


Developing the BETTSI: A tree-thinking diagnostic tool to assess individual elements of representational competence

Kristin P. Jenkins,^{1,2,3}  Louise Mead,⁴ David A. Baum,⁵ Kristy L. Daniel,⁶ Carrie J. Bucklin,⁷ E. Austin Leone,⁸ J. Phil Gibson,⁹ and Erin Naegle¹⁰

¹BioQUEST, Raymond, New Hampshire 03077

²Current Address: TIDES, University of Texas at Austin, Austin, Texas 78712

³E-mail: kristin.jenkins@austin.utexas.edu

⁴BEACON and Department of Integrative Biology, Michigan State University, East Lansing, Michigan 48823

⁵Department of Botany, University of Wisconsin, Madison, Wisconsin 53706

⁶Department of Biology, Texas State University, San Marcos, Texas 78666

⁷Department of Biology, Southern Utah University, Cedar City, Utah 84720

⁸Department of Integrative Biology, Oklahoma State University, Stillwater, Oklahoma 74078

⁹Department of Biology, University of Oklahoma, Norman, Oklahoma 73019

¹⁰Department of Biology, Columbia College, Sonoma, California 95370

Received October 12, 2021

Accepted December 29, 2021

Phylogenies are a ubiquitous visual representation of core concepts in evolutionary biology and it is important that students develop an ability to read and correctly interpret these diagrams. However, as with any representation of complex disciplinary information, learning to correctly interpret phylogenies can be challenging, requiring that a diversity of educational strategies be deployed. Representational competence is the ability to develop and effectively use abstract representations. Accurately interpreting a phylogenetic tree as a presentation of evolutionary relationships requires that students develop general representational competence as well as knowledge of specific technical aspects of tree interpretation, such as knowing the graphical components of trees and what they represent. Here, we report on the development of a basic diagnostic tool of students' representational competence and technical skills with phylogenies, the Basic Evolutionary Tree-Thinking Skills Instrument (BETTSI). This short, multiple-choice instrument was designed to provide instructors with a quick diagnostic of students' ability to read and interpret phylogenies. It has been checked for reliability and validity and provides a convenient formative and summative assessment of students' understanding of evolutionary trees.

KEY WORDS: Evolution, introductory level, phylogeny, representational competence, tree thinking.

Background

PHYLOGENETIC TREES

For over 200 years, philosophers and scientists from around the globe have grappled with the idea of how organisms change over time, what we now call evolution, and how to represent this change graphically (Malik et al. 2018; Lennox 2019). A vari-

ety of images have been used to represent these ideas including Aristotle's *Scala Naturae* and Haeckel's intricate genealogical trees (Daynat 2003; Lennox 2019). Although Haeckel coined the term "phylogeny," the underlying idea of his drawings was similar to ladder representations, which show living organisms rising in status to the pinnacle of evolution, usually envisioned

as humans. This ladder-of-progress view of evolution remains a common misconception about the evolutionary process today (Omland et al. 2008).

The term “phylogeny” or phylogenetic tree now has a very clear scientific definition that distinguishes it from these earlier images. In modern scientific discussions, phylogenetic trees are a representation of proposed evolutionary relationships, and specifically a representation of common ancestry. Darwin’s original drawing of a tree (Barrett et al. 1987) is a phylogeny in this sense, because it presents the evolution of lineages that have diverged over time from common ancestors. Phylogenetics, “the field of study concerned with inferring the evolutionary relationships of living and extinct taxa and using the information to learn about patterns and processes of evolution (Baum and Smith 2013),” is central to the modern field of evolutionary biology. This field has its roots in Willi Hennig’s conceptual framework and methodology for phylogenetic analysis known as cladistics (Hennig 1950), which emerged from much lively scientific debate in the 1960s and rapidly became established as the most widely accepted approach to investigating and representing evolutionary history (Wiley et al. 1991). Modern phylogenetics built on Hennig’s framework to allow scientists to use multiple sources of evidence, combined with sophisticated mathematical and statistical tools, to infer the evolutionary relationships of different species as well as the evolutionary history of the characteristics of those species. The emergence of molecular data added a powerful line of evidence that was particularly amenable to statistical analysis. Thanks to the development of more powerful computers, new software tools, and enhanced statistical methods (Maddison and Maddison 1989; Swofford 1989), the last few decades have seen an explosion of phylogenetic knowledge.

Advances in phylogenetic science also resulted in a greater appreciation for the challenge of interpreting, and the very real potential for misinterpreting, phylogenetic trees (Maddison 1997; O’Hara 1997; Huelsenbeck et al. 2001; Swofford et al. 2001). The concept of “tree thinking” refers to a set of skills needed to properly interpret and use phylogenetic information, including familiarity with the terminology (e.g., branch, root, node), the ability to understand how relationships are represented in a phylogenetic tree, and the ability to make predictions on the basis of those relationships. As phylogenetic trees have become a ubiquitous visual in biological research papers and textbooks, it has become clear that they are often misinterpreted (Baum et al. 2005; Omland et al. 2008). The modern use of phylogenetic trees emerged so rapidly that most researchers learned how to construct and interpret trees as they used them, without formal training. As a result, explicit instruction on reading and interpreting trees has not been codified in the biology curriculum and is often overlooked in a crowded syllabus. This can result in a great deal of confusion as students enter the classroom with strong preconcep-

tions about evolution and biases in the interpretation of visual representations of relationships.

There has been considerable progress since attention was first brought to the challenge of teaching tree thinking in a talk at the 2005 Evolution Meeting in Fairbanks, Alaska, and in a follow-up paper that year (Baum et al. 2005). Since then, many underlying cognitive and pedagogical issues associated with interpreting phylogenetic trees have been identified (see Table 1). The Education Symposium at the 2021 annual meeting of the Society for the Study of Evolution, titled “Tree Thinking: Have We Met the Challenge?,” explored diverse aspects of tree thinking, including cognitive psychology, pedagogical approaches, and classroom activities for teaching tree-thinking skills. A consistent theme during the symposium and other recent educational work is that it is essential for instructors to be aware of the specific challenges inherent in mastering tree thinking and to be able to evaluate students’ tree-thinking skills. For this reason, it is critical that we develop valid tree-thinking assessments that can then be used to rigorously evaluate the effectiveness of various instructional strategies.

Substantial research has described how students struggle with each aspect of using phylogenetic trees. Appropriate use of trees requires students to first *read* the tree correctly, understanding the basic topology of the evolutionary “map” (Halverson 2011), and then to *interpret* the tree, layering in the evolutionary concepts represented by the tree topology to provide meaning to the diagram (Halverson 2010; Maroo and Halverson 2011). Tree thinking is an application and demonstration of representational competence—how a person uses a variety of perceptions of reality to make sense of, and demonstrate understanding with, visual tools. Representational competence can change with the difficulty of the task (Barnea 2000) and is significantly context specific (Halverson and Friedrichsen 2013). Research on students’ representational competence in the area of tree thinking has identified several prior and alternative conceptions associated with tree thinking, such as student’s prior knowledge about organisms, influencing their interpretation of the evolutionary relationships represented in phylogenies (Halverson et al. 2009, 2011; Halverson 2011; Matuk and Uttal 2012; Ainsworth and Saffer 2013; Halverson and Friedrichsen 2013; Novick and Catley 2014; Schramm and Schmiemann 2019). In addition, at least seven critical elements influence representational competence development for tree thinking (Table 1). For example, researchers have demonstrated that without explicit instruction in tree reading, students who learned to read in English tend to apply a left-to-right approach to reading trees (Novick et al. 2012). Additionally, orientation matters. When trees are presented vertically, with the root at the bottom, students often use the familiar left-to-right reading approach, leading to the incorrect practice of reading across the tips (Gregory 2008). Simply tipping a tree

Table 1. Research has demonstrated seven different categories of elements that can shape the development of tree-reading skills in students.

Elements shaping students' tree-reading ability and skill development	References
Explicit instruction in tree reading	Gendron 2000; Halverson 2008; Halverson 2010; Meisel 2010; McLaurin et al. 2013; Young et al. 2013
Practice rotating and comparing phylogenetic diagrams	Maroo and Halverson 2011; Ainsworth and Saffer 2013
Orientation and style of the phylogenetic diagram	Catley and Novick 2008; Catley et al. 2010; Halverson et al. 2013
Task order	Halverson et al. 2013
Experience constructing phylogenetic diagrams	Halverson 2011
Student acceptance and/or understanding evolutionary theory	Morabito et al. 2010; Evans et al. 2012; Eddy et al. 2013; Walter et al. 2013; Gibson and Hoefnagels 2015; McCullough et al. 2020
Conceptual approach to tree-phylogenetic instruction	Meir et al. 2007; Gregory 2008; Halverson et al. 2011; Gibson and Hoefnagels 2015

horizontally can mitigate this problem (Novick et al. 2012). Instructional materials that present trees in three dimensions, or without restricting the order of the tips, also reduce tip-reading practices (Halverson 2010; McLaurin et al. 2013).

Much of the body of research exploring students' struggles with tree thinking stemmed from a quiz in Baum et al.'s (2005) paper "The tree thinking challenge," which invited scientists to evaluate their own tree-thinking skills. A Google Scholar search did not find the term "tree thinking" in the scientific literature until 1988 (O'Hara 1988). From 1988 through 2005, fewer than 600 instances of the phrase "tree thinking" were returned in a Google Scholar search. After the Tree-Thinking Challenge paper was released, however, the term is found in some 2600 items. This includes a wave of new modules and resources that have been developed to engage students with reading and constructing phylogenies (e.g., Goldsmith 2003; Halverson 2010; Baum and Jenkins 2020).

With greater understanding of the cognitive processes involved in tree thinking, the need for instruments that give precise and informative feedback about students' tree-thinking skills became evident. In response, we have developed the Basic Evolutionary Tree-Thinking Skills Instrument (BETTSI), a short, multiple-choice assessment. Our goal in developing the BETTSI was to allow rapid identification of students' mastery of tree-thinking skills and concepts so that instructors focus attention on areas in which students need more instruction, potentially freeing up class time for a deeper exploration of other aspects of evolution that can be tackled once students are fluent with phylogenies.

EARLY CONSIDERATIONS AND DIFFERENT APPROACHES FOR ASSESSING TREE THINKING

Baum et al. (2005) highlighted some of the issues associated with tree thinking and included a Tree-Thinking Quiz as a supplement. The quiz was generated by the authors as a way to "evaluate your tree-thinking skills" and consisted of a set of 10 basic questions as well as a set of more complex questions based on examples from the literature. The intention of the Tree-Thinking Quiz was to raise awareness of the challenges associated with tree thinking. Likewise, Baum and Smith (2013) wrote a series of chapter quizzes for their book, *Tree Thinking*, which were intended to reinforce learning. However, none of these prior quizzes were carefully designed to survey and diagnose specific common problems with tree thinking.

Naegle (2009) developed a more formal assessment of undergraduate biology students' tree-thinking skills, the Tree-Thinking Concept Inventory (TTCI). The TTCI consisted of 26 multiple choice items and three experiential/ability questions. Naegle established the validity of the TTCI by comparing responses on the instrument to responses collected from student interviews. Calculation of Cronbach's Alpha score ($\alpha < 0.64$) indicated that the instrument had low reliability ($\alpha < 0.7$). Furthermore, because the questions on the TTCI are multiple-choice questions with correct responses, rather than Likert-scale questions, the Cronbach's Alpha test is unable to accurately predict the consistency of responses. Later versions of the TTCI were analyzed using the *Kuder-Richardson 20* (ρ_{KR20}) method to determine reliability. The KR-20 takes into account the

dichotomous nature of multiple-choice tests, whereas Cronbach's alpha assumes all answer responses are equally correct (Bodner 1980; Kline 2005). After the first round of KR20 Testing, the TTCI was evaluated by two experts for analysis of instrument construct validity (Moskal and Leydens 2000) and the TTCI was reduced to 16 items. Additionally, the answer choices for each question were reformatted to ensure there was only one correct answer available to make Scantron-type scoring possible. However, even the modified TTCI targeted multiple conceptions, not just tree thinking, and the content was more suited to upper-level college students rather than introductory students. The TTCI was never published and is not readily available for educational use, leaving faculty without a simple way to assess students' tree-thinking ability.

Although the goal of instruments may be different, the development of any instrument should include an examination of the instrument's ability to provide valid and reliable results. Measures of validity and reliability are critical for ensuring that the instrument is precise, accurate, and consistent to build confidence that it is capturing the intended data. Evidence that an instrument provides valid results suggests the variable being measured by the instrument (in our case, concepts and skills associated with thinking about phylogenetic trees) accurately represents the construct or item of interest. Evidence that an instrument provides reliable results suggests the instrument gives consistent results when implemented under similar circumstances. There are multiple forms of evidence for reliability (e.g., stability, internal consistency, interrater reliability) and validity (e.g., content, internal and external structure, generalization). Instruments that have been vetted for reliability and validity are more useful to instructors (and education researchers), as they are more likely to provide accurate insight regarding students' understanding (Campbell and Nehm 2013). The Tree-Thinking Challenge paper has been accessed over 22,000 times since its publication, indicating a strong interest in assessing tree-thinking skills; however, the original Tree-Thinking Quiz provided no evidence of validity or reliability. Naegle's TTCI and more recent assessments (i.e., Kummer et al. 2019) have been tested for validity and reliability, but cover a range of evolutionary concepts related to tree thinking. Other tree-thinking assessments, such as Blacquiere and Hoese (2016) focused on a very specific subset of common strategies used by novice learners in identifying relationships based on identifying the most recent common ancestor. Our goal was to create a tool that faculty could use in introductory classes to identify a wider range of common problems students may have when learning tree-thinking skills. The Basic Evolutionary Tree-Thinking Skills Instrument (BETTSI) was designed to be short, easily graded, and focused on basic tree-thinking skills. Here, we describe the development, validation, and reliability testing of this diagnostic.

Methods

DEVELOPMENT OF THE BETTSI

We identified concepts and skills in the published literature important for reading and interpreting phylogenetic trees (see Table 1) and then classified these concepts as either basic or advanced according to how they related to representational competence (Halverson et al. 2011). We then selected questions from the Tree-Thinking Quiz I (Baum et al. 2005) that explicitly addressed these concepts, and refined the selected questions to focus on a single concept. We also wrote new questions for basic concepts not addressed by the original Tree-Thinking Quiz. Answer options included common alternative conceptions and frequent tree-reading errors.

For example, the first question in the Baum et al.'s (2005) Tree-Thinking Quiz is shown in Figure 1. The question asks the reader to determine whether two organisms are more closely related to one another than a third organism. In the BETTSI, this question is designed to reveal the common mistake students make of reading across the tips. The question has been modified to reduce distractors, including changing the tree to a bracket formation and adding a directional arrow for time. These changes reduce the likelihood that students are making other mistakes, such as reading along the backbone line of the diagonal tree (Novick and Cately 2007). These changes are consistent across the questions, reducing the cognitive load on students to allow them to focus on the problem of reading and interpreting the tree, rather than figuring out how to read different styles of trees, or in which direction time is moving. The exception to this is question 13 that is designed to reveal whether students understand that the style of the tree does not change the interpretation.

Results

TESTING AND REVISIONS

The BETTSI went through three rounds of revision, and each version was tested for validity and reliability (Table 2). We recruited all participants and collected data following IRB-approved guidelines from University of Southern Mississippi (CH2-13092301) and Texas State University (2016Z3809 and 20160725U2504). We established both construct and face validity for each version of the BETTSI. Construct validity was established by having five phylogenetics experts write and revise the questions to ensure that each question appropriately addressed the targeted concept. We established face validity for each instrument version, which demonstrates that readers understand what the question is asking, by testing the readability of instrument items with introductory biology students. We asked a group of students to evaluate the overall appearance, structure, and wording of the instruments. During interviews, students indicated

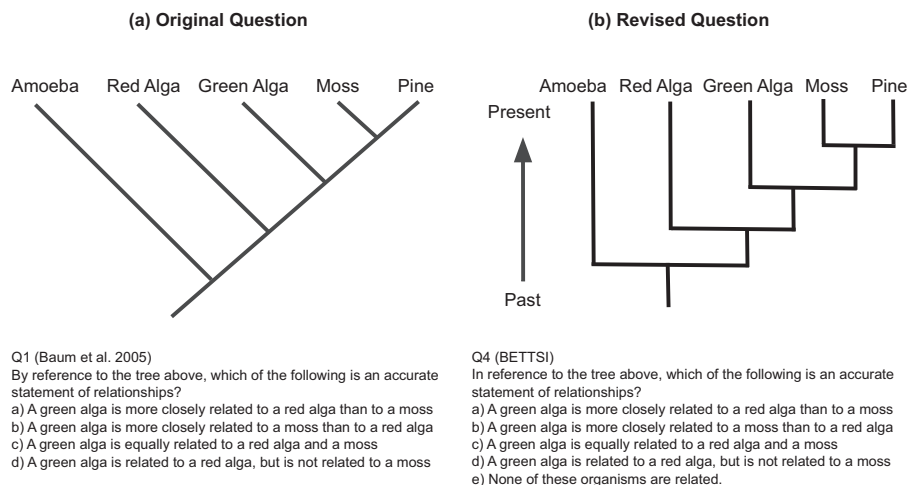


Figure 1. Comparison of a question from the Tree Reading Quiz (A: Original Question) and the modified version of the question in the BETTSI (B: Revised Question).

Table 2. Comparison of face-validity testing, and reliability testing across BETTSI versions.

Instrument version	Number of content questions	Number of students: Face validity	Number of students: Reliability	Internal consistency (ρ_{KR20})	Split-half reliability (r_{kk})
BETTSI 1.0	9	66	329	0.48	0.500
BETTSI 2.0	10	37	198	0.72	0.745
BETTSI 3.0 (Final)	11	46	89	0.75	0.789

that having a unique tree for each question increased the cognitive load of the assessment, not allowing them to focus on each question as easily. These students were not the same students whose responses were used for the reliability testing. For the reliability analysis, we ran internal consistency analyses calculating the KR-20 score for test/retest reliability to determine if the patterns in student responses for each question were correlated between the pre- and posttest. Additionally, we ran a split-half reliability analysis (Spearman-Brown Prophecy Formula) to determine if two randomly selected parts of the instrument were correlated and with responses consistent with each half.

BETTSI 1.0

Our initial instrument consisted of a nine-item, multiple-choice questionnaire to assess tree-thinking skills, and three additional items about student confidence and familiarity with reading phylogenies. We administered BETTSI 1.0 to 329 introductory biology students (129 majors and 200 nonmajors) as pre-/post-assessment with explicit instruction in tree thinking. We also established face validity by having an additional 66 students provide feedback on the wording of the questions. The BETTSI 1.0 failed both reliability ($\rho_{KR20} = 0.48$) and validity testing due to

multiple points of confusion about the diagrams and what the items were asking.

BETTSI 2.0

The second iteration of the instrument, BETTSI 2.0, consisted of a 10-item, multiple-choice questionnaire that maintained the same three confidence and familiarity questions. In several cases, we replaced the multiple-choice option of “I don’t know” with a likely alternative conception response, forcing students to select an alternative conception or the correct answer. We reordered the questions to provide better established task order (Halverson et al. 2013), specifically starting with a tree-reading skill followed by an interpretation question. We simplified all the phylogenies based on the face validity student responses to reduce cognitive load and limit inadvertent distractors. We revised questions to ensure each addressed only a single tree-thinking concept. Specifically, we divided one of the tree-comparison questions into two separate questions so that each question assessed a single concept. This additional question addressed the concept that tree orientation or format does not alter how the tree is interpreted.

Table 3. Line-item analysis difficulty and discrimination table.

Discrimination	Difficulty		
	Easy (0.85-1.0)	Medium (0.51-0.84)	Hard (0.0-0.50)
Poor (<0.1)			
Fair (0.1-0.3)		Q9	Q5
Good (>0.3)	Q14	Q4,Q7,Q8,Q12	Q6,Q10,Q11,Q13

We tested BETTSI 2.0 by administering it to 198 introductory biology students and retested the face validity of the revised instrument with an additional 37 students. The reliability tests indicated the BETTSI 2.0 was acceptable, with a $\rho_{KR20} = 0.72$ (Salkind 2010). However, we identified a few minor issues we wanted to address, including clarifying question wording to increase face validity and converting all images to black and white to improve print quality.

BETTSI 3.0

The third iteration of the BETTSI included the same three confidence and familiarity questions followed by 11 content items, each addressing one tree-thinking skill. (See Appendix Tables A1 and A2 for the complete list of concepts mapped to specific items.) In BETTSI 3.0, the assessment was modified so that students were asked two questions about most of the tree diagrams rather than offering a new tree for each question, to help reduce cognitive load. We found that BETTSI 3.0 was both reliable ($\rho_{KR20} = 0.75$) and had strong face validity (see Table 2). In addition, we ran a line-item analysis to assess the difficulty and discrimination of each item. To assess item difficulty, we calculated an Item Difficulty Index value for each item (Table 3). These values represent the proportion of participants who answered the item correctly. Difficulty index values less than 0.5 represent high difficulty, values between 0.5 and 0.85 represent medium difficulty, and values greater than 0.85 represent low difficulty. The BETTSI presents a full range of question difficulty helping to reveal a number of common misconceptions in tree thinking.

We also calculated an Item Discrimination Index value for each item (Table 3). These values represent how well each item is able to discriminate between participants who have mastered the content and those who have not. Discrimination index values less than 0.1 indicate poor discrimination, values between 0.1 and 0.3 indicate fair discrimination, and values greater than 0.3 indicate good discrimination (Salkind 2010). After determining the Item Difficulty Index and Item Discrimination Index values for each item, we tabulated where each item fell on both scales. We found that nine of the 11 items had good discrimination values—of which one had low difficulty, four had medium difficulty, and four had high difficulty. The remaining two items had fair discrimination values—of which one had medium difficulty and one

had high difficulty. Overall, a majority of our items had good discrimination values while still being appropriately challenging in difficulty.

We performed an exploratory factor analysis using SPSS 25 to identify underlying factors in the BETTSI 3.0, which we hypothesized would align with the specific concepts outlined in Appendix A. We used principal axis factoring with promax rotation on the $n = 89$ student population responses from the BETTSI 3.0. The resulting Kaiser-Meyer-Olkin (KMO) value, which indicates the measure of shared variance in the items, was 0.56 (Table 4), suggesting our items should not be divided into factors (Beavers et al. 2013). Therefore, we did not interpret the remaining factorizing results. Although we expected the BETTSI 3.0 factor structure to align with tree-thinking concepts and common alternative conceptions (Appendix A), the lack of factor structure may be explained by these elements being closely linked. An instrument with poor factor structure does not suggest a poor or limited instrument. Instead, it simply suggests that items do not segregate into underlying groups.

The BETTSI has been published as an Open Education Resource, and may be accessed at <https://qubeshub.org/publications/2106/1>. Users will find the BETTSI in .docx and .pdf formats, and a teacher guide with suggestions for implementation. Appendix A includes a table with common misconceptions and correct questions indicated. Typically, the BETTSI takes about 20 min, and can be administered in class or as homework.

Discussion

We developed a tree-thinking diagnostic tool composed of 11 questions that target both mechanical/technical and interpretive tree-thinking skills. Many of these skills fall under the umbrella of representational competence and involve students' understanding of how to interpret the evolutionary information represented in these diagrams, but students must also master the technical understanding of what the specific features of a phylogeny (i.e., nodes, branches, root) indicate and the information they contain. Through an iterative process of evaluating each question for its representation of a singular construct and face validation to ensure students were reading the question as intended, we

Table 4. Appropriateness tests for principal axis factoring.

Test		Value
Kaiser-Meyer-Olkin measure of sampling adequacy		0.557
Bartlett's test of sphericity	Approx. chi-square	127.436
	df	55
	Significance*	<0.001

*Significance calculated at $P < 0.05$.

developed an instrument with reliable and valid evidence that it assesses students' representational competence and technical skills in tree thinking. The results from the BETTSI 3.0 item analysis suggest an appropriately challenging instrument with excellent capabilities as a diagnostic tool to quickly identify tree-thinking concepts and skills in need of specific attention by an instructor.

During the development of the BETTSI, we made several improvements to the instrument. For example, early iterations included multiple phylogenies that required an unnecessary cognitive investment from students in reading and interpreting a new tree to answer each question. In the next iteration of the BETTSI, we reduced this cognitive load by asking two questions about many of the phylogenies, decreasing the total number of unique trees presented. Furthermore, the earlier versions of the BETTSI included multiple skills or alternative conceptions in a single question. For example, a common question format required the student to both evaluate the relationships among taxa and integrate the direction of time. Several changes were made in the BETTSI 3.0 to reduce distractions, reduce cognitive fatigue, and increase the focus on specific tree thinking skills. For example, phylogenies in BETTSI 3.0 are presented using right angle branches, a format that has been shown to be less confusing to students (Cately and Novick 2008), except in one question in which the ability to read different tree formats was being evaluated. Similarly, unless the question specifically evaluates students' preconceptions based on familiar organisms, phylogenies have been simplified by removing organisms. Phylogenies in the BETTSI 3.0 were simplified further by including an arrow indicating the direction of time. BETTSI 3.0 is designed to address one tree-thinking skill per question, and to reveal common errors based on the incorrect distractors chosen.

Tree thinking provides students with a conceptual structure to explore complex evolutionary topics and ideas in a relatively concrete manner. Correctly applying tree thinking involves not only the technical aspects of knowing what features of the diagram carry information (e.g., tree topology) and which do not (e.g., order of tip labels), but also being able to correctly interpret what those features signify biologically. For example, it is

one thing to understand that a line indicates a lineage, but it is a deeper level of understanding to know that it is actually a representation of many, separate, interacting populations, or that a node represents a population or subset of populations that have become genetically disconnected from others in the lineage. Our experience in teaching tree thinking has convinced us that time spent in the classroom teaching and learning tree-thinking skills supports students' deeper understanding of evolution (Gregory 2008; Gibson and Hoefnagle 2015).

Identifying and addressing potential issues in students' tree thinking up front allows more productive discussions of evolutionary concepts. We recommend using the BETTSI as a pre-assessment to evaluate how much instructional time should be committed to address different basic tree-thinking skills and identify what those specific tree-thinking skills are. We also recommend following the BETTSI with explicit instruction in tree thinking, as this approach can greatly improve students' tree-thinking skills (Gendron 2000; Halverson 2010; Meisel 2010; McLaurin et al. 2013; Young et al. 2013), and there are now many educational resources for teaching tree thinking. A few of our favorites include the Great Clade Race (Goldsmith 2003), and variations such as the Vertebrate Clade Race (Baum and Jenkins 2020), Caminicles (Gendron 2000), Pipe Cleaner Trees (Halverson 2010), and Phylocards (Gibson and Cooper 2017). These activities may be modified to fit individual classroom needs, but all involve actively engaging students in all aspects of tree thinking and provide learning opportunities by creating cognitive dissonance around common tree-thinking errors. The BETTSI can also be used as a summative assessment, to evaluate and compare the effectiveness of instructional approaches.

Tree thinking is central to understanding evolutionary biology, but it is a concept that is vulnerable to a number of misconceptions. Further, the way that information is extracted and interpreted is not necessarily intuitive and requires special instruction as with any complex figure or diagram. We have found that the BETTSI is a simple tool that can help clarify which tree thinking skills must be addressed in classes and to evaluate the impact of different teaching strategies and activities.

AUTHOR CONTRIBUTIONS

KJ, LM, DB, KD, CB, EAL, and JPG conceptualized and designed the study. LM, KD, CB, EAL, and JPG curated the data. LM, KD, CB, EAL, and JPG analyzed and interpreted the data. KJ, LM, KD, CB, EAL, JPG, and EN drafted and revised the manuscript. DB revised the manuscript. All authors have given approval of the version to be published, and have participated sufficiently to take public responsibility for appropriate portions of the content. The authors agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy of integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

This work was supported by the National Evolutionary Synthesis Center (National Science Foundation grant number 0905606) and BEACON (National Science Foundation grant number 0939454). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the students who participated in this study, NESCent, BEACON and the J.F. Crow Institute for the Study of Evolution.

DATA ARCHIVING

Requests for access to the data used to establish validity and reliability may be directed to co-author KD who will provide de-identified data as allowed under the IRB protocol. The BETTSI is published on the QUBESHUB under a Creative Commons Attribution Share Alike license: BETTSI – Basic Evolutionary Tree-Thinking Skills Instrument, <https://doi.org/10.25334/ZQY1-W289>.

LITERATURE CITED

- Ainsworth, S., and J. Saffer. 2013. Can children read evolutionary trees? *Merrill Palmer Q.* 59:221–247.
- Barnea, N. 2000. Teaching and learning about chemistry and modelling with a computer-managed modelling system. Pp. 307–324 in J. K. Gilbert and C. Boulter, eds. *Developing models in science education*. Kluwer, Dordrecht, The Netherlands.
- Barrett, P. H., P. J. Gautrey, S. Herbert, D. Kohn, and S. Smith, eds. 1987. *Charles Darwin's notebooks, 1836–1844. Transmutation notebook B*. British Museum (Natural History), Lond.
- Baum, D. A., and K. Jenkins. 2020. Vertebrate clade race. QUBES Educational Resources. <https://doi.org/10.25334/D9GX-PC46>.
- Baum, D. A., and S. D. Smith. 2013. *Tree thinking: an introduction to phylogenetic biology*. Roberts & Company Publishers, Greenwood Village, CO.
- Baum, D. A., S. D. Smith, and S. S. Donovan. 2005. The tree-thinking challenge. *Science* 310:979–980.
- Beavers, A. S., J. W. Lounsbury, J. K. Richards, S. W. Huck, G. J. Skolits, and S. L. Esquivel. 2013. Practical considerations for using exploratory factor analysis in educational research. *Pract. Assess. Res. Eval.* 18:1–13.
- Blacquiere, L. D., and Hoese, W. J. 2016. A valid assessment of students' skill in determining relationships on evolutionary trees. *Evo Edu Outreach* 9, 5. <https://doi.org/10.1186/s12052-016-0056-9>
- Bodner, G. 1980. Statistical analysis of multiple-choice exams. *J. Chem. Educ.* 57:188–190.
- Campbell, C. E., and R. H. Nehm. 2013. A critical analysis of assessment quality in genomics and bioinformatics education research. *CBE Life Sci. Educ.* 12:530–541.
- Catley, K. M., and L. R. Novick. 2008. Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. *Bioscience* 58:976–987.
- Catley, K. M., L. R. Novick, and C. K. Shade. 2010. Interpreting evolutionary diagrams: when topology and process conflict. *J. Res. Sci. Teach.* 47:861–882.
- Daynat, B. 2003. The roots of phylogeny: how did Haeckel build his trees? *Syst. Biol.* 52:515–527.
- Eddy, S. L., A. J. Crowe, M. P. Wenderoth, and S. Freeman. 2013. How should we teach tree-thinking? An experimental test of two hypotheses. *Evol. Educ. Outreach* 6:1–11.
- Evans, E. M., K. S. Rosengren, J. D. Lane, and K. L. Price. 2012. Encountering counterintuitive ideas: constructing a developmental learning progression for evolution understanding. Pp 174–199 in K. S. Rosengren, S. K. Brem E. M. Evans, and G. M. Sinatra, eds. *Evolution challenges: integrating research and practice in teaching and learning about evolution*. Oxford Univ. Press, New York.
- Gregory, T. R. 2008. Understanding evolutionary trees. *Evol. Educ. Outreach* 1:121–137.
- Gendron, R. P. 2000. The classification & evolution of Caminacules. *Am. Biol. Teach.* 62:570–576.
- Gibson, J. P., and J. T. Cooper. 2017. Botanical phylo-cards: a tree-thinking game to teach plant evolution. *Am. Biol. Teach.* 79:241–244.
- Gibson, J. P., and M. H. Hoefnagels. 2015. Correlations between tree thinking and acceptance of evolution in introductory biology students. *Evol. Educ. Outreach* 8:15.
- Goldsmith, D. W. 2003. The great clade race: presenting cladistic thinking to biology majors and general science students. *Am. Biol. Teach.* 65:679–682.
- Halverson, K. L. 2008. Using hypothetical flowering plants to develop fundamental phylogenetic tree-reading and tree-building skills. Oral presentation at the annual meeting of the National Association of Biology Teachers, Memphis, TN.
- . 2010. Using pipe cleaners to bring the tree of life to life. *Am. Biol. Teach.* 72:223–224.
- . 2011. Improving tree-thinking one learnable skill at a time. *Evol. Educ. Outreach* 4:95–106.
- Halverson, K. L., and P. Friedrichsen. 2013. Learning tree thinking: developing a new framework of representational competence. Pp. 185–201 in D. Treagust and C. Y. Tsui, eds. *Multiple representations in biological education. Models and modeling in science education*. Vol. 7. Springer, Dordrecht, The Netherlands.
- Halverson, K. L., S. K. Abell, P. M. Friedrichsen, and J. C. Pires. 2009. Testing a model of representational competence applied to phylogenetic thinking. Poster presented at the NARST Conference, April 17–21, Garden Grove, CA.
- Halverson, K. L., J. C. Pires, and S. K. Abell. 2011. Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Sci. Educ.* 95:794–823.
- Halverson, K. L., C. J. Boyce, and J. D. Maroo. 2013. Order matters: pre-assessments and student generated representations. *Evol. Educ. Outreach* 6:24.
- Hennig, W. 1950. *Grundzüge einer Theorie der phylogenetischen Systematik*. Available via <https://agris.fao.org/agris-search/search.do?recordID=US201300727806>.
- Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.
- Kline, T. J. B. 2005. *Psychological testing: a practical approach to design and evaluation*. Sage Publications, Thousand Oaks, CA.

- Kummer, T. A., C. J. Whipple, S. M. Bybee, B. J. Adams, and J. L. Jensen. 2019. Development of an evolutionary tree concept inventory. *J. Microbiol. Biol. Educ.* 20:20.2.42.
- Lennox, J. 2019. Aristotle's biology. In E. N. Zalta, ed. *The Stanford encyclopedia of philosophy*. Available via <https://plato.stanford.edu/archives/win2019/entries/aristotle-biology/>.
- Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–536.
- Maddison, W. P., and D. R. Maddison. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. *Folia Primatol.* 53:190–202.
- Malik, A. H., J. M. Ziermann, and R. Diogo. 2018. An untold story in biology: the historical continuity of evolutionary ideas of Muslim scholars from the 8th century to Darwin's time. *J. Biol. Educ.* 52:3–17.
- Maroo, J., and K. L. Halverson. 2011. Tree-Thinking: a branch of mental rotation. *Synergy* 2:53–59.
- Matuk, C., and D. Uttal. 2012. Narrative spaces in the representation and understanding of evolution. Pp. 119–144 in K. Rosengren, S. Brem, E. Evans, and G. Sinatra, eds. *Evolution challenges: integrating research and practice in teaching and learning about evolution*. Oxford Univ. Press, Cambridge, U.K.
- McCullough, E. L., L. Verdeflor, A. Weinsztok, J. R. Wiles, and S. Dorus. 2020. Exploring activities for understanding evolutionary relationships depicted by phylogenetic trees: united but diverse. *Am. Biol. Teach.* 82:333–337.
- McLaurin, D., K. L. Halverson, and C. J. Boyce. 2013. Using manipulative models to develop tree-thinking. *Biol. Int.* 54:108–121.
- Meisel, R. P. 2010. Teaching tree-thinking to undergraduate biology students. *Evol. Educ. Outreach* 3:621–628.
- Meir, E., J. Perry, J. C. Herron, and J. Kingsolver. 2007. College student's misconceptions about evolutionary trees. *Am. Biol. Teach.* 69:71–76.
- Moskal, B. M., and J. A. Leydens. 2000. Scoring rubric development: validity and reliability. *Pract. Assess. Res. Eval.* 7:10. Available via <http://pareonline.net/getvn.asp?v=7&n=10>.
- Morabito, N. P., K. M. Catley, and L. R. Novick. 2010. Reasoning about evolutionary history: post-secondary students' knowledge of most recent common ancestry and homoplasy. *J. Biol. Educ.* 44:166–174.
- Naegle, E. 2009. Patterns of thinking about phylogenetic trees: a study of student learning and the potential of tree-thinking to improve comprehension of biological concepts. Ph.D. diss. Idaho State Univ., Pocatello, ID.
- Novick, L. R., and K. M. Catley. 2007. Understanding phylogenies in biology: the influence of a Gestalt perceptual principle. *J. Exp. Psychol. Appl.* 13:197–223.
- . 2014. When relationships depicted diagrammatically conflict with prior knowledge: an investigation of student's interpretations of evolutionary trees. *Sci. Educ.* 98:269–304.
- Novick, L. R., A. Stull, and K. M. Catley. 2012. Reading phylogenetic trees: the effects of tree orientation and text processing on comprehension. *Bioscience* 62:757–764.
- O'Hara, R. J. 1988. Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Syst. Biol.* 37:142–155.
- . 1997. Population thinking and tree thinking in systematics. *Zool. Scr.* 26:323–329.
- Omland, K. E., L. G. Cook, and M. D. Crisp. 2008. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *Bioessays* 30:854–867.
- Salkind, N., ed. 2010. KR-20. In *Encyclopedia of research design*. Vol. 1. SAGE, Thousand Oaks, CA. <https://doi.org/10.4135/9781412961288.n205>.
- Schramm, T., and P. Schmiemann. 2019. Teleological pitfalls in reading evolutionary trees and ways to avoid them. *Evol. Educ. Outreach* 12:1–14.
- Swofford, D. L. 1989. PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign, IL.
- Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck, P. G. Foster, P. O. Lewis, and J. S. Rogers. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50:525–539.
- Walter, E. M., K. L. Halverson, and C. J. Boyce. 2013. Investigating the relationship between college students' acceptance of evolution and tree thinking understanding. *Evol. Educ. Outreach* 6:1–8.
- Wiley, E. O., D. Siegel-Causey, D. A. Brooks, and V. A. Funk, eds. 1991. *The Compleat cladist: a primer of phylogenetic procedures*. Museum of Natural History, University of Kansas, Lawrence, KS.
- Young, A. K., B. T. White, and T. Skurtu. 2013. Teaching undergraduate students to draw phylogenetic trees: performance measures and partial successes. *Evol. Educ. Outreach* 16:1–15.

Associate Editor: Dr. Norman
Handling Editor: Prof. Tracey Chapman

Appendices

Developing the BETTSI: A tree-thinking diagnostic tool to assess individual elements of representational competence

Table A1. Tree reading concepts.

Concept	Abbreviation in diagnostic table	Tree interpretation skill	Tree reading skill
How closely related two tip entities are is represented by distance from the most recent common ancestor (MRCA).	MRCA—most recent common ancestor	<ul style="list-style-type: none"> • Relatedness is determined by following branches back to a shared node. • Nodes represent a common ancestral population of the lineages descended from that node. 	<ul style="list-style-type: none"> • Ability to trace branches from tip to shared nodes • Recognize nodes as evolutionary events separating populations • Branches can rotate around nodes
All contemporary species have been evolving from LUCA (last universal common ancestor) for the same amount of time and are equally “evolved.”	Evolutionary connection to LUCA—Last Universal Common Ancestor	<ul style="list-style-type: none"> • Clades represent groups of related organisms. • A tree diagram shows the order in which lineages diverged from one another. • The evolutionary pathway of a tip entity can be traced back through time by following the branches. • There are no “higher” or “lower” extant organisms. • Traits are accumulated evolutionary changes. 	<ul style="list-style-type: none"> • A complete clade can be removed from the tree by cutting one branch • Recognize the representation of passage of time from root to tip • Recognize the continuity of time
Evolutionary change accumulates in a lineage over time.	Trait changes occur along branches	<ul style="list-style-type: none"> • Nodes represent evolutionary divergence of two populations by accumulation of traits. • Traits can be gain of function or loss of function. • Branches, nodes, and tips represent groups of populations, not individuals. 	<ul style="list-style-type: none"> • Changes in traits for a lineage can be indicated by marks on a branch • The way branches are drawn (e.g., circular vs. right angle branching) does not change evolutionary relationships
Evolution occurs in populations.	Population thinking		<ul style="list-style-type: none"> • A branch represents changes accumulating in populations, often leading to divergence between subpopulations • A node represents evolutionary divergence of two populations

Table A2. Categories of different tree-thinking concepts, misconceptions, and skills addressed by specific questions in the BETTSI. Correct answers are bold. Underlying issues leading to incorrect answers are diagnosed based on instructor experience and corroborated with student reflections. For all of these questions, students may make mistakes because they do not read the question correctly. This could be in part due to the unfamiliarity of the terminology and information that can be extracted from the diagrams. Tree reading concepts are summarized in Table A1.

Question	Associated figure	Common issues
<p>In reference to the tree, which of the following is an accurate statement of relationships? (Q4)</p> <p>(a) A green alga is more closely related to a red alga than to a moss</p> <p>(b) A green alga is more closely related to a moss than to a red alga</p> <p>(c) A green alga is equally related to a red alga and a moss</p> <p>(d) green alga is related to a red alga but is not related to a moss</p> <p>(e) None of these organisms are related.</p>		<ul style="list-style-type: none"> • Incorrect answers on this question may be arrived at by: • counting “steps,” nodes, or branches; • reading across the tips; • grouping by similarity; • or considering taxa to be biologically unrelated.
<p>Three students are arguing over the correct interpretation of the tree in Question 4. Which student is correct? (Q5)</p> <p>(a) Student A insists that pine is the most highly evolved living species because it evolved most recently and is more complex than the other species.</p> <p>(b) Student B says the amoeba is the most highly evolved living species because it is older than the other species.</p> <p>(c) Student C says that no living species are more highly evolved than another because all living species have been evolving for the same amount of time from their common ancestor.</p> <p>(d) None of the students are correct.</p> <p>(e) I do not know how to interpret the tree.</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • ladder thinking, • equating “complexity” with being “more evolved”; • thinking of evolution as a continuous improvement process, evolving toward “perfection,”—more recently evolved species are more highly evolved; • reading across the tips left to right with right position being “more advanced”; • confusion about the historical nature of evolution; • not considering all species trace back to a single common ancestor; • or confusion about how evolution is represented in the diagram.

(Continued)

Table A2. Continued.

Question	Associated figure	Common issues
<p>In the tree, assume that the ancestor had a long tail, ear flaps, external testes, and fixed claws. Based on the tree and assuming that all evolutionary changes in these traits are shown, what traits does a sea lion have? (Q6)</p>		<ul style="list-style-type: none"> • Incorrect answers on this question may be arrived at by: • incorrectly tracing a lineage through the diagram; • or not recognizing the meaning of the trait mark on the branch.
<p>(a) long tail, ear flaps, external testes, and fixed claws (b) short tail, no ear flaps, external testes, and fixed claws (c) short tail, no ear flaps, abdominal testes, and fixed claws (d) short tail, ear flaps, abdominal testes, and fixed claws (e) long tail, ear flaps, abdominal testes, and retractable claws</p>		
<p>Looking at the tree, two students are discussing the evolutionary relationship between sea lions, seals, and dogs. Which student do you think is correct? (Q7)</p>		<ul style="list-style-type: none"> • Incorrect answers to this question may indicate: • reading across the tips; • grouping by type/trait; • grouping by similarity or ecology; • inability to trace an evolutionary lineage through the tree; • node counting; • or confusion about how to interpret the diagram.
<p>(a) Student A says seals and sea lions are equally related to dogs because the lineages of seals and sea lions share the same common ancestor with dogs. (b) Student B says that sea lions are more closely related to dogs than seals because there are fewer trait differences between sea lions and dogs, and sea lions are next to dogs in the diagram. (c) Neither student is correct. (d) I do not know how to interpret the tree.</p>		
<p>Which of the five marks on the tree corresponds to the most recent common ancestor of taxon 3 and taxon 5? (Q8) Answer: D</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • confusion about MRCA; • confusion about common ancestry; • or reading across the tips.

(Continued)

Table A2. Continued.

Question	Associated figure	Common issues
<p>A lineage refers to the entire evolutionary history of a species or taxon. Using this definition, which image tree has correctly traced the Taxon C lineage, as indicated by the bolded thick black line. (Q9) Answer: E</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • confusion about the meaning of “lineage”; • confusion about the evolutionary connection to LUCA; • confusion about the “root”; • or confusion about the meaning of nodes and right angles in the diagram.
<p>Using the tree, which of the following is an accurate statement? (Q10)</p> <p>(a) A seal is more closely related to a horse than to a whale</p> <p>(b) A seal is more closely related to a whale than to a horse</p> <p>(c) A seal is equally related to a horse and a whale</p> <p>(d) A seal is related to a whale but is not related to a horse</p> <p>(e) None of these organisms are related</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • reading across the tips; • grouping by similarity or ecology; • confusion about tracing a lineage; • node counting, thinking in steps, or counting “evolutionary steps”; • or confusion about what the tree represents.
<p>Imagine you could travel backward through time and examine the last common ancestor of a giraffe and a hippo. What would it be? (Q11)</p> <p>(a) A giraffe</p> <p>(b) A hippo</p> <p>(c) A horse</p> <p>(d) A species that cannot be classified as any of the above.</p> <p>(e) There is no common ancestor between a giraffe and a hippo.</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • confusion about tracing lineages; • confusion about what nodes represent; • confusion about MRCA; • confusion about the evolutionary process of accumulated changes; • or reading across the tips.
<p>Which of the following trees provides different information about the evolutionary relationships among the groups? (Q12) Answer: B</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • reading across the tips; • confusion about tracing a lineage; • confusion about the concept of clades; • or confusion about MRCA.

(Continued)

Table A2. Continued.

Question	Associated figure	Common issues
<p>Which of the following trees provides different information about the evolutionary relationships among the groups? (Q13)</p> <p>(a) Tree 1 (b) Tree 2 (c) Tree 3 (d) All trees are the same. (e) All trees are different.</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • reading across the tips; • confusion about what branches and nodes represent; • confusion about tracing a lineage; • lack of awareness about rotation around nodes; • or confusion about the style versus content.
<p>Which of the following trees provides different information about the evolutionary relationships among the groups? (Q14)</p> <p>(a) Tree 1 (b) Tree 2 (c) Tree 3 (d) They are all the same. (e) They are all different</p>		<ul style="list-style-type: none"> • Incorrect answers may indicate: • confusion about the relevance of orientation, or the position of the root.