STUDENT INTERPRETATIONS OF PHYLOGENETIC TREES

IN AN INTRODUCTORY BIOLOGY COURSE

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Jonathan Andrew Dees

In Partial Fulfillment
for the Degree of
MASTER OF SCIENCE

Major Department:
Biological Sciences
Option: Biology

June 2013

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Student Interpretations of Phylogenetic Trees
in an Introductory Biology Course

By

Jonathan Andrew Dees

The Supervisory Committee certifies that this ***disquisition*** complies with

North Dakota State University's regulations and meets the accepted standards

for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Lisa Montplaisir

Co-Chair

Dr. Jennifer Momsen

Co-Chair

Dr. Julia Bowsher

Dr. Warren Christensen

Approved:

June 7, 2013	Dr. Wendy Reed
Date	Department Chair

ABSTRACT

Phylogenetic trees are a common visual representation in biology, and the most important visual representation used in evolutionary biology. Thus, phylogenetic trees have also become an important component of biology education. We sought to determine what forms of reasoning are utilized by introductory biology students to interpret taxa relatedness on phylogenetic trees, what percentage of students correctly interpret taxa relatedness, and how these results alter in response to instruction and over time. Our students demonstrated a tendency for counting synapomorphies and nodes, rather than more common misinterpretations found in current literature. Students also struggled mightily with correctly interpreting phylogenetic trees, including many who exhibited memorization of correct reasoning. Broad initial instruction achieved little for phylogenetic tree understanding. More targeted instruction on evolutionary relationships improved understanding, but to a still unacceptable level. It appears these visual representations, which can directly affect student understanding of evolution, represent a formidable challenge for instructors.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

INTRODUCTION

The first branching diagrams for illustrating evolution in the history of biology are often attributed to the much-maligned Jean-Baptiste Lamarck (Gould, 1999; Lamarck, 1809). Charles Darwin later drafted his first branching diagram in a research notebook (1837) before publishing a rooted phylogenetic tree as the only illustration in his famous work *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 nearly all disciplines of biology (Baum and Offner, 2008; Omland *et al.*, 2008), and now serve as the main framework for evaluating evidence of evolution (Baum *et al.*, 2005). Phylogenetic trees are so prevalent in biology that "tree-thinking" has been coined 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* (2011) by recommending that communication of scientific concepts through visual representations should be standard in undergraduate biology education.

Phylogenetic trees are visual representations that describe the hypothesized evolutionary relationships among taxa (Baum and Offner, 2008). More precisely, phylogenetic trees illustrate hypotheses about nested groups of taxa (monophyletic groups or clades), which are supported by shared traits known as synapomorphies (Novick and Catley, 2007). Cladograms are the simplest phylogenetic trees, in which branch lengths are arbitrary and typically determined by readability. In contrast, phylogram branches are drawn proportional to some amount of evolutionary change, such as gene divergence, while chronograms have branches that are scaled to represent absolute time (Baum and Offner, 2008; Gregory, 2008; Omland *et al.*, 2008).

1

As visual representations, cladograms are specifically classified as schematic diagrams, which generally illustrate abstract concepts rather than physical appearances of objects (iconic diagrams) or the quantitative relationships of charts and graphs (Hegarty *et al.*, 1990; Lee, 2010; Novick and Catley, 2007). Due to their abstract nature, schematic diagrams are used to describe processes that are difficult to observe, such as evolution, and depend on learned conventions for interpretation (Novick and Catley, 2007). Therefore, cladograms are abstract representations of an abstract concept that use abstract conventions, and it should not be surprising that cladograms present a big challenge to biology students and educators. Understanding cladograms and other phylogenetic trees requires learning the conventions, overcoming preconceived and naïve ideas about taxa, and interpreting the evolutionary relationships of taxa based solely on the branching patterns depicted in the schematic diagrams (Gregory, 2008; Halverson *et al.*, 2011).

As stated by AAAS in *Vision and Change in Undergraduate Biology Education: A Call to Action* (2011), phylogenetic trees are important in biology education for at least three reasons. First, and as previously mentioned, AAAS recommends that all undergraduates have experience communicating scientific concepts and interpretations using visual methods. Phylogenetic trees are visual representations that depict hypothesized relationships between taxa, and as such have the potential to communicate critical evolutionary principles and ideas to a variety of audiences (MacDonald and Wiley, 2012). Second, AAAS recommends that all undergraduates understand the scientific process and how biologists construct knowledge. Stated in a simpler way, learning biology means learning to do biology. Phylogenetic trees are tools used by biological researchers across disciplines (Baum *et al.*, 2005) and are the most important tool of evolutionary biologists, as phylogenetic trees document and organize our current knowledge about the properties of taxa

and their historical relationships (Novick and Catley, 2007). If our students are to learn biology, we must incorporate the primary tools of biologists into undergraduate biology education.

Lastly, and most importantly, AAAS recommends that all undergraduates acquire a basic understanding of evolution, which in turn is essential for understanding biological systems at all levels. Phylogenetic trees integrate evolutionary concepts and provide students with a framework for organizing knowledge about biological diversity (Baum and Offner, 2008). The benefits of phylogenetic trees are derived from their ability to provide unique insights into the patterns and processes of evolution. Visualized in this manner, evolution becomes a conceptual template for thinking and learning about biology, and not just a collection of facts (Novick and Catley, 2007). Our concept of evolution and our phylogenetic tree interpretations are connected, such that each affects the other. Stated another way, how students interpret phylogenetic trees directly impacts their understanding of evolutionary processes (Gregory, 2008; Omland *et al.*, 2008). Common phylogenetic tree misinterpretations are critical barriers to understanding evolution (Meir *et al.*, 2007), and nothing in biology makes sense without considering evolution (Dobzhansky, 1964).

Along with the broad recommendations of AAAS, the College Board (2012) specifically references phylogenetic trees in the standards for its Advanced Placement (AP) biology courses. Although detailed content standards are rare in higher education, AP biology courses offered in high school have rigorous content standards and are intended to be comparable with introductory biology courses at the college level. According to the College Board, knowledge of the structure and function of phylogenetic trees, taxa relatedness, data used for construction, and the dynamic nature of phylogenetic trees is essential for conceptual understanding of biology. Additionally, successful biology students are expected to use phylogenetic trees to make inferences about the evolutionary histories of taxa and construct phylogenetic trees from provided data.

Foundational Studies

Many research, review, and curriculum articles have been published concerning student interpretations of phylogenetic trees and other related topics. With more than 25 Google Scholar citations respectively as of May 2013, Meir *et al.* (2007), Novick and Catley (2007), and Sandvik (2008) comprise the foundational investigations on student interpretations of phylogenetic trees. Additionally, Novick and Catley (2013) is the most comprehensive research published to date in this area, but is too recent to be cited often in the literature. These projects broke new ground by establishing how difficult phylogenetic trees are for students to interpret, by providing the initial evidence for specific misinterpretations, and by comparing student abilities with different styles of phylogenetic trees. Like all research, however, these groundbreaking studies have limitations.

Sandvik (2008) had a very small sample and examined only relatively advanced students who had completed several college biology courses. Meir *et al.* (2007) and Sandvik (2008) both focused entirely on diagonal cladograms (Figures 1a-b), which is also a limitation of the present study. Although bracket cladograms (Figure 1c) are more common in research journals (Novick and Catley, 2007), diagonal cladograms are more common in biology textbooks at all levels of education (Catley and Novick, 2008). Research has demonstrated that students, especially those with less experience in biology, perform better on tasks involving bracket cladograms compared to circular cladograms (Figure 1d) and diagonal cladograms (Halverson, 2011; Halverson *et al.*, 2011; Novick and Catley, 2007; 2013). Although Meir *et al.* (2007) used various orientations of diagonal cladograms, Sandvik (2008) and the present study are further limited by using diagonal cladograms that are exclusively drawn upward from left to right (Figure 1a). Novick *et al.* (2012) demonstrated that such an orientation, although far more common in textbooks, tends to be more difficult for students to understand compared to diagonal cladograms that are drawn downward

from left to right (Figure 1b). The increased difficulty is hypothesized to stem from our learned tendency to scan text and visual representations from left to right.
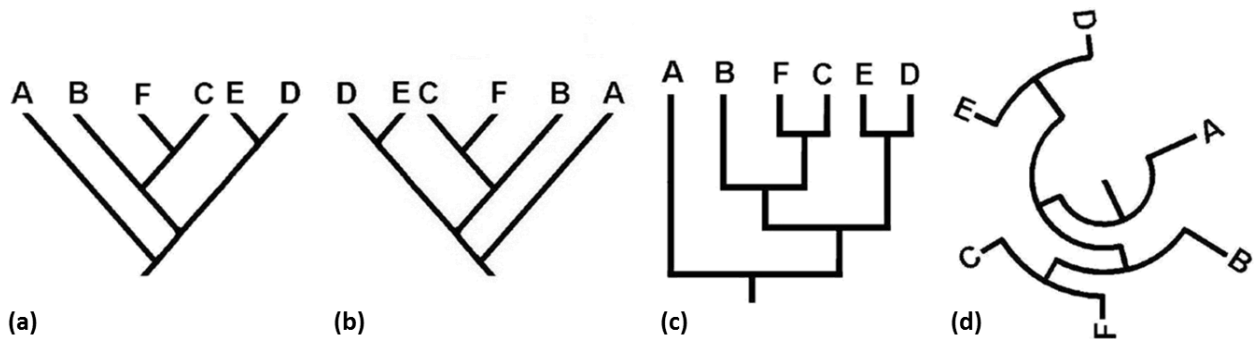
Figure 1. Cladogram styles with equivalent branching patterns (adapted from Gregory, 2008)

Two studies by Novick and Catley (2007; 2013) are distinct among the four foundational phylogenetic tree interpretation studies in that both examine diagonal and bracket cladograms for comparison. The first study accomplished this task by requiring students to translate evolutionary relationships between diagonal cladograms, bracket cladograms, and informative non-cladogram visual representations. The researchers found that translation accuracy decreased when diagonal cladograms were involved in the exercise, especially for students with little or no background in biology (Novick and Catley, 2007). Although an important aspect of diagram literacy in general, translation is not a central skill for understanding phylogenetic trees (Novick and Catley, 2013). Central skills such as evaluating taxa relatedness, identifying monophyletic groups, constructing phylogenetic trees, identifying shared characteristics due to common ancestry (homology), and making inferences based on phylogenetic trees have been recognized in the literature (Halverson, 2011; Meir *et al.*, 2007; Novick and Catley, 2013). The second Novick and Catley study (2013) examined many of the central skills, and determined that students performed better with bracket cladograms compared to diagonal cladograms. Students with stronger biology backgrounds also outperformed students with weaker biology backgrounds, regardless of phylogenetic tree style.

5

The central skills for interpreting phylogenetic trees studied by Novick and Catley (2013) and others were either characterized through qualitative research (Halverson, 2011) or identified by the expertise of researchers (Meir *et al*., 2007; Novick and Catley, 2013), and have never been quantitatively established as critical for phylogenetic tree interpretation. Because the purpose of phylogenetic trees is to illustrate evolutionary relationships, however, it would be nonsensical to suggest that evaluating taxa relatedness is not a requisite skill for interpreting phylogenetic trees. Therefore, the present study employed taxa relatedness interpretations from introductory biology students as the primary indicator of phylogenetic tree understanding.

Phylogenetic Tree Misinterpretations

The nature of phylogenetic tree misinterpretations uncovered by the foundational studies and others is diverse. Evolution in general, and specifically phylogenetic trees, are often viewed as progressive with one main branch. Extant taxa diverging from the main branch are frequently ranked on a scale from primitive to advanced, with humans or mammals generally designated as the most advanced taxon (Baum *et al.*, 2005; Gregory, 2008; Halverson, 2011; Halverson *et al.*, 2011; Meisel, 2010; Omland *et al.*, 2008), despite the lack of biological justification for ranking extant taxa, or for assuming humans are the goal of evolution (Dawkins, 2009). There is a strong tendency to confuse taxon ages with 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 often misinterpreted (Gregory, 2008; Meir *et al.*, 2007; Omland *et al.*, 2008; Perry *et al.*, 2008). Misreading time flow as horizontal on vertical phylogenetic trees (Figures 1a-c) can lead to fundamental errors, such as the conclusion that taxa on the right side of the phylogenetic tree evolved from taxa on the left side, rather than from a common ancestor. Another prevalent error is to assume that evolution happens only at nodes, or straight lines on a phylogenetic tree imply

6

no change from the ancestral state (Baum *et al.*, 2005; Gregory, 2008; Meir *et al.*, 2007; Meisel, 2010; Perry *et al.*, 2008). This misinterpretation can also lead to the false conclusion that extant taxa are descended from other extant taxa, rather than from a common ancestor.

Misinterpretations specific to the relatedness of taxa have also been characterized in the literature (Table 1). The correct method for interpreting taxa relatedness is to utilize most recent common ancestry. 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 Figures 1a-d is more closely related to taxon C than to taxon B, since taxon F and taxon C share a more recent common ancestor. An alternative method for interpreting relatedness is to employ monophyletic groups. Taxon F and taxon C in Figures 1a-d belong to a monophyletic group that does not include taxon B, and so taxon F and taxon C must be more closely related to each other than to taxon B. The misinterpretation cited most often in current literature is using the distance between taxa on phylogenetic trees to determine relatedness (branch tip proximity in Table 1, or commonly referred to as "reading the tips"). Branches can be rotated about the nodes, however, such that taxa positions are arbitrary. Using Figures 1a-d as an example, taxon C and taxon E are placed right next to each other and could be misinterpreted as closely related, even though most recent common ancestry indicates the taxa are distantly related. As previously mentioned, 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 1b as an example, one might suggest that 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 closely related due to descent from a recent common ancestor, and not because one extant taxon is descended from the other.

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

Interpretation	Brief Description and References
Most Recent Common Ancestry (Correct)	Taxa that share a more recent common ancestor are more closely related to each other than to another taxon.[a, b, c, e, f, h, i, l, m, o, q]
Monophyletic Grouping (Correct)	Taxa in the same monophyletic group are more closely related to each other than to a taxon outside of the monophyletic group.[d, e, f, l]
Counting Nodes (Incorrect)	Taxa relatedness is determined by counting nodes or branches between the taxa on phylogenetic trees.[d, e, f, g, h, m, p]
Branch Tip Proximity (Incorrect)	Taxa relatedness is determined by perceived distance between the taxa on phylogenetic trees.[a, b, d, e, f, g, h, l, m, n, p, q]
Contemporary Descent (Incorrect)	Taxa relatedness is determined by indicating that a taxon is descended from another extant taxon.[a, b, d, e, f, g, h, j, l, o, p]
External Insights (Incorrect)	Taxa relatedness is determined by knowledge which is not provided by the phylogenetic tree.[a, b, d, e, f, j, k, m]

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

Counting nodes between taxa on phylogenetic trees to determine relatedness is another misinterpretation found in the literature (Table 1). As an example, four nodes separate taxon C from taxon E, while only three nodes separate taxon B from taxon E in Figures 1a-d. The lesser number of nodes could be misinterpreted as a closer evolutionary relationship between taxon B and taxon E, even though common ancestry indicates taxon B and taxon C are equally related to taxon E. Lastly, previous research has also uncovered misinterpretations in which external (and usually naïve) knowledge of taxa is applied to determine relatedness. As an example, one might suggest that whales (mammals) are closely related to sharks (cartilaginous fish) based on similar aquatic traits, even though these taxa are very distantly related. Unfortunately, phylogenetic tree

misinterpretations are not confined to biology students, as such errors are common even among professional biologists and have appeared in the full range of journals (Omland *et al.*, 2008).

Research Contributions

Although focused only on introductory biology students and diagonal cladograms drawn upward from left to right (Figure 1a), this investigation expands and builds upon previous studies found in the literature. Meir *et al.* (2007), Novick and Catley (2007; 2013), Sandvik (2008), and all other known projects related to student interpretations of phylogenetic trees collected data via questionnaires in which students had no academic stake in their responses. An exception to this generalization is some qualitative research that used coursework and exams as secondary data sources to validate pre-test/post-test and student interview data (Halverson, 2011; Halverson *et al.*, 2011). Questionnaire data are relatively easy to obtain and useful for comparing quantitative scores, but we must be concerned with how seriously students complete questionnaires that will not affect academic standing (Sundberg, 2002). The present study was conducted in the context of an introductory biology class, and as such, all data were collected *in situ* from homework and exams in which students received significant points toward their final grades in the course. Thus, students had a definite academic stake in their responses. The biggest disadvantage to authentic *in situ* data is that collection is subject to course structure and best practices for teaching biology. The instructor must do what is best for student learning, which is not necessarily what is best for the research. For example, some of our data sets were collected from students working together on phylogenetic tree tasks, as cooperative learning has been demonstrated to promote individual academic achievement, quality of relationships between students, self-esteem and psychological health, and positive attitudes toward college and learning (Johnson *et al.*, 1998). Group activities have also been shown to increase student learning specifically for large-enrollment introductory

biology courses (Armstrong *et al.*, 2007). Mixing group and individual data is certainly not ideal for research, but our data reflect the reality of a learner-centered introductory biology course.

While most of the previous phylogenetic tree interpretation studies collected data at one point in time via a questionnaire, our investigation provided a unique opportunity to collect data from isomorphic prompts throughout a class. Such data can be used to examine learning progress as a result of instruction and feedback, as well as consistency of student understanding over time. The most common quantitative method for assessing learning is the pre-test/post-test instrument (Sundberg, 2002), and such instruments have often been used to measure student understanding of evolution by natural selection (e.g., Anderson *et al.*, 2002; Bishop and Anderson, 1990; Nehm and Reilly, 2007). As previously mentioned, our research used taxa relatedness interpretations as the main indicator of phylogenetic tree understanding, and student responses to similar questions before and after instruction can be used to assess student learning. Examining the consistency of post-instructional understanding over time is a relatively unstudied area of research, despite calls for such investigations (Handelsman *et al.*, 2004).

Research Questions

Building upon the foundational studies and other investigations on student interpretations of phylogenetic trees, our research aims to answer the following questions:

1) What forms of reasoning are utilized 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 for the first two research questions change in response to instruction and over time?

METHODS

This investigation was conducted during the second course of a two-course introductory biology sequence for science majors at a public university with very high research activity in the Midwestern United States. The large-enrollment course served students with a variety of majors (Table 2) at various stages in their academic careers (Table 3). Completion of the first course in the sequence was encouraged but was not considered a prerequisite for the second course.

Table 2. Student enrollment by major grouping

Major Grouping	No. of Students (%)
Agricultural Sciences	19 (22)
Biological Sciences	32 (36)
Pre-Professional Health	10 (11)
Natural Resources	16 (18)
Undeclared and Other	11 (13)

Table 3. Student enrollment by class

Class	No. of Students (%)
Freshmen	21 (24)
Sophomores	29 (33)
Juniors	16 (18)
Seniors	22 (25)

The instructor utilized a learner-centered approach to teaching biology in which multiple forms of active engagement were employed in lieu of passive lectures. Class 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 scientific case studies (Herreid, 1994). Modeling instruction (Brewe, 2008; Hestenes, 1987; Hmelo *et al.*, 2000; Liu and Hmelo-Silver, 2009) was the most prominent pedagogical strategy, as students frequently constructed box-and-arrow models of complex biological processes, such

as evolution, energy flow through ecosystems, and nutrient cycling. Course 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 main units: evolution, form and function, and ecology. Class meetings of 75 minutes occurred twice each week for 16 weeks, and the data for this investigation were collected from four assessments (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 the evolved characteristics necessary for plant survival on land.
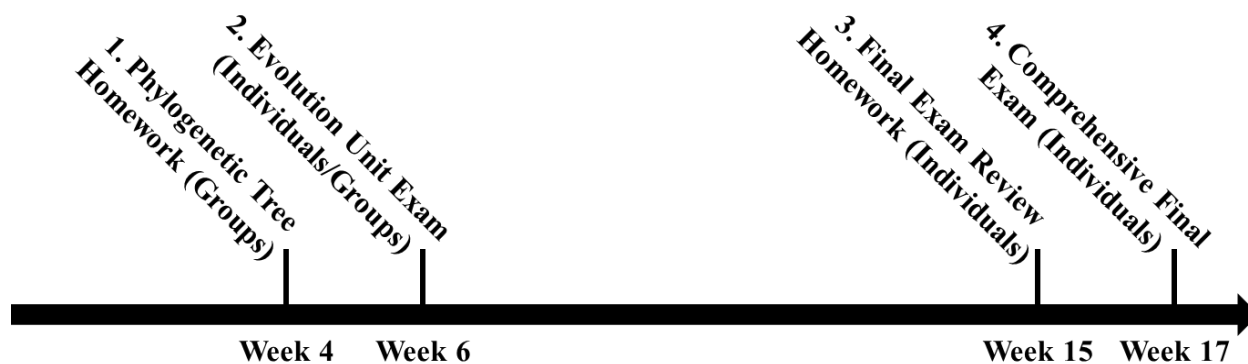
Figure 2. Timeline of data collection from four separate assessments throughout the course

Instruction and Assessment

Students worked extensively in permanent, self-selected groups of three or four students throughout the course. All quizzes and exams were pyramid assessments (Eaton, 2009) featuring individual and collaborative components (approximately 75% completed as individuals and 25% in groups by point allocation). The initial phylogenetic tree homework was completed in groups after the instructor introduced phylogenetic trees during class, and after the students completed a reading quiz on related materials in the course textbook (Freeman, 2011). The phylogenetic tree

12

introduction consisted of five challenging questions posed by the instructor and answered by the students using letter cards. The questions familiarized students with the structural characteristics of phylogenetic trees, such as nodes (representing common ancestors) and monophyletic groups, and presented the idea that relatedness is determined by common ancestry. Each question posed by the instructor was 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 the class referred to simple cladograms, in which only the branching pattern has meaning. Non-cladogram phylogenetic trees such as phylograms were briefly mentioned by the instructor, but students were never required to interact with or reason from them during the course.

The initial homework featured a series of open-ended questions and was designed around a phylogenetic tree of chordates (Figure 3). In addition to prompts about synapomorphies, recent common ancestors, and monophyletic groups, one question referring to relatedness also appeared in the homework (Figure 3). Poor group performance on this question compelled the instructor to revisit phylogenetic tree interpretations during class. The same question from the homework was presented to students and debated via directed, small-group discussions. The ensuing whole-class discussion identified an appropriate reasoning strategy for determining relatedness of taxa: most recent common ancestry. Additional class time was also allocated to examining different aspects of phylogenetic trees. The instructor utilized letter card questions and discussions to review basic structural features, investigate the idea of equivalent phylogenetic trees through branch rotations, debunk the concept of "primitive" and "advanced" taxa, and explore different forms of evidence used by evolutionary biologists to construct phylogenetic trees. In addition to revisiting the initial homework prompt, taxa relatedness was specifically targeted by the instructor through two letter card questions with small-group and whole-class discussions during class.

Florida Lancelet | Black Hagfish | Zebra Shark | Northern Pike | Fire Salamander | American Alligator | Bald Eagle | Platypus | Red Kangaroo | Western Gorilla

Feathers
Gizzard
Long Gestation
Placenta
Lactation
Claws/Nails
True Lungs
Bone Skeleton
Jaw
Cranium
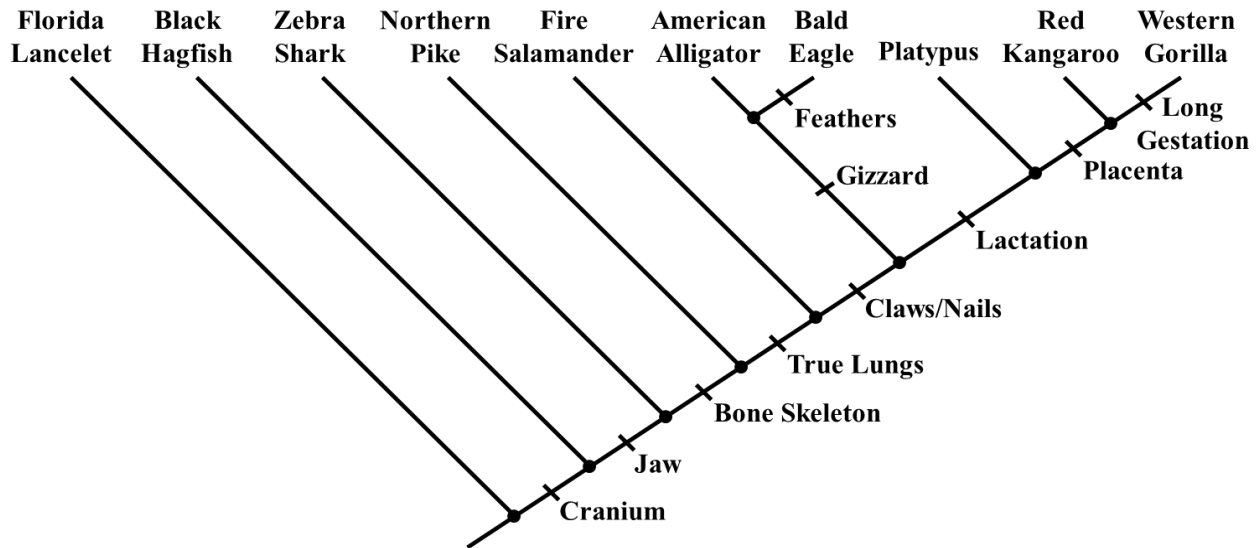
Based on the phylogenetic tree, would American alligators be more closely related to fire salamanders or red kangaroos? Explain the reasoning for your choice.

Figure 3. Phylogenetic tree from the initial homework with its taxa relatedness question

Taxa relatedness questions similar to the initial homework question appeared on both the individual and group components of the evolution unit exam two weeks after the homework was submitted. Students were provided a phylogenetic tree and relatedness prompt for the individual component (Figure 4), but the group component required students to build the phylogenetic tree from a table of morphological traits before answering a relatedness question (Figure 5). Students were never asked to construct a phylogenetic tree prior to completing this assessment. The initial homework and both unit exam relatedness questions were isomorphic, but utilized phylogenetic trees with different taxa (chordates, mammals, and plants), synapomorphies (only morphological traits), and branching patterns (topologies). The evolutionary relationships and synapomorphies illustrated in the phylogenetic trees were extracted primarily from the course textbook (Freeman, 2011) and are biologically accurate to the best of our understanding. Authenticity was of primary concern when constructing the phylogenetic trees for student use during the introductory course.
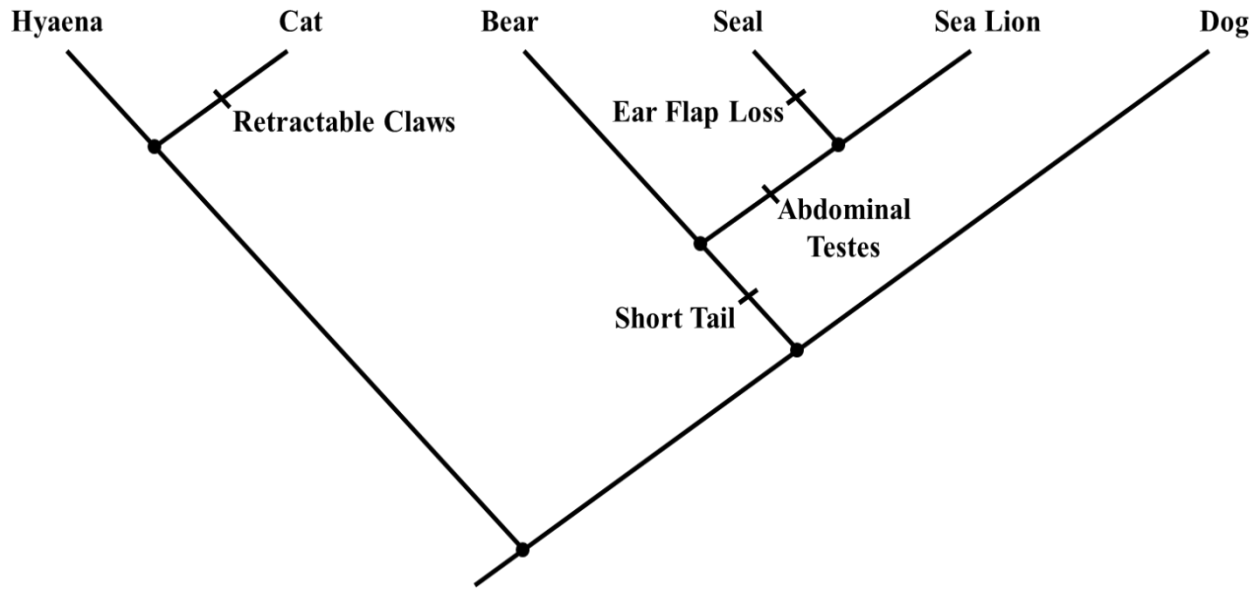
14

Figure 4. Phylogenetic tree (adapted from Baum *et al.*, 2005) from the individual component of the evolution unit exam with its taxa relatedness question

Use the morphological traits shown in Table A to construct a basic phylogenetic tree for plants (any style of phylogenetic tree is fine). Be sure to clearly label all synapomorphies.

Table A. Morphological traits of plants (X = trait possessed by plant).

		Plants					
		Lady Fern	Durum Wheat	Liverworts	Ginkgo	Green Algae	White Spruce
Traits	Seeds		X		X		X
	Needles						X
	Cuticle	X	X	X	X		X
	Flowers		X				
	Stomata	X	X		X		X
	Cones				X		X

> **Based on the phylogenetic tree of plants you created, would lady ferns be more closely related to liverworts or durum wheat? Explain the reasoning for your choice.**

Figure 5. Phylogenetic tree construction prompt from the group component of the evolution unit exam with its taxa relatedness question

15

Due to the long time lapse between the evolution unit exam and the comprehensive final exam (Figure 2), students were required to individually complete a review homework two weeks prior to the final exam that included phylogenetic trees. A taxa relatedness question appeared on the review homework (Figure 6) and then again on the final exam (Figure 7) to provide a total of five data sources for this study (two in groups and three as individuals). The relatedness prompts for the review homework and comprehensive final exam were isomorphic, but used phylogenetic trees with different taxa and branching patterns. The prompt structure was changed slightly from the initial homework and two unit exam questions to multiple-choice with open-ended reasoning. This small alteration was made to prevent students from feeling compelled to select one taxon or the other, such as on the initial homework (Figure 3), because involved taxa are equally related.
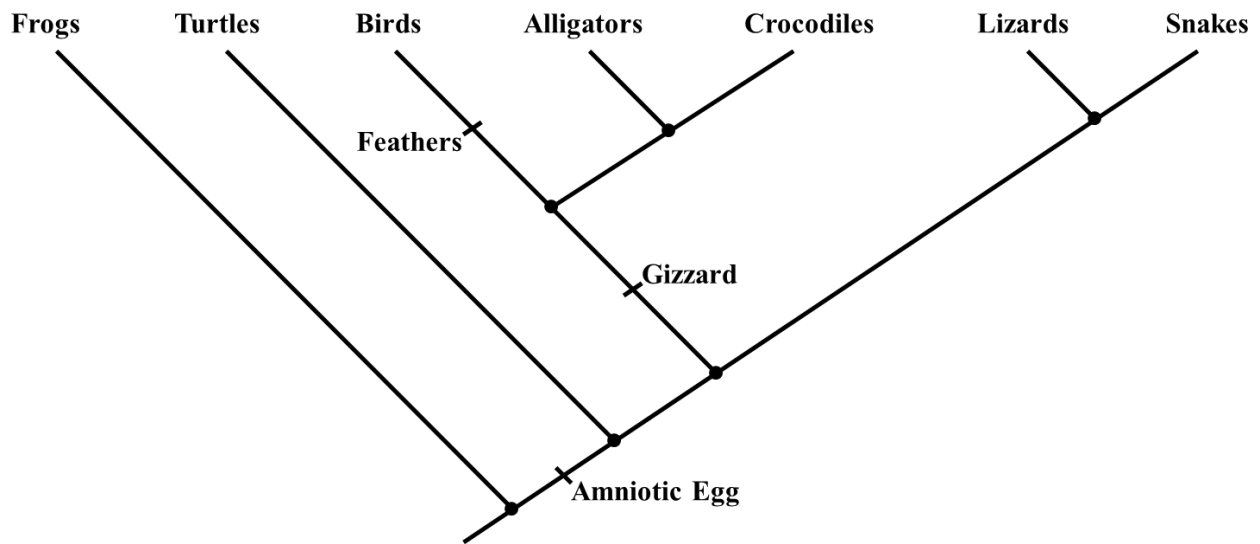
According to the phylogenetic tree, how would you describe the relatedness of lizards to crocodiles and birds? Explain the reasoning for your choice.

(a) Lizards are more closely related to crocodiles than birds.
(b) Lizards are more closely related to birds than crocodiles.
(c) Lizards are equally related to crocodiles and birds.
(d) Lizards are not related to crocodiles and birds.

Figure 6. Phylogenetic tree from the review homework with its taxa relatedness question

16

Figure 7. Phylogenetic tree (adapted from Baum *et al.*, 2005) from the individual component of the comprehensive final exam with its taxa relatedness question

Rubric Development and Coding

We began by investigating the reasoning provided by groups for the relatedness question on the initial phylogenetic tree homework (Figure 3), and identified patterns in the responses. A coding rubric was initially developed from student responses using a grounded theory approach (Glaser and Strauss, 1967), but the final rubric categories were later confirmed and refined using existing literature on misinterpretations of phylogenetic trees (Table 4). The negation reasoning and counting synapomorphies categories of the rubric are novel for this investigation, and to our knowledge, have not previously been described in the literature. Additional examples of counting synapomorphies, monophyletic grouping, and negation reasoning generated by students and used to develop the rubric can be found in the Appendix (Tables A1-A3).

17

Table 4. Rubric for taxa relatedness reasoning with examples from student responses

Rubric Category	Brief Description and References *(Student-Generated Example)*
Most Recent Common Ancestry (Correct)	Taxa that share a more recent common ancestor are more closely related to each other than to another taxon.[a, b, c, e, f, h, i, l, m, o, q] *Lizards share the same most recent common ancestor with both birds and crocodiles (supports choice "c" in Figure 6).*
Monophyletic Grouping (Correct)	Taxa in the same monophyletic group are more closely related to each other than to a taxon outside of the monophyletic group.[d, e, f, l] *The bear is in the same monophyletic group as the sea lion and the seal, thus possessing more of a close relationship than with the cat (supports choice "sea lions" in Figure 4).*
Counting Nodes (Incorrect)	Taxa relatedness is determined by counting nodes or branches between the taxa on phylogenetic trees.[d, e, f, g, h, m, p] *Lizards and birds have less nodes between them than the lizards and crocodiles (supports choice "b" in Figure 6).*
Counting Synapomorphies (Incorrect)	Taxa relatedness is determined by counting synapomorphies between the taxa on phylogenetic trees. *Bears have only one trait different with sea lions and two traits that differ with cats (supports choice "sea lions" in Figure 4).*
Branch Tip Proximity (Incorrect)	Taxa relatedness is determined by perceived distance between the taxa on phylogenetic trees.[a, b, d, e, f, g, h, l, m, n, p, q] *The salamander is closer to the alligator than the red kangaroo on the tree (supports choice "fire salamanders" in Figure 3).*
Contemporary Descent (Incorrect)	Taxa relatedness is determined by indicating that a taxon is descended from another extant taxon.[a, b, d, e, f, g, h, j, l, o, p] *The sea lion branches off from bears, and cats are on a totally different branch (supports choice "sea lions" in Figure 4).*
External Insights (Incorrect)	Taxa relatedness is determined by knowledge which is not provided by the phylogenetic tree or prompt.[a, b, d, e, f, j, k, m] *We believe this because the alligator and salamander are more closely related in characteristics, habitat, geographic location, and behaviors (supports choice "fire salamanders" in Figure 3).*
Negation Reasoning (Neutral)	Reasoning includes descriptions of how not to interpret taxa relatedness on phylogenetic trees (concurrent with other reasoning in all cases). *Bears and sea lions have a more recent common ancestor [most recent common ancestry], and you must not pay attention to the top of the tree because the branches can rotate [negation reasoning] (supports choice "sea lions" in Figure 4).*
Other Responses (Incorrect)	Reasoning did not conclusively fall into any of the above categories. *Lizards and birds share a common ancestor, or node (supports choice "b" in Figure 6).*

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

All responses from the first two assessments (initial phylogenetic tree homework and the individual and group components of the evolution unit exam) were individually numbered, and a certified random number generator was used to select a sample of 20 responses (15% of the total at that time). Two independent raters separately coded the responses, and then reached consensus for each response through discussion. Following calibration of the rubric, agreement between the raters was 94% for the remaining 258 responses from all four assessments, and all disagreements were resolved via discussion. Student responses often included more than one form of reasoning and consequently fell into multiple rubric categories, which resulted in 278 group and individual responses that generated 360 reasoning codes. The coding process was partially-blind, in which one rater was aware of group and student identities while the other rater was not. Due to the high agreement between independent raters, we do not believe rater bias was a significant issue.

The taxa relatedness questions used throughout the course required students to choose an answer and provide reasoning for their selection (Figures 3-7). Because the answer choices given by students were not always consistent with their reasoning, responses were coded a second time according to the answer choice (correct or incorrect) and the reasoning used to defend the choice (correct, incorrect, or mixture). The rubric categories of monophyletic grouping and most recent common ancestry were considered correct reasoning for determining taxa relatedness. Negation reasoning always appeared with other forms of reasoning and was considered neither correct nor incorrect, while all other categories were deemed incorrect reasoning (Table 4). This procedure allowed us to identify students who simply guessed correct answers without truly understanding the phylogenetic trees (correct answer with incorrect reasoning), and students who knew correct reasoning but did not understand how to apply it (incorrect answer with correct reasoning). Only

responses that provided correct answers with correct reasoning (monophyletic grouping or most recent common ancestry) exhibited true understanding of taxa relatedness on phylogenetic trees.

Statistical Analysis

Coding student responses for taxa relatedness questions by level of correctness resulted in a two-way table of counts with categorical variables (correctness versus data source). Answering the third research question concerning changes in student responses as a result of instruction and over time required statistical analysis, and the main test for determining the relationship between categorical variables in a two-way table of counts is the chi-squared (χ^2) statistic. However, such analysis requires random samples from two or more independent populations, or a single random sample with individuals classified according to both of two categorical variables (Moore, 2007). Thus, our experimental design with repeated measurements of understanding for the same basic taxa relatedness question over time violates the assumptions underlying the chi-squared statistic.

An alternative to the chi-squared statistic is the McNemar test, which is a non-parametric method used to determine whether the row and column frequencies of a two-by-two contingency table are equal (McNemar, 1947). This test was designed for use with matched data sets, such as repeated measurements, by assuming statistical dependence between the samples (Sun and Yang, 2008). The test examines the difference from one data set to another, and results in a chi-squared statistic with one degree of freedom. However, the McNemar test is limited to data sets with two categories (dichotomous). We used this analysis only to examine differences in how the students answered taxa relatedness questions (correct or incorrect answer choice) from one data source to another, irrespective of reasoning used to defend their choices.

A generalization of the McNemar test known as the Stuart-Maxwell test, however, can be applied to square contingency tables larger than two-by-two (Maxwell, 1970; Stuart, 1955). This

test was also designed for use with matched data sets by assuming statistical dependence for the samples. The result of the test is a chi-squared statistic with R–1 degrees of freedom, where R is the number of categories (Sun and Yang, 2008). We used this analysis to examine differences in how students answered taxa relatedness questions (correct or incorrect answer choice) combined with reasoning used to defend their choices (correct, incorrect, or mixture) from one data source to another. The combinations of two answer choice possibilities and three reasoning possibilities resulted in six categories, and hence a six-by-six contingency table with five degrees of freedom.

The McNemar and Stuart-Maxwell tests can be used to compare matched data sets from groups or matched data sets from individuals, but group responses cannot be directly compared to individual responses. Taxa relatedness responses from the initial phylogenetic tree homework (which was completed in groups) were compared to responses from the group component of the evolution unit exam, while relatedness responses from the final exam review homework (which was completed individually) were compared to responses from the individual component of the final exam. Responses from the individual component of the evolution unit exam were not used for statistical analysis, as the results were unreliable due to the poor structure of the relatedness question (see Discussion). Students who completed one assessment for a comparison but not the other were excluded from statistical analysis due to the data matching requirements of both the McNemar and Stuart-Maxwell tests. All analyses were completed using the RStudio statistical program (version 0.97.336 with the IRR extension package installed).

Institutional Review Board

This investigation was completed in compliance with all requirements of the Institutional Review Board for research with humans (protocol SM12217: Using Course Artifacts to Improve Teaching and Learning in Biology), and qualified for exempt status under federal regulations.

RESULTS

The taxa relatedness prompts for this investigation required students to choose an answer and then provide reasoning for their selection (Figures 3-7). Correct answer selection (Figure 8) ranged from very low on the initial phylogenetic tree homework completed by groups (8.3%) to moderate on the group component of the evolution unit exam (69.6%). The extremely high result for the individual component of the evolution unit exam (95.5%) is an aberration due to the poor structure of the relatedness question (see Discussion). McNemar testing shows performance was significantly different between the initial phylogenetic tree homework completed by groups and the group component of the evolution unit exam (χ^2=13.07, df=1, p<0.001). However, there was no significant difference in performance between the final exam review homework completed by individuals and the individual component of the final exam (χ^2=0.06, df=1, p=0.814).
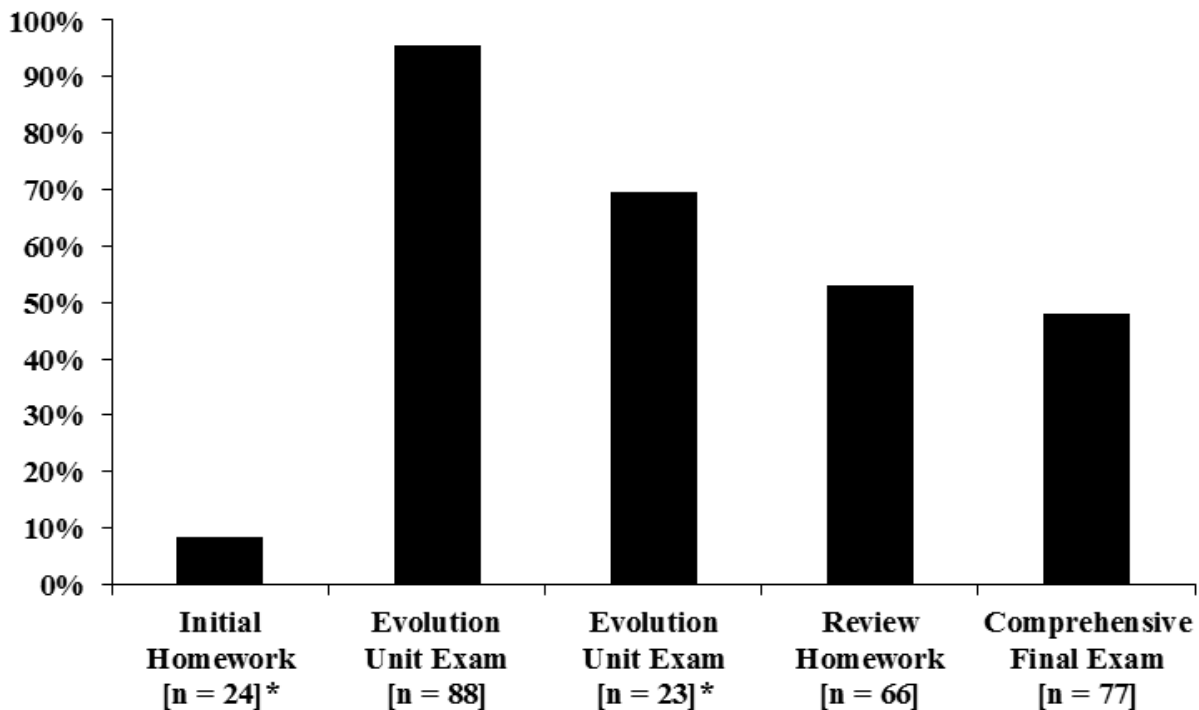

Figure 8. Percentage of responses with the correct answer selected for all five data sources

The reasoning rubric (Table 4) was used to code reasoning from all five data sources into one or more categories (Table 5). Groups performed poorly on the initial homework, as only two groups used most recent common ancestry or monophyletic grouping reasoning. By far the most common form of incorrect reasoning was counting synapomorphies, followed by counting nodes, branch tip proximity, and external insights. Taxa relatedness responses from both the individual and group components of the evolution unit exam collected two weeks after the initial homework were greatly improved in terms of reasoning. Counting nodes and synapomorphies remained the most common types of incorrect reasoning, while branch tip proximity and external insights both virtually disappeared from responses. The reasoning of individuals persisted from the individual component of the evolution unit exam to the final exam review homework submitted nine weeks later, although the use of monophyletic grouping reasoning decreased substantially (22% to 8% of responses). Individual responses from the comprehensive final exam differed somewhat from the preceding review homework. Most recent common ancestry, contemporary descent, branch tip proximity, and external insights all increased as counting synapomorphies decreased to only two responses on the final exam, where the phylogenetic tree did not contain synapomorphies.

Correct interpretations of taxa relatedness on phylogenetic trees require both knowledge of correct reasoning, as demonstrated by student explanations of their reasoning (Table 5), and understanding how to apply correct reasoning, as demonstrated by selecting the correct answers for taxa relatedness questions (Figure 8). The results from the second coding scheme for answer choice (correct or incorrect) and reasoning used to support the answer choice (correct, incorrect, or mixture) identify the combinations of knowledge and understanding (Table 6). Although two groups selected the correct answer, and two other groups offered correct or mixed reasoning on the initial homework, not a single group provided a completely correct response. The evolution

unit exam exhibited the highest rates of completely correct taxa relatedness responses (58% for the individual component and 57% for the group component), although results for the individual component are most likely inflated due to the poor structure of the taxa relatedness question (see Discussion). Stuart-Maxwell testing shows performance was significantly different between the initial homework and the group component of the evolution unit exam (χ^2=17.5, df=4, p=0.002). Unlike the other four taxa relatedness prompts, the group component of the evolution unit exam required students to build the phylogenetic tree from provided data (Figure 5), and 22 of the 23 groups constructed a phylogenetic tree that was sufficient to correctly answer the ensuing taxa relatedness question (i.e., contained accurate evolutionary relationships for the involved taxa).

Table 5. Reasoning used by students to determine taxa relatedness for all five data sources

Rubric Category	Initial Homework [n = 24]*	Evolution Unit Exam [n = 88]	Evolution Unit Exam [n = 23]*	Review Homework [n = 66]	Comprehensive Final Exam [n = 77]
Most Recent Common Ancestry	*1 (4)*	*51 (58)*	*17 (74)*	*40 (61)*	*54 (70)*
Monophyletic Grouping	*1 (4)*	*19 (22)*	*7 (30)*	*5 (8)*	*7 (9)*
Counting Nodes	6 (25)	19 (22)	2 (9)	12 (18)	14 (18)
Counting Synapomorphies	12 (50)	11 (13)	6 (26)	6 (9)	2 (3)
Branch Tip Proximity	5 (21)	1 (1)	1 (4)	1 (2)	7 (9)
Contemporary Descent	0 (0)	2 (2)	1 (4)	1 (2)	4 (5)
External Insights	4 (17)	1 (1)	0 (0)	1 (2)	3 (4)
Negation Reasoning	0 (0)	5 (6)	0 (0)	3 (5)	3 (4)
Other Responses	1 (4)	6 (7)	1 (4)	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). *Responses were submitted by permanent groups of three or four students.					

Table 6. Coding results for answer choice (correct or incorrect) and reasoning used by students to defend the answer choice (correct, incorrect, or mixed) for all five data sources

Rubric Category	Initial Homework [n = 24]*	Evolution Unit Exam [n = 88]	Evolution Unit Exam [n = 23]*	Review Homework [n = 66]	Comprehensive Final Exam [n = 77]
Correct Answer Correct Reasoning	0 (0)	51 (58)	13 (57)	31 (47)	29 (38)
Correct Answer Mixed Reasoning	0 (0)	14 (16)	3 (13)	1 (1)	3 (4)
Correct Answer Incorrect Reasoning	2 (8)	19 (22)	0 (0)	3 (5)	5 (6)
Incorrect Answer Correct Reasoning	1 (4)	0 (0)	1 (4)	6 (9)	16 (21)
Incorrect Answer Mixed Reasoning	1 (4)	1 (1)	3 (13)	4 (6)	8 (10)
Incorrect Answer Incorrect Reasoning	20 (84)	3 (3)	3 (13)	21 (32)	16 (21)
*Responses were submitted by permanent groups of three or four students.					

The relatedness questions on the final exam review homework and comprehensive final exam were completed as individuals and had lower rates of completely correct responses (47% and 38%, respectively). Stuart-Maxwell testing shows the performance difference between the last two assessments was borderline significant (χ^2=10.9, df=5, p=0.052). After answering very similar relatedness questions in class and on multiple previous assessments (including a review homework two weeks before), only 38% of the students provided a completely correct response to the taxa relatedness question on the final exam. An additional 31% of the students selected an incorrect answer despite offering correct or at least mixed forms of reasoning.

Only 24% of the students offered correct answers coupled with correct reasoning for both the review homework and final exam relatedness prompts, which were completed as individuals. Academic majors, class distinctions, and final course grades for these 19 students were compared to course enrollment as a whole (Table 7). Statistical analysis could not be performed in this case due to the small number of students who offered completely correct responses to both questions.

Table 7. Comparison of all students and those who offered correct responses

Major Grouping	Course Enrollment No. of Students (%)	Correct Responders No. of Students (%)
Agricultural Sciences	19 (22)	3 (16)
Biological Sciences	32 (36)	4 (21)
Pre-Professional Health	10 (11)	3 (16)
Natural Resources	16 (18)	6 (31)
Undeclared and Other	11 (13)	3 (16)
Class	**Course Enrollment No. of Students (%)**	**Correct Responders No. of Students (%)**
Freshmen	21 (24)	5 (26)
Sophomores	29 (33)	7 (37)
Juniors	16 (18)	3 (16)
Seniors	22 (25)	4 (21)
Final Course Grade	**Course Enrollment No. of Students (%)**	**Correct Responders No. of Students (%)**
A	19 (24)	10 (53)
B	30 (37)	7 (37)
C	19 (24)	1 (5)
D	9 (11)	1 (5)
F	3 (4)	0 (0)

DISCUSSION

Student responses to questions concerning relatedness of taxa on phylogenetic trees were collected from five data sources throughout an introductory biology course. Two of the questions were completed by groups of three or four students (Figures 3 and 5), while three questions were completed by individuals (Figures 4, 6, and 7). The prompts asked students to choose an answer and then provide reasoning for their selection, and responses were coded according to reasoning strategy (Table 5) and correctness (Table 6). Only responses that provided correct answers along with correct reasoning exhibited true understanding of taxa relatedness on phylogenetic trees.

Individual Component of the Evolution Unit Exam

As previously disclosed, taxa relatedness responses from the individual component of the evolution unit exam were not used for correctness analysis, as the outcomes were unreliable due to the structure of the question (Figure 4). According to the phylogenetic tree on the exam, bears are more closely related to sea lions than cats, based on monophyletic grouping and most recent common ancestry. Some forms of incorrect reasoning, such as branch tip proximity and external insights, lead students to choose cats instead of sea lions. However, the more common incorrect strategies of counting synapomorphies and nodes lead students to choose sea lions as the answer. Therefore, an astonishing outcome of 95.5% correct answers (Figure 8) may not reflect student understanding, but rather the poor structure of the question in relation to the phylogenetic tree.

Because nearly all students selected the correct answer to the taxa relatedness question on the individual component of the evolution unit exam, regardless of their reasoning, we are unable to distinguish students with true understanding from students who simply memorized most recent common ancestry or monophyletic grouping as an appropriate response for relatedness questions. Guessing was still distinguishable when students chose the correct answer and provided incorrect

reasoning, but the overall outcome of 58% completely correct responses (Table 6), the highest of any data source, is likely inflated and unreliable. Although student responses from the individual component of the evolution exam are not trusted for correctness analysis, we included the data for two important reasons. First, reasoning given by students, regardless of their answer choices, provided valuable insights into student thinking and contributed to development of the reasoning rubric (Table 4). Second, the question is a cautionary tale by providing an example of what not to do when assessing student understanding of phylogenetic trees. Instructors must carefully design evaluation items, such that responses provide an accurate reflection of student understanding.

Research Question 1: Student Reasoning

Relatedness of taxa is understood by biologists in terms of most recent common ancestry, similar to the family trees of humans (Baum *et al.*, 2005). Following the initial phylogenetic tree homework in which all students struggled, about two out of every three students were aware that most recent common ancestry somehow determined relatedness. The percentage from individual assessments varied from 58% on the individual component of the evolution unit exam to 70% on the final exam (Table 5), although far fewer students wielded that knowledge correctly (Table 6). Student usage of the alternative correct reasoning, monophyletic grouping, was a very interesting outcome of this study. Its usage was especially frequent on the individual and group components of the evolution unit exam (22% and 30%, respectively). Monophyletic groups were discussed at length during the introductory course, but neither the instructor nor the textbook (Freeman, 2011) directly suggested using monophyletic groups to determine taxa relatedness. Students generated this form of reasoning on their own, either spontaneously or from outside materials. Examples of monophyletic grouping reasoning utilized by students can be found in the Appendix (Table A1).

28

Counting synapomorphies and nodes were by far the two most popular forms of incorrect reasoning used by students to determine taxa relatedness on phylogenetic trees (Table 5), which was an unexpected result for this investigation. Branch tip proximity, contemporary descent, and external insights dominate the literature (see Table 4 for references), yet these forms of reasoning were far less common in responses. Determining taxa relatedness by counting synapomorphies has not previously been described in the literature to our knowledge, but proved to be a worthy instructional adversary. Two students even attempted to use this reasoning for the comprehensive final exam relatedness question, which did not include synapomorphies (Figure 7). The students suggested the three taxa involved in the question are equally related to each other (which is true) because there are no trait differences between them. This is a completely illogical statement, and a testament to the persistence of phylogenetic tree misinterpretations. Additional examples of the counting synapomorphies form of reasoning generated by students can be found in the Appendix (Table A2). The existence and strength of this reasoning form presents an educational dilemma.

A previous investigation demonstrated that labeled synapomorphies on phylogenetic trees support comprehension of evolutionary relationships (Novick *et al.*, 2010). The researchers used translation tasks between two phylogenetic tree styles (diagonal and bracket), and students were significantly more accurate when synapomorphies were present. The investigators suggested that synapomorphies improve performance for translation exercises 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 facilitate identification of points along continuous lines where hierarchical levels begin. From a biological perspective, synapomorphies facilitate identification of monophyletic groups, which are preserved during translation from one style of phylogenetic tree to another. Although synapomorphies seem to be helpful for translating

29

between phylogenetic trees, they also appear to be problematic for interpretations within a single phylogenetic tree, as our students often misused them to determine taxa relatedness. In one case, synapomorphies act as guide markers, while in another case, synapomorphies act as distractors. This apparent conflict in the usefulness of synapomorphies for phylogenetic tree translation and interpretation exercises warrants further investigation.

Research Question 2: Student Correctness

Because the main purpose of phylogenetic trees is to illustrate evolutionary relationships, our investigation utilized taxa relatedness interpretations as the primary indicator of phylogenetic tree understanding. The relatedness prompts for this study required students to choose an answer and then provide reasoning for their selection (Figures 3-7). Measuring understanding of students was achieved by combining results from answer choices (Figure 8) and reasoning used to defend choices (Table 5). Only responses that provided correct answers coupled with correct reasoning exhibited true understanding and the ability to correctly interpret phylogenetic trees. Discounting the unreliable individual component of the evolution unit exam, less than half of the individuals in the introductory biology course demonstrated such an ability (Table 6).

With the ability to pool knowledge and understanding, we expected permanent groups of students to outperform individuals. Discounting the initial homework in which students had little experience with phylogenetic trees, and the unreliable question from the individual component of the evolution unit exam, groups did fare somewhat better than individuals. More than half of the groups (57%) provided a completely correct response to the taxa relatedness prompt on the group component of the evolution unit exam, while 47% and 38% of individual students provided such responses on the final exam review homework and final exam, respectively. This result is based on only three sets of data, but aligns with our expectations for cooperative learning. However, an

alternative explanation is that students performed better on the group component of the evolution unit exam due to the act of building a phylogenetic tree before responding to the taxa relatedness question (Figure 5). Phylogenetic tree construction could have forced students to focus more on relationships between taxa than would otherwise be the case when given the diagram, or perhaps the greatly increased time on task required for the exercise led to improved student performance. This alternative explanation cannot be ruled out, and determining the effects of phylogenetic tree construction tasks on student understanding warrants further research.

Other combinations of relatedness question answers and reasoning also provide valuable insights into student understanding. Correct answers with incorrect reasoning indicate students who simply guessed correct answers without truly understanding phylogenetic trees. Keeping in mind that the individual component of the evolution unit exam is unreliable, as nearly all of the students (95.5%) chose the correct answer regardless of reasoning due to poor question structure, guessing correctly was relatively rare during this investigation (ranging from 0% to 8%). On the other hand, incorrect answers coupled with correct reasoning indicate students who memorized appropriate reasoning for a taxa relatedness question but did not understand how to apply it to a phylogenetic tree. This outcome was far more common during our investigation, ranging from 4% earlier in the course to 21% on the final exam. An additional 4-13% of individuals or groups provided incorrect responses with mixed reasoning (including correct and incorrect reasoning), which also indicate some degree of memorization without true understanding. Shallow learning strategies are 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 the frequency, type, and student use of feedback (Hattie and Timperley, 2007).

Academic major did not seem to play a big role in student understanding of phylogenetic trees, as the proportions of students by major offering correct answers with correct reasoning for the review homework and final exam taxa relatedness questions were very similar to the overall enrollment proportions (Table 7). However, fewer students majoring in biological sciences were successful (21%) than would be expected by enrollment (36%), while natural resources students were somewhat more successful (31%) than would be expected by enrollment (18%). Collegiate experience did not seem to play any role at all in student understanding, as the proportions were virtually identical for successful students and enrollment as a whole by year in school (Table 7). Final course grade does appear to be predictive for student understanding of phylogenetic trees, however, as a disproportionate number of students who demonstrated understanding received an A for the course (53%), compared to 24% of students overall (Table 7). This result could reflect the importance of understanding evolution in general for success in the course. One of the three major units was devoted to evolution, and the other major units required some understanding of evolution, as the theory connects all of biology (Dobzhansky, 1964). Because phylogenetic tree interpretations and understanding of evolution affect each other (Gregory, 2008; Omland *et al.*, 2008), it seems likely that students who correctly interpreted phylogenetic trees also exhibited better understanding of evolution, and thus performed at a higher level throughout the course.

Research Question 3: Instructional and Time Effects

The initial phylogenetic tree homework was completed by permanent groups of students after the instructor introduced phylogenetic trees and after the students completed a reading quiz on related materials in the course textbook (see Methods). This original exposure clearly did not generate understanding of phylogenetic trees, as only 8% of the groups utilized correct forms of reasoning, and no groups provided a completely correct response to the taxa relatedness prompt.

The instructor intervention after poor group performances on the initial homework, which used the same phylogenetic tree and relatedness question from the homework (Figure 3) and focused more on the relatedness concept than prior instruction, had a large effect. Knowledge of correct reasoning greatly increased on the group component of the evolution unit exam (Table 5), while completely correct responses dramatically increased from 0% to 57%. The overall distributions of correctness (Table 6) also changed significantly between the two data sources (p=0.002).

The positive result from the instructor intervention has several implications for teaching and learning about phylogenetic trees. First, interpreting phylogenetic trees is far from intuitive and necessitates explicit training, which agrees with previous conclusions (Novick and Catley, 2013; Sandvik, 2008). The initial instructional approach of introducing basic characteristics of phylogenetic trees and then allowing students to make inferences on their own did not produce understanding. However, targeted training through active learning exercises for various aspects of phylogenetic trees, including taxa relatedness, had a sizable effect on student understanding. Second, phylogenetic trees are not to be taken lightly by introductory biology instructors. Time on task is an important factor that affects learning (Bransford *et al.*, 2000), and considering the significance of phylogenetic trees for understanding the central biological concept of evolution (Gregory, 2008; Omland *et al.*, 2008), significant class and out-of-class time should be devoted to these visual representations. Finally, feedback plays an important role in learning (Hattie and Timperley, 2007), and targeted feedback concerning phylogenetic trees, combined with iterative instruction, seemed to promote student understanding, at least to some extent.

The taxa relatedness question on the final exam review homework (nine weeks after the evolution unit exam) proved to be the first reliable marker of individual student understanding. Although the individual component of the evolution unit exam is unreliable for gauging student

understanding, we can assert that reasoning strategies used by students changed little during the long time lapse between the unit exam and the final exam review homework (Table 5). The only notable difference between the two coding distributions was the monophyletic grouping form of correct reasoning, which decreased from 22% to 8% of student responses. Using monophyletic groups to determine relatedness of taxa was a popular strategy for the evolution unit exam (22% and 30% of responses for the individual and group components, respectively), but its usage was not common for the initial homework, final exam review homework, or final exam. Four of the five taxa relatedness questions used during this investigation involved animals, so taxa selection for assessment does not appear to be a factor. The only notable difference between prompts was the multiple-choice format of the final exam review homework and final exam (Figures 6 and 7). This slight alteration is unlikely to have caused such a disparity in monophyletic grouping usage, however, so the reason for its popularity during the evolution unit exam remains unknown.

As the first reliable marker of individual understanding in regards to phylogenetic trees, the final exam review homework revealed that less than half (47%) of the students in the course provided a completely correct response to the taxa relatedness question. This outcome indicates very poor understanding, especially in light of students having access to class notes and virtually all other resources for the homework. Reasoning strategies differed somewhat on the final exam compared to the preceding review homework (Table 5), and the incidence of completely correct responses decreased from 47% to 38%, although the overall distributions of correctness (Table 6) only marginally altered (p=0.052). The moderate differences between the review homework and final exam could be due to a variety of factors. In contrast to the review homework and previous questions, the final exam prompt utilized a phylogenetic tree without synapomorphies (Figure 7). Students did not have access to resources during the final exam as they did while completing the

34

homework, and eleven students submitted a final exam who did not submit a review homework. All three factors probably contributed somewhat to the differences in reasoning and correctness distributions between the final exam review homework and final exam.

The taxa relatedness question on the comprehensive final exam (Figure 7) provided our investigation with its most demoralizing outcome. After broad initial instruction, more targeted instruction, ample feedback, and experiencing the same basic taxa relatedness question on four previous occasions, only 38% of the students provided a completely correct response to the taxa relatedness question on the high-stakes final exam. Yet, 70% of students referenced most recent common ancestry in their reasoning strategy. This result is both evidence for the difficulties that students have interpreting phylogenetic trees, and evidence for the rampant attempts of students to memorize information for an exam without true understanding.

Research Conclusions

The importance of understanding phylogenetic trees in biology is unquestionable, yet this study and others have demonstrated that students, and sometimes professionals, struggle mightily with interpreting these visual representations. Broad initial instruction appeared to do little good for phylogenetic tree understanding, as not a single group of students could accurately determine taxa relatedness on the initial homework. More targeted instruction on evolutionary relationships greatly improved understanding, but to a still unacceptable level. Only 57% of groups provided a completely correct response to the relatedness question on the group component of the evolution unit exam, and results were even worse on later attempts by individuals. Continued feedback and loss of points on assessments after the initial spike from targeted instruction appeared to do little in the way of promoting learning. The students in the course knew without a doubt that at least one phylogenetic tree would appear on the comprehensive final exam and had ample experience

working with them, yet only 38% could accurately answer a simple taxa relatedness question. It appears these common visual representations, which can directly affect student understanding of evolution (Gregory, 2008; Omland *et al.*, 2008), represent a formidable challenge for instructors.

Even though branch tip proximity, external insights, and contemporary descent forms of reasoning dominate the phylogenetic tree misinterpretation literature, our students demonstrated a strong preference for counting synapomorphies and nodes to determine taxa relatedness during this investigation. These very algorithmic forms of reasoning also proved to be the most resistant to correction, as branch tip proximity and external insights practically disappeared following the initial phylogenetic tree homework and targeted instruction, while counting synapomorphies and nodes persisted. An apparent conflict in the usefulness of synapomorphies on phylogenetic trees was also discovered for translation and interpretation exercises, and this dilemma remains to be resolved by further research. Additionally, students showed a definite propensity for memorizing responses without truly understanding the information, as 70% of students were aware that most recent common ancestry somehow determined taxa relatedness on phylogenetic trees, yet barely more than half of those students (38%) could accurately use that knowledge. This phenomenon occurred despite the active engagement, learner-centered, and non-lecture approach to teaching biology used by the instructor, which encouraged conceptual learning rather than memorization.

Aligning with our expectations of cooperative learning, groups outperformed individuals on phylogenetic tree tasks, although this outcome is based on limited data. Finally, our research further justifies the recommendations of Halverson *et al.* (2011) that multiple-choice assessment items are clearly insufficient for capturing student understanding of phylogenetic trees. Student responses to our two-part taxa relatedness questions often contained reasoning explanations that disagreed with answer selections. The answer selections of students alone provided us with little

information about student understanding, whereas the coupling of answer choices and reasoning used to defend the answer choices proved to be powerful. We also caution that phylogenetic tree assessments must be carefully constructed, regardless of assessment format, as demonstrated by the unreliable results from the question on the individual component of the evolution unit exam.

Research Limitations

All research has limitations, and this investigation presents no exception. All educational researchers must be careful not to extrapolate the results of their studies to situations outside the context of their research. The results of this investigation reflect the understanding of students in an introductory biology course at one institution, taught by an instructor with a learner-centered approach that heavily emphasized conceptual learning, active engagement (virtually no lectures), cooperative learning, and higher-order cognitive skills. Our results do not necessarily reflect the phylogenetic tree understanding of students with significantly different educational experiences.

This investigation is further limited by the nature of the assessment items used to collect data. We employed only cladograms (in which branch lengths have no meaning) that were drawn in a diagonal style and upward from left to right. It has been argued that students tend to struggle more with diagonal phylogenetic trees compared to the bracket style (Novick and Catley, 2007; 2013), and 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, it is entirely possible that students would have performed better on our tasks if we had simply used a different style and orientation of phylogenetic tree. Because the main purpose of phylogenetic trees is to illustrate evolutionary relationships, we also utilized 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 just as important. Additionally, each phylogenetic tree used for the investigation had a unique branching pattern (topology). Although topology would not affect the taxa relatedness interpretations of experts, this may not be the case for novices. These assessment item limitations are currently being investigated through additional studies (see next subsection).

Unlike nearly all previous research on student interpretations of phylogenetic trees, data for this investigation were collected in the context of a biology course, and as such, students had an academic stake in their responses. Authentic *in situ* data provide a powerful depiction of how students understand information in a real classroom setting, but such data also come with a price. The instructor must do what is best for student learning, which is not necessarily what is best for the investigation. Some of our data sets were collected from students working together on tasks, as cooperative learning has been shown to result in a whole host of positive effects on students (Johnson *et al.*, 1998). Mixing group and individual data is not ideal from a research standpoint, nor is collecting some data from homeworks and other data from exams for comparison, but our data necessarily reflected the reality of a learner-centered introductory biology course.

Finally, the statistics available to analyze our data are limited. The mixing of group and individual data, along with repeated and dependent measurements, eliminates the possibility of using most common statistical methods. The McNemar and Stuart-Maxwell tests provided some useful information concerning how student performances changed between assessments, but the remaining analysis is observational. We may at some point apply Bayesian statistical methods to this and future data, but such methods are beyond the scope of this thesis.

Future Research

This investigation provided preliminary information regarding how introductory biology students interpret phylogenetic trees. This foundational knowledge will drive future research on

phylogenetic tree education, primarily by addressing some of the research limitations. To address assessment item limitations, we will analyze data from introductory biology courses to determine if our students show different aptitudes with diagonal and bracket phylogenetic trees, as claimed by Novick and Catley (2007; 2013). The study will also investigate skills other than determining taxa relatedness, such as identifying monophyletic groups and utilizing parsimony to distinguish homology from convergent evolution. Through a second project, we will examine phylogenetic tree construction abilities of introductory biology students, and the relationship of those abilities to interpretation skills. The project will determine if phylogenetic tree construction is cognitively more complex than phylogenetic tree interpretation and necessary for "tree-thinking", as claimed by Halverson (2011). Preliminary analysis indicates that neither claim is true for our students. In the future, investigations could also be designed to examine whether phylogenetic tree topology has a significant effect on student interpretations. Data on various topologies could be gathered from a large number of biology students, but this idea would probably best be examined through split-plot assessments with different topologies randomly assigned to students within a course.

Finally, whenever our research is discussed with others, the first question asked is always the same: how should phylogenetic trees be taught? To date, the only pedagogical tools available are published in teaching journals and are primarily anecdotal. Our line of research and literature that has accumulated over the last decade will be synthesized to design, test, and disseminate the first evidence-based curriculum for phylogenetic trees. Considering the critical importance of the visual representations for thinking about evolution (Gregory, 2008; Meir *et al.*, 2007; Omland *et al.*, 2008), and the fundamental importance of evolution for understanding biology (Dobzhansky, 1964), such research could have a significant impact on biology education.

REFERENCES

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

Anderson, D., Fisher, K., and Norman, G. (2002). Development and Evaluation of the Concept Inventory of Natural Selection. Journal of Research in Science Teaching, 39(10), 952-978.

Armstrong, N., Chang, S., and Brickman, M. (2007). Cooperative Learning in Industrial-Sized Biology Classes. CBE: Life Sciences Education, 6, 163-171.

Baum, D.A., Smith, S.D., and Donovan, S.S.S. (2005). The Tree-Thinking Challenge. Science, 310, 979-980.

Baum, D.A. and Offner, S. (2008). Phylogenies and Tree-Thinking. American Biology Teacher, 70(4), 222-229.

Bishop, B.A. and Anderson, C.W. (1990). Student Conceptions of Natural Selection and Its Role in Evolution. Journal of Research in Science Teaching, 27, 415-427.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. (1956). Taxonomy of Educational Objectives Handbook 1: Cognitive Domain. New York: Longman.

Bransford, J.D., Brown, A.L., Cocking, R.R., and the National Research Council Committee on Developments in the Science of Learning (2000). How People Learn: Brain, Mind, Experience, and School. Washington, D.C.: National Academy Press.

Brewe, E. (2008). Modeling Theory Applied: Modeling Instruction in Introductory Physics. American Journal of Physics, 76(12), 1155-1160.

Catley, K.M. and Novick, L.R. (2008). Seeing the Wood for the Trees: Analysis of Evolutionary Diagrams in Biology Textbooks. BioScience, 58(10), 976-987.

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

Crowe, A., Dirks, C., and Wenderoth, M.P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to Enhance Student Learning in Biology. CBE: Life Sciences Education, 7, 368-381.

Darwin, C. (1837). Transmutation of Species. Notebook B, 36.

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, United Kingdom: Murray.

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

Dobzhansky, T. (1964). Biology, Molecular and Organismic. American Zoologist, 4(4), 443-452.

Eaton, T.T. (2009). Engaging Students and Evaluating Learning Progress Using Collaborative Exams in Introductory Courses. Journal of Geoscience Education, 57(2), 113-120.

Elby, A. (1999). Another Reason that Physics Students Learn by Rote. American Journal of Physics, 67(7), S52-S57.

Freeman, S., O'Connor, E., Parks, J., Cunningham, M., Hurley, D., Haak, D., Dirks, C., and Wenderoth, M.P. (2007). Prescribed Active Learning Increases Performance in Introductory Biology. CBE: Life Sciences Education, 6, 132-139.

Freeman, S. (2011). Biological Science (Fourth Edition). San Francisco: Benjamin Cummings.

Glaser, B. and Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine.

Gould, S.J. (1999). Branching Through a Wormhole. Natural History, 108(2), 24-27, 84-89.

Gregory, T.R. (2008). Understanding Evolutionary Trees. Evolution: Education and Outreach, 1, 121-137.

Halverson, K.L. (2011). Improving Tree-Thinking One Learnable Skill at a Time. Evolution: Education and Outreach, 4, 95-106.

Halverson, K.L., Pires, C.J., and Abell, S.K. (2011). Exploring the Complexity of Tree Thinking Expertise in an Undergraduate Systematics Course. Science Education, 95, 794-823.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S., and Wood, W. (2004). Scientific Teaching. Science, 304, 521-522.

Hattie, J. and Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81-112.

Hegarty, M., Carpenter, P.A., and Just, M. (1990). Diagrams in the Comprehension of Scientific Texts. In R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson (Editors), Handbook of Reading Research: Volume 2, 641-668. New York: Longman.

Herreid, C.F. (1994). Case Studies in Science: A Novel Method of Science Education. Journal of College Science Teaching, 23, 221-229.

Hestenes, D. (1987). Toward a Modeling Theory of Physics Instruction. American Journal of Physics, 55(5), 440-454.

Hmelo, C.E., Holton, D.L., and Kolodner, J.L. (2000). Designing to Learn About Complex Systems. Journal of the Learning Sciences, 9(3), 247-298.

Johnson, D.W., Johnson, R.T., and Smith, K. (1998). Cooperative Learning Returns to College: What Evidence Is There That It Works? Change, 30, 26-35.

Lamarck, J.B. (1809). Philosophie Zoologique. Paris, France.

Lee, V.R. (2010). Adaptations and Continuities in the Use and Design of Visual Representations in U.S. Middle School Science Textbooks. Int. Journal of Science Education, 32(8), 1099-1126.

Liu, L. and Hmelo-Silver, C.E. (2009). Promoting Complex Systems Learning Through the Use of Conceptual Representations in Hypermedia. Journal of Research in Science Teaching, 46(9), 1023-1040.

Lyman, F.T. (1981). The Responsive Classroom Discussion: The Inclusion of All Students. In A. Anderson (Editor), Mainstreaming Digest, 109-113. College Park: University of Maryland Press.

MacDonald, T. and Wiley, E. (2012). Communicating Phylogeny: Evolutionary Tree Diagrams in Museums. Evolution: Education and Outreach, 5, 14-28.

Maxwell, A.E. (1970). Comparing the Classification of Subjects by Two Independent Judges. British Journal of Psychiatry, 116, 651-655.

McNemar, Q. (1947). Notes on the Sampling Error of the Difference Between Correlated Proportions or Percentages. Psychometrika, 12, 153-157.

Meir, E., Perry, J., Herron, J.C., and Kingsolver, J. (2007). College Students' Misconceptions About Evolutionary Trees. American Biology Teacher, 69(7), e71-e76.

Meisel, R.P. (2010). Teaching Tree-Thinking to Undergraduate Biology Students. Evolution: Education and Outreach, 3, 621-628.

Momsen, J.L., Long, T., Wyse, S., and Ebert-May, D. (2010). Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. CBE: Life Sciences Education, 9, 435-440.

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., and Grosz, N. (2013). Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses. CBE: Life Sciences Education, 12, 239-249.

Moore, D.S. (2007). Basic Practice of Statistics (Fourth Edition). New York: W.H. Freeman.

Morabito, N.P., Catley, K.M., and Novick, L.R. (2010). Reasoning About Evolutionary History: Post-Secondary Students' Knowledge of Most Recent Common Ancestry and Homoplasy. The Journal of Biological Education, 44(4), 166-174.

Nehm, R.H. and Reilly, L. (2007). Biology Majors' Knowledge and Misconceptions of Natural Selection. BioScience, 57(3), 263-272.

Novick, L.R. and Catley, K.M. (2007). Understanding Phylogenies in Biology: The Influence of a Gestalt Perceptual Principle. Journal of Experimental Psychology: Applied, 13(4), 197-223.

Novick, L.R., Catley, K.M., and Funk, D. (2010). Characters are Key: Effect of Synapomorphies on Cladogram Comprehension. Evolution: Education and Outreach, 3, 539-547.

Novick, L.R., Shade, C.K., and Catley, K.M. (2011). Linear Versus Branching Depictions of Evolutionary History: Implications for Diagram Design. Cognitive Science, 3, 536-559.

Novick, L.R., Stull, A.T., and Catley, K.M. (2012). Reading Phylogenetic Trees: The Effects of Tree Orientation and Text Processing on Comprehension. BioScience, 62(8), 757-764.

Novick, L.R. and Catley, K.M. (2013). Reasoning About Evolution's Grand Patterns: College Students' Understanding of the Tree of Life. Am. Education Research Journal, 50(1), 138-177.

Omland, K.E., Cook, L.G., and Crisp, M.D. (2008). Tree Thinking for All Biology: The Problem with Reading Phylogenies as Ladders of Progress. BioEssays, 30, 854-867.

Perry, J., Meir, E., Herron, J., Maruca, S., and Stal, D. (2008). Evaluating Two Approaches to Helping College Students Understand Evolutionary Trees Through Diagramming Tasks. CBE: Life Sciences Education, 7, 193-201.

Pungente, M.D. and Badger, R.A. (2003). Teaching Introductory Organic Chemistry: Blooming Beyond a Simple Taxonomy. Journal of Chemical Education, 80(7), 779-784.

Sandvik, H. (2008). Tree Thinking Cannot Be Taken for Granted: Challenges for Teaching Phylogenetics. Theory in Biosciences, 127, 45-51.

Smith, K.A. (2000). Going Deeper: Formal Small-Group Learning in Large Classes. New Directions for Teaching and Learning, 81, 25-46.

Stuart, A. (1955). A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. Biometrika, 42, 412-416.

Sun, X. and Yang, Z. (2008). Generalized McNemar's Test for Homogeneity of the Marginal Distributions. Proceedings of the SAS Global Forum 2008. Cary, NC: SAS Institute Inc.

Sundberg, M.D. (2002). Assessing Student Learning. Cell Biology Education, 1, 11-15.

Tanner, K., Chatman, L.S., and Allen, D. (2003). Approaches to Cell Biology Teaching: Cooperative Learning in the Science Classroom. Cell Biology Education, 2, 1-5.

Tomanek, D. and Montplaisir, L. (2004). Students' Studying and Approaches to Learning in Introductory Biology. Cell Biology Education, 3, 253-262.

APPENDIX

Table A1. Selected student-generated examples of monophyletic grouping reasoning

Data Source	Student-Generated Example
Evolution Unit Exam Individuals (Figure 4)	Sea lions. When you circle the smallest monophyletic group with the bear, sea lion is included but not the cat.
Evolution Unit Exam Individuals (Figure 4)	Sea lions. The most recent common ancestor between the bears and cat has all species as descendants but the descendants of the common ancestor of bears and sea lions are only the bear, sea lion, and seal.
Evolution Unit Exam Individuals (Figure 4)	The bear is more closely related to a sea lion because they share the most recent common ancestor that I have labeled as B [*most recent common ancestry*]. If you perform a one-snip test at B, the bear and sea lion fall off while the cat would not [*monophyletic grouping*].
Evolution Unit Exam Individuals (Figure 4)	Sea lions. If a one-snip test was performed, the bear and sea lion are part of the same monophyletic group. Even though the cat would be part of a monophyletic group also, it would then be more of a broad group including all the animals.
Evolution Unit Exam Groups (Figure 5)	Durum wheat. Because the monophyletic group that contains durum wheat and lady fern is smaller than the group that contains liverwort [*monophyletic grouping*]. The monophyletic group that includes the durum wheat has lady fern as its ancestor [*contemporary descent*].
Evolution Unit Exam Groups (Figure 5)	Durum wheat is our answer. Lady ferns and durum wheat share the more recent common ancestor [*most recent common ancestry*]. Also, if you did a one-snip test at their most recent common ancestor, lady ferns and durum wheat fall off together [*monophyletic grouping*].
Review Homework Individuals (Figure 6)	Lizards are equally related to both of them because the most recent common ancestor between lizards and either birds or crocodiles is at the node I circled [*most recent common ancestry*]. The monophyletic group [referring to the node mentioned in the previous line] contains both birds and crocodiles [*monophyletic grouping*].
Review Homework Individuals (Figure 6)	Lizards are more closely related to crocodiles [incorrect] because if you did a one-snip test at the node before the birds, the crocodile will be in the same monophyletic group as the lizard.
Final Exam Individuals (Figure 7)	Seals are equally related to both horses and whales because the most recent common ancestor between them is the same spot [*most recent common ancestry*]. One could even perform a one-snip test to show they are in the same monophyletic group [*monophyletic grouping*].
[*italics*] indicates reasoning code and [text] indicates incorrect answer or response clarification	

Table A2. Selected student-generated examples of counting synapomorphies reasoning

Data Source	Student-Generated Example
Initial Homework Groups (Figure 3)	The fire salamander [incorrect]. Nearness of the branches [*branch tip proximity*] and fewer branch/nodes between alligator and salamander [*counting nodes*], and the lesser amount of synapomorphies needed to connect these two species [*counting synapomorphies*] is why.
Initial Homework Groups (Figure 3)	Fire salamander [incorrect] because there is only two different traits between the salamander and alligator, claws or nails and gizzard, but there is three different traits between an alligator and kangaroo.
Initial Homework Groups (Figure 3)	Fire salamander [incorrect] because they share having cranium, jaw, bone skeleton, and lungs. The alligator differs only in having claws and a gizzard, whereas the kangaroo differs in three characteristics.
Evolution Unit Exam Individuals (Figure 4)	Sea lions. To get to a common ancestor of bears and cats, there are 3 divergence points and 2 differing labeled traits. To get to a common ancestor of bears and sea lions, there is 1 divergence point and only 1 labeled trait [*counting nodes and counting synapomorphies*].
Evolution Unit Exam Groups (Figure 5)	Liverworts [incorrect]. Because there are two ancestors between just like the durum wheat [*counting nodes*], but durum wheat has seeds and flowers which lady ferns do not have. The only thing lady ferns have that liverworts do not is stomata [*counting synapomorphies*].
Review Homework Individuals (Figure 6)	Lizards are more closely related to crocodiles [incorrect]. Only one mentionable difference between each other, and birds have two.
Review Homework Individuals (Figure 6)	Lizards are more closely related to crocodiles than birds [incorrect]. From going to lizards and crocodiles, there is only development of the gizzard. But when going to a bird, there is another development of feathers, which differentiates it from the croc and lizards.
Review Homework Individuals (Figure 6)	Lizards are more closely related to crocodiles than birds [incorrect]. All 3 organisms have amniotic eggs, the only difference between the lizards and crocs are gizzards. Lizards are not closely related to birds because birds have both gizzards and feathers, and lizards do not.
Review Homework Individuals (Figure 6)	Lizards are more closely related to crocodiles [incorrect] because the birds are avian and crocs, alligators, and lizards are reptiles [*external insights*]. Crocs have gizzards, but they are still more closely related to lizards because they have no feathers [*counting synapomorphies*].
Final Exam Individuals (Figure 7)	Seals are equally related to both horses and whales because there are no detectable trait changes that occurred on the phylogenetic tree.
[*italics*] indicates reasoning code and [text] indicates incorrect answer or response clarification	

Table A3. Selected student-generated examples of negation reasoning

Data Source	Student-Generated Example
Evolution Unit Exam Individuals (Figure 4)	Sea lions. You cannot read a tree by its tips [*negation reasoning*] and you need to look and find out when the most recent common ancestor comes about. The most recent common ancestor for the sea lion and bear comes about more recent than the most recent common ancestor of the bear and cat, meaning the bear must be more closely related to the sea lion [*most recent common ancestry*].
Evolution Unit Exam Individuals (Figure 4)	I think sea lions because looking at the tree you can find the common ancestor between a bear and sea lion more recently than you find the common ancestor between the bear and the cat [*most recent common ancestry*]. It does not matter what order they are in on the tips of the phylogenetic tree [*negation reasoning*].
Evolution Unit Exam Individuals (Figure 4)	Sea lions. Bears and sea lions share a more recent common ancestor [*most recent common ancestry*]. You are never supposed to pay any attention to the top of the phylogenetic tree, because you can rotate the branches [*negation reasoning*].
Review Homework Individuals (Figure 6)	Lizards are equally related to crocodiles and birds. They all share the same most recent common ancestor [*most recent common ancestry*], and you can spin the birds and crocs so they are both equally distant from the lizards [*negation reasoning*].
Final Exam Individuals (Figure 7)	The seal, whale, and horse are equally related because they all share the same initial ancestor (labeled A) [*most recent common ancestry*] and distance between them does not matter [*negation reasoning*].
[*italics*] indicates reasoning code and [text] indicates incorrect answer or response clarification	