

Further Effects of Phylogenetic Tree Style on Student Comprehension in an Introductory Biology Course

Jonathan Dees,[†] Caitlin Bussard,[‡] and Jennifer L. Momsen^{§*}

[†]Department of Plant Biology, University of Georgia, Athens, GA 30602; [‡]College of Dentistry and Dental Clinics, University of Iowa, Iowa City, IA 52242; [§]Department of Biological Sciences, North Dakota State University, Fargo, ND 58108

ABSTRACT

Phylogenetic trees have become increasingly important across the life sciences, and as a result, learning to interpret and reason from these diagrams is now an essential component of biology education. Unfortunately, students often struggle to understand phylogenetic trees. Style (i.e., diagonal or bracket) is one factor that has been observed to impact how students interpret phylogenetic trees, and one goal of this research was to investigate these style effects across an introductory biology course. In addition, we investigated the impact of instruction that integrated diagonal and bracket phylogenetic trees equally. Before instruction, students were significantly more accurate with the bracket style for a variety of interpretation and construction tasks. After instruction, however, students were significantly more accurate only for construction tasks and interpretations involving taxa relatedness when using the bracket style. Thus, instruction that used both styles equally mitigated some, but not all, style effects. These results inform the development of research-based instruction that best supports student understanding of phylogenetic trees.

INTRODUCTION

Phylogenetic trees are powerful tools that facilitate thinking about biological phenomena from an evolutionary perspective (“tree thinking”; O’Hara, 1988; Gregory, 2008). Although phylogenetic trees are often viewed simply as visual representations of hypothesized evolutionary relationships among taxa, these diagrams are also the main analytical tool used by biologists to assess evidence of evolution (Baum *et al.*, 2005; Novick and Catley, 2007). Further, phylogenetic trees provide an efficient framework to organize our growing knowledge of biological diversity (Thanukos, 2009; Wiley, 2010; Baum and Smith, 2013). Because evolution is a unifying theory and core concept in biology (Dobzhansky, 1973; American Association for the Advancement of Science, 2011; Next Generation Science Standards Lead States, 2013; College Board, 2015), and due to advancements in phylogenetic inference and DNA-sequencing technologies (Omland *et al.*, 2008), phylogenetic trees have become increasingly important across the life sciences (Baum and Offner, 2008). As a result, learning to interpret and reason from phylogenetic trees is now an essential component of biology education (O’Hara, 1997; Lents *et al.*, 2010; Meisel, 2010; Novick and Catley, 2016).

Despite the significance of phylogenetic trees, students at all levels routinely struggle to interpret these diagrams (Meir *et al.*, 2007; Halverson *et al.*, 2011; Catley *et al.*, 2013; Novick and Catley, 2013; Blacchiere and Hoese, 2016), even after explicit instruction (Phillips *et al.*, 2012; Smith *et al.*, 2013; Dees *et al.*, 2014). Student difficulties with phylogenetic trees have been attributed to a number of factors, starting with abstractness. As a type of schematic diagram, phylogenetic trees present abstract information that requires learned rules and conventions for correct interpretation (Novick and Catley, 2007). In other words, understanding phylogenetic trees is not intuitive, and students must be taught how to extract information from these diagrams (Sandvik,

James J. Smith, *Monitoring Editor*

Submitted April 20, 2017; Revised January 2, 2018; Accepted January 9, 2018

CBE Life Sci Educ June 1, 2018 17:ar17

DOI:10.1187/cbe.17-03-0058

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

© 2018 J. Dees *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>). “ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

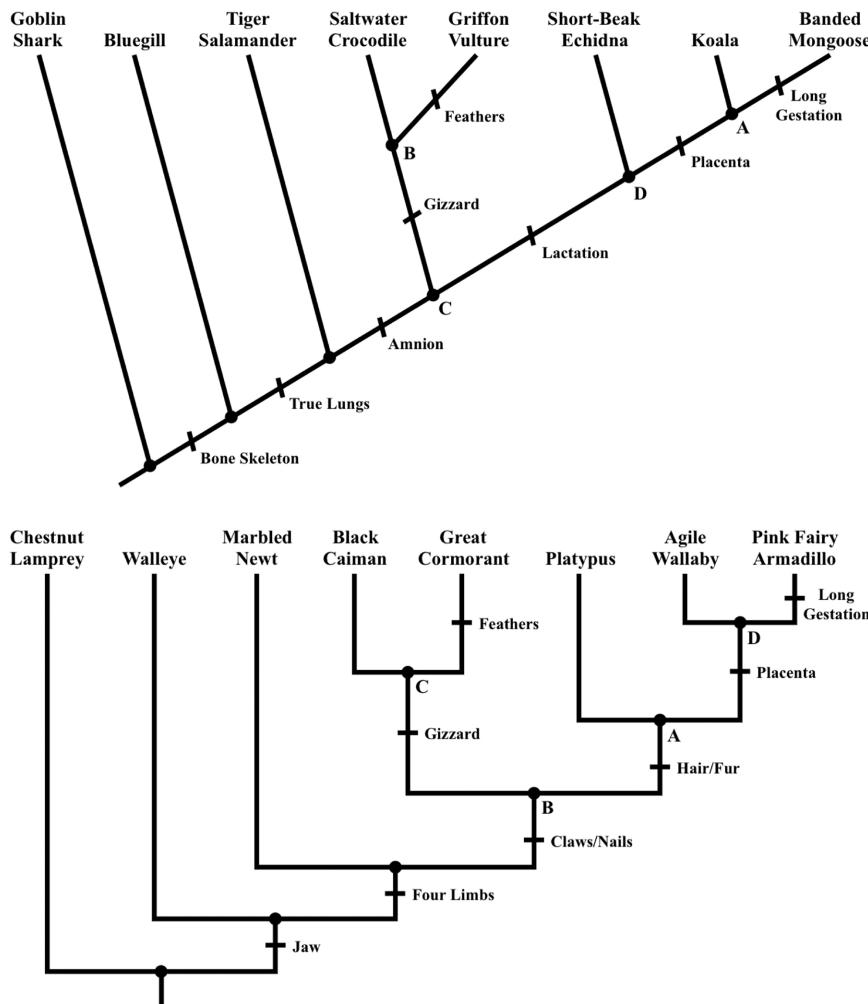


FIGURE 1. Equivalent diagonal (top) and bracket (bottom) phylogenetic trees that are the same size and have the same branch pattern but involve different taxa and some different traits.

2008; Eddy *et al.*, 2013). The gestalt perceptual principles of good continuation and spatial proximity have also been shown to negatively impact students, especially for phylogenetic trees drawn in the diagonal style (Figure 1) and when interpreting taxa relatedness (Novick and Catley, 2007, 2013). Finally, student interpretations of phylogenetic trees and student conceptions of evolution are interrelated, such that each affects the other (Gregory, 2008; Omland *et al.*, 2008). Thus, misinterpretations of phylogenetic trees impede student understanding of evolution (Meir *et al.*, 2007), and conversely, misconceptions about evolution also lead to student difficulties with phylogenetic trees.

One factor that can be controlled by instructors and that has been observed to influence how students interpret phylogenetic trees is style (Baum and Offner, 2008; Halverson *et al.*, 2011). Two styles of phylogenetic tree that contain equivalent information, diagonal and bracket, commonly appear in textbooks, journals, and other resources (Figure 1; Catley and Novick, 2008). However, to our knowledge, only three studies have explicitly examined effects of style on student understanding of

phylogenetic trees (Novick and Catley, 2007, 2013; Dees *et al.*, 2017). An initial study by Novick and Catley (2007) used translation tasks to detect differences in how students perceived diagonal and bracket phylogenetic trees. Students were asked to convert visual representations of evolution, including diagonal and bracket phylogenetic trees, from one representation to another while maintaining the same evolutionary relationships among taxa. Accuracy was significantly lower for translations involving diagonal phylogenetic trees, and this style effect was more pronounced for students with less experience in biology.

In a later study, Novick and Catley (2013) used a suite of interpretation tasks to further examine effects of style on student understanding of phylogenetic trees. For example, students were asked to evaluate taxa relatedness, recognize monophyletic and nonmonophyletic groups, and identify traits shared by taxa due to common ancestry. Across nearly all tasks, accuracy was significantly lower when students interpreted diagonal phylogenetic trees, and this style effect was often found, regardless of background in biology. Finally, Dees *et al.* (2017) examined effects of style on student interpretations and construction of phylogenetic trees by collecting data in the context of an introductory biology course. Before instruction on phylogenetic trees, students were asked to complete a number of interpretation tasks for both styles that were similar to those used by Novick and Catley (2013). Students also constructed a phylogenetic tree in the style of their choice from data provided to them.

For most interpretation tasks, accuracy was again significantly lower for diagonal phylogenetic trees. Students who constructed diagonal phylogenetic trees were also significantly less accurate compared with those who used the bracket style for the construction task.

Although three previous studies provided multiple lines of evidence indicating students had more difficulties with diagonal phylogenetic trees compared with the bracket style, each investigation had important limitations that warrant further research. Novick and Catley (2007, 2013) used surveys to collect data from students who were mostly recruited as volunteers from psychology, education, and biology courses. From a motivational perspective, students may not take surveys as seriously as course work that affects their academic standing (Sundberg, 2002). In addition, neither study included construction tasks, which are common instructional activities for phylogenetic trees (e.g., Gendron, 2000; Goldsmith, 2003; Julius and Schoenfuss, 2006; Burks and Boles, 2007; Lents *et al.*, 2010; Eddy *et al.*, 2013; Bokor *et al.*, 2014; Lampert and Mook, 2015). Finally, both studies were conducted by the same researchers at

the same institution, which limits the robustness of the claims due to potential experimenter bias (Makel and Plucker, 2014).

Dees *et al.* (2017) addressed some of these limitations by obtaining data through course work, by examining both interpretations and construction of phylogenetic trees, and by providing data from another institution. However, data from Dees *et al.* (2017) were collected only from introductory biology students before instruction on phylogenetic trees. Further, students were asked to construct one phylogenetic tree in the style of their choice, resulting in a between-student comparison of construction accuracy for diagonal and bracket phylogenetic trees rather than a stronger within-student comparison.

The most notable limitation of the three preceding studies is that none of them directly investigated the impact of instruction on mitigating style effects. The majority of introductory textbooks now use bracket phylogenetic trees exclusively, and it is probable that students, if they have any experience with phylogenetic trees at all, have had less experience with the diagonal style. As a result, style effects observed in the literature may simply reflect student exposure to one style of phylogenetic tree over another. If so, we might expect such style effects to decrease or disappear following instruction that incorporated both styles of phylogenetic tree equally. Thus, the goal of the present study was to further explore style effects by gathering data that addressed the limitations of previous studies by satisfying the following criteria: they 1) were obtained through course work in biology, 2) included interpretations and construction of phylogenetic trees, 3) supported within-student comparisons of performance across styles, and 4) were collected before, after, and long after unbiased instruction that integrated diagonal and bracket phylogenetic trees equally. These data allowed us to address the following research questions:

1. Do introductory biology students demonstrate differential interpretation abilities for diagonal and bracket phylogenetic trees before, after, and long after instruction?
2. Do introductory biology students demonstrate differential construction abilities for diagonal and bracket phylogenetic trees before, after, and long after instruction?

METHODS

Data for this study were collected in the context of an introductory biology course for science and related majors at a large, public university in the midwestern United States. There were no prerequisites for enrollment, and the course served students ($n = 83$) at various stages in their academic programs (30% freshmen, 41% sophomores, 18% juniors, and 11% seniors). Content started with inheritance (weeks 1–3) and progressed through evolution and biodiversity (weeks 4–8), form and function of plants and animals (weeks 9–12), and ecology (weeks 13–15). Instruction on phylogenetic trees occurred toward the end of the evolution and biodiversity unit during the seventh week of class. Although phylogenetic trees appeared in later textbook chapters (e.g., animal physiology), students were not asked to interact with phylogenetic trees during subsequent instruction or assessments until the last week of the course, when students completed a series of review activities during class to prepare for the comprehensive final exam (week 17).

Instruction was learner-centered and emphasized collaboration (Johnson *et al.*, 1998; Tanner *et al.*, 2003; Armstrong *et al.*, 2007) by having assigned groups of three or four students build and evaluate conceptual models (Dauer *et al.*, 2013; Bray Speth *et al.*, 2014; Long *et al.*, 2014), discuss clicker questions (Caldwell, 2007; Freeman *et al.*, 2007; Perez *et al.*, 2010), and construct scientific arguments (Driver *et al.*, 2000). Classes were observed, and instructional materials and assessments were collected to document instruction throughout the course.

Instrument Design

We developed a series of four instruments to measure effects of phylogenetic tree style on student comprehension before (pre-instructional homework), after (post-instructional homework and unit exam), and long after instruction (review activity for the final exam). Each instrument contained a diagonal phylogenetic tree and an equivalent bracket phylogenetic tree that were the same size and had the same branch pattern but involved different taxa and traits (see Figure 1). Isomorphic interpretation tasks accompanied each diagonal and bracket phylogenetic tree such that accuracy could be compared across styles. These interpretation tasks were modified from a previous study (Dees *et al.*, 2017) and based largely on the essential tree-thinking skills proposed by Novick and Catley (2013). Specifically, students were asked to identify the most recent common ancestor of taxa, recognize monophyletic groups, determine whether extant taxa are descended from other extant taxa (“contemporary descent”; Dees *et al.*, 2014), and evaluate taxa relatedness.

Students were also asked to construct phylogenetic trees from provided data, either in a specified style or in the style of their choice. The instruments that were assigned as homework included two construction tasks, one for each style, which resulted in equivalent phylogenetic trees. Because the unit exam and review activity for the final exam were completed during class and subject to time constraints, these instruments contained a single construction task that allowed students to use the style of their choice. To reduce context effects, in which student reasoning about evolution varies for different taxa and traits (Nehm and Ha, 2011), phylogenetic trees used for interpretation tasks exclusively involved animals (e.g., Figure 1), while all construction tasks involved plants. The four instruments used for this investigation are available in the Supplemental Material.

Data Collection

To reduce style bias, it was essential that any resource used by students introduce both phylogenetic tree styles concurrently. As a result, students were never asked to read about phylogenetic trees in the textbook for the course (Urry *et al.*, 2014), which only used the bracket style. Instead, before formal classroom instruction on phylogenetic trees during the seventh week of class, students were asked to watch a short screencast (just under 2 minutes) that was posted on the course management system. Notably, the screencast showed a diagonal phylogenetic tree and an equivalent bracket phylogenetic tree side by side, such that one style was not introduced first or favored over the other style. Similar to a broad textbook introduction, the screencast simply described the purpose of phylogenetic trees and defined a few essential terms (e.g., nodes and branches),

without explaining how to interpret or construct the diagrams. After the screencast was posted, each student was randomly assigned either the diagonal or bracket section of the pre-instructional homework. Once students submitted the first homework, they were assigned the opposite section of the homework. This distribution method was used to control for order effects, in which student responses are impacted by the sequence of assessment items (Halverson *et al.*, 2013; Federer *et al.*, 2015).

Instruction on phylogenetic trees began shortly after both the diagonal and bracket sections of the pre-instructional homework were submitted by students. Similar to the earlier screencast, students were shown a pair of equivalent diagonal and bracket phylogenetic trees in a side-by-side manner during initial instruction. Subsequent instructional activities for interpretations involved one style or the other, but overall, an equal number of diagonal and bracket phylogenetic trees were used by the instructor. When instructional activities included construction tasks, students were allowed to use the style of their choice. Verification feedback (i.e., labeling responses as correct or incorrect; Marsh *et al.*, 2012) was provided for the pre-instructional homework and submitted instructional activities. The post-instructional homework was distributed to students using the same method as the pre-instructional homework, and verification feedback was provided before the unit exam.

One week after instruction on phylogenetic trees, students completed a unit exam during class that assessed understanding of speciation, biodiversity, and phylogenetic trees. The section of the unit exam devoted to phylogenetic trees was structured the same as the instruments that were deployed as homework, except only one construction task was included due to time constraints. To control for order effects, each student received one of two versions of the unit exam, which varied only in the sequence of assessment items. A diagonal phylogenetic tree and associated interpretation tasks preceded an equivalent bracket phylogenetic tree and associated interpretation tasks in version A, while the order was reversed in version B. The single construction task, which allowed students to use the style of their choice, appeared after the two sets of interpretation tasks in both versions of the unit exam. Answer feedback (i.e., providing correct answers without explaining why answers are correct or incorrect; Marsh *et al.*, 2012) was given to students 1 week after the unit exam in the form of a grading rubric that was posted on the course management system.

Finally, during the last week of class and 8 weeks after the unit exam, students participated in various review activities to prepare for the comprehensive final exam. To investigate style effects long after instruction on phylogenetic trees, data had to be collected without students preparing in advance. Thus, the last instrument was deployed as one of the review activities rather than as part of the final exam. The instrument was structured the same as the section of the unit exam that was devoted to phylogenetic trees. Two versions of the instrument that varied only in the sequence of assessment items were also created and distributed in the same manner as the unit exam to control for order effects. Students completed the review activity during class without access to resources, which concluded data collection for this investigation. Although phylogenetic trees also appeared on the final exam, the associated assessment items were not designed for this study.

Data Coding

Student responses to interpretation and construction tasks were coded using the methods outlined in an earlier investigation (Dees *et al.*, 2017). Tasks that involved identifying the most recent common ancestor of taxa required a multiple-choice answer, and responses were coded as correct or incorrect. Tasks that involved recognizing a monophyletic group had multiple correct answers, and responses were again coded as correct or incorrect. Tasks that involved determining whether extant taxa are descended from other extant taxa (“contemporary descent”; Dees *et al.*, 2014) required a “yes” or “no” answer with reasoning. Answers and reasoning were each coded as correct or incorrect, wherein correct reasoning stated or implied that extant taxa evolved from a common ancestor rather than one evolving from the other. Tasks that involved evaluating taxa relatedness required a multiple-choice answer with reasoning. Answers were coded as correct or incorrect, while a published rubric was used to code student reasoning as correct, incorrect, or mixed (Dees *et al.*, 2014). Correct reasoning cited most recent common ancestry or monophyletic groups as criteria for determining taxa relatedness, while incorrect reasoning typically referred to the number of nodes or traits between taxa, relative distance between taxa, or information that was not provided by phylogenetic trees. Students often included multiple forms of reasoning in their responses, and in some cases used mixed reasoning that contained both correct and incorrect criteria for evaluating taxa relatedness.

Student responses to construction tasks were coded for accuracy as correct, adequate, or incorrect using a published rubric (Dees and Momsen, 2016). Phylogenetic trees that included one or more major errors, such as incorrect relatedness and incorrect traits, were considered incorrect. Student responses that included only minor errors, such as extra nodes and empty branches, were coded as adequate. Major and minor errors were differentiated based on whether or not the errors impeded students from interpreting taxa relatedness or traits possessed by taxa. Finally, phylogenetic trees with no major or minor errors were considered correct. All student responses to interpretation and construction tasks that were collected for this investigation were coded by two independent raters with greater than 94% agreement (kappa coefficient greater than 0.86; Cohen, 1960).

Statistical Analyses

For each instrument, we analyzed student responses to isomorphic interpretation tasks associated with equivalent diagonal and bracket phylogenetic trees as paired, categorical data. In the case of dichotomous categories (e.g., correct or incorrect), we used an exact version of the McNemar test, which is suitable for small sample sizes, accounts for the paired nature of our data, and generates within-student comparisons of performance across styles (McNemar, 1947; Rufibach, 2011). For style effects, the null hypothesis of the McNemar test is that an equal number of students switched categories in one direction (e.g., incorrect to correct) as in the opposite direction from one style of phylogenetic tree to the other style (McDonald, 2014). In the case of trichotomous categories (e.g., correct, incorrect, or mixed), we used the Stuart-Maxwell extension of the McNemar test (Stuart, 1955; Maxwell, 1970; Sun and Yang, 2008). Order effects within

each instrument and changes in student performance between instruments were investigated using the same statistics but different variables of interest (e.g., instrument as the variable rather than style of phylogenetic tree).

Student responses to construction tasks were analyzed in the same manner as interpretation tasks, with the exception of data from the unit exam and review activity for the final exam. Due to time constraints, these two instruments included one construction task that allowed students to use the style of their choice rather than a construction task for each style of phylogenetic tree. Therefore, student responses to these construction tasks had to be analyzed as unpaired, categorical data. We used the Fisher exact test, which is suitable for small sample sizes and generates between-student comparisons of performance across styles. In this situation, the null hypothesis of the Fisher exact test is that accuracy was independent of the style used by students to construct phylogenetic trees (Fisher, 1934). Finally, we used the exact binomial test to determine whether students chose either style significantly more than the other style for construction tasks on the unit exam and review activity for the final exam. For this scenario, the null hypothesis of the exact binomial test is that students constructed an equal number of diagonal and bracket phylogenetic trees (McDonald, 2014).

RESULTS

Data were collected through a pre-instructional homework ($n = 74$), a post-instructional homework ($n = 75$), a unit exam ($n = 81$), and a review activity for the final exam ($n = 72$). Some students elected not to submit their homework or attend class when the review activity was completed, resulting in smaller sample sizes compared with the unit exam. In addition, two students withdrew from the course ($n = 83$) after the pre-instructional homework and before the unit exam. No order effects were observed for any task on any instrument (i.e., whether students received tasks for diagonal or bracket phylogenetic trees first did not significantly impact accuracy; all $p > 0.26$). Accuracy increased significantly from the pre-instructional homework to the post-instructional homework for all interpretation and construction tasks across both styles (all $p < 0.04$). Further, accuracy did not change significantly from the post-instructional homework to the unit exam and final exam review activity for any interpretation or construction task across both styles (all $p > 0.12$).

Interpretations

On the pre-instructional homework, students were significantly more accurate when interpreting bracket phylogenetic trees for three tasks: identifying the most recent common ancestor of taxa, recognizing monophyletic groups, and determining whether extant taxa are descended from other extant taxa (“contemporary descent”; Table 1). These significant differences in accuracy were no longer detected after instruction that balanced the use of diagonal and bracket phylogenetic trees and did not re-emerge during the unit exam or final exam review activity. For interpretations concerning contemporary descent, students were asked to provide reasoning for their answers. Although students’ answers were consistently more accurate than their reasoning, the patterns of answers and reasoning were similar when comparing diagonal and bracket phylogenetic trees across all four instruments.

In contrast, there was no significant difference in accuracy for evaluating taxa relatedness on the pre-instructional homework, although accuracy was extremely low for both diagonal and bracket phylogenetic trees (Table 1). However, students were significantly more accurate when evaluating taxa relatedness on bracket phylogenetic trees following instruction, and this difference persisted through the unit exam and final exam review activity. Interpretations concerning taxa relatedness required students to provide reasoning for their answers, and the patterns of student answers and reasoning were similar across all four instruments. Specific forms of reasoning used by students to evaluate taxa relatedness are available in Supplemental Table S1. Note that students seemed to abandon the idea of using branch tip proximity to determine taxa relatedness rather quickly after instruction, whereas counting nodes or synapomorphies remained common misinterpretations throughout the course for both styles of phylogenetic tree.

Construction

Across all four instruments, before and after classroom instruction, students were significantly more accurate when constructing bracket phylogenetic trees (Figure 2). However, this difference in accuracy disappeared for each instrument when adequate phylogenetic trees were considered correct, indicating the adequate category was responsible for the discrepancy between styles. In other words, diagonal phylogenetic trees contained far more minor errors, but the occurrence of major errors was similar across styles. Specific major and minor errors found in phylogenetic trees constructed by students are available in Supplemental Table S2. In addition, as with all tasks in this investigation, accuracy increased significantly from the pre-instructional homework to the post-instructional homework for construction tasks. Improvement was driven mostly by students switching from incorrect to adequate phylogenetic trees when using the diagonal style, whereas improvement was driven mostly by students switching from incorrect to correct phylogenetic trees when using the bracket style. Finally, note that students constructed a single phylogenetic tree in the style of their choice during the unit exam and final exam review activity, and students overwhelmingly chose the diagonal style for both instruments (79 and 78% diagonal phylogenetic trees, respectively; $p < 0.001$ vs. an equal distribution of diagonal and bracket phylogenetic trees for each instrument).

DISCUSSION

Building from previous studies, we investigated effects of style on student interpretations and construction of phylogenetic trees in the context of an introductory biology course for science and related majors. In contrast to prior research, this study supported within-student comparisons of performance across styles and included data collected from course materials before, after, and long after unbiased instruction that integrated diagonal and bracket phylogenetic trees equally. These data allowed us to explore the interplay of instruction and representation style on student interpretations and construction of phylogenetic trees for the first time. Our results indicate such instruction eliminated some, but not all, style effects that favored the bracket style, which suggests diagonal phylogenetic trees may not be suitable for introductory-level biology courses.

TABLE 1. Correct student responses for all interpretation tasks and instruments with comparisons of accuracy across phylogenetic tree styles

Style	Pre-HW (<i>n</i> = 74)	Post-HW (<i>n</i> = 75)	Unit exam (<i>n</i> = 81)	Final review (<i>n</i> = 72)
Most recent common ancestor				
Diagonal	73%	95%	93%	92%
Bracket	86%	95%	98%	94%
Comparison	<i>p</i> = 0.02	<i>p</i> = 1.00	<i>p</i> = 0.22	<i>p</i> = 0.75
Monophyletic group				
Diagonal	54%	88%	93%	92%
Bracket	68%	91%	96%	93%
Comparison	<i>p</i> = 0.04	<i>p</i> = 0.69	<i>p</i> = 0.38	<i>p</i> = 1.00
Contemporary descent: answer				
Diagonal	73%	97%	95%	94%
Bracket	89%	100%	99%	97%
Comparison	<i>p</i> < 0.01	<i>p</i> = 0.50	<i>p</i> = 0.38	<i>p</i> = 0.50
Contemporary descent: reasoning				
Diagonal	53%	81%	75%	74%
Bracket	72%	85%	80%	78%
Comparison	<i>p</i> < 0.01	<i>p</i> = 0.51	<i>p</i> = 0.39	<i>p</i> = 0.55
Taxa relatedness: answer				
Diagonal	11%	39%	49%	46%
Bracket	15%	55%	59%	60%
Comparison	<i>p</i> = 0.58	<i>p</i> = 0.02	<i>p</i> = 0.04	<i>p</i> < 0.01
Taxa relatedness: reasoning				
Diagonal	5% ^a	36% ^a	42% ^a	40% ^a
Bracket	8% ^a	53% ^a	58% ^a	50% ^a
Comparison	<i>p</i> = 0.55 ^b	<i>p</i> = 0.02 ^b	<i>p</i> < 0.01 ^b	<i>p</i> = 0.03 ^b

^aMixed reasoning was also found in <10% of student responses.

^b*p* values were derived from a Stuart-Maxwell test due to trichotomous categories (correct, incorrect, or mixed reasoning). All other *p* values were derived from an exact version of the McNemar test.

Interpretations

Before classroom instruction on phylogenetic trees, students were significantly more accurate with the bracket style for most interpretation tasks, including identifying the most recent common ancestor of taxa, recognizing monophyletic groups, and determining whether extant taxa are descended from other extant taxa (“contemporary descent”). These differences in accuracy were mitigated by instruction that balanced the use of diagonal and bracket phylogenetic trees and did not re-emerge during the course.

In contrast, student interpretations of taxa relatedness on phylogenetic trees exhibited a different pattern. Before instruction, there was no significant difference in accuracy across styles due to a floor effect. The vast majority of students simply did not know how to evaluate taxa relatedness, and thus the style of phylogenetic tree did not impact student responses. Following instruction, however, students were significantly more accurate when evaluating taxa relatedness on bracket phylogenetic trees across all three post-instructional instruments. This difference included both answers and reasoning, as students used somewhat different forms of reasoning for each style of phylogenetic tree (Supplemental Table S1). However, accuracy for evaluating taxa relatedness was quite low for both styles even after instruction, which aligns with previous studies on

student understanding of taxa relatedness (Phillips *et al.*, 2012; Smith *et al.*, 2013; Dees *et al.*, 2014). In addition, student reasoning when evaluating taxa relatedness changed over time. Students quickly abandoned using branch tip proximity to determine taxa relatedness, but counting nodes and synapomorphies remained common misinterpretations throughout the course. This outcome mirrors the results of previous studies (Perry *et al.*, 2008; Dees *et al.*, 2014), suggesting that counting features of phylogenetic trees to evaluate taxa relatedness may be a robust, if inappropriate, misinterpretation that is more resistant to instruction than others. Overall, instruction that balanced the use of diagonal and bracket phylogenetic trees was not beneficial for style effects in regard to evaluating taxa relatedness, and our students were typical in their struggles with evaluating taxa relatedness on phylogenetic trees in general.

Construction

The majority of diagonal and bracket phylogenetic trees constructed by students were correct or adequate in terms of accuracy across all four instruments. However, bracket phylogenetic trees were also significantly more accurate than diagonal phylogenetic trees across instruments due to a much lower incidence of minor errors (e.g., extra nodes and empty branches;

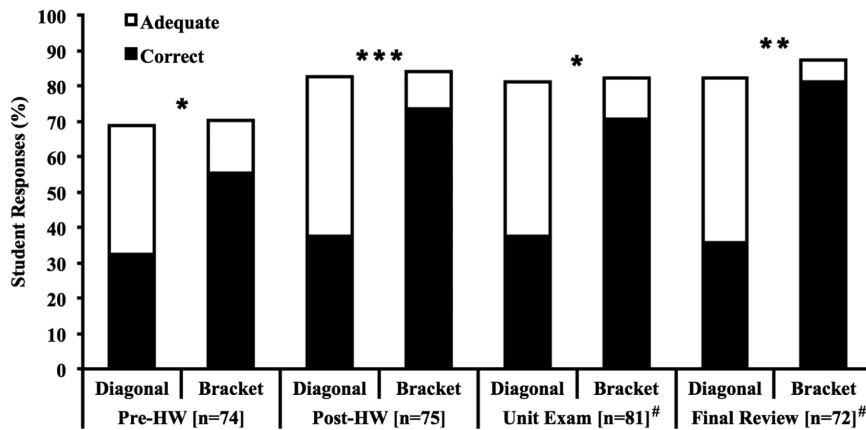


FIGURE 2. Accuracy of phylogenetic trees constructed by students with comparisons across styles for all instruments. #Students constructed one phylogenetic tree in the style of their choice during the unit exam (64 diagonal, 17 bracket) and final exam review activity (56 diagonal, 16 bracket), resulting in between-student rather than within-student comparisons of accuracy across styles for those instruments. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Supplemental Table S2). Further, improvement from the pre-instructional homework to the post-instructional homework was driven mostly by students switching from incorrect to adequate for diagonal phylogenetic trees and from incorrect to correct for the bracket style. Although the minor errors observed in student-constructed phylogenetic trees should not hinder performance on our interpretation tasks, such errors could be indicative of other misinterpretations. For example, extra nodes and empty branches may reflect the common belief among students that evolutionary changes occurred only at nodes (Baum *et al.*, 2005; Meir *et al.*, 2007; Gregory, 2008). Thus, in some cases, students may have intentionally included more minor errors when constructing diagonal phylogenetic trees, and these errors are not trivial.

Alternatively, students may simply be hastier when constructing diagonal phylogenetic trees and inadvertently include more minor errors. Diagonal phylogenetic trees contain about one-third the number of lines as equivalent bracket phylogenetic trees. Thus, when resulting phylogenetic trees for construction tasks are not known in advance, the diagonal style is simpler and much faster for trial-and-error approaches. We hypothesize that simplicity and speed are the primary reasons why students consistently preferred to construct diagonal phylogenetic trees when allowed to use the style of their choice during this study and two previous investigations (Dees and Momsen, 2016; Dees *et al.*, 2017). Therefore, the speed and ease of using the diagonal style for construction tasks may have led students to inadvertently include more minor errors (i.e., sloppiness).

Implications and Future Directions

The present study is novel in documenting the persistence of style effects after instruction that integrated diagonal and bracket phylogenetic trees equally. Further, it confirms prior research on style effects with an independent population of students who had a vested interest in learning to interpret and construct phylogenetic trees. As a result of our research and that

of others, we join Novick and Catley (2007, 2013) in recommending that introductory biology instructors use only the bracket style for instruction on phylogenetic trees and as visual representations of evolution in general. However, diagonal and bracket phylogenetic trees are both commonly used by biologists (Catley and Novick, 2008), which necessitates that biology majors gain familiarity with diagonal phylogenetic trees in their upper-division course work. Given that significant style effects were observed for some tasks after instruction that integrated diagonal and bracket phylogenetic trees equally, it is likely these style effects will persist in upper-division courses without instructional interventions. Unfortunately, we are unaware of specific pedagogy that has successfully mitigated all style effects for student interpretations and construction of phylogenetic trees, and it is unlikely such pedagogy will be developed by instructors without first determining why these style effects exist.

One intriguing hypothesis, supported by evidence, is that students visually perceive diagonal and bracket phylogenetic trees differently. Novick and Catley (2007) used translation exercises (i.e., converting one visual representation of evolution to another while retaining the same information) to demonstrate that students often interpret lines of diagonal phylogenetic trees as single entities, whether accurate or not. In the diagonal phylogenetic tree of Figure 1, for example, the line from node A to koalas is a single branch. However, students may also interpret the line from node C to saltwater crocodiles as a single branch rather than two branches. In contrast, it is more apparent in the equivalent bracket phylogenetic tree of Figure 1 that two branches occur between node B and black caimans. Thus, the hierarchical structure of monophyletic groups within phylogenetic trees could be obscured by the diagonal style, and as a result, student understanding may be impeded.

Alternatively, diagonal phylogenetic trees could disproportionately encourage misinterpretations of evolution as a ladder of progress from “lower” to “higher” organisms. This hypothesis emerges from classroom observations of introductory biology students constructing a phylogenetic tree of large groups of vertebrates (e.g., amphibians and mammals) in the style of their choice. Because branches can be rotated around nodes on phylogenetic trees without changing relationships, taxa can appear in almost any order along the branch tips (Baum and Offner, 2008). However, when our students used the diagonal style to construct a phylogenetic tree of vertebrates during class, mammals almost invariably appeared in the rightmost position. Conversely, we did not observe any discernible pattern when students used the bracket style, as mammals appeared in a random location along the branch tips. Therefore, it is possible that diagonal phylogenetic trees reinforce the common misinterpretation of evolution as a ladder of progress toward a goal, which is generally humans and other mammals (Gregory, 2008).

Consequently, students could disproportionately focus on irrelevant features when interpreting diagonal phylogenetic trees, such as the number of nodes between taxa or the proximity of branch tips (Supplemental Table S1). We believe that the hypothesis put forth by Novick and Catley (2007) is one critical driver of differences in student performance across phylogenetic tree styles. However, other factors likely contribute to these style effects, and future research should explore alternative hypotheses.

In addition to style effects, other variables may influence student understanding of phylogenetic trees. For example, equivalent phylogenetic trees can be drawn in a vertical, horizontal, or even circular orientation. To our knowledge, only one study has investigated effects of orientation on student comprehension. Phillips *et al.* (2012) found no significant difference in accuracy for two tasks, identifying monophyletic groups and evaluating taxa relatedness, between horizontal and vertical phylogenetic trees drawn only in the bracket style. Further, most phylogenetic trees in textbooks and other instructional resources are not scaled for time or degree of divergence (i.e., chronograms and phylograms), and it is unknown whether scaled phylogenetic trees would help or hinder student comprehension. Future research should explore variables other than style, such as orientation and scaling, that could impact student understanding of phylogenetic trees, either alone or in combination with other variables. Once we determine which variables affect student interpretations and construction of phylogenetic trees and why such variables are important, we can design and evaluate research-based instruction that best promotes student learning.

ACKNOWLEDGMENTS

This research was conducted in compliance with Institutional Review Board regulations (protocol SM12217) and was funded by a STEM education fellowship from North Dakota State University and the National Science Foundation (DRL-1420321).

REFERENCES

- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.
- Armstrong, N., Chang, S. M., & Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. *CBE—Life Sciences Education*, 6, 163–171.
- Baum, D. A., & Offner, S. (2008). Phylogenies & tree-thinking. *American Biology Teacher*, 70, 222–229.
- Baum, D. A., & Smith, S. D. (2013). *Tree thinking: An introduction to phylogenetic biology*. Greenwood Village, CO: Roberts.
- Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree-thinking challenge. *Science*, 310, 979–980.
- Blacquiere, L. D., & Hoese, W. J. (2016). A valid assessment of students' skill in determining relationships on evolutionary trees. *Evolution Education Outreach*, 9, ar5.
- Bokor, J. R., Landis, J. B., & Crippen, K. J. (2014). High school students' learning and perceptions of phylogenetics of flowering plants. *CBE—Life Sciences Education*, 13, 653–665.
- Bray Speth, E., Shaw, N., Momsen, J., Reinagel, A., Le P, Taqieddin R., & Long, T. (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE—Life Sciences Education*, 13, 529–539.
- Burks, R. L., & Boles, L. C. (2007). Evolution of the chocolate bar: A creative approach to teaching phylogenetic relationships within evolutionary biology. *American Biology Teacher*, 69, 229–237.
- Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE—Life Sciences Education*, 6, 9–20.
- Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience*, 58, 976–987.
- Catley, K. M., Phillips, B. C., & Novick, L. R. (2013). Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. *Research in Science Education*, 43, 2327–2348.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- College Board. (2015). *AP Biology: Course and exam description*. New York.
- Dauer, J. T., Momsen, J. L., Bray Speth, E., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50, 639–659.
- Dees, J., Freiermuth, D., & Momsen, J. L. (2017). Effects of phylogenetic tree style on student comprehension in an introductory biology course. *American Biology Teacher*, 79, 729–737.
- Dees, J., & Momsen, J. L. (2016). Student construction of phylogenetic trees in an introductory biology course. *Evolution Education Outreach*, 9, ar3.
- Dees, J., Momsen, J. L., Niemi, J., & Montplaisir, L. (2014). Student interpretations of phylogenetic trees in an introductory biology course. *CBE—Life Sciences Education*, 13, 666–676.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35, 125–129.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287–312.
- Eddy, S. L., Crowe, A. J., Wenderoth, M. P., & Freeman, S. (2013). How should we teach tree-thinking? An experimental test of two hypotheses. *Evolution Education Outreach*, 6, ar13.
- Federer, M. R., Nehm, R. H., Opfer, J. E., & Pearl, D. (2015). Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. *Research in Science Education*, 45, 527–553.
- Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Edinburgh, UK: Oliver and Boyd.
- Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., ... Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE—Life Sciences Education*, 6, 132–139.
- Gendron, R. P. (2000). The classification and evolution of Caminalcules. *American Biology Teacher*, 62, 570–576.
- Goldsmith, D. W. (2003). The great clade race: Presenting cladistic thinking to biology majors and general science students. *American Biology Teacher*, 65, 679–682.
- Gregory, T. R. (2008). Understanding evolutionary trees. *Evolution Education Outreach*, 1, 121–137.
- Halverson, K. L., Boyce, C. J., & Maroo, J. D. (2013). Order matters: Pre-assessments and student generated representations. *Evolution Education Outreach*, 6, ar24.
- Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Science Education*, 95, 794–823.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change*, 30, 26–35.
- Julius, M. L., & Schoenfuss, H. L. (2006). Phylogenetic reconstruction as a broadly applicable teaching tool in the biology classroom: The value of data in estimating likely answers. *Journal of College Science Teaching*, 35, 40–45.
- Lampert, E., & Mook, J. (2015). Modeling with nonliving objects to enhance understanding of phylogenetic tree construction. *American Biology Teacher*, 77, 587–599.
- Lents, N. H., Cifuentes, O. E., & Carpi, A. (2010). Teaching the process of molecular phylogeny and systematics: A multi-part inquiry-based exercise. *CBE—Life Sciences Education*, 9, 513–523.

- Long, T. M., Dauer, J. T., Kostelnik, K. M., Momsen, J. L., Wyse, S. A., Bray Speth, E., & Ebert-May, D. (2014). Fostering ecoliteracy through model-based instruction. *Frontiers in Ecology and the Environment*, *12*, 138–139.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*, 304–316.
- Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory*, *20*, 645–653.
- Maxwell, A. E. (1970). Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry*, *116*, 651–655.
- McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, MD: Sparky House.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153–157.
- Meir, E., Perry, J., Herron, J. C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. *American Biology Teacher*, *69*, 71–76.
- Meisel, R. P. (2010). Teaching tree-thinking to undergraduate biology students. *Evolution Education Outreach*, *3*, 621–628.
- Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*, 237–256.
- Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.
- Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, *13*, 197–223.
- Novick, L. R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *American Educational Research Journal*, *50*, 138–177.
- Novick, L. R., & Catley, K. M. (2016). Fostering 21st-century evolutionary reasoning: Teaching tree thinking to introductory biology students. *CBE—Life Sciences Education*, *15*, ar66.
- O'Hara, R. J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Systematic Zoology*, *37*, 142–155.
- O'Hara, R. J. (1997). Population thinking and tree thinking in systematics. *Systematic Zoology*, *26*, 323–329.
- Omland, K. E., Cook, L. G., & Crisp, M. D. (2008). Tree thinking for all biology: The problem with reading phylogenies as ladders of progress. *BioEssays*, *30*, 854–867.
- Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., & Cooper, S. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE—Life Sciences Education*, *9*, 133–140.
- Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D. (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE—Life Sciences Education*, *7*, 193–201.
- Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2012). Teaching tree thinking to college students: It's not as easy as you think. *Evolution Education Outreach*, *5*, 595–602.
- Rufibach, K. (2011). Assessment of paired binary data. *Skeletal Radiology*, *40*, 1–4.
- Sandvik, H. (2008). Tree thinking cannot taken for granted: Challenges for teaching phylogenetics. *Theory in Biosciences*, *127*, 45–51.
- Smith, J. J., Cheruvelil, K. S., & Auvenshine, S. (2013). Assessment of student learning associated with tree thinking in an undergraduate introductory organismal biology course. *CBE—Life Sciences Education*, *12*, 542–552.
- Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, *42*, 412–416.
- Sun, X., & Yang, Z. (2008). Generalized McNemar's test for homogeneity of the marginal distributions. In SAS Users Group International (Eds), *SAS Global Forum* (Paper 382-2008). Cary, NC: SAS Institute.
- Sundberg, M. D. (2002). Assessing student learning. *Cell Biology Education*, *1*, 11–15.
- Tanner, K., Chatman, L. S., & Allen, D. (2003). Approaches to cell biology teaching: Cooperative learning in the science classroom—beyond students working in groups. *Cell Biology Education*, *2*, 1–5.
- Thanukos, A. (2009). A name by any other tree. *Evolution Education Outreach*, *2*, 303–309.
- Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R. B., & Reece, J. B. (2014). *Campbell biology in focus*. Boston, MA: Pearson Education.
- Wiley, E. O. (2010). Why trees are important. *Evolution Education Outreach*, *3*, 499–505.