

## Article

# Student Interpretations of Phylogenetic Trees in an Introductory Biology Course

Jonathan Dees,\* Jennifer L. Momsen,\* Jarad Niemi,<sup>†</sup> and Lisa Montplaisir\*

\*Department of Biological Sciences, North Dakota State University, Fargo, ND 58108; <sup>†</sup>Department of Statistics, Iowa State University, Ames, IA 50011

Submitted January 17, 2014; Revised August 20, 2014; Accepted September 5, 2014  
Monitoring Editor: Kenneth Miller

Phylogenetic trees are widely used visual representations in the biological sciences and the most important visual representations in evolutionary biology. Therefore, phylogenetic trees have also become an important component of biology education. We sought to characterize reasoning used by introductory biology students in interpreting taxa relatedness on phylogenetic trees, to measure the prevalence of correct taxa-relatedness interpretations, and to determine how student reasoning and correctness change in response to instruction and over time. Counting synapomorphies and nodes between taxa were the most common forms of incorrect reasoning, which presents a pedagogical dilemma concerning labeled synapomorphies on phylogenetic trees. Students also independently generated an alternative form of correct reasoning using monophyletic groups, the use of which decreased in popularity over time. Approximately half of all students were able to correctly interpret taxa relatedness on phylogenetic trees, and many memorized correct reasoning without understanding its application. Broad initial instruction that allowed students to generate inferences on their own contributed very little to phylogenetic tree understanding, while targeted instruction on evolutionary relationships improved understanding to some extent. Phylogenetic trees, which can directly affect student understanding of evolution, appear to offer introductory biology instructors a formidable pedagogical challenge.

## INTRODUCTION

The first branching diagrams for illustrating evolution are often attributed to Jean-Baptiste Lamarck (Lamarck, 1809, 1815; Gould, 1999). Charles Darwin later drafted his first branching diagram in a research notebook (1837) before publishing a phylogenetic tree as the only illustration in *On the Origin of Species by Means of Natural Selection* (1859) to describe descent with modification, or what is now known as evolution. Phylogenetic trees have since become increasingly essential in virtually all disciplines of biology (Baum and Offner, 2008;

Omland *et al.*, 2008) and now function as the most important tool for evaluating evidence of evolution (Baum *et al.*, 2005). Phylogenetic trees are so prevalent in the biological sciences that “tree thinking” has been coined as a term to describe the ability to conceptualize evolutionary relationships among taxa (Meisel, 2010). Consequently, learning to interpret phylogenetic trees has also become an essential component of biology education. The American Association for the Advancement of Science (AAAS) formalized this idea in *Vision and Change in Undergraduate Biology Education: A Call to Action* (AAAS, 2011) by recommending that communication of scientific concepts through visual representations be standard in undergraduate biology education.

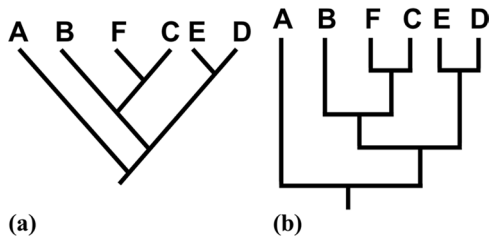
Phylogenetic trees are important tools for organizing knowledge of biological diversity, and they communicate hypothesized evolutionary relationships among nested groups of taxa (monophyletic groups) that are supported by shared traits known as *synapomorphies* (Novick and Catley, 2007). As visualizations, phylogenetic trees are a type of schematic diagram that illustrates abstract concepts rather than appearances of objects (iconic diagrams) or quantitative relationships (charts and graphs; Hegarty *et al.*, 1991; Novick

DOI: 10.1187/cbe.14-01-0003

Address correspondence to: Jennifer L. Momsen (jennifer.momsen@ndsu.edu).

© 2014 J. Dees *et al.* CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Non-commercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

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



**Figure 1.** The two most common phylogenetic tree styles with equivalent branching patterns and taxa relatedness: (a) diagonal and (b) bracket style (adapted from Gregory, 2008).

and Catley, 2007; Lee, 2010). Because of this abstract nature, schematic diagrams are used to describe processes that are difficult to observe, such as evolution, and are governed by learned conventions for interpretation (Novick and Catley, 2007). Thus, it should not be surprising that these diagrams present a learning challenge in introductory biology. Interpreting phylogenetic trees requires learning conventions, overcoming prior and often naïve knowledge of taxa, and interpreting evolutionary relationships based solely on branching patterns depicted in the diagrams (Gregory, 2008; Halverson *et al.*, 2011).

**Phylogenetic Tree Misinterpretations**

Traditionally, most recent common ancestry is used to interpret taxa relatedness. Taxa that share a more recent common ancestor must be more closely related to each other than to another taxon with a less recent common ancestor. For example, taxon F in Figure 1 is more closely related to taxon C than to taxon B, because taxon F and taxon C share a more recent common ancestor. An alternative method for interpreting taxa relatedness is using monophyletic groups. Taxon F and taxon C in Figure 1 belong to a monophyletic group that does not include taxon B, which again indicates taxon F and taxon C are more closely related to each other than to taxon B.

Misinterpreting taxa relatedness, is, however, quite common (Table 1 and references therein). The most common misinterpretation related to taxa relatedness is using distance between taxa on phylogenetic trees to determine relatedness, that is, branch tip proximity in Table 1, often referred to as “reading the tips.” Branches can be rotated about the nodes, however, such that taxa positions are arbitrary. Using Figure 1 as an example, taxon C and taxon E are placed next to each other and could be misinterpreted as closely related, even though most recent common ancestry and monophyletic groups indicate the taxa are distantly related.

In addition, extant taxa are sometimes misinterpreted as descended from other extant taxa, and this general error can also affect taxa-relatedness interpretations (contemporary descent in Table 1). Using Figure 1a as an example, one might suggest taxon F is closely related to taxon C (which is true), because taxon F is descended from taxon C. Taxon F and taxon C are not closely related because one extant taxon is descended from the other, but rather due to descent from a recent common ancestor. Counting nodes between taxa on phylogenetic trees to establish relatedness is another misinterpretation documented in the literature. As an example, four nodes separate taxon C from taxon E in Figure 1, while

**Table 1.** Phylogenetic tree interpretations for taxa relatedness described in the literature

Interpretation	Brief description and references
<i>Most recent common ancestry</i>	<i>Taxa sharing a more recent common ancestor are more closely related to each other than to another taxon.</i> <sup>a-c, e, f, h, i, l, m, o, q, r</sup>
<i>Monophyletic grouping</i>	<i>Taxa in the same monophyletic group are more closely related to each other than to a taxon outside the monophyletic group.</i> <sup>d-f, l</sup>
Branch tip proximity	Taxa relatedness is determined by perceived distance between taxa on phylogenetic trees. <sup>a, b, d-f, g, h, l-n, p, q</sup>
Contemporary descent	Taxa relatedness is determined by indicating a taxon is descended from another extant taxon. <sup>a, b, d-f, g, h, j, l, o, p</sup>
Counting nodes	Taxa relatedness is determined by counting nodes or branches between taxa on phylogenetic trees. <sup>d-f, g, h, m, p</sup>
External insights	Taxa relatedness is determined by knowledge that is not provided by phylogenetic trees. <sup>a, b, d-f, j, k, m, r</sup>

Italics indicate correct phylogenetic tree interpretations for taxa relatedness.

<sup>a</sup>Baum *et al.*, 2005. <sup>b</sup>Baum and Offner, 2008. <sup>c</sup>College Board, 2012. <sup>d</sup>Gregory, 2008. <sup>e</sup>Halverson, 2011. <sup>f</sup>Halverson *et al.*, 2011. <sup>g</sup>Meir *et al.*, 2007. <sup>h</sup>Meisel, 2010. <sup>i</sup>Morabito *et al.*, 2010. <sup>j</sup>Novick and Catley, 2007. <sup>k</sup>Novick *et al.*, 2010. <sup>l</sup>Novick *et al.*, 2011. <sup>m</sup>Novick and Catley, 2013. <sup>n</sup>Novick *et al.*, 2012. <sup>o</sup>Omland *et al.*, 2008. <sup>p</sup>Perry *et al.*, 2008. <sup>q</sup>Sandvik, 2008. <sup>r</sup>Smith *et al.*, 2013.

three nodes separate taxon B from taxon E. A smaller number of nodes could be misinterpreted as a closer evolutionary relationship between taxon B and taxon E, yet most recent common ancestry and monophyletic groups indicate taxon B and taxon C are equally related to taxon E. Finally, prior studies uncovered misinterpretations in which external (and usually naïve) knowledge of taxa is applied to determine relatedness. For example, one might suggest whales (mammals) are closely related to sharks (cartilaginous fish) based on similar aquatic environments and traits, but these taxa are very distantly related.

Misinterpreting phylogenetic trees is not restricted to errors related to taxa relatedness; phylogenetic tree misinterpretations are diverse and pose significant barriers to understanding evolution (Meir *et al.*, 2007). For example, evolution is often viewed as progressive and directional. Such a viewpoint may be reinforced by phylogenetic trees: extant taxa diverging from a main branch on a phylogenetic tree are often ranked from primitive to advanced, with humans or mammals generally designated as the most advanced taxon (Baum *et al.*, 2005; Gregory, 2008; Omland *et al.*, 2008; Sandvik, 2008, 2009; Meisel, 2010; Halverson, 2011; Halverson *et al.*, 2011). This ranking occurs despite a lack of biological justification for either ranking extant taxa or for assuming humans are the main goal of evolution (Dawkins, 2009). In addition, there is a tendency for learners to confuse taxon and lineage ages when using phylogenetic trees (Gregory, 2008; Omland *et al.*, 2008), and the relative flow of time from the root to the terminal nodes of extant taxa is frequently misinterpreted (Meir *et al.*, 2007; Gregory, 2008; Omland *et al.*,

2008; Perry *et al.*, 2008). Misreading time as flowing horizontal on vertical phylogenetic trees (Figure 1, a and b) can lead to fundamental errors, such as concluding that extant taxa on the right side of the diagram evolved from extant taxa on the left side, rather than from a common ancestor. A final prevalent error is assuming evolution occurs only at nodes, that is, assuming straight lines on phylogenetic trees imply no changes from ancestral states (Baum *et al.*, 2005; Meir *et al.*, 2007; Gregory, 2008; Perry *et al.*, 2008; Meisel, 2010).

### Research Contributions

Building on previous phylogenetic tree interpretation research, the present study aims to answer the following questions: 1) What forms of reasoning are used by introductory biology students to interpret taxa relatedness on phylogenetic trees? 2) What percentage of introductory biology students correctly interpret taxa relatedness on phylogenetic trees? 3) How do results from the first two research questions change over time in an introductory biology course?

Many of the previous investigations on student interactions with phylogenetic trees collected data using onetime, ungraded questionnaires (for exceptions, see Halverson, 2011; Halverson *et al.*, 2011; Eddy *et al.*, 2013). Although such data are useful, students may not treat questionnaires that will not affect their academic standing as seriously or thoughtfully as assessments that contribute to final course grades (Sundberg, 2002). Onetime questionnaires are also limited to examining student understanding at a single point in time. The present study sought to capture the knowledge and reasoning of students in an authentic classroom environment. Data were collected in situ from homework and exams in which students received substantial points toward their final grades in an introductory biology course. This investigation further provided a unique opportunity to acquire data from isomorphic questions over the course of a semester. Such data can be used to examine learning progress as a result of specific instruction and feedback, as well as consistency of student phylogenetic tree understanding over time.

## METHODS

This investigation was conducted during the second course of a two-course introductory biology series for science majors at a large, public university with very high research activity (Carnegie Foundation, 2013) in the midwestern United

**Table 2.** Course enrollment by major group

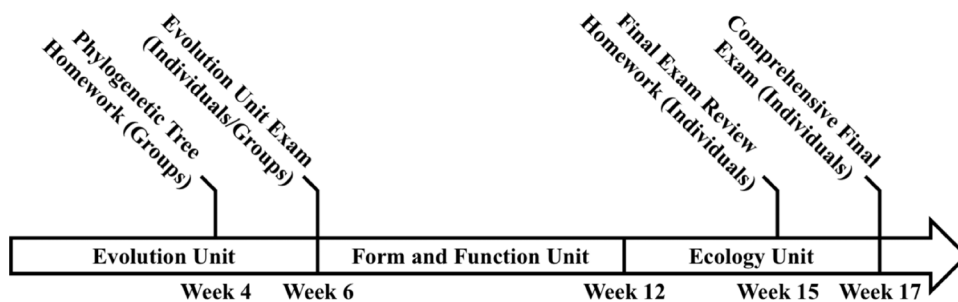
Major group	Students (%)
Agricultural sciences	19 (22)
Biological sciences	32 (36)
Natural resources	16 (18)
Pre-professional health	10 (11)
Undeclared and other	11 (13)

States. The large-enrollment course ( $n = 88$ ) served students pursuing a number of majors (Table 2) at various stages in their academic careers (24% freshmen, 33% sophomores, 18% juniors, and 25% seniors). The first course in the introductory series focused on cell biology and included little or no exposure to phylogenetic trees. Although recommended, completion of the first course was not a prerequisite for the second course.

### Course Context

The instructor used a learner-centered approach to teaching biology, in which multiple forms of active engagement were used in place of passive lectures. Course activities included letter card questions (Freeman *et al.*, 2007), collaborative learning groups (Smith, 2000; Tanner *et al.*, 2003), small-group and whole-class discussions, think-pair-share sessions (Lyman, 1981), and case studies (Herreid, 1994). Model-based instruction (Hestenes, 1987; Hmelo *et al.*, 2000; Brewster, 2008; Liu and Hmelo-Silver, 2009) was a prominent pedagogical strategy, as students frequently constructed box-and-arrow models of complex biological processes, such as evolution, nutrient cycles, and energy flow through ecosystems. Students worked in permanent, self-selected groups of three or four individuals on nearly all aspects of the course, including pyramid exams (Eaton, 2009) with individual and group components (75 and 25% of points, respectively). Learning objectives, instruction, and assessments largely targeted higher-order cognitive skills of analysis, synthesis, and evaluation (Bloom *et al.*, 1956; Crowe *et al.*, 2008; Momsen *et al.*, 2010, 2013).

The introductory biology course included three primary units: evolution, form and function, and ecology (Figure 2). Although most prominent during the evolution unit, phylogenetic trees were used throughout the course when appropriate. For example, phylogenetic trees appeared in the form and function unit to help students visualize and reason about evolved traits required for plant survival on land.



**Figure 2.** Timeline of primary course units and data collection from assessments.

### Instruction and Data Collection

Two homework assignments and two exams were the data sources for this study (Figure 2). The initial phylogenetic tree homework was completed in groups soon after phylogenetic trees were introduced as part of the evolution unit. The introduction consisted of a series of questions posed by the instructor and answered by students using letter cards. The questions familiarized students with structural characteristics of phylogenetic trees, such as nodes (represent common ancestors) and monophyletic groups, and presented the idea that taxa relatedness is determined by common ancestry. Letter card questions were followed by small-group and whole-class discussions until the entire class established the correct answer using appropriate reasoning. All phylogenetic tree questions used during class and for assessments referred to cladograms, in which only branching patterns have meaning. Chronograms (which show absolute time) and phylograms (which show amount of change) were briefly mentioned by the instructor, but students were never required to interact with or reason from them during the course (for further descriptions of phylogenetic tree types, see Baum and Offner, 2008; Omland *et al.*, 2008).

The initial phylogenetic tree homework featured a short series of open-ended questions designed around a phylogenetic tree of chordates. In addition to prompts about recent common ancestors, synapomorphies, and monophyletic groups, one question regarding taxa relatedness appeared on the group homework (Figure 3). Poor group performance for this question compelled the instructor to revisit phylogenetic tree interpretations during class. The question was presented to students again and debated through directed, small-group discussions. A subsequent whole-class discussion acknowledged most recent common ancestry as an appropriate reasoning strategy for determining taxa relatedness on phylogenetic trees. After the initial homework was revisited during class, taxa relatedness was specifically targeted through two additional letter card questions. Instruction specific to phylogenetic trees and evolutionary relatedness occurred across three consecutive course meetings, ending in week 5. We therefore include each student's

average attendance across these 3 d in subsequent analysis as a reflection of the potential impact of instruction on student reasoning with phylogenetic trees.

Phylogenetic trees and taxa-relatedness questions similar to the initial homework were placed on three subsequent assessments, which followed the end of instruction by 1, 10, and 12 wk, respectively (Figure 2). Such prompts were included on both the individual and group components of the evolution unit exam in which students completed the individual component before the group component (Supplemental Figures S1 and S2). A phylogenetic tree was provided for the individual component, but the group component required students to construct a phylogenetic tree from data before answering a taxa-relatedness question. Students were never asked to construct phylogenetic trees before completing the evolution unit exam. A phylogenetic tree and taxa-relatedness questions were also placed on the review homework 2 wk before the final exam (Figure S3) and on the individual component of the final exam (Figure S4). The prompt structure for the review homework and final exam was changed slightly from a two-choice prompt with open-ended reasoning to a four-choice prompt with open-ended reasoning. This alteration was made for several reasons. First, students had seen several taxa-relatedness questions throughout the semester; to avoid retest concerns, we created prompts that were familiar to students but offered a somewhat new opportunity to interpret relatedness. Second, the multiple-choice foils prevented students from feeling obligated to select one taxon or the other, providing students with the option to identify taxa as equally related or unrelated. In both the review homework and final exam, the taxa involved were equally related. The phylogenetic tree on the final exam was also the only phylogenetic tree used as part of this investigation that did not include labeled synapomorphies.

### Rubric Development and Coding

The initial rubric for coding student responses to taxa-relatedness questions was developed using a grounded theory

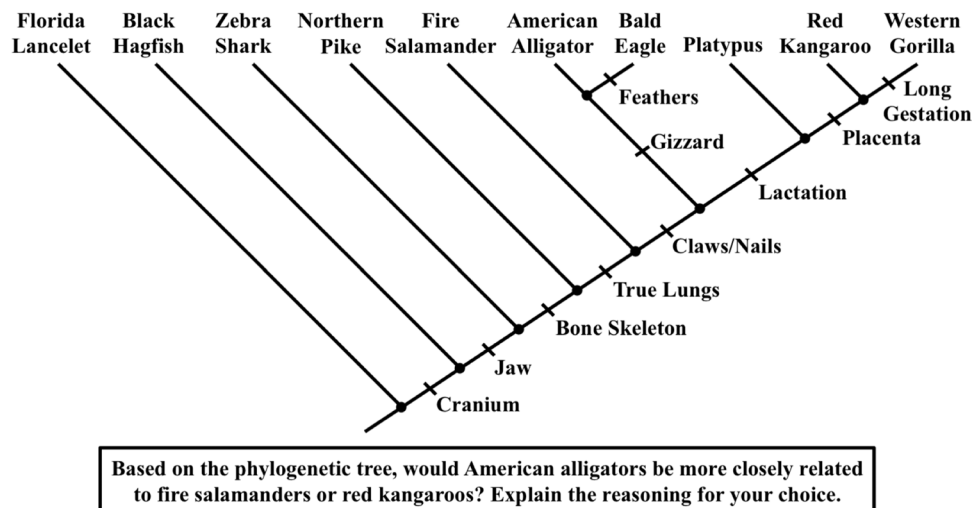


Figure 3. Phylogenetic tree and taxa-relatedness question from the initial homework.

approach (Glaser and Strauss, 1967). This reflected the nature of the project as developing in real time in response to classroom experiences and student learning difficulties.

Existing literature on phylogenetic tree interpretations (Table 1) was then used to confirm and refine some categories for the final rubric (Supplemental Material) and to identify two new reasoning strategies. Specifically, we found evidence that students determine relatedness by counting synapomorphies (taxa relatedness is determined by counting synapomorphies between the taxa on phylogenetic trees) and by using negation reasoning (reasoning includes descriptions of how not to interpret taxa relatedness on phylogenetic trees; in all cases, this reasoning occurs concurrently with other reasoning; see the Supplemental Material). In addition, we found evidence of students using monophyletic grouping (taxa in the same monophyletic group are more closely related to each other than to a taxon outside the monophyletic group) to reason about relatedness. While some research has identified monophyletic grouping as a possible reasoning approach, no one has provided evidence to show that students actually use monophyletic grouping.

For training the raters, all responses from the initial homework and both components of the evolution unit exam were numbered, and a random number generator was used to select 20 initial responses (15% of the total at the time). Two independent raters coded the initial responses and reached consensus through discussion. Following rubric calibration, agreement between the two raters was 94% for the remaining 258 responses from all four assessments, and disagreements were resolved through discussion. Student responses often included more than one form of reasoning and consequently fell into multiple rubric categories, resulting in 360 total reasoning codes assigned to 278 group and individual responses. Coding was partially blind, in which one rater was aware of group and individual identities while the second rater was not. Due to high agreement between independent raters, we do not believe rater bias was a significant issue for this investigation.

The taxa-relatedness questions used throughout the course required students to choose an answer and provide reasoning for their selection. Because answers selected by students were not always consistent with their reasoning, responses were coded again for answer (correct or incorrect) and reasoning used to support the answer (correct, incorrect, or mixed, *i.e.*, a mix of correct and incorrect reasoning). The categories of most recent common ancestry and monophyletic grouping were considered correct reasoning, while negation reasoning always appeared with other forms of reasoning and was considered neither correct nor incorrect. All other rubric categories were deemed incorrect reasoning for taxa relatedness. This coding procedure identified students who guessed correct answers (correct answer with incorrect reasoning), and students who memorized correct reasoning without understanding its application (incorrect answer with correct reasoning). Only responses with both correct answers and correct reasoning demonstrated understanding of taxa relatedness on phylogenetic trees.

### Statistical Analyses

Following the suggestion of Theobald and Freeman (2014), we constructed statistical models to test various hypotheses

regarding student reasoning about phylogenetic trees. To assess hypotheses related to reasoning and answer selection, we constructed statistical models that accounted for variables affecting reasoning and answer selection. In addition, random effects were used to capture repeated measurements on the same groups and individuals on multiple assessments. Specifically, mixed-effect ordinal logistic-regression models were used to analyze taxa-relatedness reasoning, while mixed-effect logistic-regression models were used to analyze correct answers. For group reasoning, group assignment was modeled as a random effect, and assessment was a fixed effect. For individual reasoning, student was modeled as a random effect, while assessment, class attendance, year in school, and academic major were fixed effects. For group correctness, group assignment was modeled as a random effect, and assessment and reasoning (correct, incorrect, or mixed) were fixed effects. For individual correctness, student was modeled as a random effect, while reasoning, assessment, class attendance, year in school, and academic major were fixed effects. *F*-tests were used to determine significance of batches of explanatory variables (*e.g.*, major), while *t* tests were used to determine significance of individual explanatory variables. Additional details of the statistical analyses (*e.g.*, odds ratios) are available in the Supplemental Material.

## RESULTS

To assess student understanding of phylogenetic trees, we performed four separate analyses. Referencing the analyses by their response variable, the analyses are group reasoning, individual reasoning, group correctness, and individual correctness. “Group” indicates the responses are collectively from a group of students, while “individual” indicates responses from individual students. “Reasoning” indicates the response is the ordinal trichotomized reasoning variable: correct, mixed, and incorrect. “Correctness” indicates the response is whether the binary answer was correct or not. The following sections report both summary statistics for reasoning (Table 3) and correctness (Table 4) along with model-based analyses.

### Group Reasoning

Table 3 shows that group performance was poor on the initial homework: only two groups applied correct reasoning, while 12 groups used the incorrect reasoning of counting synapomorphies and six used the incorrect reasoning of counting nodes. On the subsequent assignment, 17 groups used most recent common ancestry and seven groups used monophyletic grouping. Counting synapomorphies was still prominent with six groups. Two groups used counting nodes.

For testing statistical significance of improved reasoning from the initial homework assessment to the evolution unit exam assessment, a mixed-effect ordinal logistic-regression model was used with the ordinal trichotomized reasoning as the response, the group as a random effect, and assessment as a fixed effect (see “Group Reasoning” in the Supplemental Material). A significant improvement in reasoning was seen from the initial homework to the evolution unit exam ( $t(21) = 4.51, p = 0.0002$ ).

**Table 3.** Reasoning used by students to determine taxa relatedness from all four assessments

Category	Group assessment <sup>a</sup>		Individual assessment		
	Initial homework (n = 24)	Evolution unit exam (n = 23)	Evolution unit exam (n = 88)	Review homework (n = 66)	Final exam (n = 77)
<i>Most recent common ancestry</i>	1 (4)	17 (74)	51 (58)	40 (61)	54 (70)
<i>Monophyletic grouping</i>	1 (4)	7 (30)	19 (22)	5 (8)	7 (9)
Counting nodes	6 (25)	2 (9)	19 (22)	12 (18)	14 (18)
Counting synapomorphies	12 (50)	6 (26)	11 (13)	6 (9)	2 (3)
Branch tip proximity	5 (21)	1 (4)	1 (1)	1 (2)	7 (9)
Contemporary descent	0 (0)	1 (4)	2 (2)	1 (2)	4 (5)
External insights	4 (17)	0 (0)	1 (1)	1 (2)	3 (4)
Negation reasoning <sup>b</sup>	0 (0)	0 (0)	5 (6)	3 (5)	3 (4)
Other responses	1 (4)	1 (4)	6 (7)	9 (14)	8 (10)

Values are the number of responses that received a particular code with percentage of responses in parentheses. Note: total number of codes is greater than the total number of responses and the percentages will sum to more than 100% because responses often fell into multiple categories. Italics indicate correct forms of reasoning.

<sup>a</sup>Responses were submitted by permanent groups of three or four students.

<sup>b</sup>Negation reasoning is neither correct nor incorrect.

### Individual Reasoning

Apart from a decrease in monophyletic grouping, individual reasoning largely persisted from the evolution unit exam through the review homework 9 wk later (Table 3). Individuals were less likely than groups to use correct reasoning on the evolution unit exam, and counting nodes was more common than counting synapomorphies among individuals. Reasoning varied somewhat between the review homework and final exam. Most categories increased as counting synapomorphies decreased to only two responses on the final exam, in which the phylogenetic tree did not include synapomorphies.

A mixed-effect ordinal logistic-regression model was used with the ordinal trichotomized reasoning as the response; the student as a random effect; and attendance, assessment, year in school, and major as fixed effects (see "Individual Reasoning" in the Supplemental Material). Reasoning was significantly related to attendance ( $t(140) = 2.23, p = 0.03$ ) but not significantly related to assessment ( $F(2140) = 0.96,$

$p = 0.38$ ), year in school ( $F(3140) = 0.74, p = 0.53$ ), and major ( $F(5140) = 0.77, p = 0.45$ ).

### Group Correctness

Interpreting taxa relatedness on phylogenetic trees requires both knowledge of correct reasoning, as indicated by student reasoning descriptions (Table 3), and application of correct reasoning, as indicated by selecting correct answers for taxa-relatedness questions. Coding for answer (correct or incorrect) and reasoning used to support the answer (correct, incorrect, or mixed) revealed six different combinations of knowledge and application (Table 4). Two groups selected the correct answer on the initial phylogenetic tree homework, and two other groups offered correct or mixed reasoning, but not a single group provided a correct answer coupled with correct reasoning. Responses from the group component of the evolution unit exam were greatly improved, although only 57% provided correct answers coupled with

**Table 4.** Answers and reasoning used by students to support answers from all four assessments

Category	Initial homework (n = 24) <sup>a</sup>	Evolution unit exam (n = 23) <sup>a</sup>	Evolution unit exam (n = 88)	Review homework (n = 66)	Final exam (n = 77)
Correct answer, correct reasoning	0 (0)	13 (57)	51 (58)	31 (47)	29 (38)
Correct answer, mixed reasoning	0 (0)	3 (13)	14 (16)	1 (1)	3 (4)
Correct answer, incorrect reasoning	2 (8)	0 (0)	19 (22)	3 (5)	5 (6)
Incorrect answer, correct reasoning	1 (4)	1 (4)	0 (0)	6 (9)	16 (21)
Incorrect answer, mixed reasoning	1 (4)	3 (13)	1 (1)	4 (6)	8 (10)
Incorrect answer, incorrect reasoning	20 (84)	3 (13)	3 (3)	21 (32)	16 (21)

Values are number of responses that received a particular code with percentage of responses in parentheses.

<sup>a</sup>Responses were submitted by permanent groups of three or four students.

correct reasoning. Unlike other taxa-relatedness questions, groups were required to build a phylogenetic tree from data for the group component of the evolution unit exam (Figure S2), and all but one group constructed a phylogenetic tree that was sufficient to correctly answer the question (i.e., contained accurate evolutionary relationships according to provided data).

For group correctness, a logistic-regression model was used with answer correctness as the response, group assignment as a random effect, and assessment and reasoning (correct, incorrect, or mixed) as fixed effects (see “Group Correctness” in the Supplemental Material). We tested the hypothesis that improved reasoning leads to improved answer selection, after controlling for assessment, and found marginal significance ( $F(2,20) = 3.09, p = 0.07$ ). Looking at specific differences within this general hypothesis, we found a significant improvement in answer selection between groups who had correct reasoning versus those with incorrect reasoning ( $t(20) = 2.28, p = 0.03$ ), marginally significant improvement between correct versus mixed reasoning ( $t(20) = 1.80, p = 0.09$ ), and an insignificant difference between mixed versus incorrect reasoning ( $t(20) = 0.64, p = 0.53$ ).

### Individual Correctness

Taxa-relatedness questions completed by individuals had lower rates of correct answers coupled with correct reasoning compared with the prior group component of the evolution unit exam (Table 4), excluding the individual component of the evolution unit exam due to poor question structure (see *Discussion*). After answering similar taxa-relatedness questions in class and on assessments (including a review homework 2 wk earlier), only 38% of students provided both a correct answer and correct reasoning for the final exam taxa-relatedness question. An additional 31% of students chose an incorrect answer despite offering correct or mixed forms of reasoning.

For individual correctness, a logistic-regression model was used with answer correctness as the response; student as a random effect; and reasoning, assessment, class attendance, year in school, and academic major as fixed effects (see “Individual Correctness” in the Supplemental Material). We tested the hypothesis that improved reasoning leads to improved answer selection, after controlling for assessment, class, major, and attendance, and found significant differences ( $F(2139) = 13.93, p < 0.0001$ ). Looking at specific comparisons within this general hypothesis, we found a significant improvement in answer selection between individuals who had correct reasoning versus those who had incorrect reasoning ( $t(139) = 5.12, p < 0.0001$ ), significant improvement between correct versus mixed reasoning ( $t(139) = 2.52, p = 0.01$ ), but an insignificant difference between mixed versus incorrect reasoning ( $t(139) = 1.29, p = 0.20$ ). We also found a marginally significant improvement for those who attended class ( $t(139) = 1.78, p = 0.08$ ), but no significant effect of year in school ( $F(3139) = 0.60, p = 0.61$ ) or major ( $F(5139) = 1.22, p = 0.30$ ).

## DISCUSSION

Phylogenetic trees are an essential component of undergraduate biology education that remain difficult for students to

interpret. Our in situ research documents common reasoning patterns used by students in an introductory biology course. Counting synapomorphies and nodes between taxa on phylogenetic trees were the most common forms of incorrect reasoning for determining taxa relatedness. Students independently generated an alternative form of correct reasoning using monophyletic groups, but the popularity of this approach decreased over time. Further, after multiple learning opportunities, including broad instruction on phylogenetic trees, targeted instruction for evolutionary relationships, textbook readings, and homework, slightly more than half of groups and less than half of individuals provided correct answers coupled with correct reasoning for taxa-relatedness questions. Many students appeared to have memorized correct reasoning without understanding its application, and of the variables we measured, attendance was the only predictor of student performance on taxa-relatedness questions. This investigation has important implications for instruction and research on student interpretations of phylogenetic trees.

### Individual Component of the Evolution Unit Exam

As previously mentioned, results from coding student responses for answer (correct or incorrect) and reasoning (correct, incorrect, or mixed) are problematic for the individual component of the evolution unit exam due to flawed question structure (Figure S1). According to the phylogenetic tree and using most recent common ancestry or monophyletic grouping reasoning, bears are more closely related to sea lions than cats. Some incorrect reasoning, such as branch tip proximity and external insights (rare among our students), led to incorrectly choosing cats instead of sea lions. However, incorrect strategies of counting synapomorphies and nodes (most common among our students) led to correctly choosing sea lions. Thus, nearly all students (96%) selected the correct answer regardless of reasoning (Table 4), which reflects the flawed question structure rather than student understanding. Although responses from the individual component of the evolution unit exam are unreliable for determining student understanding of taxa relatedness, we included the results for two important reasons. First, student reasoning alone, regardless of its application, provided valuable insights into how students approach phylogenetic trees and aided development of the taxa-relatedness reasoning rubric (Supplemental Material). Second, the flawed prompt is an informative example of what not to do when assessing student understanding of phylogenetic trees.

### Student Reasoning

Taxa relatedness is understood by biologists in terms of most recent common ancestry, similar to family trees of humans (Baum *et al.*, 2005). Following the initial phylogenetic tree homework in which all groups struggled, a majority of students were aware that most recent common ancestry determines taxa relatedness (Table 3), although far fewer students applied the reasoning correctly (Table 4). Use of the alternative correct reasoning, monophyletic grouping, was a novel outcome for this study. Monophyletic groups were discussed at length during the course and in relation to phylogenetic trees, but neither the instructor nor the textbook (Freeman, 2011) directly suggested using monophyletic

groups to determine taxa relatedness. Our students generated this alternative reasoning on their own, either spontaneously or from outside materials. Over time, however, students used this reasoning less frequently, perhaps in response to direct feedback on the unit exam that highlighted most recent common ancestry reasoning.

While branch tip proximity, contemporary descent, and external insights are the most common forms of incorrect reasoning cited in the literature (Table 1), these forms of reasoning were rather uncommon in responses from our students. Counting synapomorphies, and counting nodes were by far the most common forms of incorrect reasoning used by our students to determine taxa relatedness (Table 3). Determining taxa relatedness by counting synapomorphies has not previously been described in the literature to our knowledge but proved to be a persistent approach. Two students even applied this reasoning on the final exam, in which the phylogenetic tree did not include synapomorphies (Figure S4). Both students suggested that seals are equally related to horses and whales (which is correct), because there are no trait differences between the three taxa. Such responses are illogical and demonstrate the persistence of incorrect reasoning strategies.

The existence and frequency of synapomorphy counting among students presents a pedagogical dilemma. A previous investigation concluded that labeled synapomorphies on phylogenetic trees encourage comprehension of evolutionary relationships (Novick *et al.*, 2010). Investigators used translation exercises between two common phylogenetic tree styles (Figure 1), and students were significantly more accurate when synapomorphies were present. The researchers suggested that labeled synapomorphies improve translation performance due to a combination of cognitive psychology and biological understanding. Phylogenetic trees are constructed from nested groups of taxa, and from a cognitive perspective, synapomorphies help identify points along continuous lines where hierarchical levels begin. From a biological viewpoint, synapomorphies help identify common ancestors and monophyletic groups, which are maintained during translation from one style of phylogenetic tree to another. Although useful for translating between phylogenetic trees, synapomorphies are problematic for interpreting a single phylogenetic tree, as our students often misused them to determine taxa relatedness. In one case, synapomorphies act as guides, while in another case, synapomorphies act as distractors. This apparent conflict between phylogenetic tree translation and interpretation tasks warrants further investigation.

As cited in the literature and supported by the present study, introductory biology students use many forms of incorrect reasoning when interpreting phylogenetic trees, especially before instruction. Thus, it was not surprising that attendance during targeted, active-learning instruction on evolutionary relatedness was a significant predictor of correct taxa-relatedness reasoning. Interpreting phylogenetic trees is an ability acquired through instruction and practice rather than an intuitive ability (Sandvik, 2008). If formal instruction is the most important factor for understanding phylogenetic trees, it should not be surprising that year in school and major were not correlated with correct taxa-relatedness reasoning. Phylogenetic trees are difficult for introductory biology students to interpret without instruction, regardless of their college experience or interest in biology.

### Student Correctness

Because the purpose of phylogenetic trees is to visually represent evolutionary relationships, taxa-relatedness questions exemplify tree thinking (Novick and Catley, 2013). This investigation used such prompts to measure understanding of phylogenetic trees by combining results from answers and reasoning used to support answers. Responses that provided both correct answers and correct reasoning demonstrated understanding of taxa relatedness on phylogenetic trees. After the initial homework and targeted instruction on evolutionary relationships, and disregarding the individual component of the evolution unit exam (unreliable for correctness coding), approximately half of the students demonstrated such understanding across multiple assessments (Table 4).

With the ability to share interpretations, groups were expected to outperform individuals on taxa-relatedness questions. Although based on only three data sets (excluding the initial phylogenetic tree homework, which was completed before targeted instruction on evolutionary relationships, and the unreliable individual component of the evolution unit exam), results align with expectations for cooperative work (Table 4). However, another explanation is that students performed better on the group component of the evolution unit exam versus the review homework and final exam due to building the phylogenetic tree before answering the taxa-relatedness question (Figure S2). Phylogenetic tree construction could have required students to concentrate on taxa relationships or simply forced students to spend more time on task, and this alternative explanation cannot be ruled out. Because the only two studies examining benefits of phylogenetic tree construction before interpretation disagree with each other (Halverson, 2011; Eddy *et al.*, 2013), determining effects of phylogenetic tree construction on student understanding warrants further investigation.

Some answer and reasoning combinations other than correct answers with correct reasoning offer valuable insights into student understanding of phylogenetic trees. Correct answers coupled with incorrect reasoning indicate students who guessed correctly without understanding phylogenetic trees. Disregarding the individual component of the evolution unit exam, guessing correctly was rare during this investigation (Table 4). On the other hand, incorrect answers with correct reasoning indicate students who memorized taxa-relatedness reasoning but did not understand how to apply the reasoning to phylogenetic trees. This outcome of incorrect answer with correct reasoning was far more common, ranging from 4% earlier in the course to 21% on the final exam. Another 4–13% of responses provided incorrect answers with mixed reasoning, which also indicates some degree of memorization without understanding. Shallow learning strategies are very common in the sciences (Elby, 1999; Pungente and Badger, 2003; Tomanek and Montplaisir, 2004) and can be attributed at least in part to assessment practices (Momsen *et al.*, 2010, 2013) and frequency, type, and student use of feedback (Hattie and Timperley, 2007).

Similar to correct reasoning, attendance was a significant predictor of choosing correct answers for taxa-relatedness questions, while major and year in school were not significant factors. Because attendance has been demonstrated to strongly influence student performance in biology courses in general (Moore *et al.*, 2003; Freeman *et al.*, 2007), it should not be surprising to see the same trend for phylogenetic trees.



However, we caution that attendance is not a direct proxy for the effects of instruction. While attendance may reflect time on task, it may also reflect student motivation. That is, in addition to consistent class attendance, motivated students may also study regularly outside class and have an intrinsic interest in evolution that results in better performance. Further, as expected, reasoning was critical for correct answer selection, as correct reasoning was associated with correct answers when compared with incorrect reasoning for both groups and individuals. No significant difference in correct answer selection was found for groups and individuals using mixed versus incorrect reasoning. Indeed, it seems that mixed reasoning is literally a mixed bag but more often results in students choosing incorrect answers.

### *Changes in Student Reasoning Over Time*

The initial phylogenetic tree homework was completed by groups after phylogenetic trees were introduced using a series of letter card questions. This original exposure clearly did not generate understanding of phylogenetic trees, as only two groups used correct or mixed reasoning for the taxa-relatedness question, and not a single group offered a correct answer with correct reasoning (Table 4). Over time and following additional learning opportunities, including targeted instruction focused more on evolutionary relationships than prior instruction had been, there was significant improvement for the group component of the evolution unit exam (57% correct answers with correct reasoning).

Encouraging results from the additional instruction have implications for teaching and learning about phylogenetic trees. First, interpreting phylogenetic trees is far from intuitive and requires explicit instruction, which agrees with previous studies (Sandvik, 2008; Novick and Catley, 2013). The initial instructional approach of introducing characteristics of phylogenetic trees and allowing students to generate inferences on their own did not produce understanding. However, targeted instruction through active-learning exercises for various aspects of phylogenetic trees, including taxa relatedness, had a sizable impact on student understanding. Second, phylogenetic trees should not be underestimated by introductory biology instructors. Time on task is a critical factor for learning (National Research Council, 2000), and considering the importance of phylogenetic trees for understanding evolution (Gregory, 2008; Omland *et al.*, 2008), sufficient class time must be devoted to these visual representations. Finally, feedback also plays a significant role in learning (Hattie and Timperley, 2007), and targeted feedback combined with iterative instruction seemed to promote student understanding, at least to some extent. The significant learning gain from no groups to 57% of groups providing correct answers with correct reasoning resulted from a single iteration of instruction and feedback, suggesting that additional iterations targeting evolutionary relationships could be beneficial for student understanding.

The taxa-relatedness question on the review homework (Figure S3) proved to be the first valid indicator for individual understanding of phylogenetic trees. Although the individual component of the evolution unit exam is unreliable for measuring student understanding, we can assert that reasoning strategies used by individuals changed little during the 9-wk time lapse between the evolution unit exam and

review homework (Table 3). The only apparent difference between coding distributions was monophyletic grouping reasoning, which decreased from 22 to 8% of responses. Although the review homework used a multiple-choice format, this slight alteration is unlikely to have caused such a difference in use of monophyletic grouping reasoning. Instruction could have guided students toward using most recent common ancestry reasoning, but all formal instruction occurred before the evolution unit exam. The origin and popularity of monophyletic grouping reasoning during the individual and group components of the evolution unit exam (22 and 30% of responses, respectively) remains unknown. However, the decrease in this reasoning could be attributed to feedback from the instructor following the evolution unit exam: student study habits may have included reviewing the posted rubric, which highlighted most recent common ancestry reasoning. As the first valid measurement of individual understanding, the review homework indicated less than half of students (47%) provided a correct answer with correct reasoning for the taxa-relatedness prompt (Table 4). This result indicates poor understanding, especially in light of students having access to class notes, previous taxa-relatedness questions, and other resources for the homework.

Reasoning strategies were not considerably different for the final exam compared with the previous review homework (Table 3), although counting synapomorphies nearly disappeared, as expected, due to a lack of synapomorphies on the phylogenetic tree (Figure S4). After initial instruction, targeted instruction on evolutionary relationships, feedback, and exposure to the same basic taxa-relatedness question on four previous occasions, only 38% of students provided a correct answer with correct reasoning for the taxa-relatedness question on the high-stakes final exam (Table 4). However, 70% of responses referenced most recent common ancestry (Table 3). This outcome is evidence for the extreme difficulties introductory biology students have interpreting phylogenetic trees and for students memorizing reasoning without understanding its application.

### *Conclusions and Limitations*

Phylogenetic tree misinterpretations are obstacles to understanding evolution (Meir *et al.*, 2007), and nothing in biology makes sense without considering evolution (Dobzhansky, 1964). Yet our study and others have repeatedly shown that biology students struggle with interpreting these visual representations. Broad initial instruction that encouraged students to generate inferences on their own did not impact phylogenetic tree understanding. Following a single iteration of targeted instruction on evolutionary relationships, student interpretations significantly improved, but to just more than half of groups and less than half of individuals providing correct answers and reasoning for taxa-relatedness questions. This improvement may be due to instruction but may also reflect students studying outside class. Still, these diagrams, which can directly affect student understanding of evolution (Gregory, 2008; Omland *et al.*, 2008), represent a formidable teaching challenge for introductory biology instructors.

Groups outperformed individuals on taxa-relatedness interpretations, although this result is based on limited data and could have been confounded by phylogenetic tree construction. Because the two known studies regarding benefits

of phylogenetic tree construction before interpretation disagree (Halverson, 2011; Eddy *et al.*, 2013), this idea warrants further investigation. Finally, the present study supports a recommendation of Halverson *et al.* (2011) that multiple-choice assessments are insufficient for capturing student understanding of phylogenetic trees. Responses to our two-part taxa-relatedness questions often contained answers that disagreed with reasoning. Answers alone provided limited information about student understanding, as did reasoning alone. The format of answers combined with supportive reasoning, however, proved to be powerful.

We recognize this research is limited by the nature of the assessment items used to collect data. We used only cladograms (in which branch lengths have no meaning) that were drawn in a diagonal style, and upward from left to right (Figure 1a). It has been argued that student performance improves when bracket phylogenetic trees (Figure 1b), compared with the diagonal style (Novick and Catley, 2007, 2013), are used; it has also been argued that students perform better with diagonal phylogenetic trees that are drawn downward from left to right rather than upward from left to right (Novick *et al.*, 2012). Thus, students might have performed better on our exercises if we had used a different style and orientation of phylogenetic tree. Because the major purpose of phylogenetic trees is to illustrate evolutionary relationships, we used taxa-relatedness interpretations as the primary indicator of phylogenetic tree understanding. However, it could be argued that other skills, such as identifying monophyletic groups and constructing phylogenetic trees from provided data, are equally as important. Finally, in an effort to avoid retesting bias, each phylogenetic tree used for the investigation contained a unique branching pattern (topology) and different taxa (chordates, plants, reptiles, and two phylogenetic trees of mammals) that could have affected interpretations.

Despite these limitations, the present study provides *in situ* data collected in the context of an introductory biology course for science majors. The results provide a powerful depiction of how students understand phylogenetic trees in a classroom setting and demonstrate that students often memorize information in the absence of conceptual understanding, even when instruction specifically discouraged shallow learning strategies. Thus, phylogenetic trees remain a formidable challenge—for both learners and instructors alike.

## ACKNOWLEDGMENTS

This research was conducted in compliance with Institutional Review Board regulations (IRB protocol SM12217) and was funded by the National Science Foundation (DUE-0833268) and a STEM Education Fellowship from North Dakota State University. We are grateful to Julia Bowsher, Warren Christensen, Jeff Boyer, Erika Offerdahl, and Stuart Haring for comments on earlier versions of the manuscript.

## REFERENCES

American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.

Baum DA, Offner S (2008). Phylogenies and tree-thinking. *Am Biol Teach* 70, 222–229.

Baum DA, Smith SD, Donovan SSS (2005). The tree-thinking challenge. *Science* 310, 979–980.

Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*, New York: Longman.

Brewe E (2008). Modeling theory applied: modeling instruction in introductory physics. *Am J Phys* 76, 1155–1160.

Carnegie Foundation (2013). *The Carnegie Classifications of Institutions of Higher Education*. <http://classifications.carnegiefoundation.org> (accessed 28 December 2013).

College Board (2012). *AP Biology: Course and Exam Description*, New York.

Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.

Darwin C (1837). *Notebook B: Transmutation of Species*. <http://darwin-online.org.uk/content/frameset?itemID=CUL-DAR121.&viewtype=side&pageseq=1>.

Darwin C (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London: Murray.

Dawkins R (2009). *The Greatest Show on Earth: Evidence for Evolution*, New York: Free Press.

Dobzhansky T (1964). Biology, molecular and organismic. *Am Zool* 4, 443–452.

Eaton TT (2009). Engaging students and evaluating learning progress using collaborative exams in introductory courses. *J Geosci Educ* 57, 113–120.

Eddy SL, Crowe AJ, Wenderoth MP, Freeman S (2013). How should we teach tree-thinking? An experimental test of two hypotheses. *Evol Educ Outreach* 6, 13.

Elby A (1999). Another reason that physics students learn by rote. *Am J Phys* 67, S52–S57.

Freeman S (2011). *Biological Science*, 4th ed., San Francisco: Benjamin Cummings.

Freeman S, O'Connor E, Parks JW, Cunningham M, Hurley D, Haak D, Dirks C, Wenderoth MP (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* 6, 132–139.

Glaser BG, Strauss AL (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine.

Gould SJ (1999). Branching through a wormhole. *Nat Hist* 108, 24–27, 84–89.

Gregory TR (2008). Understanding evolutionary trees. *Evol Educ Outreach* 1, 121–137.

Halverson KL (2011). Improving tree-thinking one learnable skill at a time. *Evol Educ Outreach* 4, 95–106.

Halverson KL, Pires CJ, Abell SK (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Sci Educ* 95, 794–823.

Hattie J, Timperley H (2007). The power of feedback. *Rev Educ Res* 77, 81–112.

Hegarty M, Carpenter PA, Just MA (1991). Diagrams in the comprehension of scientific texts. In: *Handbook of Reading Research* vol. 2, ed. R Barr, ML Kamil, PB Mosenthal, and PD Pearson, New York: Longman.

Herreid CF (1994). Case studies in science: a novel method of science education. *J Coll Sci Teach* 23, 221–229.

Hestenes D (1987). Toward a modeling theory of physics instruction. *Am J Phys* 55, 440–454.

Hmelo CE, Holton DL, Kolodner JL (2000). Designing to learn about complex systems. *J Learn Sci* 9, 247–298.

- Lamarck JB (1809). *Philosophie Zoologique*, Paris.
- Lamarck JB (1815). *Histoire Naturelle des Animaux sans Vertèbres*, Paris.
- Lee VR (2010). Adaptations and continuities in the use and design of visual representations in US middle school science textbooks. *Int J Sci Educ* 32, 1099–1126.
- Liu L, Hmelo-Silver CE (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *J Res Sci Teach* 46, 1023–1040.
- Lyman FT (1981). The responsive classroom discussion: the inclusion of all students. In: *Mainstreaming Digest*, ed. A Anderson, College Park: University of Maryland Press.
- Meir E, Perry J, Herron JC, Kingsolver J (2007). College students' misconceptions about evolutionary trees. *Am Biol Teach* 69, 71–76.
- Meisel RP (2010). Teaching tree-thinking to undergraduate biology students. *Evol Educ Outreach* 3, 621–628.
- Momsen J, Offerdahl E, Kryjevskaja M, Montplaisir L, Anderson E, Grosz N (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE Life Sci Educ* 12, 239–249.
- Momsen JL, Long TM, Wyse S, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- Moore R, Jensen M, Hatch J, Duranczyk I, Staats S, Koch L (2003). Showing up: the importance of attendance for academic success in introductory science courses. *Am Biol Teach* 65, 325–329.
- Morabito NP, Catley KM, Novick LR (2010). Reasoning about evolutionary history: post-secondary students' knowledge of most recent common ancestry and homoplasy. *J Biol Educ* 44, 166–174.
- National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School*, Washington, DC: National Academies Press.
- Novick LR, Catley KM (2007). Understanding phylogenies in biology: the influence of a Gestalt Perceptual Principle. *J Exp Psychol Appl* 13, 197–223.
- Novick LR, Catley KM (2013). Reasoning about evolution's grand patterns: college students' understanding of the tree of life. *Am Educ Res J* 50, 138–177.
- Novick LR, Catley KM, Funk DJ (2010). Characters are key: the effect of synapomorphies on cladogram comprehension. *Evol Educ Outreach* 3, 539–547.
- Novick LR, Shade CK, Catley KM (2011). Linear versus branching depictions of evolutionary history: implications for diagram design. *Top Cogn Sci* 3, 536–559.
- Novick LR, Stull AT, Catley KM (2012). Reading phylogenetic trees: the effects of tree orientation and text processing on comprehension. *BioScience* 62, 757–764.
- Omland KE, Cook LG, Crisp MD (2008). Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *Bioessays* 30, 854–867.
- Perry J, Meir E, Herron JC, Maruca S, Stal D (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE Life Sci Educ* 7, 193–201.
- Pungente MD, Badger RA (2003). Teaching introductory organic chemistry: blooming beyond a simple taxonomy. *J Chem Educ* 80, 779–784.
- Sandvik H (2008). Tree thinking cannot be taken for granted: challenges for teaching phylogenetics. *Theory Biosci* 127, 45–51.
- Sandvik H (2009). Anthropocentrism in cladograms. *Biol Philos* 24, 425–440.
- Smith JJ, Cheruvilil KS, Auvenshine S (2013). Assessment of student learning associated with tree thinking in an undergraduate introductory organismal biology course. *CBE Life Sci Educ* 12, 542–552.
- Smith KA (2000). Going deeper: formal small-group learning in large classes. *New Dir Teach Learn* 81, 25–46.
- Sundberg MD (2002). Assessing student learning. *Cell Biol Educ* 1, 11–15.
- Tanner K, Chatman LS, Allen D (2003). Approaches to cell biology teaching: cooperative learning in the science classroom—beyond students working in groups. *Cell Biol Educ* 2, 1–5.
- Theobald R, Freeman S (2014). Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE Life Sci Educ* 13, 41–48.
- Tomanek D, Montplaisir L (2004). Students' studying and approaches to learning in introductory biology. *Cell Biol Educ* 3, 253–262.