

RESEARCH ARTICLE

Open Access



A valid assessment of students' skill in determining relationships on evolutionary trees

Luke D. Blacquiere and William J. Hoese* 

Abstract

Background: Evolutionary trees illustrate relationships among taxa. Interpreting these relationships requires developing a set of “tree-thinking” skills that are typically included in introductory college biology courses. One of these skills is determining relationships among taxa using the most recent common ancestor, yet many students instead use one or more alternate strategies to determine relationships. Several alternate strategies have been well documented and these include using superficial similarity, proximity at the tips of a tree, or the fewest intervening nodes in the tree to group taxa.

Results: We administered interviews ($n = 16$) and pencil-and-paper questionnaires ($n = 205$), and constructed a valid and reliable assessment that measured how well students determined relationships among taxa on an evolutionary tree. Our questions asked students to consider a focal taxon and identify which of two additional taxa is most closely related to it. We paired the use of most recent common ancestor with one of three alternative strategies (i.e., similarity, proximity, or node-counting) to explicitly test students' understanding of the relationships among the taxa on each tree.

Conclusions: Our assessment enables us to identify students who are effectively distracted by an alternative strategy, those who use the most recent common ancestor inconsistently, or who are guessing in order to determine relationships among taxa. Our 18-question tool (see Additional file 1) can be used for formative assessment of student understanding of how to interpret relationships on evolutionary trees. Because our assessment tests for the same skill throughout, students who answer incorrectly, even once, likely have an incomplete understanding of how to determine relationships on evolutionary trees and should receive follow-up instruction.

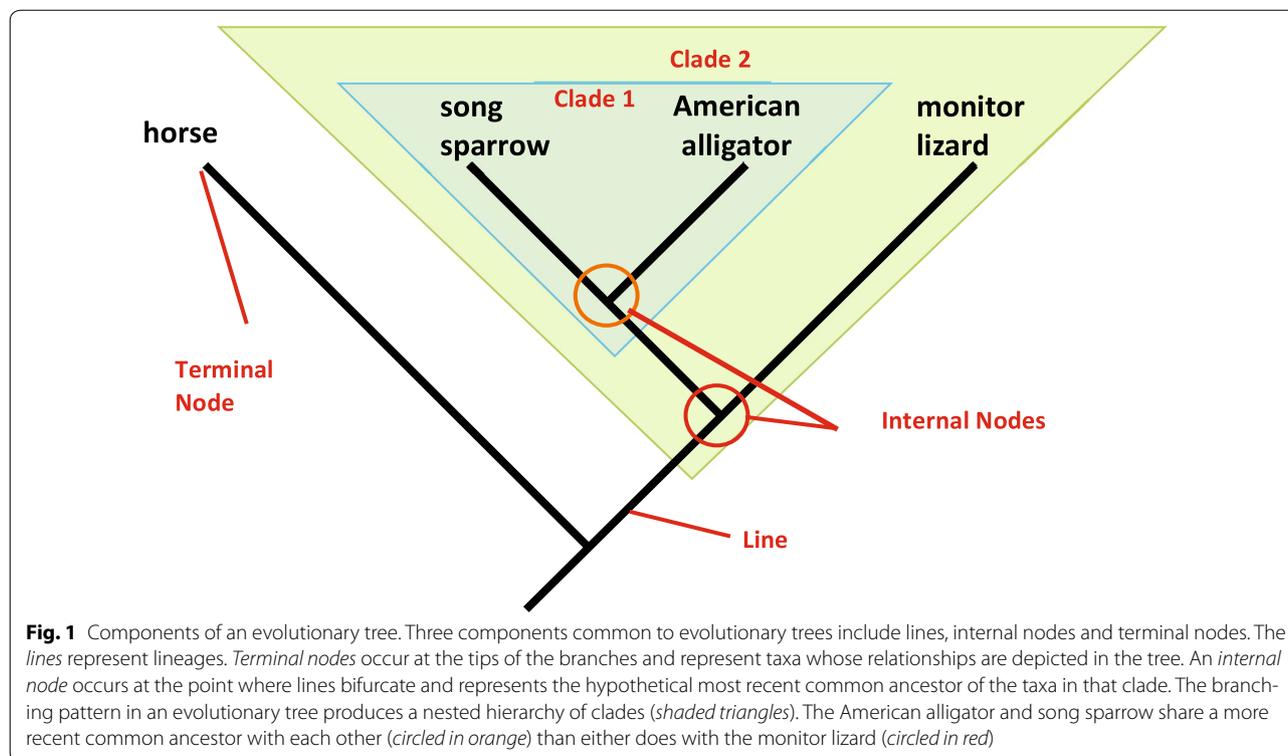
Keywords: Evolutionary trees, Tree-thinking, Cladogram

Background

Evolutionary trees, or cladograms, are branching diagrams that depict hypotheses about the relative relationships among taxa (Fig. 1). Evolutionary trees represent the visual illustration of Charles Darwin's (1859) central claim that species are related and the diversity of species is a result of descent with modification from common ancestors. Biologists use cladograms as a tool to examine evolutionary patterns of relationships among taxa and to

test hypotheses about these relationships (O'Hara 1988; Baum et al. 2005). Cladograms contain the following three major components: lines, internal nodes, and terminal nodes (Fig. 1) (Hennig 1966). The lines represent lineages of taxa. Internal nodes occur where branches split and they represent the hypothetical common ancestors of the lineages that follow; lineages splitting from an internal node are evolutionarily independent of one another. Terminal nodes occur at the ends of lines and represent taxa whose relationships are depicted in the tree. At the base of the tree is the root; time proceeds from the root to the ends of the lines. A common ancestor and all of its descendants is a monophyletic group,

*Correspondence: bhoese@fullerton.edu
California State University Fullerton, 800 N, State College Blvd, Fullerton,
CA 92831, USA



or clade, and an evolutionary tree is a series of clades arranged in a nested hierarchy (Hennig 1966; Thanukos 2009).

Interpreting evolutionary trees requires a skill-set called “tree-thinking” (O’Hara 1988). Tree-thinking is the ability to accurately interpret the relationships depicted in an evolutionary tree (O’Hara 1997; Baum et al. 2005; Baum and Offner 2008). Although many skills comprise tree-thinking (see O’Hara 1997; Baum et al. 2005), using the most recent common ancestor (MRCA) to determine relationships on a cladogram is fundamental (Hennig 1966; Novick and Catley 2013). Consider three taxa: the two taxa that are most closely related share a more recent common ancestor with each other than either does with the third taxon (i.e., they are members of a clade that does not include the third taxon) (see Fig. 1). Using MRCA to determine relationships enables students to decipher the information presented in an evolutionary tree (Meisel 2010; Novick and Catley 2013).

Although understanding how to interpret cladograms is an essential skill for identifying evolutionary relationships, problems arise as students learn to examine these diagrams (see Baum et al. 2005; Catley 2006; Meir et al. 2007; Gregory 2008; Omland et al. 2008; Sandvik 2008; Smith and Cheruvilil 2009; Morabito et al. 2010; Novick et al. 2011). The difficulties students encounter when interpreting evolutionary trees is varied. Students with

limited prior knowledge of evolutionary trees often use superficial similarity or shared habitats to determine relationships (Halverson et al. 2011). Students who have been introduced to evolutionary trees, but who have not yet mastered them, often incorrectly ascribe meaning to components of the tree that provide no useful information about the relationships of the taxa (Gregory 2008). These include implying evolutionary progression from left to right across the terminal nodes (Sandvik 2008; Novick et al. 2012), using the number of internal nodes separating taxa to determine relationships (Meir et al. 2007; Halverson et al. 2011), and determining relationships based on how close together terminal taxa are to one another (Novick and Catley 2013; Catley et al. 2013). We focused our investigation on three of these commonly reported incorrect alternative strategies: proximity, similarity, and node counting.

Sometimes students incorrectly equate proximity with relatedness; taxa that are closer to one another along the branch tips are thought to be more closely related than taxa that are more distant across the branch tips (Baum et al. 2005; Meir et al. 2007; Gregory 2008; Novick and Catley 2013; Catley et al. 2013). Reading trees as ladders of progression where each taxon evolves from the one to the left of it has been suggested as a contributing factor to the use of this incorrect strategy (Baum et al. 2005; Omland et al. 2008).

Superficial similarity is sometimes used as an alternative strategy to determine relationships among taxa (Baum et al. 2005). Although morphological similarity may provide cues to relatedness, two distantly related taxa may resemble each other due to convergence (i.e., homoplasy) or the retention of shared ancestral form (i.e., symplesiomorphy). A classic example of convergent similarity includes dolphins, which resemble sharks yet share a more recent common ancestor with other mammals. To illustrate retention of a shared ancestral form an American alligator looks more similar to a monitor lizard than a song sparrow, yet the alligator is more closely related to the sparrow than the lizard (Padian and Chiappe 1998) (Fig. 1).

Students who use the node counting strategy interpret relationships by counting the number of internal nodes separating taxa; taxa with fewer internal nodes between them are thought to be more closely related than taxa with more internal nodes separating them. This strategy arises from the false notion that internal nodes are the only place where evolution occurs (Baum et al. 2005; Meir et al. 2007; Gregory 2008) and the fewer evolutionary changes (i.e., nodes) separating taxa the closer they are related to one another.

We developed a valid and reliable assessment (see Additional file 1) that measured whether students could determine relationships among taxa on an evolutionary tree. Our tree-thinking questions asked students to consider a focal taxon and identify which of two additional taxa is most closely related to it. We paired the use of MRCA with one of three common alternative strategies (i.e., proximity, similarity, or node-counting) to test students' understanding of the relationships among the taxa on each tree. Our assessment enabled us to distinguish between students who were effectively distracted by an alternative strategy from those who accurately determined evolutionary relationships on evolutionary trees.

Methods

We developed our assessment working with students in the first-semester biology course for majors, *Evolution and Biodiversity*, at California State University, Fullerton (CSUF). CSUF is a large (~37,000 students), comprehensive, Master's granting, and Hispanic-Serving Institution, with 56.7 % female, and 43.3 % male students. CSUF serves a diverse population of students, and over 50 % are the first in their families to receive a college degree; within the College of Natural Sciences and Mathematics the ethnic composition of the students included 32 % Hispanic, 31 % Asian, 23 % white, and 2 % African American (CSUF Institutional Research and Analytical Studies for fall 2012). Students entering this course typically had completed one or two high school biology courses.

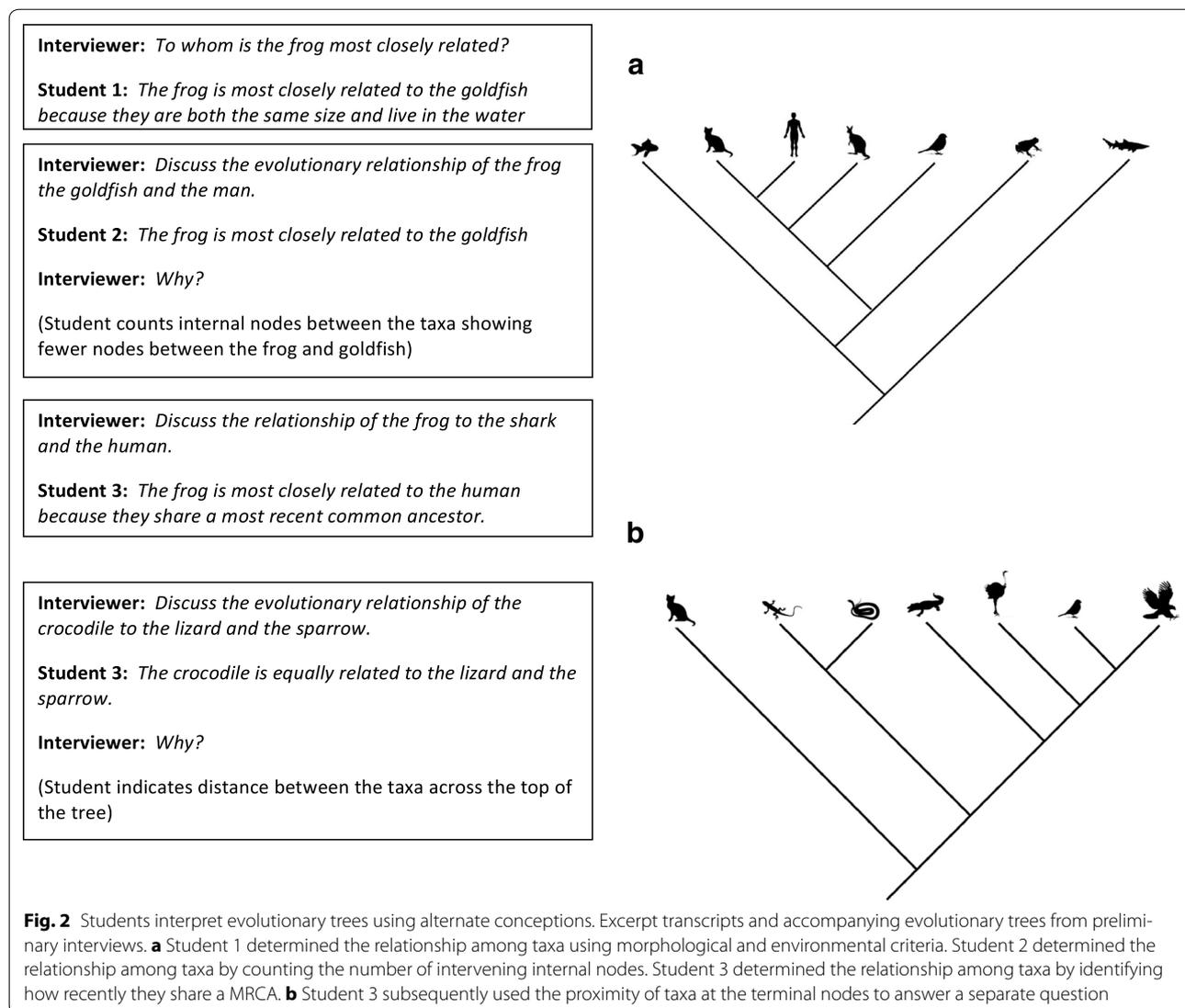
Student participation was voluntary and confidential. Student grades were not affected by participation; no penalty was assessed for non-participation. Students were apprised of the research procedures, objectives and goals and signed an informed consent form. Research was completed in compliance with California State University, Fullerton Institutional Review Board IRB HSR# 10-0397 and IRB HSR# 12-0160. Students under 18 years of age were not included in the research. Students who participated in interviews were given \$10.00 gift cards to the university bookstore or USB flash drives to compensate them for their time.

Preliminary interviews

We video-recorded preliminary one-on-one interviews using scripted open-ended questions to assess students' understanding of the components and key concepts of evolutionary trees (see Fig. 1). Questions were exploratory in nature. We presented students with sample evolutionary trees and asked them to describe the significance of the parts (lines, internal nodes, terminal nodes), the direction in which time moved, and to interpret relationships that were represented on the trees. The interviews (n = 21) began before students received instruction about evolutionary trees and ended after instruction on evolutionary trees had concluded in the course. We used information from these preliminary interviews to evaluate prior knowledge about evolutionary trees, document the strategies that were used to interpret these trees, and guide us in developing the full assessment (see examples in Fig. 2). We identified three key patterns in the preliminary interviews that informed the development of the assessment.

First, prior to receiving instruction about evolutionary trees, most students, except those who completed AP Biology in high school, did not use an evolutionary framework to interpret relatedness on cladograms. Instead, students used environmental cues to interpret cladograms (Fig. 2a) or treated the cladogram as a food web. This finding led us to focus on a post-instruction assessment because prior to instruction many students were unable to recognize or solve phylogenetic problems (see also Halverson et al. 2011). When students do not use an evolutionary framework to interpret cladograms, their answers do not provide insight into their ability to reason about evolutionary relationships.

Second, the preliminary interviews confirmed the use of alternative strategies to interpret relationships on cladograms (see Gregory 2008) and showed students regularly used three strategies, similarity, proximity and node counting (Fig. 2). We used these same three strategies (i.e., similarity, proximity, and node counting) as distracters and paired them against the correct scientific



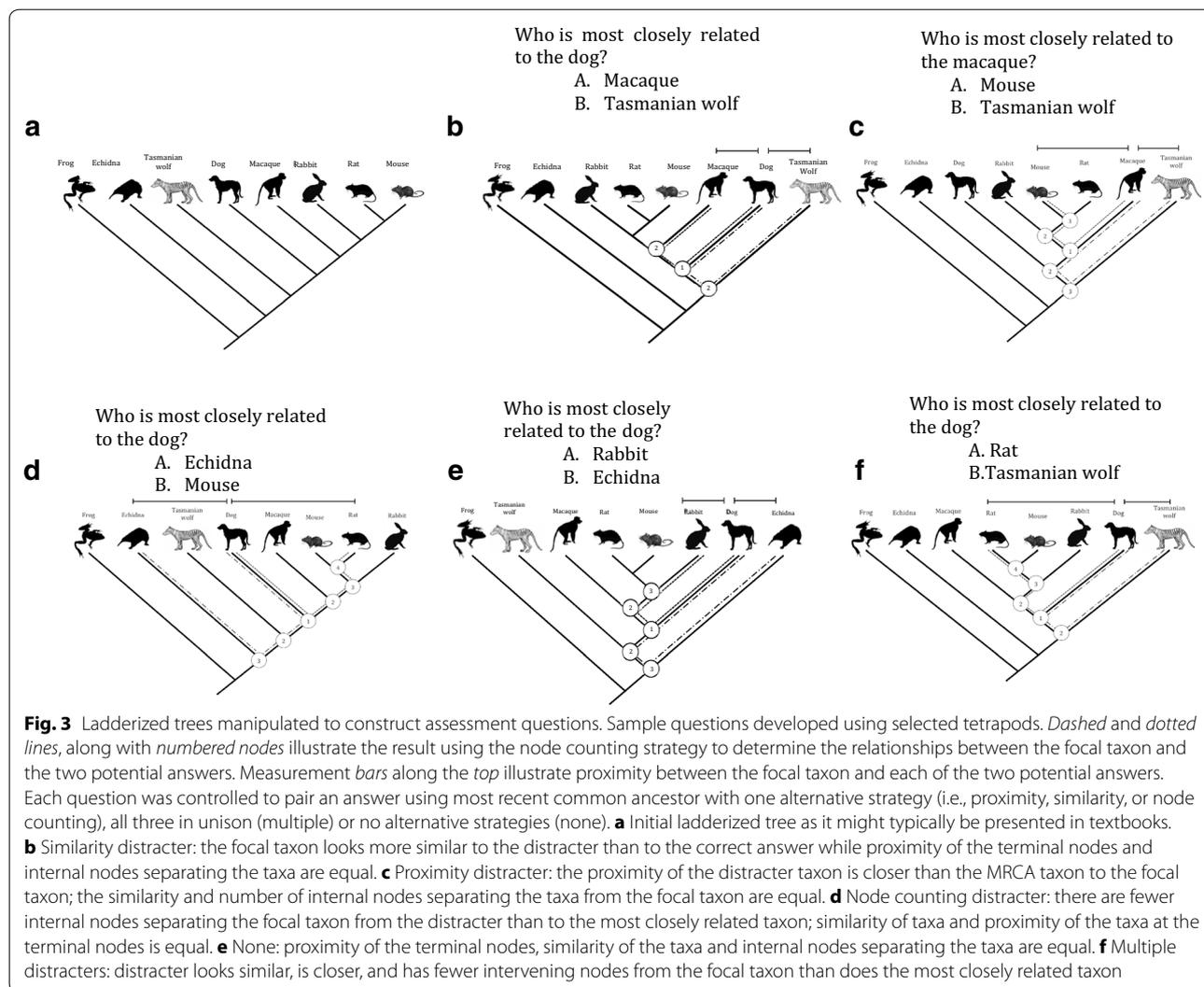
response (i.e., MRCA) to make an authentic, rigorous assessment.

Last, our findings demonstrated that individual students used multiple strategies to interpret trees. When individuals were asked to interpret the relationships of taxa on different cladograms, they did not consistently use the same strategy (see Fig. 2a versus b) suggesting to us that student interpretation strategies were flexible. Because strategies were used inconsistently, they do not meet the criteria of a misconception (Wandersee et al. 1994); while they are common across our population, they are not strongly held or stable (Hammer 1996).

Assessment development

We manipulated ladderized trees (Gregory 2008) typical in textbooks (Catley and Novick 2008) (Fig. 3a) in order to pose questions that presented two answers to choose

from, the distracter and the correct answer. Questions on the assessment asked students to determine which of two taxa was most closely related to a focal taxon. The correct answer was the taxon that shared a most recent common ancestor with the focal taxon. Paired with the correct response was a distracter based on one alternative strategy (i.e., proximity, similarity or node counting), all three alternative strategies, or that enabled students to use none of the alternative strategies to make their selection. Questions designed to exhibit one alternative strategy controlled for the use of the other two strategies in selecting an answer (Fig. 3b–d). For example, a question that used the similarity-based distracter controlled for the other two strategies (i.e., the two taxa were equidistant from the focal taxon and had the same number of intervening nodes between them and the focal taxon) (Fig. 3b). In questions designed with the distracter that



had none of the three alternative strategies, students could not use proximity, similarity or node counting strategies to determine the answer (Fig. 3e). Questions with a distracter exhibiting all three alternative strategies enabled students to use of any of the three alternative strategies in making the incorrect choice (Fig. 3f). As a result of the design, a binomial question with a forced-choice between the correct scientific strategy and an alternative strategy, incorrect answers demonstrated that students were not using MRCA to determine the relationship between taxa on each question. Because there was a 50 % chance of answering correctly on any one question, we developed an 18-question assessment that enabled us to minimize the likelihood that a student who answered all questions correctly was not using the most recent common ancestor to determine relatedness. The probability of randomly answering all 18 questions correctly is $(0.5)^{18}$.

We designed our questions using accurate representations of scientific hypotheses about the relationships of our taxa. We included taxa, from urban and suburban areas, likely to be familiar to our student population; we also designed questions with large, charismatic mega-fauna, and questions with several less-familiar taxa. A diversity of taxonomic groups including: vertebrates, mollusks, insects, echinoderms and plants (Table 1) were used in the questions. Evolutionary trees designed to include the similarity strategy required selection of focal taxa that had either homoplasy or a symplesiomorphy with the distracter.

Additional aspects of the assessment design contributed to an authentic evaluation of student understanding. Questions included trees with five, six, seven and eight taxa (Table 2) and we varied the direction of the correct answer relative to the focal taxon (i.e., to the left or right) (Table 3). All questions in this assessment used

Table 1 Number and distribution of taxa groups used for questions in the instrument

Taxa group	Number of questions	Question #'s
Vertebrate	9	1, 2, 5, 9, 11, 12, 15, 16, 17
Echinoderm	2	7, 14
Mollusk	2	8, 10
Insect	4	4, 6, 13, 18
Plant	1	3

Table 2 Number and distribution of evolutionary tree topologies used with questions in the instrument

Number of taxa topology	Number of questions	Question #'s
5	1	16
6	3	3, 7, 17
7	7	4, 6, 8, 9, 10, 11, 14
8	7	1, 2, 5, 12, 13, 15, 18

Table 3 Number and distribution of the direction of the correct answer for questions in the instrument

Direction of correct answer relative to focal taxon	Number of questions	Question #'s
Right	6	2, 4, 9, 13, 15, 18
Left	12	1, 3, 5, 6, 7, 8, 10, 11, 12, 14, 16, 17

diagonal-format evolutionary trees with the orientations of the root either on the right or on the left (Fig. 4) (Novick et al. 2012). Questions were systematically

ordered so questions using the same alternative strategy as a distracter and questions using the same group of taxa were not ordered consecutively (see Table 4). The two multiple alternative strategy questions were placed at the end of the instrument. Pictures used on the evolutionary trees were obtained from previous studies (Baum et al. 2005; Gregory 2008) and from the internet, and some of these were modified to be silhouettes.

Validity

After developing the 18 questions and accompanying evolutionary trees we piloted the assessment in an interview format with students in the first-semester biology course for majors, *Evolution and Biodiversity* (n = 16) and in a pencil and paper format with graduate student teaching assistants (n = 4) and faculty members that teach the *Evolution and Biodiversity* course (n = 3). Interviews in *Evolution and Biodiversity* showed two students who answered all questions correctly. These students used MRCA to determine evolutionary relationships on all 18 questions, while others, who did not answer all questions correctly, used a mix of strategies to determine evolutionary relationships. Faculty members were consulted about the content of the assessment after completion. We used the pilot to verify content validity and confirm that the interpretation of the questions and evolutionary trees with distracters (the choices connected to alternative strategies) were as intended in the question design.

Assessment administration

The instrument was administered to students in a pencil-and-paper format (n = 205) during the lab portion of the course. Each student was given an assessment booklet and individuals recorded their answers on a separate

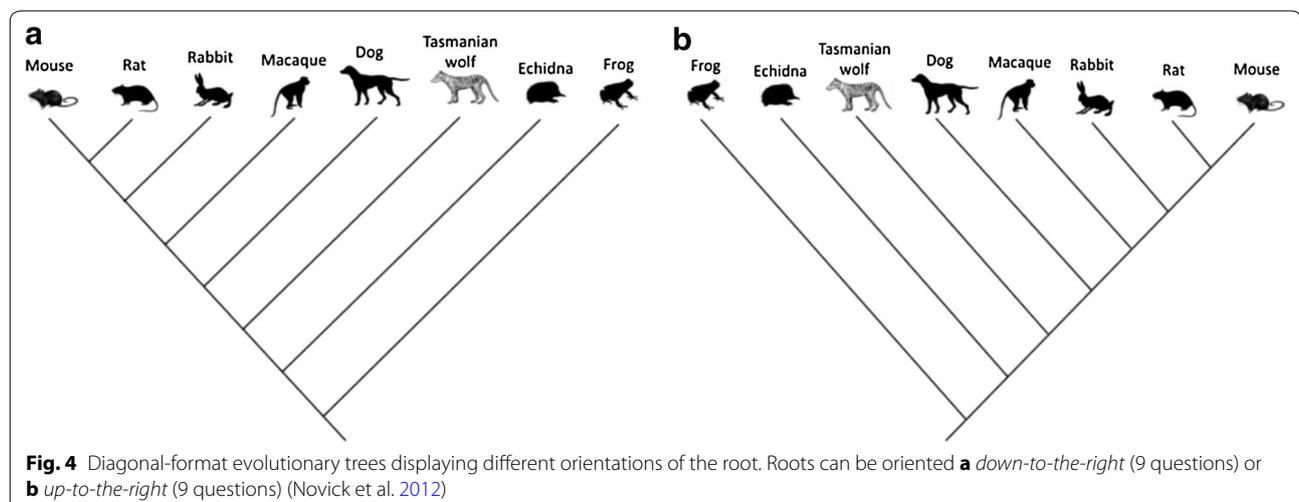


Table 4 Number and distribution of questions with three, zero, and multiple common alternative conceptions

Alternative conception	Number of questions	Question #'s
Proximity	4	3, 6, 12, 15
Similarity	4	2, 5, 9, 13
Node counting	3	1, 8, 14
No alternative conception	5	4, 7, 10, 11, 16
Multiple alternative conceptions	2	17, 18

answer sheet. Students were given unlimited time to complete the test, and typically took less than 20 min.

Analysis of assessment

Questions were analyzed for difficulty and discrimination. We measured the difficulty of each question by the proportion of students who answered the question correctly. Questions on the assessment were created with the goal of discriminating between students that use MRCA to determine relationships on evolutionary trees and students who do not. We used the point biserial method, finding the correlation between performance on an individual question and the instrument as a whole, to calculate discrimination values. Questions with good discrimination values separate students who exhibit mastery of the concept being assessed from students who do not. Discrimination values of 0.40 or higher are described as very good questions (Ebel and Frisbie 1986). The reliability of the instrument, a measure of the internal consistency, was determined using Cronbach's alpha (Cronbach 1951). We used a threshold of 0.60 for Cronbach's alpha, above which indicates strong internal consistency (Gronlund 1993).

On a separate day, the students completed Lawson's Classroom Test of Scientific Reasoning (Lawson 1978). Student scores were compared with their score on the instrument to evaluate if scientific reasoning ability was correlated with their performance on our assessment.

Results and discussion

Validity

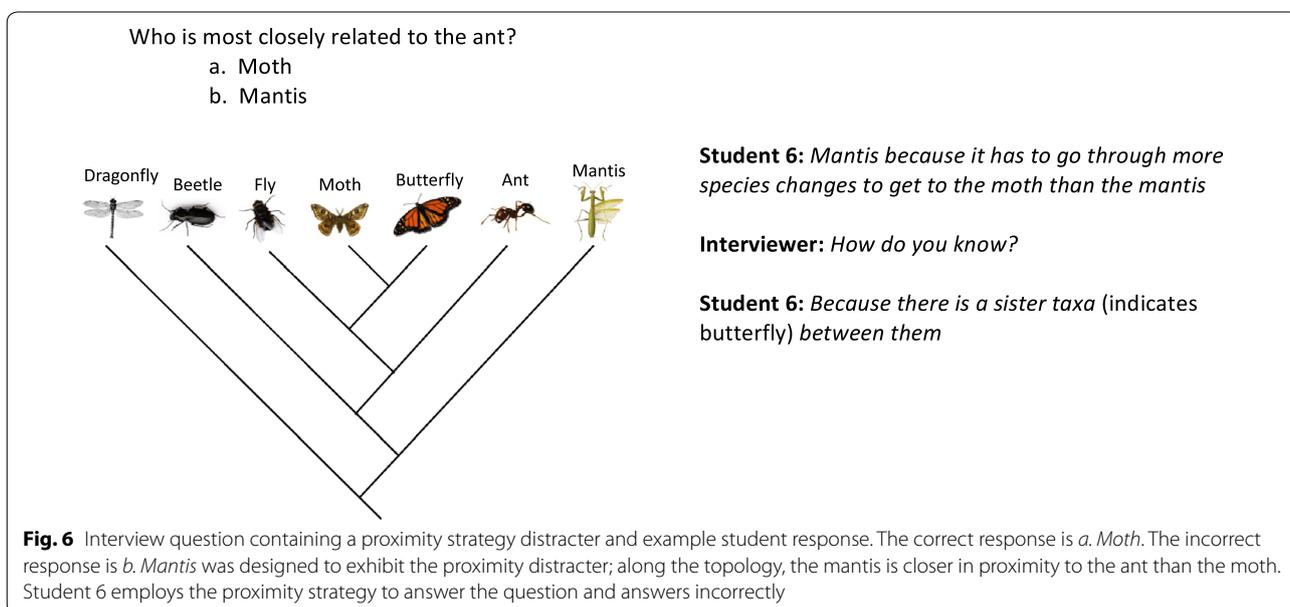
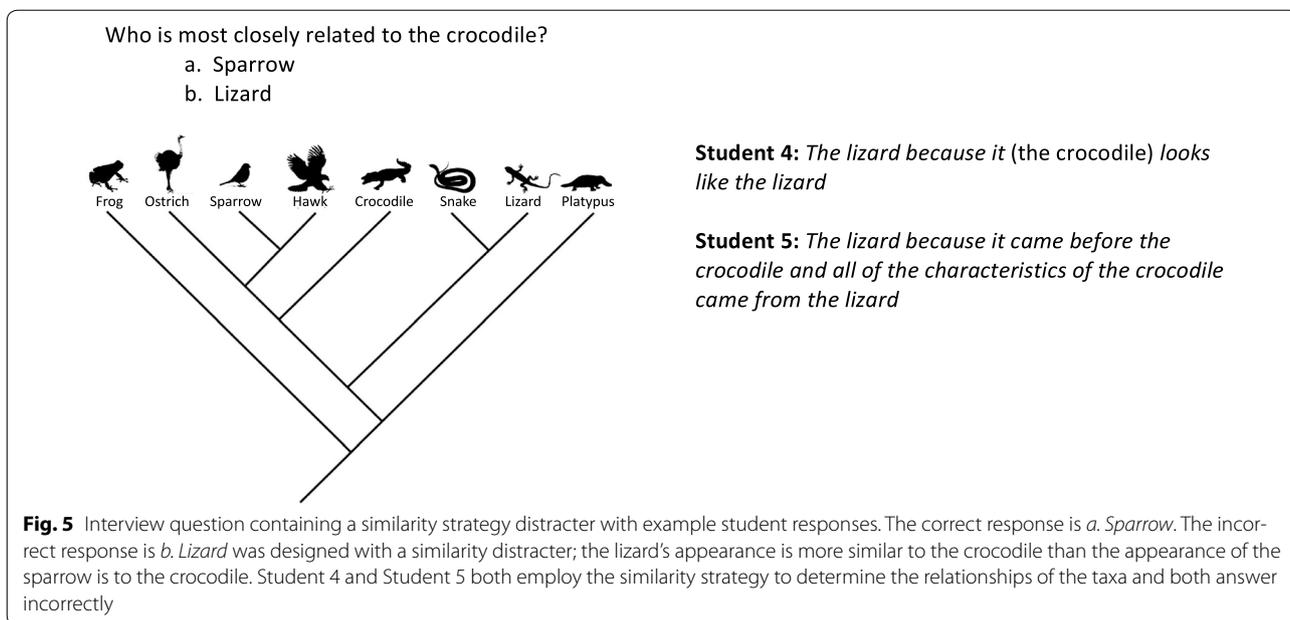
The pilot interviews demonstrated that students understood the directive of the questions. Students employing an incorrect strategy, included as a distracter on the assessment, interpreted the distracter responses as intended in the design of the assessment (Figs. 5, 6, 7). Student responses in the pilot interviews established that students who consistently used MRCA to interpret the relationships on evolutionary trees consistently answered the questions correctly (Fig. 8) and students that did not use MRCA to interpret relationships did not consistently answer questions correctly (Figs. 5, 6, 7).

Because of the assessment design, students using an alternate strategy to interpret the relationships on evolutionary trees were not expected to answer all questions incorrectly. The assessment included eleven questions containing one incorrect strategy as distracter and five questions containing an unknown or no incorrect strategy as a distracter. Students who approached a question using a strategy that had been controlled for a particular question (a strategy not incorporated as a distracter) could not use that strategy to arrive at an answer. When students could not identify a clear answer using their determined strategy, they often voiced confusion during interviews and admitted guessing in order to answer the question. When guessing on a single question, students had an equal chance in answering that question correctly or incorrectly. For example, answer choices for a question with a proximity-based distracter had the same number of internal nodes between them and the focal organism, therefore students who used node counting to determine relationships were not able to distinguish between the two choices using this strategy. Given two choices to answer the question, students have a 0.50 probability of answering correctly.

Faculty members (content experts) verified the content validity of the assessment. They verbally affirmed that the assessment tested the ability to use MRCA to interpret relationships on an evolutionary tree, and confirmed that distracters were appropriate for each question (especially for taxa included for similarity-based distracters). The validity of the test was also investigated using the group difference method (Cronbach and Meehl 1955). Because professors (content experts) and graduate student teaching associates had the construct, the ability to determine relationships on evolutionary trees, whereas many of the students in the *Evolution and Biodiversity* course did not have the construct (as verified by the interviews), professors scoring higher than students provide evidence that the assessment has construct validity. The scores of the faculty ($n = 3$) mean 0.98 with standard error (SE) 0.019 and graduate student teaching associates ($n = 4$) mean 0.96 with SE 0.0266 were higher than the scores of students in *Evolution and Biodiversity* ($n = 205$) mean 0.64 with SE 0.020 verifying the construct validity of the assessment.

Item analysis

The difficulty values for the questions on the assessment ranged from 0.46 to 0.80 (Table 5). The preferred level of difficulty for a two-response multiple-choice question is 0.75 (Thompson and Levitov 1985) and questions with difficulty values falling in the range of 0.30–0.70 are best for providing information about the differences between student understanding (Kaplan and Saccuzzo 1997).



The suite of questions in the assessment have appropriate levels of difficulty to discriminate between students who accurately determined relationships among taxa on evolutionary trees and students who did not accurately determine relationships among taxa on evolutionary trees. While difficulty varies between some questions within a distracter category, the focus of the study was to develop questions to determine whether or not a student accurately interpreted relationships among taxa on evolutionary trees. We did not explore the underlying reasons why disparity existed between some questions

within distracter categories. The reasons for the disparity among questions could include differences in the structure, topology and taxa included on the tree.

A two-way sign test showed significant differences in student performance between all distracter categories ($p < 0.05$) except between the following distracter categories: similar—none, proximity—node counting, and multiple—node counting (Table 5).

We investigated several factors that could influence student performance on questions that were unrelated to either conceptual understanding or the alternative

Who is most closely related to the snake?
 a. Hawk
 b. Platypus

Student 7: (Student counts internal nodes) *Platypus. It is the closest*

Interviewer: *How are you measuring how close they are?*

Student 7: *The nodes (student points at internal nodes). Three to the platypus. Five to the hawk.*

Fig. 7 Interview question containing a node counting strategy distracter and example student response. The correct response is *a. Hawk*. The incorrect response demonstrates a node counting distracter; counting internal nodes between taxa, fewer internal nodes separate the platypus and the snake than separate the hawk and the snake. Student 7 employs the node counting strategy to answer the question and answers incorrectly

Who is most closely related to the ant?
 a. Moth
 b. Mantis

Student 8: *The moth*

Interviewer: *Why?*

Student 8: *Because they (ant and moth) share a more recent common ancestor than the ant and the mantis*

Fig. 8 Interview question and example student response demonstrating the use of MRCA to determine the relationship of taxa. The correct response is *a. Moth* because it shares a more recent common ancestor with the ant than the mantis. Student 8 determines the relationship of the taxa using MRCA and answers the question correctly

conceptions tested. Using a two-sample t test we found no significant difference in student performance (mean number of correct responses per question \pm standard deviation) (1) on the first eight questions (139.5 ± 19.54) versus the second eight questions (128 ± 11.31) of the assessment ($p = 2.72$, $t = 1.14$, $df = 14$), (2) when the correct answer was to the right of the focal taxon (139.5 ± 24.86) or to the left of the focal taxon (126.75 ± 19.03) ($p = 0.301$, $t = 1.10$, $df = 8$), or (3) when trees were drawn with up-to-the-right orientation of the root (129.44 ± 21.81) or down-to-the-right orientation of the root (132.56 ± 22.00) ($p = 0.767$, $t = -0.0301$, $df = 16$).

Discrimination values on our assessment ranged from 0.42 to 0.76 (Table 6). Discrimination values of 0.40 or higher are described as very good questions (Ebel and Frisbie 1986), thus we had strong discrimination across the assessment. These results indicate that the instrument discriminates between students who accurately interpret relationships on an evolutionary tree from those that use an alternative strategy.

The reliability of the instrument is the degree the instrument produces consistent results. Cronbach's alpha, which measures internal consistency, was used to estimate reliability. Internal consistency estimates the extent to which items that measure the same construct

Table 5 Two-way sign test comparing student performance on distracter categories

	Proximity	Similarity	Node counting	None
Proximity (60 ± 3.42 %)				
Similarity (71 ± 3.16 %)	p < 0.001 (t = 4.89)			
Node counting (57 ± 3.46 %)	p = 0.356 (t = -1.50)	p < 0.001 (t = 6.04)		
None (69 ± 3.23 %)	p = 0.0012 (t = -5.41)	p = 0.9327 (t = 1.10)	p < 0.001 (t = -6.29)	
Multiple (53 ± 3.49 %)	p = 0.001 (t = 2.94)	p < 0.001 (t = 6.50)	p = 0.2211 (t = 1.49)	p < 0.001 (t = 6.19)

Left column: distracter categories with average performance ± standard error (205 students). Cells include p value and t-statistic (df = 204)

Table 6 Difficulty and discrimination values for the questions in the instrument

Question	Distracter	Difficulty	Discrimination
1	Node counting	0.50	0.55
2	Similarity	0.79	0.42
3	Proximity	0.66	0.54
4	None	0.80	0.53
5	Similarity	0.70	0.56
6	Proximity	0.70	0.57
7	None	0.61	0.67
8	Node counting	0.67	0.56
9	Similarity	0.60	0.60
10	None	0.77	0.59
11	None	0.60	0.69
12	Proximity	0.53	0.76
13	Similarity	0.77	0.49
14	Node counting	0.54	0.73
15	Proximity	0.52	0.75
16	None	0.66	0.70
17	Multiple	0.46	0.61
18	Multiple	0.60	0.67

have similar results. The internal consistency of the items was excellent (Cronbach’s alpha of 0.90).

Students who answered incorrectly were effectively distracted by our question design, but students who answered correctly had a 50 % chance of getting each individual question right simply by guessing. We, therefore, compared the distribution of our student scores (n = 205) on the assessment with the expected distribution of correct answers based on random chance (i.e., binomial distribution of scores on an 18 question assessment for a population of 205); student performance followed a similar pattern to the predicted distribution of scores except at the extremes (Fig. 9). For a sample of

205 participants less than one individual is predicted to answer all 18 questions correctly simply by chance; thus we concluded that the 43 students who answered all questions correctly likely used MRCA to determine evolutionary relationships. It is possible that students who answered 15–17 of the 18 questions correctly also understood the fundamentals of how to read relationships in evolutionary trees, but since our questions focus on a single concept, students who answer incorrectly, even once, may not fully understand how to read these trees.

Student scores on the instrument were significantly and positively correlated with scores on Lawson’s Classroom Test of Scientific Reasoning (r = 0.31, p < 0.001); students with higher scientific reasoning scores performed better on our assessment. Scientific reasoning is composed of inquiry, experimentation, evidence evaluation, inference and argumentation (Zimmerman 2007) a skill set that applies to evolutionary tree interpretation. While we did not measure learning gains, our results are consistent with other studies that have found a positive correlation between scientific reasoning abilities and student gains in learning science (Coletta and Phillips 2005).

We expected that students who were using one of the three alternate strategies consistently would experience some cognitive dissonance when they encountered a question that did not enable them to use that strategy (e.g., using node counting strategy on a question where the number of nodes between the focal taxon and the two choices were the same). Yet, students rarely recognized that if their strategy was correct it should work on all questions, and if their strategy wasn’t working then it was not a valid way to approach any of the questions. Students often switched strategies throughout the assessment, indicating that these strategies were not deeply seated misconceptions (see Wandersee et al. 1994), but rather, alternate approaches that should be relatively easily dispelled with additional training. Recently we have used this assessment as a diagnostic and training tool with our graduate teaching associates and undergraduate supplemental instruction leaders. The assessment has been very effective in helping us identify instructors who have problems interpreting relationships among taxa on an evolutionary tree; with relatively little additional training they master this skill fairly quickly. We find, anecdotally, that rather than learning to determine relationships in a gradual manner, students typically experience a “light-bulb” moment when they understand how to read these trees.

Our novel question design can also be adapted and used by instructors to develop their own questions using this binary, forced choice model to test for one or more alternate conceptions while controlling for the use of other strategies.

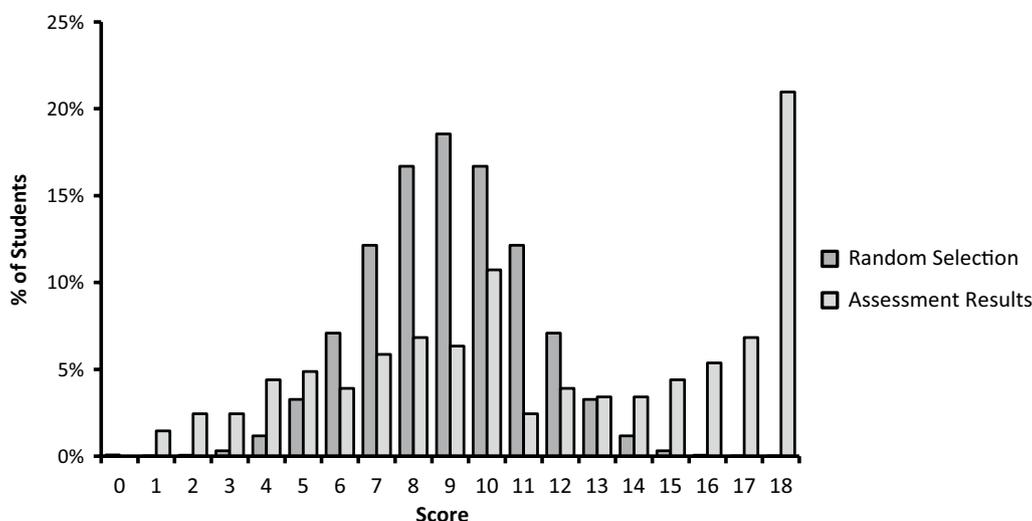


Fig. 9 Comparison of student scores on our evolutionary tree assessment containing 18 questions ($N = 205$) with the predicted binomial distribution of scores in a class of 205 if answers were randomly selected ($n = 18$, $p = 0.5$)

Conclusions

Understanding relationships of taxa on evolutionary trees (a fundamental component of tree-thinking) is a difficult skill for students to master. We developed an assessment to measure students' aptitude in interpreting taxa relationships on evolutionary trees to inform instructors about the students' level of understanding and provide students with feedback about their own understanding.

To provide an accurate and effective measure of students' aptitude, questions on the assessment were designed with authenticity. First, all of the evolutionary trees include accurate representations of scientific hypotheses about relationships of taxa. Second, a variety of taxonomic groups were represented in the evolutionary trees. Third, a variety of tree structures were included. Many different branching patterns both ladderized and non-ladderized were incorporated and the number of taxa along the topology was varied. Fourth, common alternative conceptions were used as distracters. The combination of these four design features results in a rigorous test of students' ability to interpret relationships among taxa on evolutionary trees.

The analysis of the assessment reported demonstrates that students understood the directive of the question. Content and construct validity was verified by content experts and the group difference method, respectively. The reliability, determined by Cronbach's alpha, was excellent. The difficulty and discrimination values of questions indicate that the instrument discriminates between students who interpret relationships on an evolutionary tree using how recently taxa share a

common ancestor and students that use an alternative strategy.

Additional file

Additional file 1. Valid, 18 question, assessment used to measure student ability to determine relationships on evolutionary trees. The assessment is composed of 18 forced-choice binomial questions with the following distracter types: proximity, node-counting, similarity, none, or multiple (see "Methods" and Table 6).

Abbreviation

MRCA: most recent common ancestor.

Authors' contributions

Conceptual idea was first identified by WH. WH and LB designed and developed questions. Data were collected and analyzed by LB. The manuscript was written together by LB and WH. Both authors read and approved the final manuscript.

Acknowledgements

Jennifer Burnaford gave invaluable feedback about the design of the instrument.

Sean Walker aided in the statistical analysis. Shayna Foreman provided key suggestions to question design. The authors thank Asha Mada, Austin Xu, Bryce Renfeldt, Hetal Raval, Tejal Petal and the other members of the Hoese lab for their feedback on questions. Research supported by NSF DUE 0633262 to W. J. Hoese. This project benefitted from conversations at the Tree Reasoning in Evolution Education (TREE) workshop at the National Evolutionary Synthesis Center.

Competing interests

The authors declare that they have no competing interests.

Received: 6 January 2016 Accepted: 14 April 2016

Published online: 23 May 2016

References

- Baum DA, Smith SD, Donovan SS. The tree-thinking challenge. *Science*. 2005;310(5750):979–80.
- Baum DA, Offner S. Phylogenies & tree-thinking. *Am Biol Teach*. 2008;70(4):222–9.
- Catley KM. Darwin's missing link—a novel paradigm for evolution education. *Sci Educ*. 2006;90(5):767–83.
- Catley KM, Novick LR. Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. *Bioscience*. 2008;58(10):976–87.
- Catley KM, Phillips BC, Novick LR. Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: an entry point to understanding evolution. *Res Sci Educ*. 2013;43(6):2327–48.
- Coletta VP, Phillips JA. Interpreting FCI scores: normalized gain, reinstruction scores, and scientific reasoning ability. *Am J Phys*. 2005;73(12):1172–9.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281.
- Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle of life. London: Murray; 1859.
- Ebel RL, Frisbie DA. Essentials of educational measurement. Englewood Cliffs: Prentice-Hall; 1986.
- Gregory TR. Understanding evolutionary trees. *Evol Educ Outreach*. 2008;1(2):121–37.
- Gronlund NE. How to make achievement tests and assessments. 5th ed. Boston: Allyn & Bacon; 1993.
- Halverson KL, Pires JC, Abell SK. Exploring the complexity of tree thinking expertise in an undergraduate plant systematics course. *Sci Educ*. 2011;95(5):794–823.
- Hammer D. Misconceptions or p-prims: how may alternative perspectives of cognitive structure influence instructional perceptions and intentions. *J Learn Sci*. 1996;5(2):97–127.
- Hennig W. Phylogenetic systematics. Urbana: University of Illinois Press; 1966.
- Kaplan RM, Saccuzzo DP. Psychological testing: principles, applications, and issues. 4th ed. Pacific Grove: Brooks/Cole; 1997.
- Lawson AE. The development and validation of a classroom test for formal reasoning. *J Res Sci Teach*. 1978;15(1):11–24.
- Meir E, Perry J, Herron JC, Kingsolver J. College students' misconceptions about evolutionary trees. *Am Biol Teach*. 2007;69(7):e71–6.
- Meisel RP. Teaching tree-thinking to undergraduate biology students. *Evol Educ Outreach*. 2010;3(4):621–8.
- Morabito NP, Catley KM, Novick LR. Reasoning about evolutionary history: post-secondary students' knowledge of most recent common ancestry and homoplasy. *J Biol Educ*. 2010;44(4):166–74.
- Novick LR, Shade CK, Catley KM. Linear versus branching depictions of evolutionary history: implications for diagram design. *Top Cogn Sci*. 2011;3(3):536–59.
- Novick LR, Stull AT, Catley KM. Reading phylogenetic trees: the effects of tree orientation and text processing on comprehension. *Bioscience*. 2012;62(8):757–64.
- Novick LR, Catley KM. Reasoning about evolution's grand patterns college students' understanding of the tree of life. *Am Educ Res J*. 2013;50(1):138–77.
- O'Hara RJ. Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Syst Biol*. 1988;37(2):142–55.
- O'Hara RJ. Population thinking and tree thinking in systematics. *Zool Scr*. 1997;26(4):323–9.
- Omland KE, Cook LG, Crisp MD. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *BioEssays*. 2008;30(9):854–67.
- Padian K, Chiappe LM. The origin and early evolution of birds. *Biol Rev*. 1998;73(1):1–42.
- Sandvik H. Tree thinking cannot taken for granted: challenges for teaching phylogenetics. *Theory Biosci*. 2008;127(1):45–51.
- Smith JJ, Cheruvilil KS. Using inquiry and tree-thinking to "March through the animal phyla": teaching introductory comparative biology in an evolutionary context. *Evol Educ Outreach*. 2009;2(3):429–44.
- Thanukos A. A name by any other tree. *Evol Educ Outreach*. 2009;2(2):303–9.
- Thompson B, Levitov JE. Using microcomputers to score and evaluate test items. *Coll Microcomput*. 1985;3(2):163–8.
- Wandersee JH, Mintzes JJ, Novak JD. Research on alternative conceptions in science. In: Gabel DL, editor. Handbook of research on science teaching and learning. New York: MacMillan; 1994. p. 177–210.
- Zimmerman C. The development of scientific thinking skills in elementary and middle school. *Dev Rev*. 2007;27(2):172–223.

Submit your manuscript to a SpringerOpen® 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
