

RESEARCH ARTICLE

Open Access

# Teaching undergraduate students to draw phylogenetic trees: performance measures and partial successes

Aimee K Young, Brian T White\* and Tara Skurtu

## Abstract

**Background:** An in-depth understanding of the process and products of evolution is an essential part of a complete biology education. Phylogenetic trees are a very important tool for understanding evolution and presenting evolutionary data. Previous work by others has shown that undergraduate students have difficulty reading and interpreting phylogenetic trees. However, little is known about students' ability to construct phylogenetic trees.

**Methods:** This study explores the ability of 160 introductory-level biology undergraduates to draw a correct phylogenetic tree of 20 familiar organisms before, during and after a General Biology course that included several lectures and laboratory activities addressing evolution, phylogeny and 'tree thinking'. Students' diagrams were assessed for the presence or absence of important structural features of a phylogenetic tree: connection of all organisms, extant taxa at branch termini, a single common ancestor, branching form, and hierarchical structure. Diagrams were also scored for how accurately they represented the evolutionary relationships of the organisms involved; this included separating major animal groups and particular classification misconceptions.

**Results:** Our analyses found significant improvement in the students' ability to construct trees that were structured properly, however, there was essentially no improvement in their ability to accurately portray the evolutionary relationships between the 20 organisms. Students were also asked to describe their rationale for building the tree as they did; we observed only a small effect on this of the curriculum we describe.

**Conclusions:** Our results provide a measure, a benchmark, and a challenge for the development of effective curricula in this very important part of biology.

**Keywords:** Assessment, Phylogenetic trees, Undergraduate education

## Background

Improving students' understanding of the underlying evolutionary processes that provide a framework for thinking about living organisms is an important goal of biology teachers and education researchers worldwide. As part of this, many researchers argue that evolutionary processes cannot be fully understood unless students are able to read phylogenetic trees and interpret the evolutionary relationships depicted therein (O'Hara 1997, Baum et al. 2005, Baum and Offner 2008, Omland et al. 2008, Perry et al. 2008). Phylogenetic trees are one of the

most important tools that evolutionary biologists use to record and synthesize information, explain phenomena and predict relationships among organisms (Novick and Catley 2007). For this reason, Baum et al. (2005) recommend that evolution education include clear and explicit instruction on building trees as well as on reading relationships and traits depicted in phylogenetic trees. Although a wide range of research has identified students' difficulties with interpreting these trees (O'Hara 1997, Lopez et al. 1997, Meir et al. 2007, Novick and Catley 2007, Halverson et al. 2011, Perry et al. 2008, Sandvik 2008), only a few have explored students' abilities to construct them (Staub et al. 2006, Halverson et al. 2011, Halverson 2011). Our research has focused on students'

\* Correspondence: [brian.white@umb.edu](mailto:brian.white@umb.edu)  
Biology Department, University of Massachusetts Boston, 100 Morrissey  
Boulevard, Boston, MA 02125, USA

abilities to construct phylogenetic trees from a set of familiar organisms. We distributed a survey to 160 undergraduate introductory biology students at the University of Massachusetts Boston to determine how well they could depict the evolutionary relationships among these organisms. The design of the study allowed us to measure the effect of a laboratory study targeting tree-building skills on students' abilities to draw trees accurately. The structure of their trees and classification of organisms were scored in comparison to scientifically accepted trees. In addition, we identified students' rationales for creating their phylogenetic tree.

### **Students' prior knowledge**

Students come to the classroom with significant pre-existing knowledge about the natural world and this knowledge informs their learning of evolutionary concepts. Two lines of research have investigated this in detail: investigations of folkbiology and naïve biology as well as investigations of students' understanding of phylogeny.

Folkbiology reflects how people understand the natural world and infer relationships among living things without formal instruction (Lopez et al. 1997, Coley et al. 1999, Hatano and Inagaki 1999, Medin and Atran 1999, Atran et al. 2004). Cobern et al. (1999) found that, even when students have formal instruction, they rely more heavily on their personal experiences with the natural world when asked about scientific concepts. Folkbiology is often informed by rich experiences with nature and can be influenced by one's culture, location and prior knowledge (Atran 1999, Coley et al. 1999, Diamond and Bishop 1999, Hatano and Ingaki 1999, Ross et al. 2003, Medin and Atran 2004). Folk taxonomy, a subdiscipline of folkbiology, refers to the hierarchical nature of folkbiological classification (Atran et al. 2004) and tends to be culturally universal and resistant to change (Atran 1999). Both biological classification and folkbiological classification rely on direct contact and experience with plants and animals in the natural environment (Medin and Atran 1999). By contrast, naïve biology demonstrates a lack of experience with the natural world and is usually associated with urban populations (Hatano and Ingaki 1999, Atran et al. 1999). Both naïve biology and folkbiology are ways of interacting with and thinking about the natural world without the influence of modern science (Hatano and Ingaki 1999). Researchers can uncover the criteria students use for classification and the groups of organisms that students create by observing how naïve biology and folkbiology help students make predictions about relationships among organisms.

Several studies have been conducted to compare the classification systems used by American undergraduates, who do not need to rely on ecological knowledge in today's

industrialized world, and members of an Itzaj-Mayan culture, who are still dependent on the environment for their survival (Lopez et al. 1997, Atran 1999, Coley et al. 1999, Atran et al. 2004). When developing classification systems for familiar animals, the American students used morphology, behavior and size to inform their classification decisions whereas the Itzaj considered ecological factors, morphology and behavior (Lopez et al. 1997). Coley et al. (1999) found that the Itzaj-Mayan organized mammals using the animals' habitat and behavior whereas Northeastern University undergraduates grouped mammals based on morphology, specifically the organism's size. Both the Itzaj and American undergraduates considered diet when grouping organisms and separated herbivores from carnivores. The American students used domestication as a criterion for grouping organisms whereas the Itzaj put a greater emphasis on habitat.

Educational research has worked towards uncovering how accurately students can classify organisms, what type of reasoning they use when classifying, and how they illustrate these relationships among organisms (O'Hara 1997, Meir et al. 2007, Halverson et al. 2011, Halverson 2011). Halverson et al. (2011) examined students' abilities to read and construct phylogenetic trees and reported that these two skills are independent of each other. Some students were able to explain relationships in phylogenetic trees very clearly but no students could build an accurate tree (Halverson et al. 2011). Furthermore, students showed greater gains in tree reading than in tree building (Halverson 2011). Ecological and morphological reasoning were employed by many students trying to interpret and predict evolutionary relationships. In this study, ecological reasoning included knowledge about geographic location and habitat. For example, students using ecological knowledge explained that aquatic animals belonged in the same group (Halverson et al. 2011). Additionally, students used morphological reasoning by explaining that certain organisms are related because of common physical appearances (Halverson et al. 2011). Halverson et al. (2011) also found that students were generally hesitant to use trees to solve problems and were clearly uncomfortable when building phylogenetic trees.

Other research has identified specific misconceptions relating to college students' abilities to read phylogenetic trees. Meir et al. (2007) found that most of the college-level biology students did not have sufficient skills for reading phylogenetic trees appropriately.

### **Educational interventions to improve tree thinking and their evaluation**

The foregoing research has highlighted a set of frequent and persistent misconceptions held by students at a range of levels. In response to these challenges, educa-

tors have developed a wide range of activities targeted to increase students' tree thinking skills. Although some of these approaches have been shown to be effective learning tools, others have not been rigorously evaluated and some may even create to misconceptions of their own. Catley and Novick (2008) examined evolutionary diagrams in 31 popular biology textbooks and located over 500 cladograms. Over half of the diagrams found were 'ladders'. In an earlier study, Novick and Catley (2007) found that ladder diagrams are much more difficult for students to understand than tree diagrams (Figure 1). When interpreting ladder diagrams, students are prone to practice common misconceptions such as using node counting and looking at tip proximity to determine relationships. Furthermore, many of the diagrams were poorly described, ambiguous, and did not depict evolutionary relationships clearly. Finally, about 20% of evolutionary diagrams in the textbooks were neither cladograms nor trees (Catley and Novick 2008) and did not show the importance of common ancestry, branching morphology and shared derived characters.

Particular activities can also confuse students on how to interpret relationships in a phylogenetic tree. Some activities may only strengthen commonly held misconceptions. For example, many educators use an activity that requires students to classify nuts and bolts to show how relationships among organisms are determined. Nickels and Nelson (2005) point out that these activities can lead to crucial misconceptions about phylogeny because there are always many different equally plausible ways that the nuts and bolts can be organized. This contrasts with biological classification that typically leads to one (or very few) most parsimonious taxonomy(ies).

Computer programs have been implemented to help students work with phylogenetic trees to gain a better understanding of macroevolution. Students who had used a computer program, EvoBeaker (described by Perry et al. 2008), were compared with students who had received a traditional lecture that included a tree-building activity called the Great Clade Race (Goldsmith 2003). In EvoBeaker, students observe how evolutionary trees are constructed, predict evolutionary relationships

and observe the effects of adjusting the parameters of a simulation on the resulting phylogenetic tree. In Goldsmith's study, students' understanding of phylogeny and tree thinking was assessed by measuring the frequency of important misconceptions and the acquisition of the key tree-reading skills. Students in both treatment groups significantly improved; however, there was no significant difference in learning between the EvoBeaker and the Great Clade Race groups. Others have developed a museum exhibit that consists of an interactive game for teaching tree building (Horn et al. 2012). They found that visitors were engaged with the game and discussed its scientific aspects while playing; the visitors' understanding of phylogeny was not measured.

Our study uses an open-ended survey - *The Diversity of Life Survey* - to explore students' ability to construct phylogenetic trees from a set of 20 familiar organisms. Students were surveyed before, during and after a freshman-level General Biology course that included specific activities designed to increase students' understanding of evolution. Our results show some increase in students' abilities with significant room for improvement. Going forward, our measures and results provide a baseline for further work in this area.

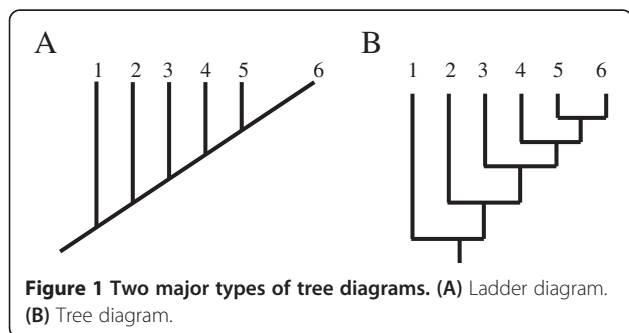
## Methods

### Course content

Biology 112 is a second semester introductory biology course required for all students seeking a Bachelor of Science degree in Biology or Biochemistry. Students enrolled in the premedical program are also required to take Biology 112. The course consists of 3 hours of lectures and 3 hours of laboratory studies each week. The lecture material presented in Biology 112 is similar to that found in many college-level introductory biology courses. The course covers a range of topics including evolution, plant and animal diversity, physiology, and ecology. A weekly laboratory component corresponding to the lectures explores biological concepts through computer simulations, manipulation of specimens, observations and dissections. *Biology* (Campbell and Reece 2005) is the required text for the course.

Prior to the distribution of the *Diversity of Life Survey*, the lecturer discussed processes of natural selection and population genetics. The pre-survey was then distributed. After the surveys were collected, the professor began a series of lectures on speciation and phylogeny. Examples of student phylogenetic trees from the pre-survey were used to discuss the importance of showing common ancestry, placing extant organisms at the tips of branches, using hierarchical organization and mapping traits on phylogenetic trees.

Several laboratory sessions further emphasized elements of evolution and phylogeny. During the first



**Figure 1** Two major types of tree diagrams. (A) Ladder diagram. (B) Tree diagram.

laboratory session of the semester, students visited the Harvard Museum of Natural History in Cambridge, MA. The laboratory activities included finding answers to specific questions regarding classification, convergent evolution, skeletal morphology, analogous structures and relationships between different marine mammals. The subsequent *Skulls and Evolution* laboratory session required students to examine different types of mammalian skulls and correlate this with their evolutionary relationships. This laboratory session is similar to the one described by Nelson and Nickels (2001) and includes skulls from each mammalian order with a specific focus on hominids and marine mammals. In the *Molecular Phylogeny* laboratory session, students used a computer program to construct phylogenetic trees based on the amino acid sequence of cytochrome c. They then used this tool to answer a series of phylogenetic questions (for example, 'is a bat a bird or a mammal?'); for the laboratory report, they devised their own phylogenetic question and answered it by building and using an appropriate tree.

The last laboratory session of the semester, the *Phylogenetic Collection Lab* (White 2009), was the treatment laboratory activity in our study design. Our goal was to determine if this particular semester-long laboratory series, which involved in-depth exploration of phylogeny, significantly improved students' abilities to use phylogenetic trees to classify organisms. Over the course of the semester, students had collected physical specimens of organisms from 16 different phyla of their choosing. They brought these in to the *Phylogenetic Collection Lab* session to compare and contrast organisms from the same and different phyla. The class then generated a massive phylogenetic tree indicating the kingdom, phylum, class, order, family, genus and species of the organisms they had collected. This exercise showed students the diversity of organisms found in the natural world as well as ways to map the relationships between organisms using phylogenetic trees. This laboratory session also forced students to research many phyla and discover what characteristics are unique to those phyla as well as the types of species found in each phylum.

These studies were approved by the Institutional Review Board of the University of Massachusetts, Boston in accordance with the University's policies governing research on human subjects.

#### **Diversity of life survey**

The *Diversity of Life Survey* asked students 'to design a tree that will help orient visitors' in a hypothetical museum using the 20 organisms provided in the survey (Figure 2). Further directions stated, 'your tree should include all the groups of organisms listed below and communicate the ways they are evolutionarily related to one another. On the next page, draw a tree diagram to show the relationships between these organisms.' The names

and pictures of 20 organisms were provided on the survey followed by a blank sheet of paper for students to draw their phylogenetic trees. The organisms were beetles, bats, whales, squirrels, snails, humans, butterflies, lizards, birds, horseshoe crabs, crocodiles, millipedes, sharks, rats, leeches, sea stars, fishes, jellyfish, spiders and turtles. These particular organisms were chosen because they are familiar to most students, represent a range of taxa, and have differing levels of relatedness. The last section of the survey asked the students three open-ended questions: how they organized the tree; how they determined relatedness between organisms; and how they represented similarities and differences in their tree representation.

#### **Survey administration**

The pre-survey was distributed to all students on the sixth day of class. Students were told to return the survey within three days for class credit. The instructor made it very clear that class credit would be awarded regardless of the quality of answers. It was also strongly emphasized that students should not refer to any outside resources when completing the survey. The post-survey, which was identical to the pre-survey, was administered at two different time points. In both cases, students completed the post-survey entirely during a laboratory session. Half of the students took the post-survey before starting the *Phylogenetic Collection Lab* (*No Lab* group). The other half of students completed the post-survey directly after the *Phylogenetic Collection Lab* (*Lab* group). This allowed us to separate the effect of the *Phylogenetic Collection Lab* from the other parts of the course. A significant difference between the pre- and post-surveys indicates a probable effect of the course as a whole, with (*Lab* group) or without (*No Lab* group) the *Phylogenetic Collection Lab*. Because this design does not include a no-treatment control, it is important to note that differences between the pre- and post-surveys could also be due to events outside of the class and even to students' increasing familiarity with the survey itself. Finally, as long as the scores on the pre-surveys in both the *Lab* and *No Lab* groups do not differ significantly, a significant difference between the post-surveys of the two groups indicates an incremental effect of the *Phylogenetic Collection Lab* in addition to the preceding lectures and laboratory sessions.

Students who did not complete both a pre-survey and a post-survey were excluded from the analysis. There were 222 students enrolled in Biology 112. Of those students, 160 (72.1%) completed both the pre- and post-survey; the other 62 students failed to complete either or both of the surveys. All surveys were randomized and numbered for analysis.























This survey is designed to see how well you understand the diversity of living organisms. There is no right or wrong answer; you will receive full credit for whatever you write. We are most interested in **your** understanding of these important biological issues. Please do not consult any outside sources (textbook, www, other people, etc.) when completing this survey!!

## Organizing Diversity Exercise

Assume that you are working for a natural history museum like the Harvard Museum of Natural History, only smaller. Your museum has specimens of the following 20 groups of organisms in its collection. Your task is to design a tree that will help orient visitors to the collection.

Your tree should include all the groups of organisms listed below and communicate the ways they are evolutionarily related to one another. On the next page, draw a tree diagram to show the relationships between these organisms.

You should use the numbers or the organism names to identify each group in your tree. Please include additional text and graphics that you think will help visitors understand how you have organized these groups of organisms. There is no right or wrong answer to this task, but it is important that you are able to explain the logic behind your approach.

1  beetles	2  bats	3  whales	4  squirrels	5  snails
6  humans	7  butterflies	8  lizards	9  birds	10  horseshoe crabs
11  crocodiles	12  millipedes	13  sharks	14  rats	15  leeches
16  sea stars	17  fishes	18  jellyfish	19  spiders	20  turtles

**Figure 2** Diversity of Life Survey.

A rubric was designed to identify structural components of the tree, classification of the organisms, classification misconceptions and rationale for organization. Two independent scorers used the rubric to measure students' abilities to create phylogenetic trees; both scorers scored all surveys. Surveys were numbered and assessed blindly; the scorers did not know whose survey they were scoring or whether they were evaluating a pre- or post-survey. The overall agreement for all categories was 90.3% between the two independent scorers. Because agreement was high, survey scores were combined by choosing an individual student's pre- and post-survey scores randomly from either scorer.

The structure of each student's tree was evaluated based on the presence of five features essential to a proper phylogenetic tree. Figure 3 shows five typical trees illustrating these features. Some of these parallel the categories described by Halverson (2011); these similarities are noted below.

1. *All organisms connected*: Because all life descended from a single common ancestor, a correct tree should show connections between all organisms. Student trees were scored as having all organisms

connected (Figure 3A,C,D) or not (Figure 3B,E). Those that did not were typically organized into several separate groups, sometimes with humans separate from the other animals.

2. *Extant species at the ends of branches*: Because all organisms in the survey are currently living, they should only be placed at the ends of branches. Student trees were scored as having this property only if all organisms were at the ends of branches (Figure 3A,C); trees with any extant species on internal nodes were scored as lacking this feature (Figure 3B,D); surveys lacking this feature would be scored by Halverson (2011) as having 'taxa along branches'.
3. *Single common ancestor*: As in (1), student trees should indicate the common ancestor of all living things. Students could indicate this in one of two ways: the tree could have branched out from a single root or the student could have written the phrase 'common ancestor' on the originating branch. Trees showing at least one of these features were scored as having a single common ancestor (Figure 3A,C); trees without an obvious indication of a common ancestor or those that used one of the



(See figure on previous page.)

**Figure 3 Sample student trees.** The students' work is shown in black and white. The individual animals in completed surveys were highlighted by the investigators to facilitate scoring as follows: vertebrates are blue (mammals) and green (non-mammal vertebrates); invertebrates are yellow (arthropods) and pink (non-arthropod invertebrates). Individual surveys are discussed in the text. Each of the drawings in this figure exemplifies several important features from our rubric. Key features are highlighted below; see text for details. **(A)** This drawing shows a branching structure and a single level of hierarchy; it also shows a single common ancestor. **(B)** This drawing is not a single tree and extant taxa are sometimes at internal nodes; furthermore, it appears to show hybridization among groups. **(C)** This is a correctly structured tree. **(D)** This drawing includes a loop or network where there are multiple paths from one organism to another. **(E)** This drawing has essentially none of the important features of a phylogenetic tree.

survey organisms as the common ancestor were scored as lacking this feature (Figure 3B,D,E).

- 4. Branching morphology:** Students should be able to use branching morphology to clearly depict evolutionary relationships between organisms. Diagrams that included branches were scored as having this feature (Figure 3A,B,C). All other representations, such as lists without any branching lines or networks where branches were intertwined or where there were multiple paths from one node to another (Figure 3D, see especially spiders-lizards-crocodiles-humans-squirrels-rats-spiders; also E) were scored as lacking this feature (Novick and Hurley 2001). In Halverson (2011), trees in these four categories (phylogenetic diagram, segregated organisms, simple progressive tree, and taxa along branches) would be considered to have this feature; the other six (flow chart, dichotomous key, ecological web, pictures, lists and none) would not.
- 5. Hierarchy:** Students should recognize that there are levels of similarity and divergence in organisms, organizing their tree to reflect this hierarchy. In a tree diagram, hierarchical relationships are represented by nodes where one line enters and multiple lines leave (Novick and Catley 2007). Student trees were scored for hierarchy on a three-level scale (0 = no hierarchy; 1 = one-level hierarchy; 2 = two or more-level hierarchy). Student trees that indicated that all organisms were equally related were scored as not showing any hierarchical relationships (Figure 3B,D,E). Surveys where organisms were organized into groups were scored as having one level of hierarchy (Figure 3A). Diagrams including two or more levels of hierarchy - groups within groups - were scored as having two or more levels of hierarchy (Figure 3C). Hierarchy level 0 roughly corresponds to Halverson (2011)'s 'single progressive tree' category.

**Total structure score:** These five structural elements were combined to give an overall structure score. This score was computed by adding one point for each of the first four elements to the hierarchy score. Thus, the structure score ranged from 0 (no proper structural

elements) to 6 (all elements and two or more hierarchical levels).

#### Classification of organisms

In addition to having the proper structural elements, a correct tree must also indicate the correct phylogenetic relationships among the organisms present. We therefore developed a second rubric to identify the degree to which students correctly classified the 20 organisms in the survey. This process occurred at two levels of resolution. First, trees were scored based on how well the diagram communicated the separation between the invertebrate organisms and the vertebrate organisms. The second step assessed how well the tree communicated a key distinction within each of the higher-level groups.

A completely correct tree would clearly communicate the separation between the nine invertebrates and ten vertebrates in the survey; each misplaced organism indicates that the student does not know the group to which that organism belongs and/or does not know how to communicate this using a tree. The wide variety of student answers made it challenging to develop a consistent scheme for this evaluation. Most responses could be categorized as a tree, a list or a network. For trees, a dividing line, or 'best split', was inserted on the phylogenetic tree in a position to segregate the greatest number of invertebrates from the greatest number of vertebrates. If a student wrote out lists of organisms instead of drawing a tree, then the list with the maximum number of invertebrates was used for the invertebrate count and the list with the maximum number of vertebrates was used for the vertebrate count. If there was a single list of organisms, then a division between invertebrates and vertebrates could not be determined. If the student drew a network, a distinction could not be made between invertebrates and vertebrates because there are multiple pathways connecting organisms and the survey was not included in this part of the analysis (Figure 3A,B,D,E).

If the division between invertebrates and vertebrates could be established on a survey, we proceeded to look within these groups. Within the invertebrate group, as determined by the invertebrate/vertebrate split previously described, a division between arthropods and non-arthropod invertebrates was created. Distinguishing

between these two groups within the student's phylogenetic tree used the same procedure for distinguishing between the invertebrates and vertebrates. A completely correct split would indicate five arthropods on one side of the division and four non-arthropod invertebrates on the other side; this would have an error score of 0. A completely random grouping would yield an error score of 4. A similar procedure was used for the division between the five mammals and the six non-mammal vertebrates in the survey; the maximum error here would be 5. The overall error score was the sum of these two values; thus, the total error score cannot exceed 9.

The tree shown in Figure 3C is the only tree in the figure that can be scored using this part of the rubric. It would be scored as follows. First, the best vertebrate/invertebrate split would be placed on the line next to the word 'invertebrates'. This would yield two smaller trees, one on the left with all nine invertebrates and one on the right with all 11 vertebrates - a completely correct division. Next, within the invertebrates, the best split corresponds to breaking the invertebrate sub-tree along the line that leads to 'arthropods'. This leaves four of five arthropods on one side and four of four non-arthropod invertebrates on the other side - this yields an invertebrate error score of 1 (millipedes are placed incorrectly in non-arthropod invertebrates). The vertebrates are clearly split into mammals and non-mammals by breaking the line to mammals for a mammal/non-mammal vertebrate error score of 0. Thus, the overall error score for this tree is 1.

### Classification misconceptions

In addition to determining the number of classification mistakes in a given tree, we also explored certain common specific classification errors that correspond to common misconceptions. Many students hold the conception that common physical characteristics are evidence of evolutionary relationships (Lopez et al. 1997, Coley et al. 1999, Halverson et al. 2011). If used appropriately, morphological characters can reveal evolutionary relationships but the students assessed did not often make the crucial distinction between homologous structures indicating common ancestry and analogous structures indicating convergent evolution. Students also often assumed that organisms living in the same environment or habitat were closely related (Lopez et al. 1997, Coley et al. 1999, Halverson et al. 2011). Based on these prevailing misconceptions, we looked for the following misplacements on students' trees: locating whales on the same terminal branch as fishes or sharks (Figure 3A,B,D,E), bats with birds (Figure 3A,B,D,E), bats with butterflies (Figure 3A,B,E), or birds with butterflies (Figure 3A,B,E).

### Rationale for organization

The survey also contained three short answer questions that explored the students' rationale when constructing their trees. The questions asked how they organized the tree, how they determined relatedness, and how they depicted similarities and differences in their tree representation. Student responses were examined for key words referring to morphology, habitat, taxonomy and diet expressed in a reasonable context. For example, if a student wrote, 'I put the whales and the sharks together because they both live in the ocean and have a large tail used for locomotion', that answer would be scored as including both morphology (tail) and habitat (ocean).

### Statistical procedures

Our analyses were conducted using PASW 18.0 and Excel 2004 (version 11.5.8) with particular tests chosen based on the nature of the data involved. Scores for correct structural components of a phylogenetic tree and common classification misconceptions were measured on a present/absent scale. For all such binary data, the differences between the *No Lab* and *Lab* groups were analyzed using a chi-square test; within-group repeated measure comparisons were completed using a McNemar change test. Scores indicating the levels of hierarchy present in the surveys were measured on a 0 to 2 scale. Comparisons between the scores of the *No Lab* and *Lab* groups for hierarchy levels were calculated using a  $2 \times 3$  chi-square test; within-group comparisons were completed using a related-samples sign test. The total structure score was measured on a 0 to 6 scale; the number of classification mistakes was measured on a 0 to 9 scale. A Wilcoxon-Mann-Whitney *U* test was used to compare these scores between the *No Lab* and *Lab* groups; within-group comparisons were calculated using a Wilcoxon signed-rank test.

### Results

Our survey was administered using a design that separated the effects of lectures and time from the effect of the *Phylogenetic Collection Lab* on students' ability to build phylogenetic trees. The pre-survey and post-survey were completed by 160 students. Of those students, 78 took the post-survey after the *Phylogenetic Collection Lab* (the '*Lab*' group) and 82 took the survey before the *Phylogenetic Collection Lab* (the '*No Lab*' group). By comparing the pre- and post-surveys of the *Lab* and *No Lab* groups, we could measure the combined effect of time, the course as a whole, and students' experience with the survey. By comparing the post-survey scores for the *Lab* and *No Lab* groups, we could determine the incremental effect of the laboratory exercise on students' abilities to construct phylogenetic trees. These data are summarized in Table 1.



**Table 1 Summary of results**

<b>I: Structure</b>											
a) All organisms connected				b) All at ends of branches				c) Single common ancestor			
	Pre	Post	<i>P</i>		Pre	Post	<i>P</i>		Pre	Post	<i>P</i>
No Lab	0.71	0.87	<0.001	No Lab	0.30	0.78	<0.001	No Lab	0.35	0.58	0.001
Lab	0.69	0.91	0.001	Lab	0.31	0.82	<0.001	Lab	0.20	0.55	<0.001
<i>P</i>	-	0.46		<i>P</i>	-	0.85		<i>P</i>	-	0.22	
d) Branching form				e) Level of hierarchy				f) Total structure score			
	Pre	Post	<i>P</i>		Pre	Post	<i>P</i>		Pre	Post	<i>P</i>
No Lab	0.68	0.95	<0.001	No Lab	1.50	1.66	0.082	No Lab	3.6	4.8	<0.001
Lab	0.70	0.95	<0.001	Lab	1.51	1.65	0.13	Lab	3.4	4.9	<0.001
<i>P</i>	-	0.87		<i>P</i>	-	0.26		<i>P</i>	-	0.51	
<b>II: Classification mistakes</b>											
g) Total mistakes				h) Vertebrate sub-groups				i) Invertebrate sub-groups			
	Pre	Post	<i>P</i>		Pre	Post	<i>P</i>		Pre	Post	<i>P</i>
No Lab	6.5	6.5	0.77	No Lab	3.7	3.5	0.33	No Lab	2.8	3.0	0.44
Lab	6.4	5.5	0.10	Lab	3.5	2.8	0.05	Lab	2.9	2.7	0.56
<i>P</i>	-	0.25		<i>P</i>	-	0.54		<i>P</i>	-	0.42	
<b>III: Specific misconceptions</b>											
j) Whale-shark-fish				k) Bat-bird				l) Bat-butterfly			
	Pre	Post	<i>P</i>		Pre	Post	<i>P</i>		Pre	Post	<i>P</i>
No Lab	0.60	0.49	0.12	No Lab	0.55	0.45	0.23	No Lab	0.34	0.12	<0.001
Lab	0.64	0.40	<0.001	Lab	0.58	0.38	0.015	Lab	0.27	0.13	0.037
<i>P</i>	-	0.36		<i>P</i>	-	0.47		<i>P</i>	-	0.70	
m) Butterfly-bird											
	Pre	Post	<i>P</i>								
No Lab	0.30	0.15	0.012								
Lab	0.31	0.13	0.006								
<i>P</i>	-	0.86									
<b>IV: Rationale for organization</b>											
n) Diet				o) Habitat				p) Taxonomy			
	Pre	Post	<i>P</i>		Pre	Post	<i>P</i>		Pre	Post	<i>P</i>
No Lab	0.05	0.06	1.00	No Lab	0.46	0.52	0.92	No Lab	0.54	0.68	0.089
Lab	0.08	0.03	0.44	Lab	0.59	0.47	0.22	Lab	0.58	0.74	0.042
<i>P</i>	-	0.109		<i>P</i>	-	0.068		<i>P</i>	-	0.99	
q) Morphology											
	Pre	Post	<i>P</i>								
No Lab	0.79	0.65	0.089								
Lab	0.85	0.62	0.007								
<i>P</i>	-	0.88									

Each smaller table compares the four samples described in Figure 2. Numbers less than one are fractions of total number of students; those greater than one are levels or total scores. *P*-values at the ends of rows reflect pre-post comparisons; those below columns reflect comparisons of *Lab* and *No Lab* groups.

We individually examined each structural component of a proper phylogenetic tree to determine if students improved their ability to create correctly structured trees as a result of the *Phylogenetic Collection Lab*. Overall, there were no significant differences between the post-

surveys of the *Lab* and *No Lab* groups; this shows no significant incremental effect of the *Phylogenetic Collection Lab*. However, several of the structural features did show significant pre- to post-survey changes in both the *Lab* and *No Lab* groups; this shows a significant effect of

time, the course as a whole, and students' experience with the survey.

The *total structure score* (Table 1f) showed a significant increase in both groups from around 3.5 to about 4.8 out of a maximum of 6. Most individual components of this score significantly improved in both groups. A large majority of student responses to the pre-survey included the elements *all organisms connected* (Table 1a) and *branching form* (Table 1d); by the post-survey, virtually all included these. By contrast, less than half of the pre-survey responses included *all at ends of branches* (Table 1b) or *single common ancestor* (Table 1c); on the post-survey, both had increased significantly with a substantial majority showing a *single common ancestor* and an even larger majority showing *all at ends of branches*. Finally, the average *level of hierarchy* (Table 1e) remained at approximately 1.5, showing no significant change in any comparison.

To measure the changes in students' abilities to properly use their trees to classify organisms in the survey, we measured the number of classification mistakes in the pre- and post-surveys. Overall, there were no significant changes in these measures; thus, neither the *Phylogenetic Collection Lab* nor the other parts of the course had any impact on student's abilities to properly classify the organisms. The average *total mistakes* (Table 1g) was between 5.5 and 6.5 out of a maximum of nine; this corresponded to 2.8 to 3.5 out of five maximum possible mistakes in classifying vertebrate sub-groups (Table 1h) and 2.7 to 3.0 out of a maximum of four mistakes in classifying invertebrate sub-groups (Table 1i). Only a very small number of students - between two and seven, depending on the group - constructed trees with no classification errors at all; none of these differences are significant (data not shown).

To measure any changes in four specific classification misconceptions, we examined each student's survey for close proximity of any of the following: whales with sharks and fishes, bats with birds, bats with butterflies, and birds with butterflies. Analysis of these data gave mixed results with some misconceptions showing a significant decrease in one group or the other - indicating an effect of the course in one or both groups - although there were no significant differences between the post-surveys of the *Lab* and *No Lab* groups - indicating that there was no incremental effect of the *Phylogenetic Collection Lab*. On the pre-surveys, about 33% of the students mistakenly grouped *bat-butterfly* (Table 1l) and *butterfly-bird* (Table 1m); in both *Lab* and *No Lab* groups, this dropped in the post-survey to roughly 13%. By contrast, the most common misconceptions, *whale-shark-fish* (Table 1j) and *bat-bird* (Table 1k) were present in more than half of the pre-surveys; this dropped slightly but significantly in the *Lab* group only.

To determine if a significant number of students changed their rationale for classification, we analyzed students' responses to the three short answer survey questions. Because we pooled each student's responses to the three questions, the total response percentage can be higher than 100%. There were few significant changes in the rationales used by the students indicating little or no effect of the curriculum on the students' rationale. On the pre-surveys, the most common rationales were *morphology* (Table 1q) followed by *taxonomy* (Table 1p) and *habitat* (Table 1o); *diet* (Table 1n) was used by only about 5% of the students. Students in the *Lab* group showed a small and marginally significant increase in the use of *taxonomy* and a small but significant decrease in their use of *morphology*. These suggest that the combination of lectures and laboratory exercises had a small but significant effect on these categories of students' rationales.

## Discussion

Many of the activities in Biology 112, and especially the *Phylogenetic Collection Lab*, were designed to familiarize students with the diversity and classification of living things, the relationships among organisms, and how to diagram those relationships using phylogenetic trees. We hypothesized that students taking the post-survey after the *Phylogenetic Collection Lab* would perform better on the *Diversity of Life Survey* than students taking the post-survey before the *Lab*. After scoring all of the surveys, we found no effect of the *Phylogenetic Collection Lab* on student responses. Specifically, the structure and classification of students' trees did not improve as a result of this laboratory session. Additionally, students did not show significantly fewer misconceptions or change their rationale as a result of the session.

Students' performance on the survey in both the *No Lab* and *Lab* groups did, however, improve over the course of the semester by incorporating accurate structural features of a tree and showing fewer misconceptions. However, these findings need to be interpreted with caution. As there was no non-intervention control for the *No Lab* group, it is not appropriate to assume that students improved solely as a result of the lecture and other laboratory exercises. It is important to bear in mind that students often score higher on the post-survey because they are seeing it for a second time and are consequently more familiar with it. As a result, we must be careful to attribute students' improvement over the semester to the combined effects of time, lectures and experience with the survey.

### Structural features of a phylogenetic tree

Most of the prior research on phylogenetic trees shows the difficulties students face when reading trees.

Halverson (2011) explored undergraduate students' abilities to build trees and found improvements in some structural features following a semester-long upper-division plant systematics course. Our research builds on this and has identified specific challenges students have when building phylogenetic trees and will help educators understand which structural characteristics of a tree are easily adopted and which are more difficult for students to incorporate.

Students in the *No Lab* and *Lab* groups incorporated more structural features of a phylogenetic tree in their post-survey responses than in their pre-survey responses. Specifically, we observed significant improvement over the semester in students drawing a tree with all organisms connected, placing extant species at the tips of branches, including a single common ancestor and using branching morphology. Thus, the total structure score also showed significant improvement over the semester. Almost all the students (95%) used branching morphology in their post-surveys and about 90% of students connected all organisms in one cohesive tree. About 80% of students placed extant species at the tips of branches and around 55% showed evidence of a single common ancestor in the post-survey. These findings are similar to Halverson's (2011), where 49% of students showed branching form in the pre-survey and 70% showed this in the post-survey. Interestingly, only 7% of the participants in Halverson's study Halverson et al. (2011) included extant taxa on internal nodes; this decreased to 0% on the post-survey. Our students also showed a significant decrease but with overall higher frequency (70% pre and 20% post); differences between the two results may be due to the different education levels of the students involved.

In our study, students were introduced to phylogenetic trees at the beginning of the semester and worked with trees during many of the laboratory sessions throughout the semester. Each time, the teaching assistants encouraged students to use appropriate structural components. This may explain why students in both groups showed improvement on including all organisms in one tree, placing extant species at the tips of branches and using branching morphology.

Just over half of the students showed a single common ancestor in their phylogenetic tree, indicating significant room for improvement. This may be explained in part by our standard of measurement. The two independent scorers were very conservative when reviewing students' trees for common ancestors and, therefore, may have missed more subtle expressions of this idea. Interviews would have been helpful to clarify this issue. Another reason for the absence of a common ancestor in students' trees may be a lack of understanding or belief that all organisms originated from a single common ancestor.

Other studies have found that freshman undergraduate biology students' understanding of common ancestry does not often include a common ancestor for all organisms (White and Yamamoto 2012).

The only structural component where we observed no improvement was in the levels of hierarchy students used in their trees. We measured three levels of hierarchy (0, 1 or 2) and the average level of hierarchy in the pre-survey was fairly high for both groups (over 1.5 levels); therefore, there was not much room for improvement. Also, the independent scorers counted the levels of hierarchy in each student's phylogenetic tree; they did not record if those levels of hierarchy were used appropriately. Although we are encouraged that students are using hierarchical features in their phylogenetic trees, further research is needed to determine if students are using it correctly.

#### **Correct classification of organisms**

Despite significant emphasis on phylogeny and classification throughout the course, students' classification of organisms in the *No Lab* and *Lab* groups did not improve over the semester. Additionally, there was no significant effect of the *Phylogenetic Collection Lab* on this measure. In the post-survey, students made between five and seven total classification mistakes out of nine. On average, students made about three out of four mistakes when classifying the invertebrates and three to four out of five mistakes when classifying the vertebrates. The lack of improvement over the semester for both groups is surprising, especially for students in the *Lab* group who investigated the classification of organisms and used these relationships to infer phylogenetic relationships directly before taking the survey. This result suggests that even a well-designed course targeting this material is not sufficient for students to assimilate these concepts.

Urban students, like the ones at University of Massachusetts Boston, generally have fewer intimate experiences with the environment and have an inarticulate framework for classifying organisms. Because we provided images of the 20 organisms in the survey, some students may have organized unfamiliar organisms based solely on physical appearance. Researchers conducting similar classification studies provide only the name of an organism to avoid biasing the students (Lopez et al. 1997, Atran 1999). We provided pictures to remind students of the type of organisms included in the analysis. A parallel study might be considered in the future to determine if students classify organisms differently in the *Diversity of Life Survey* when pictures are not provided.

Lopez et al. (1997) found that Itzaj-Mayan people, with extensive ecological knowledge and experience with the environment, could correctly differentiate among smaller

mammals when American undergraduates could not. In the *Diversity of Life Survey*, three of the mammals and all of the invertebrates could be considered smaller organisms. According to Lopez et al. (1997), our students may have a more difficult time classifying these smaller organisms, which explains their poor overall performance on the survey. Our urban students do not have the same ecological knowledge and experience with the natural environment that enabled the Itzaj to organize smaller animals. This also suggests that the selection of organisms in the survey may influence students' abilities to classify living things using a phylogenetic tree.

### Classification misconceptions

The *Phylogenetic Collection Lab* did not have a significant impact on students' classification mistakes. The two most prevalent misconceptions were placing whales with sharks or fishes and placing birds with bats. This may be an indication that students are considering analogous structures when determining relatedness among organisms. Whales, sharks and fishes have similar anatomy that allow for survival in the marine habitat. These analogous structures result from convergent evolution, not common descent. Similarly, the wings of bats and birds are analogous structures. In this case, the development of four limbs in both bats and birds is a homologous character, but the actual wings have developed through convergent evolution. Students specifically examined the structure of bat and bird wings at the Harvard Museum of Natural History in the beginning of the semester. Although wings provide the ability to fly, the anatomy of bat and bird wings are quite different. It is this difference that students were asked to explain after visiting the museum. Apparently, this exercise was not sufficient to convince many students that the sole presence of wings is not always a useful character to use when constructing a phylogeny.

Students may have also considered habitat when showing these particular misconceptions in their trees. It is obvious that whales, sharks and fishes share the same marine environment. This naïve ecological reasoning is common in non-scientists but is not a scientifically accepted criterion for determining relatedness and classification among organisms. As discussed in the next section, a considerable number of students used the organisms' habitats to infer phylogenetic affinities. This rationale did not significantly decrease over the semester and may explain why a large number of students still place the whales with the sharks and fishes in the post-survey in both the *No Lab* and *Lab* groups.

A significant number of students in both groups reduced the frequency of placing bats with butterflies and birds with butterflies over the course of the semester. There are two plausible explanations for this. First, in the pre-survey, students possibly grouped the bats, birds

and butterflies together due to the presence of wings and/or their ability to fly. Given that there was no improvement in the occurrence of the bats and birds together, students must have thought there was something unique about butterflies. Thus, students may have abandoned their previous rationale that the ability to fly or the sole presence of wings informs classification. Alternatively, students could have considered the size of the organisms in the post-survey and separated the butterfly for being much smaller than bats and birds.

### Rationale for organization

Ideally, in the absence of molecular data, students should use taxonomy and homologous structures to inform their classification decisions while building the phylogenetic tree. A less desirable rationale for organizing organisms would be habitat or diet. Although previous research indicated that American undergraduates use diet quite extensively when inferring relationships among animals (Lopez et al. 1997), we did not observe this same result. The *No Lab* and *Lab* groups did not change their rationale from diet or habitat over the course of the semester. Less than 10% of students considered diet in their post-survey whereas roughly 50% of students still considered habitat. Neither of these rationales is appropriate for generating phylogenetic trees at this level of resolution. One explanation may be that the *Phylogenetic Collection Lab* might have inadvertently reinforced the students' use of habitat when creating trees. Prior to coming to the laboratory session, students had to find 16 physical representatives of different phyla. Information on geographic location and habitat were required for each representative. This may have communicated to students that geographic location and habitat are important for classification.

The use of morphology as a rationale to infer relationships among organisms has been observed in many studies (Lopez et al. 1997, Coley et al. 1999, Atran 1999, Halverson et al. 2011). In our study, the use of morphology as a rationale for organizing the phylogenetic tree significantly decreased for both the *No Lab* and *Lab* groups over the semester. Just over 60% of students used morphology in the post-survey in both groups. There was no difference between the two groups; thus, there was no effect of the *Phylogenetic Collection Lab*. Using morphology is not necessarily negative if students are referring to homologous structures when creating their phylogenetic tree. However, because the students' answers tended to be very brief, we were not able to determine if they were using morphology correctly or not.

The use of taxonomy as a rationale increased over the semester for the *Lab* group. Almost 74% of students in this group considered taxonomic classification when organizing their phylogenetic tree. Students in the *No Lab*



group did not significantly improve over the semester. Less than 60% of students in the *No Lab* group used taxonomic language in the short answer responses. We found no effect of the *Phylogenetic Collection Lab* as measured by comparing the post surveys for both groups. It is likely that the discrepancy between the pre-post comparison and the comparison of the two post-surveys is due to differences in the power of the two statistical tests used. As a result, these differences should be interpreted with caution. The use of taxonomic knowledge is the most scientifically reasonable rationale. The *Phylogenetic Collection Lab* directs students to scientifically accepted taxonomy when building the evolutionary tree, so use of this language was expected to increase in the *Lab* group. The fact that there is no difference between groups indicates that students might not understand taxonomic language or might not appreciate how taxonomy is determined and employed in biology.

## Conclusions

The use of phylogenetic trees should be a major focus in biology education. Students will not be able to understand evolutionary processes clearly until they are able to understand and construct phylogenetic trees (O'Hara 1997, Baum et al. 2005, Baum and Offner 2008, Omland et al. 2008, Perry et al. 2008). To fully understand phylogenetic trees, it is necessary to be able to construct them (Meir et al. 2007).

This study has shown that a course that included many lectures and laboratory sessions targeting this material was insufficient to bring about the desired changes in students' understanding. Students have deeply rooted classification systems that they have used throughout their life. Urban students, in particular, may inform these classification frameworks with the knowledge they have gained through limited experiences with the natural environment. In order to understand and adopt scientifically accepted classification, students need to see the limitations in their conceptions (Posner et al. 1982). More specifically, they need to believe that phylogenetic trees showing classification are logical, comprehensible and fruitful (Posner et al. 1982). Especially in evolution education, students may be employing prior knowledge that constricts their ability to fully understand evolutionary concepts and the use of trees (Klaassen and Lijnse 1996). This prior knowledge is often embedded with misconceptions that are reflected by inconsistent reasoning patterns (Klaassen and Lijnse 1996). This was observed in our study when students used multiple rationales to explain their classification decisions. The use of vernacular language further complicates science education by using very specific terms (Klaassen and Lijnse 1996). For example,

relatedness in evolution has a very specific meaning that is distinguished from its lay definition. Also, the common names of organisms, like starfish or jellyfish, can be misleading because neither starfish nor jellyfish are closely related to fish. Without directly teaching vocabulary and addressing student misconceptions, evolution education will not improve. Because evolutionary biologists and educational researchers believe that understanding phylogenetic trees is essential to evolution education, educators need to find ways of explicitly teaching how to read and construct trees.

Researchers have argued that students should be introduced to phylogenetic trees as early as elementary school (Catley et al. 2005, Novick and Catley 2007). Starting in grades 3 to 5 (ages 8 to 11), students should be learning that shared derived characters show relatedness among organisms (Catley et al. 2005). Phylogenetic trees should be used to depict this relatedness and to make comparisons between organisms. In grades 6 to 8 (ages 11 to 14), students should be learning how to convert information from a Venn diagram into a tree. Also in these grades, students should be able to explain the difference between analogous structures and homologous structures (Catley et al. 2005). With this strong foundation in diagramming and explaining evolutionary relationships, students will have a more coherent framework for thinking about biology and, consequently, have a stronger understanding of science.

Our results suggest that the *Phylogenetic Collection Lab* could be modified for future students in ways that might increase its educational impact. Currently, students need to research the geographic location and habitat of the organisms brought into class; this may have convinced students that these descriptors are important for determining classification. Alternatively, students could describe the characters that designate each organism as belonging to a particular phylum. During the laboratory session, students should identify homologous and analogous characters among the organisms before constructing the phylogenetic tree. It should be clearly communicated that only homologous structures are considered in constructing evolutionary trees because they imply common descent. Students should understand why homology is more informative phylogenetically than analogy. Additionally, the teaching assistants should describe the structure of an evolutionary tree and model how to use these structural components to show relatedness and classification. Students could be assigned to generate phylogenetic trees with a given set of organisms and to describe their rationale as they are organizing the tree in a laboratory report. This would allow time for students to work and struggle with evolutionary trees before taking the post-survey. These modifications

would hopefully result in the improvement of students' abilities to generate trees.

Our results show that undergraduate biology students have difficulty constructing phylogenetic trees to express evolutionary relationships. Researchers and educators have been creating and testing new methodologies and pedagogical approaches to help students understand phylogenetic trees. Whereas many of these studies have focused on students' abilities to read phylogenetic trees, it was the goal of our research to determine how our students build phylogenetic trees to show relationships among organisms and if a particular laboratory practical had any effect on their abilities to do this accurately. Our research will help educators understand which mistakes students are likely to make when building phylogenetic trees as well as the types of mistakes that are alleviated with a typical college biology course. Many components, such as including a common ancestor and inferring relationships among organisms, may require more explicit instruction for students to fully understand the process of building phylogenetic trees. Furthermore, our survey and results have provided a benchmark for students' understanding and the effects of one curriculum; they can also be used to measure the effects of other educational interventions. Identifying limitations in education and evaluating the effectiveness of instruction are vital practices for the success of education. This type of research ensures that students will be receiving the best possible education.

Finally, we have developed an electronic version of this survey and are exploring automated feedback and scoring. Those interested should contact the corresponding author for details.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

The survey and study design was developed by BW. The survey scoring rubric was developed jointly by all three authors. Surveys were scored by AY and TS. AY conducted the data analysis. BW and AY participated in the drafting and editing of the manuscript. All authors read and approved the final manuscript.

Received: 13 March 2013 Accepted: 13 March 2013

Published: 24 April 2013

#### References

- Atran, S. (1999). Itzaj Maya folkbiological taxonomy: cognitive universals and cultural particulars. In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Atran, S, Medin, D, & Ross, N. (2004). Evolution and devolution of knowledge: a tale of two biologies. *Royal Anthropological Institute*, *10*, 395–420.
- Baum, DA, & Offner, S. (2008). Phylogenies and tree-thinking. *The American Biology Teacher*, *70*(4), 222–229.
- Baum, DA, Smith, SD, & Donovan, SSS. (2005). The tree-thinking challenge. *Science*, *310*, 979–980.
- Campbell, NA, & Reece, JB. (2005). *Biology* (7th ed.). San Francisco: Pearson.
- Catley, K, Lehrer R, & Reiser, B. (2005). *Tracing a prospective learning progression for developing understanding of evolution. Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievement*. Washington, DC: National Academies of Sciences.
- Catley, KM, & Novick, LR. (2008). Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. *BioScience*, *58*, 976–987.
- Coburn, WW, Gibson, AT, & Underwood, SA. (1999). Conceptualizations of: an interpretive study of 16 ninth graders' everyday thinking. *Journal of Research in Science Teaching*, *36*, 541–564.
- Coley, JD, Medin, DL, Proffitt, JB, Lynch, EB, & Atran, S. (1999). Inductive reasoning in folkbiological thought. In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Diamond, J, & Bishop, KD. (1999). Ethno-ornithology of the Ketengban people, Indonesian New Guinea. In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Goldsmith, DW. (2003). The great clade race: presenting cladistic thinking to biology majors and general science students. *The American Biology Teacher*, *65*(9), 679–682.
- Hatano, G, & Inagaki K. (1999). A developmental perspective on informal biology. In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Halverson, KL, Pires, JC, & Abell, SK. (2011). Exploring the complexity of tree thinking expertise in an undergraduate plant systematics course. *Science Education*, *95*, 794–823.
- Halverson, K. L. (2011). Improving tree-thinking one learnable skill at a time. *Evolution: Education and Outreach*, *4*, 95–106.
- Horn, MS, Leong, ZA, Block, F, Diamond, J, Evans, EM, Phillips, B, et al. (2012). Of BATs and APes: an interactive tabletop game for natural history museums. In *Proceedings of the ACM conference on human factors in computing systems (CHI '12)* (pp. 1–10). Austin, Texas: ACM Press.
- Klaassen, CWJM, & Lijnse, PL. (1996). Interpreting students' and teachers' discourse in science classes: an underestimated problem? *Journal of Research in Science Teaching*, *33*(2), 115–134.
- Lopez, A, Atran, S, Coley, JD, Medin, DL, & Smith, EE. (1997). The tree of life: universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, *32*, 251–295.
- Medin, D, & Atran, S. (1999). Introduction. In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Medin, D, & Atran, S. (2004). The native mind: biological categorization and reasoning in development and across cultures. *Psychological Review*, *111*(4), 960–983.
- Meir, E, Perry, J, Herron, JC, & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. *The American Biology Teacher*, *69*(7), e71–e76.
- Nelson, CE, & Nickels, MK. (2001). Using humans as a central example in teaching undergraduate biology labs. *Tested Studies for Laboratory Teaching: Association for Biology Laboratory Education*, *22*, 332–65.
- Nickels, MK, & Nelson, CE. (2005). Beware of nuts and bolts: putting evolution into the teaching of biological classification. *The American Biology Teacher*, *67*(5), 283–289.
- Novick, LR, & Catley, KM. (2007). Understanding phylogenies in biology: the influence of a gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, *13*(4), 197–223.
- Novick, LR, & Hurley, SM. (2001). To matrix, network, or hierarchy, that is the question. *Cognitive Psychology*, *42*(2), 158–216.
- O'Hara, RJ. (1997). Population thinking and tree thinking in systematics. *Zoologica Scripta*, *26*(4), 323–329.
- Omland, KE, Cook, LG, & Crisp, MD. (2008). Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *BioEssays*, *30*, 854–867.
- Perry, J, Meir, E, Herron, JC, Maruca, S, & Stal, D. (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE Life Sciences Education*, *7*, 193–201.
- Posner, GJ, Strike, KA, Hewson, PW, & Gertzog, WA. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education*, *66*(2), 211–227.
- Ross, N, Medin, D, Coley, JD, & Atran, S. (2003). Cultural and experiential differences in the development of folkbiological induction. *Cognitive Development*, *18*, 25–47.
- Sandvik, H. (2008). Tree thinking cannot be taken for granted: challenges for teaching phylogenetics. *Theory in Biosciences*, *127*, 45–51.

- Staub, NL, Pauw, PG, & Pauw, D. (2006). Seeing the forest through the trees: helping students appreciate life's diversity by building the tree of life. *The American Biology Teacher*, 68(3), 149–151.
- White, BT. (2009). Exploring the diversity of life with the phylogenetic collection lab. *The American Biology Teacher*, 71(3), 157–161.
- White, BT, & Yamamoto, S. (2012). Freshman undergraduate biology students' difficulties with the concept of common ancestry. *Evolution Education and Outreach*, 4, 680–687.

doi:10.1186/1936-6434-6-16

**Cite this article as:** Young et al.: Teaching undergraduate students to draw phylogenetic trees: performance measures and partial successes. *Evolution: Education and Outreach* 2013 **6**:16.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---