To test this hypothesis, we drew the dinosaur and insect cladograms in the same format, thereby untangling the two factors. The ANOVA yielded a significant effect of cladogram,  $F(1, 59) = 39.33$ ,  $p < .001$ ,  $MSE = 0.12$ , partial  $\eta^2 = .40$ , with greater accuracy on the insect than the dinosaur inference question ( $M = 0.67$  vs.  $M = 0.27$ , respectively), as predicted.

*Evidence cited* We used the same coding scheme that Novick and Catley (2013) devised for college students, except that we added CONN, defined earlier. We presented the data separately for the dinosaur and insect questions because students’ responses differed as a function of both accuracy and taxa (see Table 4).

As for the other questions, the best evidence is to write that the two named taxa share a MRCA. The next best response is to write that the two taxa share the character because they are most closely related (REL). Students also reasoned about how the taxa were connected to one another (CONN), a less sophisticated explanation: e.g., (a) “because the hissing cockroach shares a bracket with the harvest [*sic*] termite” and (b) “because it is in the same bracket as the hawk, which means they could be relatives and share traits.”

Students also provided categorical justifications (CATEG), stating for example that hawk is most likely to share warm-bloodedness with birds because “A hawk is a bird” and “The Hawk in [*sic*] the only non-reptile on the page. Its [*sic*] also a bird.” Students also relied on their prior knowledge (KNOW). For example, one student wrote that hawk is most likely also to be warm-blooded “because birds fly and so do hawks.” Students also referenced morphological characters (MORPH) of other taxa even when those characters are not shared with the named taxon from the question. For example, when asked to identify the taxon that is most likely to digest cellulose like termites, one student identified *butterfly* because “it has coiled mouth parts” even though this character is not shared with termites. These justifications were coded as least sophisticated.

For both cladograms, students who answered correctly provided more scientifically valid explanations (MRCA, REL, and SC) than did students who answered incorrectly. However, replicating the results for the clade and evolutionary relatedness questions, MRCA justifications were very rare overall. Students who answered correctly also referenced the degree of connectedness among taxa more often than did students who answered incorrectly.

Although prior knowledge (KNOW) explanations were given with similar frequency for incorrect responses to the two questions, other justifications for such responses differed according to the question asked. Nearly half of students who made an incorrect inference for the dinosaur question provided a categorical justification (CATEG), whereas only about a third of students gave such a justification for an incorrect inference for the insect question. These responses suggest that high school students consider birds to be a different kind of thing than dinosaurs even though in fact birds are dinosaurs and therefore reptiles (e.g., Lee et al. 2004).

In contrast, students were more likely to give MORPH explanations for the insect than the dinosaur question. It appears that students tried to identify a character on the cladogram that seemed relevant for digesting cellulose. Students may have been less likely to give such responses for the dinosaur question because they could not identify any characters on the cladogram that would support warm-bloodedness. Similar data and interpretations regarding character inferences in college students were reported by Novick and Catley (2013).

### Depicting Ancestry Relationships on Cladograms

To evaluate students’ general understanding of how cladograms depict ancestry, we calculated a global mapping error score by examining each student’s written responses for all questions for which an explanation was requested. Students received a score of 1 if they provided any explanation stating that one terminal taxon is the ancestor or descendant of another terminal taxon. For example, for an evolutionary relatedness question, one student wrote that “koalas are closely related to kangaroos which evolved from brush tailed possums.” Another student wrote that the bracketed taxa in the placental mammals cladogram was a clade (a correct response) because “the rock hyrax is the common ancestor of the manatee+elephant.” A score of 0 was given to students who never made such statements. This misunderstanding of how ancestry is mapped onto cladograms was common, with 32 % of students providing at least one such response.

### Discussion

This and our previous study (Novick and Catley 2013) show largely congruent patterns of errors and misunderstandings among both undergraduates and high school students. Given

that across the US tree thinking has not been a consistent part of high school or undergraduate evolution curricula, or even of high school biology teachers' preparation (Catley 2006), and is sparsely represented in the international standards examined, these results are not surprising. In the sections that follow, we discuss the consistent errors and misunderstandings evidenced by our high school student sample and then consider implications for improving evolution education at the high school level.

### Patterns of Errors and Misunderstandings

Our tree-thinking questions assessed five key skills that can be grouped into three broad categories that roughly track levels in Bloom's (1956) Taxonomy. Skills I and II (identifying shared characters and taxa), at the lowest level, require the ability to read information directly off the cladograms. Skills III and IV (evaluating clades and relative evolutionary relatedness) further require the ability to evaluate relationships among taxa on a cladogram. Finally, skill V requires the ability to make an inference about a new character given the relationship information depicted on a cladogram. Although students' average accuracy levels for these three categories track these cognitive levels (means of 0.65, 0.60, and 0.45, respectively), the more interesting story is that the variability in accuracy across questions within a skill, due to our experimental manipulations, demonstrates the importance of several other factors as well.

Our results provide evidence for three such factors: (a) *linguistic complexity*—when the skill II question was worded negatively rather than positively, accuracy was significantly reduced; (b) *relevant general knowledge*—when students were able to leverage their general knowledge of hierarchies to reason about evolutionary relatedness (skill IV) rather than having to understand how cladograms, a special type of hierarchy, represent such relationships, accuracy was enhanced; and (c) *the power of misconceptions*—when the relationships depicted among a set of taxa presented anomalous data, accuracy was impaired. Indeed, the lowest accuracy level ( $M=0.27$ ) across all our questions was found for the dinosaur inference question (skill V), which required students to reason with the anomalous information that birds are dinosaurs. Most students were unwilling to accept that conclusion of modern biology (based on the evidence provided).

Students were asked to provide a written explanation for their answers to the questions assessing skills III, IV, and V. With the exception of the evolutionary relatedness question for which students could respond based on their general knowledge of hierarchies ( $M=0.84$ ), accuracy on these questions was low, ranging from a mean of 0.27 to a mean of 0.62. It is not surprising, therefore, that students' explanations generally were devoid of reference to appropriate tree-thinking concepts such as most recent common ancestry and nested clades. Instead, they tended to rely on inappropriate concepts such as overall similarity, specific knowledge about the taxa learned previously (but not depicted in the cladograms), the idea that terminal taxa on a cladogram are ancestors and descendants of each other rather than of unnamed ancestors whose presence is indicated by characters, and a variety of other, idiosyncratic factors. The most positive news from the justifications is that in some cases students attended to the connections among the taxa in the cladogram and the appropriate shared characters. Although such responses reflect far from mature understanding, their presence suggests that these may be appropriate entry points in instruction for scaffolding students' rudimentary hierarchical knowledge toward a deeper understanding of the structure of cladograms.

## Implications for High School Instruction

These data leave little doubt that high school students do not understand the crucial concept of most recent common ancestry and have weak or nonexistent conceptual frameworks to help them with the task of recognizing clades. Consequently it is critical that any tree-thinking curriculum be research-based and situated in a learning progression that introduces elements in a developmentally timely manner (Catley et al. 2005). General principles that emerge from this study are the following: (a) Students need to understand the vital importance of and crucial difference between most recent common ancestry and common ancestry, and (b) they need to be taught that cladograms provide an authoritative source of evidence that should be weighted more heavily than superficial or ecological similarities among taxa and should override their prior folkbiological knowledge. Further, teachers should seek out opportunities to frontload other elements of the curriculum, such as when teaching genetics or the structure and function of cells, both typically taught prior to evolution, with an eye to foreshadowing the implication of these concepts for understanding evolution writ large. For example, meiosis is critical for creating the variation that ultimately leads to the patterns of speciation their students will explore using phylogenetic trees. Teachers should also strive to use current and real world examples whenever possible to illustrate the power and utility of using trees to address important societal issues [e.g., American Museum of Natural History (AMNH) 2002; Futuyma 2004; Hillis 2004; Yates et al. 2004].

A further use of phylogenies and tree thinking assessments is to generate and support classroom discussions. One can easily imagine situations when counterintuitive relationships, perhaps that birds are reptiles or that mushrooms are more closely related to kangaroos than they are to gum trees, are in direct conflict with students' folkbiological knowledge and provide a rich source of cognitive dissonance. Exploration of what processes might account for the patterns of relationships depicted in cladograms are far ranging and need to be framed in terms of macroevolutionary processes such as deep time, biogeography, adaptation, radiation, and extinction.

A basic tree thinking curriculum should provide an understanding of (a) the science behind the construction of cladograms, (b) how to read and interpret the information conveyed in cladograms, and (c) the utility and inferential power of these diagrams in the real world.

As first steps towards this goal, we are currently developing high school curriculum materials based around our research-based undergraduate instructional unit (Novick et al. 2010b) that has been used effectively in introductory and upper level biology classes in several colleges across the USA (unpublished data). The questions used in the present assessment will inform the development of an instrument to assess the effectiveness of the curriculum. Given the pervasive use of cladograms in the biological sciences, we look forward to the time when introducing grade school students to tree thinking representations and the science behind them is as commonplace as introducing those students to mathematical concepts and their graphical representations such as graphs, histograms, and charts.

**Acknowledgments** We thank Randi Neff our collaborating high school teacher for her advice and insight and Jena Gladden, our research assistant, for her help collecting data. The research reported here was

supported by the Institute of Education Sciences, US 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 US Department of Education.

## Appendix

Cladogram showing relationships among 10 placental mammal taxa (Fig. 1 in Novick and Catley 2013):

- P1. List all taxa (might be one or more) that evolved from an ancestor that had a single-chambered stomach.
- P2. Which taxon—giraffes or elephants—is the closest evolutionary relation to horses?

\_\_\_\_\_  
Explain your answer:

- P3. A *clade* is a group that includes the most recent common ancestor of the group and all descendants of that ancestor. Do the bracketed taxa labeled “A” form 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.*)

Cladogram showing relationships among 7 marsupial mammal taxa (Fig. 2 in Novick and Catley 2013):

- M1. What character was possessed by the most recent common ancestor of honey possums and brush tailed possums?
- M2. Which taxon—koalas or bandicoots—is the closest evolutionary relation to brush tailed possums? \_\_\_\_\_

\_\_\_\_\_  
Explain your answer:

- M3. A *clade* is a group that includes the most recent common ancestor of the group and all descendants of that ancestor. Do the bracketed taxa labeled “A” form 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.*)

Cladogram showing relationships among 10 insect taxa (Fig. 4 in Novick and Catley 2013):

- I1. What character was possessed by the most recent common ancestor of fleas and butterflies?
- I2. Which taxon—dragonfly or butterfly—is the closest evolutionary relation to the walking stick? \_\_\_\_\_

\_\_\_\_\_  
Explain your answer:

- I3. Given that termites digest cellulose, which taxon is most likely to share this character?

\_\_\_\_\_  
Explain your answer:

Cladogram showing relationships among 10 dinosaur taxa (Fig. 3 in the present manuscript):

- D1. List all taxa (might be one or more) that did not evolve from an ancestor that had a reduced number of finger bones.
- D2. Which taxon—*Plateosaurus* or *Diplodocus*—is the closest evolutionary relation to *Sellosaurus*? \_\_\_\_\_  
Explain your answer:
- D3. Given that birds are warm-blooded, which taxon is most likely to share this character?  
\_\_\_\_\_  
Explain your answer:

## References

- American Museum of Natural History (AMNH). (2002). *Assembling the Tree of Life: Harnessing Life's History to Benefit Science and Society*. Brochure produced for the National Science Foundation based on three NSF Tree of Life workshops held in 1999 and 2000 at Yale University, the University of California Davis, and the University of Texas Austin. [http://www.phylo.org/sub\\_sections/outreach/outreach\\_b.php](http://www.phylo.org/sub_sections/outreach/outreach_b.php). Accessed 30 Aug 2010.
- Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree thinking challenge. *Science*, 310, 979–980.
- Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook I: The cognitive domain*. New York: David McKay.
- Campbell, A., Otrell-Cass, K. (2011). Teaching evolution in New Zealand's schools—Reviewing changes in the New Zealand science curriculum. *Research in Science Education*, 41(3), 441–451.
- Catley, K. M. (2006). Darwin's missing link: A new paradigm for evolution education. *Science Education*, 90(5), 767–783.
- Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience*, 58(10), 976–987.
- Catley, K., 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.
- Catley, K. M., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. *Journal of Research in Science Teaching*, 47(7), 861–882.
- 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*. Oxford: Oxford University Press.
- Darwin, C. (1859). *On the origin of species by means of natural selection* (1st ed.). London: John Murray.
- 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.
- 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.
- Hennig, W. (1966). *Phylogenetic systematics*. Urbana: 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: Oxford University Press.
- Lee, M. S. Y., Reeder, T. W., Slowinski, J. B., & Lawson, R. (2004). Resolving reptile relationships: Molecular and morphological Markers. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 451–467). New York, NY: Oxford 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.
- Ministry of Education. (2007). *The New Zealand Curriculum for English-medium teaching and learning in years 1–13*. Wellington: Learning Media.

- 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.
- National Research Council. (1996). *National Science Education Standards*. Washington: National Academies Press.
- National Research Council. (2012). *A framework for K-12 Science education: Practices, crosscutting concepts, and core ideas*. Washington: National Academies Press.
- Nickels, M. K., & Nelson, C. E. (2005). Beware of nuts & bolts: Putting evolution into the teaching of biological classification. *The American Biology Teacher*, 67, 283–289.
- North Carolina Standard Course of Study. (2011). <http://www.ncpublicschools.org/curriculum/science/scos/2004/23biology>. Accessed Retrieved 18 Jan 2011
- 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. (2013) Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *American Educational Research Journal*, 50(1), 138–177.
- Novick, L. R., Catley, K. M., & Funk, D. J. (2010a). Characters are key: The effect of synapomorphies on cladogram comprehension. *Evolution: Education and Outreach*, 3(4), 539–547.
- Novick, L. R., Catley, K. M., & Schreiber, E. G. (2010b). *Understanding cladograms: An introduction to tree thinking*. Nashville: Department of Psychology and Human Development, Vanderbilt University. Unpublished instructional booklet.
- Novick, L. R., Catley, K. M., & Funk, D. J. (2011a). Inference is bliss: Using evolutionary relationship to guide inferences about biological categories. *Cognitive Science*, 35(2), 712–443.
- Novick, L. R., Shade, C. K., & Catley, K. M. (2011b). Linear versus branching depictions of evolutionary history: Implications for design. *Topics in Cognitive Science*, 3, 536–559.
- O'Hara, R. J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Systematic Zoology*, 37, 142–155.
- 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.
- Sandvik, H., (2008). Tree thinking cannot taken for granted: Challenges for teaching phylogenetics. *Theory in Biosciences*, 127, 45–51. <http://www.springerlink.com/content/eu62420p381402xr/>. Accessed 19 Mar 2008.
- Thanukos, A. (2009). A name by any other tree. *Evolution: Education & Outreach*, 2, 303–309.
- 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.