2017-03-01

Assessing and Improving Student Understanding of Tree-Thinking

Tyler A. Kummer
Brigham Young University

Assessing and Improving Student Understanding of Tree-Thinking

Tyler A Kummer

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Jamie Jensen, Chair
Clinton Whipple
Seth Bybee
Randall Davies
Byron Adams

Department of Biology

Brigham Young University

ABSTRACT

Assessing and Improving Student Understanding of Tree-Thinking

Tyler A Kummer
Department of Biology, BYU
Doctor of Philosophy

Evolution is the unifying theory of biology. The importance of understanding evolution by those who study the origins, diversification and diversity life cannot be overstated. Because of its importance, in addition to a scientific study of evolution, many researchers have spent time studying the acceptance and the teaching of evolution. Phylogenetic Systematics is the field of study developed to understand the evolutionary history of organisms, traits, and genes. Tree-thinking is the term by which we identify concepts related to the evolutionary history of organisms. It is vital that those who undertake a study of biology be able to understand and interpret what information these phylogenies are meant to convey.

In this project, we evaluated the current impact a traditional study of biology has on the misconceptions students hold by assessing tree-thinking in freshman biology students to those nearing the end of their studies. We found that the impact of studying biology was varied with some misconceptions changing significantly while others persisted. Despite the importance of tree-thinking no appropriately developed concept inventory exists to measure student understanding of these important concepts. We developed a concept inventory capable of filling this important need and provide evidence to support its use among undergraduate students. Finally, we developed and modified activities as well as courses based on best practices to improve teaching and learning of tree-thinking and organismal diversity. We accomplished this by focusing on two key questions. First, how do we best introduce students to tree-thinking and second does tree-thinking as a course theme enhance student understanding of not only tree-thinking but also organismal diversity. We found important evidence suggesting that introducing students to tree-thinking via building evolutionary trees was less successful than introducing the concept via tree interpretation and may have in fact introduced or strengthened a misconception. We also found evidence that infusing tree-thinking into an organismal diversity course not only enhances student understanding of tree-thinking but also helps them better learn organismal diversity.

Keywords: tree-thinking, evolutionary trees, phylogeny, organismal diversity education, evolution misconceptions

ACKNOWLEDGEMENTS

I would like to thank Dr. Jamie Jensen for more help and guidance than I could have ever hoped for. I would like to thank each member of my committee for the time and effort they put in to help me through this process. I would like to think Dr. Clinton Whipple in particular for his guidance in this project and for teaching me so much about tree-thinking. I would also like to thank my fellow graduate students, faculty, and the support staff at BYU for all their valuable help.

I would not have been able to complete this project without the support of my family. I give thanks to my father for always supporting my family and me as I furthered my education. I also give thanks to my mother for the support she has given. She has been my editor, audience, and counselor as I have spent nearly a decade studying biology. Finally, I give thanks to my wife for her patience and being my partner in all of this.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Darwin described evolution as 'descent with modification' (1859). The 'modification' portion of his statement has been extensively described and it serves as a cornerstone of most evolution units in introductory biology courses. As a result of this emphasis, most biology students can quote "Survival of the Fittest" and give examples of natural selection with relative ease. The misconceptions concerning the modification process have also been extensively described and several concept inventories exist to assess student understanding (e.g., the Conceptual Inventory of Natural Selection (Nehm & Schonfeld, 2008). However, the 'descent' portion of this statement has received much less attention and is often omitted or only briefly covered in an introductory curriculum (Catley, 2006). This is an unfortunate oversight given the importance of understanding this component of evolution. It is this descent that is depicted in phylogenetic trees, like the one shown in Figure 1. Understanding what this tree represents is referred to as 'tree-thinking'.

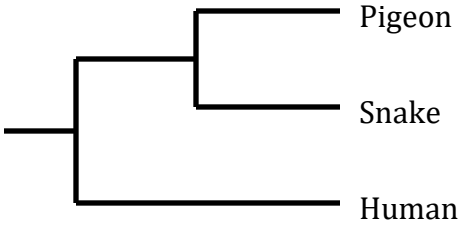


Figure 1. Sample Evolutionary Tree.

Evolutionary tree-thinking is a vital skill in understanding the diversity of life and myriad biological phenomena. Evolutionary trees are used for a wide variety of purposes ranging from tracking the spread of infectious disease to determining how to target limited resources in the conservation of biodiversity (Thanukos, 2010). Evolutionary trees have both practical importance (medical use and conservation) and conceptual importance in helping us understand

how evolution has and does occur. The importance of tree-thinking has been emphasized by the following comparison: "Just as beginning students in geography need to be taught how to read maps, so beginning students in biology should be taught how to read trees and to understand what trees communicate" (O'Hara, 1997). Phylogenetic systematics, the reconstruction of evolutionary history, has infiltrated nearly every field of life sciences. Biological literacy now requires that students be able to interpret evolutionary trees (Baum, Smith, & Donovan, 2005).

Over the last decade, how students think about and learn tree-thinking concepts has drawn significant attention from researchers. We have seen important gains in our understanding. Two important review articles have summarized much of these findings (Gregory, 2008; Meisel, 2010). These reviews were written to educators with the goal of helping educators who may be unfamiliar with evolutionary trees understand their importance, how they are to be interpreted, and what common misconceptions may impact students' ability to understand and interpret evolutionary trees. In addition to these two reviews, an entire issue of the journal *Evolution: Education and Outreach* was devoted to tree-thinking and the teaching of evolutionary trees in 2010. This issue provides readers with important background information about the development of phylogenetic systematics as a discipline and its role in science and society (Brooks, 2010; Dominici & Eldredge, 2010; Eldredge, 2010; Thanukos, 2010; Wiley, 2010), methods for teaching evolutionary tree concepts (Goldstein, 2010; Kumala, 2010a, 2010b; McLennan, 2010a; Novick, Catley, & Funk, 2010), and common pitfalls encountered as students learn tree-thinking (McLennan, 2010b; Meikle & Scott, 2010). These articles provided a substantial resource for educators who wished to better understand and incorporate tree-thinking into their courses.

These important resources, as well as others, have made a significant impact in improving evolution education but there are still important areas which have not been adequately addressed such how and if tree-thinking misconceptions resolve over the course of an undergraduate study of biology, how to teach evolutionary trees, and how to assess student understanding, i.e., developing concept inventory of tree-thinking. While these areas have not been neglected completely by researchers important aspects of them have been unaddressed or lack clear evidence to draw a conclusion.

We hoped to further our understanding of evolution education by attempting to answer these important questions. We proposed to do this by 1) Identifying the prevalence and persistence of commonly held tree-thinking misconceptions among undergraduates studying biology at the beginning and end of their college career, 2) Developing and refining inquiry-based activities and methods that will improve student understanding of tree-thinking concepts at both the course scale and activity scale and provide supporting evidence, and 3) Developing a tree-thinking concept inventory.

LITERATURE REVIEW

Learning Objectives and Assessments

Learning outcomes are recommended to be the starting point for the design of a course because they target the assessments and instruction that will help students meet the goals the instructor(s) saw as most important (Wiggins & McTighe, 2005). Learning outcomes are the intended set of tasks or abilities an instructor expects his or her students to be able to accomplish as a result of instruction. Deciding on a set of learning outcomes is often a difficult and political process as stakeholders each seek to help their vision of a course come to fruition.

The importance and difficulty of coming to a consensus on learning outcomes is also seen in science education research. If two researchers approach research questions with differing learning objectives it will potentially become difficult to compare and contrast the results produced, particularly when those findings appear to be in conflict. As a community of researchers and educators, we need to decide which concepts are related to tree-thinking and which of those concepts are vital for students to know in order to truly understand the idea of descent. A set of widely accepted learning outcomes will allow researchers to better understand how tree-thinking concepts relate to one another from the learners' perspective and to evaluate and target pedagogical interventions. Novick & Catley (2012) put forward a suggested set of core concepts, which can serve as a starting point for tree-thinking learning outcomes. These concepts consist of five key skills/components:

1. *Identify a character shared by two or more taxa due to inheritance from their most recent common ancestor*

This outcome requires that a student understand that the tree graphic depicts which characters a given taxon or taxa have and that the character was inherited from a common ancestor (Catley, Phillips, & Novick, 2013). For example, according to the tree in Figure 2, which character(s) did crocodile and goat inherit from their most recent common ancestor (MRCA)? Based on the characters mapped on the tree the answer would be that vertebrae, 4 limbs, and amniotic egg were the characters that both crocodile and goat

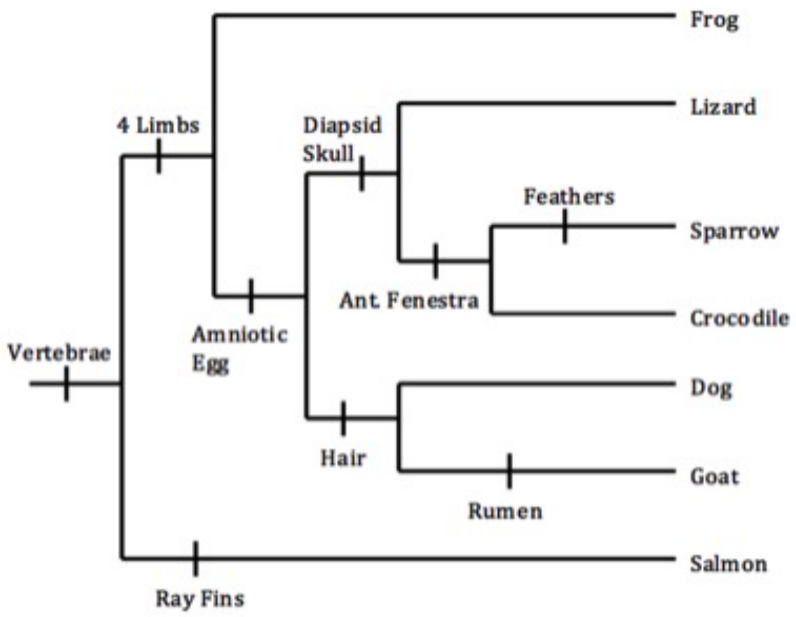


Figure 2. Sample evolutionary tree with characters mapped.

inherited from their last common ancestor. This is because the common ancestor had each of these characters before the lineages that led to crocodile and goat diverged from one another. A learning objective that would go along with this learning objective is to ask students to identify all the characters a taxon from the tree would have. This would expand what is expected of the student by requiring them to interpret an entire lineage from beginning to end. Without the ability to interpret which characters have been passed on from common ancestors students are not able to make inferences about the evolution of these characters and taxa, which makes the mapping of characters on a tree uninformative.

2. *Identify a set of taxa based on character information provided*

Students need to be able to distinguish between characters that reflect natural (based on evolutionary history) groups and those that do not, e.g., convergent characters. In evolutionary tree terminology, students should be able to identify a clade when given a synapomorphy. For

example, referring to the tree in Figure 2 which taxa have a Diapsid Skull? Using the MRCA that had a diapsid skull and looking at all of its descendant taxa, the lizard, crocodile, and sparrow would form a group that is defined by having a diapsid skull, i.e., this is their synapomorphy. Others have proposed an extension of this learning outcome by asking that the student not just identify groups based on a synapomorphy but actually place a new taxon into the tree based on synapomorphy (Eddy, Crowe, Wenderoth, & Freeman, 2013).

3. Understand the concept of a clade

Understanding the concept of a clade is critical to proper interpretation of groups based on evolutionary history. Monophyletic clades are groups that reflect the evolutionary history of the taxa that comprise them, while polyphyletic or paraphyletic groups do not reflect any meaningful history. Being able to identify groups that do and do not reflect evolutionary history allows evolutionary trees to become an important tool for creating meaningful classifications of biological diversity. Meaningful classifications stand in contrast to arbitrary classifications in which the groupings provide only the information that one bases them on (which is not to say they lack utility). Alternatively, meaningful classifications (e.g., The Periodic Table of Elements) provide information beyond the information used to create them. Classifying organisms based on monophyly allows for inferences and predictions to be made about the characters and evolution of the taxa involved. Another way to phrase this learning outcome might be to distinguish between and identify monophyletic, paraphyletic and polyphyletic groups using an evolutionary tree. This would expand what is asked of students to include all clade types and require them to understand the differences between them.

4. *Evaluate relative evolutionary relatedness*

Determining the evolutionary relatedness of organisms and genes is a vital tool for answering many biological questions. Proper interpretation of evolutionary relationships allows for inferences ranging from biogeography to gene duplication. Students must be able to compare the relatedness of taxa in order to make necessary and important biological inferences with evolutionary trees. Evaluating the evolutionary relatedness between species is complicated in multiple ways. First, a node can be rotated and still depict the exact same evolutionary relatedness. The order in which the taxa are shown in the tree changes but the relationships do not. A second complication is differing styles of trees (diagonal, bracket, and circular). Different styles give students challenges when they encounter a style with which they are unfamiliar. An additional alteration that can make comparing evolutionary relationships results difficult results when a new taxon is grafted on a tree or a taxon is pruned (removed) from a tree. Students should be able to distinguish between trees that depict rotated nodes or different styles but the same relationships and those that actually depict different relationships.

5. *Use evidence of most recent common ancestry to support an inference regarding a shared character*

Making inferences about character changes or gene function is another valuable tool that evolutionary trees give researchers. Making these inferences allows characters to be mapped on the tree and cases of homology and analogy to be distinguished (Eddy et al., 2013). This has important implications when determining the evolution of a character and taxa.

Eddy et al. also proposed a set of six learning outcomes which are similar to the core concepts defined previously (2013). One key difference is that these learning outcomes explicitly describe tasks students are expected to complete using the core concepts, see Table 1.

Table 1. Potential learning outcomes focused on specific tasks.

Potential Learning Outcomes
• Use trees to determine ancestor-descendant relationships and degrees of relatedness among taxa • Map where particular traits evolved on the branches of trees and diagnose homoplasy • Use shared, derived characters to place taxa on a tree • Recognize that traits do not necessarily evolve in a progressive manner • Recognize that a species cannot be considered higher or lower than others • Recognize that extant traits can be considered basal, but that extant species cannot.

Three learning outcomes that are unique from the five core concepts are directly tied to teleological reasoning that often makes its way into a student's understanding of evolution also known as Ladder Thinking. Viewing evolutionary descent as a process by which some organisms advance via evolution to higher states of complexity while others remain in simple forms is a fundamental misunderstanding that many students of evolution hold (Gee, 2002; Meikle & Scott, 2010; Omland, Cook, & Crisp, 2008). There are no advanced or primitive species, only species that have evolved in response to differing evolutionary forces. It is common to mistakenly identify organisms with derived characters as more evolved. But it is the characters themselves that are derived/ancestral not the organisms (Baum & Smith, 2013).

Teleology, the idea that all phenomena have a purpose or goal, is the underlying concept that drives this misconception with the addition of anthropocentrism (i.e., seeing humans as the

central element of existence). Teleological reasoning is the basis of numerous misconceptions in every field of science and has proven extremely difficult to overcome with educational intervention (Kampourakis, 2007). Using teleological reasoning to make inferences about the processes and patterns of evolution represents a fundamental misunderstanding of evolution. This misunderstanding of evolution is not confined to tree-thinking but the concepts related to evolutionary trees readily reveal it as being held by a student. We believe any set of learning outcomes targeting tree-thinking should include outcomes addressing Ladder Thinking (teleology).

A learning outcome that also might be appropriate to include is asking students to analyze a dataset to identify the most parsimonious evolutionary tree given the data. We suggest this as a potential learning outcome due to the significant number of published learning activities that appear to target this concept. It is interesting that both lists of learning outcomes discussed exclude a learning outcome about this concept. It is possible that this was done because knowing how to go from data to a tree is unnecessary to have an acceptable level of tree-thinking ability (Halverson, 2011). This will be addressed in more detail when we discuss methods for teaching tree-thinking.

The learning outcomes described previously cover both character and taxa evolution. If a student demonstrates these basic skills it would be evidence that they can accurately interpret and use the information conveyed in phylogenetic trees as well as evaluate the evidence supporting one evolutionary tree over another. This would indicate that they have reached a minimum standard of literacy in tree-thinking (Catley et al., 2013).

Appropriate and widely accepted learning outcomes will aid in addressing the absence of a much-needed concept inventory for tree-thinking. The lack of a concept inventory has resulted

in an assessment being developed for each individual study of tree-thinking or adapting a published assessment for their needs. Baum published two such assessments as supplementary material in 2005. Unfortunately, the article itself makes no mention or claims of evidence for reliability or validity in relation to the assessments. We have no set of learning outcomes to associate with the items and no evidence that the items have appropriate difficulty and discrimination. The lack of evidence makes the assessments unsuitable as a concept inventory.

A second assessment was published as part of a dissertation titled, the Tree-thinking Concept Inventory (TTCI) (Neagle, 2009). Without publication two shortcomings are present: a lack of accessibility and a lack of clarity as to the content assessed by the TTCI. Because the instrument is not readily available to researchers, its use is limited. In addition, evidence of validity and reliability are not readily available. This will keep researchers from adopting its use even if they get a copy of the assessment because they are uncertain of its utility for their population of students. Researchers also raised questions regarding the accuracy of the TTCI content. This required the alteration of the TTCI by the researchers who used it in their study (Walter, Halverson, & Boyce, 2013). These issues limit the usefulness of the TTCI as a concept inventory.

Two other assessments have been published but are not usable as concept inventories for tree-thinking. One was appropriately developed and has excellent item characteristics; unfortunately, its scope is limited to evaluating evolutionary relatedness between taxa (Blacquiere & Hoese, 2016). Being so focused makes it appropriate to use in some circumstances but not as a concept inventory for tree-thinking as a whole. The second assessment is an open answer assessment consisting of three items titled the Phylogeny Assessment Tool (PhAT). The PhAT consists of two items asking students to evaluate two

competing evolutionary trees given a data set while the third item asks students to compare the evolutionary relatedness of two taxa (Smith, Cheruvelil, & Auvenshine, 2013). The PhAT's open response format, limited content coverage, and lack of evidence for reliability and validity prevent it from meeting the standards of a concept inventory. While both of these assessments were productively used in the research they were designed for they are not suitable candidates to be tree-thinking concept inventories.

Misconceptions

Misconceptions commonly held by students about evolutionary trees have been widely studied and described (Dees, Momsen, Niemi, & Montplaisir, 2014; Gregory, 2008; Meir, Perry, Herron, & Kingsolver, 2007; Meisel, 2010). These misconceptions can be categorized into two primary groups: Misconceptions related to interpreting the graphic representation and misconceptions that are based on a fundamental misunderstanding of evolutionary principles.

Two commonly held misconceptions relating to the nature of the graphic are using the proximity of the terminal ends of the tree to determine relatedness and using the number of nodes between each taxon to determine relatedness. Meir et al. (2007) showed that among undergraduates 24% held the proximity misconception and 38% held the node-counting misconception. The proximity misconception is based on a naïve understanding of what the graphic represents. Students mistakenly view the graphic as a list in which proximity forms a group. In Figure 1, a student using this naïve understanding might incorrectly conclude that 'Human' is more closely related to 'Snake' than to 'Pigeon' simply because it appears next to 'Snake' on the tree. It appears that this misconception is overcome in a typical undergraduate biology course. For example, in one study less than 9.3% of students still held the

misconception after participating in one of two different activities intended to teach tree-thinking (Perry, Meir, Herron, Maruca, & Stal, 2008).

The second graphic-based misconception is using the number of nodes between two taxa to determine relatedness, or Node Counting. Students count the number of nodes when comparing relatedness and determine that the species with the fewest nodes between them are most related. Students with this misconception may interpret the nodes correctly as common ancestors but assume that having fewer ancestors depicted on the tree between two species makes them more closely related. Alternatively, some students interpret the nodes as characters that specifically caused a split in the lineage; therefore, the fewer the traits dividing two taxa, the closer their relationship. The latter interpretation includes a misunderstanding of evolutionary processes that possibly lies in incorrectly applying concepts from natural selection. To illustrate this misconception, compare Figure 1 to Figure 3. 'Crocodile' has been added as a sister group to 'Pigeon' in Figure 3. A student with the first interpretation of node counting may have concluded that a Human is equally related to 'Pigeon' and 'Snake' using Figure 1, but if presented with Figure 3, would conclude that 'Human' is more closely related to 'Snake' than 'Pigeon' due to the extra common ancestor of 'Pigeon' and 'Crocodile'. A student with the second interpretation of Node Counting would come to the same conclusions but for the reason that the 'Pigeon/'Crocodile' clade has gained an additional character that makes them evolutionarily farther away from a Human than a Snake.

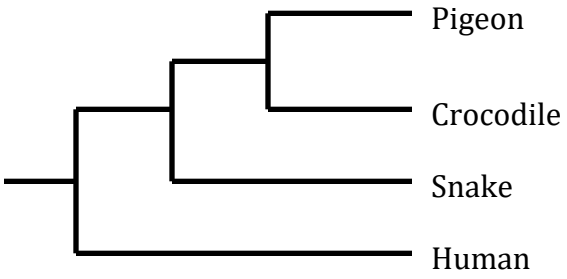


Figure 3. Sample evolutionary tree with an additional taxon.

It has been shown that mapping synapomorphies onto a tree improves student interpretation (Novick et al., 2010). It is possible that this strategy does so by helping students who view nodes as lineage splitting traits to recognize that nodes are not representations of traits but of the most recent common ancestor of the taxa that diverged from that point.

A third misconception relating to the nature of the graphic is incorrectly mapping time on the tree (Meir et al., 2007). Students are not able to determine the direction time flows and this impacts what they might infer about the evolutionary history of the taxa depicted. For example, a student may view time progressing in Figure 4 from left to right, rather than from bottom to top. This would make 'Human' the most primitive and 'Pigeon' the most advanced.

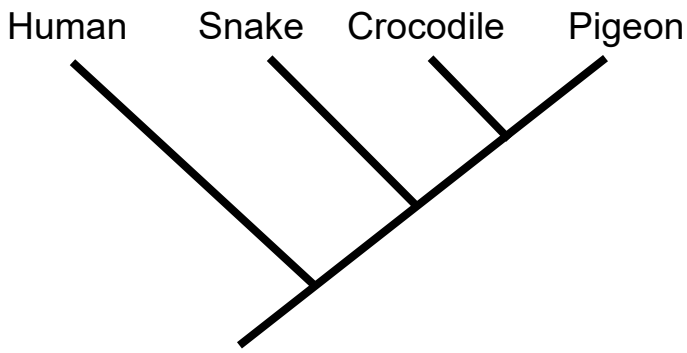


Figure 4. Sample evolutionary tree in the diagonal form.

Evolutionary trees can come in a variety of formats with the three most common being bracket, diagonal, and circular. It has been found that the format can have a significant impact

on the ability of students to understand trees (Novick & Catley, 2007). Using a bracket format (Figure 3) for the evolutionary tree results in better student understanding when compared to a diagonal format (Figure 4). Not only have bracket trees been shown to be easier to understand, even the direction of the diagonal moving from left to right can negatively impact student understanding (Novick, Stull, & Catley, 2012). Diagonal trees also lend themselves to greater anthropocentrism that is difficult even for experts to exclude from the trees they use (Sandvik, 2008). Despite the common use of bracket format trees in scientific literature, the vast majority of textbook illustrations of trees use the diagonal format (Catley & Novick, 2008). Because students are likely to encounter a variety of tree formats in their studies it is important that they can transfer their tree-thinking skills between formats.

Graphic-related misconceptions are important and prevalent misconceptions to address, but they are also easier to address, as we just need to teach students the skills necessary to interpret a graphic, rather than having to invoke a paradigm shift in their fundamental understanding of evolution. While graphic-related misconceptions can be easy to overcome, misconceptions based on theoretical misunderstandings of the process of evolution appear more challenging.

Mistaking superficial similarity for evolutionary relatedness is based on the intuitive idea that physical similarity between two organisms or taxa is driven by relatedness. While for most characters similarity does, in fact, reflect relatedness, the exceptions are not uncommon. The causes of these exceptions are well studied and understood. One common process that could lead to similarity differing from relatedness is convergent evolution. Convergent evolution is the process by which similar features are favored by natural selection in a given environment resulting in organisms independently evolving a similar feature. An example of this would be

the fusiform shape of a shark and that of a dolphin. This body shape was not inherited from a common ancestor that passed it on to both groups but instead was favored by selective processes at play in an aquatic environment. Relying on similarity that independently evolved in sharks and dolphins mistakenly results in students often misidentifying the shark and dolphin as being more closely related than they really are, i.e., grouping dolphins as sister taxon to sharks rather than to other mammals.

A second reason that similarity may not reflect relatedness is when one member of a group has derived a different form of a character while others have retained the ancestral form. A classic example of this is found in birds, having largely lost ancestral scales and replaced them with feathers. This led to the mistaken and long-held assumption that birds were a distinct lineage apart from other reptiles. And based on the presence of scales (i.e., a similarity), all reptiles were grouped together as sister taxa, when in fact, many reptiles (e.g., crocodiles) are more closely related to birds than they are to other reptiles. Essentially two closely related groups no longer appear to be closely related because one of the groups has derived a very different trait while the other now appears to be more similar to other groups that share the ancestral trait. This similarity between the groups that have the ancestral form is different than the similarity caused by convergent evolution. The similarity in this case, is in fact inherited from a common ancestor. It can be much harder to distinguish when this type of similarity is not reflecting common ancestry. It requires using additional characters, whether morphological or molecular, and sophisticated tree building methods to produce an evolutionary tree. Without the use of additional evidence, it is easy to mistake ancestral similarity as reflecting relatedness.

When students encounter information in an evolutionary tree that disagrees with their preconceived notion of physical similarity they tend to favor their prior conception over what the

evolutionary tree is depicting, hence the manifestation of this misconception (Catley, Novick, & Shade, 2010; (Novick et al., 2010). Educational activities and interventions are needed to help students overcome prior conceptions of relatedness. The goal should be to help students see how the evolutionary processes described above resulted in similarity that does not reflect the evolutionary history of the organisms. Interventions that address this misconception would not only improve student understanding of evolutionary trees but it would also deepen their understanding of the processes themselves.

The final misconception we will discuss is Ladder Thinking, the basis for one of our previously described learning outcomes. "Progressive" evolution is a commonly held misconception that is not an accurate view of evolution (Gee, 2002; Omland et al., 2008). Ladder Thinking can be manifest in many ways by students. As mentioned previously viewing some organisms as advanced while others as primitive is a common way that students demonstrate this misconception. To satisfy this assumption, students must erroneously believe that evolution of the "primitive" taxon has stopped at some point while other taxa continued to progress, when in reality, evolutionary forces continue to act upon all lineages as long as they are extant. For example, when viewing Figure 3, a student would assume that the human reached its current form as it is today long ago and has since stopped evolving while its sister taxon continued to evolve splitting into the other species on the tree. Similarly, this misconception can manifest in the way that students interpret the number of nodes depicted in the evolutionary tree. Students may see those taxa with the greatest number of nodes preceding them as being more evolved than those with fewer nodes leading to them. It again implies that evolution stopped or paused for some but did not for others.

A second way this misconception is seen is when students view one taxon as giving rise to another taxon in the tree. In this case, the students are seeing extant taxa in the tree as having parent-to-offspring relationships instead of sister taxa relationships. This implies that one extant taxon gave rise to the more "advanced" taxon as they evolved superior characters. In Figure 3, a student with this misconception might assume that the human gave rise to the snake that then gave rise to the crocodile, and so on.

Progression in evolution, or goal driven evolution, is a misconception that can even be found in life science graduate students as well as undergraduates (Catley et al., 2010; Gregory & Ellis, 2009). This implies that it is not a misconception that is being addressed by the average course of study in undergraduate education. Helping students with this misconception is not the primary goal of any published lesson plans of which we are aware. A lesson or even a course theme aimed at helping students confront this misconception and the underlying reasoning is needed. Understanding this key concept is not just a benefit to evolutionary tree-thinking but also to evolutionary theory in general.

Methods of Teaching

A lot of effort has been spent on identifying what things students do wrong (Halverson, Pires, & Abell, 2011), what misconceptions inhibit correct understanding (Gregory, 2009), and what components of tree-thinking are particularly difficult for students (e.g. style or orientation of the tree) (Catley & Novick, 2008; Catley et al., 2010; Novick et al., 2011; Novick et al., 2012). However, relatively few pedagogical solutions to these problems have been put to use in authentic classroom situations with evidence supporting their effectiveness.

The primary goal of the majority of published lessons related to tree-thinking is to teach students to build or identify the appropriate evolutionary tree for a given data set. For example, Goldsmith uses a race analogy and asks students to build a map of the race which is compared to phylogeny (2003). Davenport, Milks, and Tassell ask students to use data sets to evaluate the accuracy of two evolutionary trees (2015a). Julius and Schoenfuss ask students to build a data set (from skulls) and use existing data sets to create a phylogeny of vertebrates (2005). Kumala asks students to build a data set and phylogeny for gummy candies (2010a). Kuzoff, Kemmeter, McKinnon, and Thompson ask students to use molecular data to build a vertebrate phylogeny and then map when certain characters would have evolved (2009). Lents, Cifuentes, and Carpi ask students to build a phylogeny of primates using molecular data (2010), Singer, Hagen, and Sheehy ask students to build a mammal phylogeny based on ecological, morphological, and molecular data (2001). The rationale behind these activities is that by understanding how an evolutionary tree is made and being able to use evidence to make one, students will understand what an evolutionary tree theoretically represents.

An alternative pedagogical approach is to focus on analyzing the information depicted in a tree that has already been built (e.g., Davenport, Milks, & Tassell, 2015b; Halverson, 2010; Kumala, 2010b). It has even suggested that asking students to actually build trees inhibits their ability to analyze an existing evolutionary tree (Halverson, 2011). This hypothesis would suggest that the way many students are introduced to tree-thinking is, in actuality, inhibiting their ability to overcome misconceptions.

This hypothesis was put to the test when a group of students taught by tree building was compared to a group of students taught by tree analysis, doing a direct and controlled comparison (Eddy et al., 2013). They found that students in the tree building group performed better on a

common assessment of tree-thinking than did the students in the tree analysis group. The common assessment had no tree building and instead focused on proper reading and interpretation of trees. The results of this study do not support Halverson's (2011) claim that tree building inhibits tree analysis, and suggest instead that the best way to introduce evolutionary tree-thinking is to begin with tree building rather than with tree analysis.

While tree building alone was found to be a more effective approach to introducing evolutionary tree-thinking, there is also evidence showing that students struggle with certain aspects of tree analysis. In particular, students struggle with interpreting evolutionary relationships within a given tree. Researchers found that on a final exam only 38% of students could correctly answer evolutionary relationship questions and use correct reasoning; interestingly, 21% of students used the correct reasoning but selected the incorrect answer (Dees et al., 2014). This indicates that students can memorize correct reasoning patterns, but still not understand how to apply them to novel situations. Other research has also shown that proper interpretation of evolutionary relationships is something that students continue to struggle with even after significant instruction on evolutionary trees (Young, White, & Skurtu, 2013).

Introduction

Darwin defined evolution as descent with modification. When evolution is taught in many university courses, the mechanisms of evolution are emphasized. While understanding the mechanisms of evolution is critical to the study of biology, all too often, the descent portion of Darwin's statement is neglected in undergraduate studies. This process of descent is most often depicted in primary literature and in textbooks as branching trees. Understanding how to interpret the information conveyed in these trees is an important skill that is used in nearly every field of research in biology (Baum et al., 2005).

A considerable amount of research has been done to help educators understand how students learn about evolutionary trees. Some researchers have attempted to identify and characterize common student misconceptions related to tree-thinking (Gregory, 2008; Halverson et al., 2011). Misconceptions range from naïve interpretations due to a lack of familiarity with this style of graphical representation to fundamentally flawed conceptions of how descent and evolution occur. A second area of research has focused on how students interpret different forms of evolutionary trees and which forms are most easily understood (Catley & Novick, 2008). A third area of research has focused on how to improve instruction related to tree-thinking (Eddy et al., 2013; Phillips, Novick, Catley, & Funk, 2012). Despite all of this worthwhile study, little research has been done to determine how prevalent these misconceptions are among biology undergraduate students and how these misconceptions change as students progress in their studies.

We selected four major misconceptions as the focus of our study, each of which has been identified as commonly held (Baum & Smith, 2013). Previous studies on the prevalence of

misconceptions among college students primarily focused on misconceptions that are based on unfamiliarity with the graphic (e.g., not understanding what the bifurcation means, not recognizing the axis of time, equating a straight line with no change) (Meir et al., 2007). While misconceptions based on the graphic are worthy of study, misconceptions based on the theoretical underpinnings of evolution are more concerning and perhaps more difficult for students to overcome. We chose to focus on two misconceptions related to reading a graphic, and two related to the fundamental underpinnings of evolutionary theory.

The first misconception related to reading a graphic we refer to as Reading the Tips. Students with this misconception use the proximity of one tip to another to determine relatedness. The closer two taxa are in the tree the more related they are. The second misconception related to reading a graphic we refer to as Node Counting. Students with this misconception use the number of nodes between two taxa to determine relatedness. The fewer the number of nodes between two taxa the more related they are. The first misconception related to evolutionary theory we called Ladder Thinking. Ladder Thinking can be manifested in many ways, but the common thread is teleological-based reasoning. One example of Ladder Thinking is stating that one extant group evolved or "advanced" up the tree by acquiring more complex traits and becoming another extant group that is in the tree. Another way this misconception manifests is when a student states that one group of organisms is more evolved than another "lower" in the tree. While the phrasing is different the implication is the same. The organisms were the same and one advanced through evolution while the other remained primitive. The second misconception related to evolutionary theory we refer to as Similarity Equals Relatedness. Students with this misconception determine relatedness based on how similar the

physical traits are between various groups in the tree. For example, the more physical traits two groups share, the more closely related they are.

The purpose of this study was to determine both the prevalence and persistence of these four misconceptions among biology undergraduate students. We used a 20-question assessment to measure these misconceptions and compared the proportion of students who held these misconceptions in an introductory biology course to a senior level capstone course.

Methods

Ethics Statement

The hosting university's Institutional Review Board reviewed the design of this study and gave approval for use of human subjects. We obtained written consent from all participants.

Subjects

Subjects came from a highly selective large private institution in the United States. The student population was highly homogenous in terms of culture and ethnicity. The students from this university performed in the 96th percentile of all universities on the evolution section of the ETS Biology Field Exam and in the 99th percentile specifically on the Population Genetics and Evolution Assessment Indicator (AI7). This exam is administered at the end of a capstone evolution course, usually in the senior year, and is used by the university to evaluate the effectiveness of life sciences programs. To address the issue of how prevalent and how persistent tree-thinking misconceptions are among undergraduates we recruited participants from two undergraduate courses at the host institution: a traditionally freshman level course and a senior capstone course, each described below. We selected these courses to represent the totality

of students' educational progression as students in the intro course will be required to take the senior capstone course as part of their program of study, and vice versa.

We recruited 76 students from 6 sections of an introduction to biology for life science majors course (INTRO). This course is the first course in the curricula of several life science majors at the university. We selected this course to assess misconceptions at the beginning of an undergraduate life science major. Subjects were offered extra credit as an incentive to participate in the study and were recruited with a classroom announcement. 71% of students in the INTRO course participated in the survey.

We recruited 39 students from two sections of an evolution course (EVO) that is intended as a capstone course to be taken by students nearing the end of their undergraduate studies. These students receive considerable instruction on tree-thinking, including a lab designed to teach the basics of phylogenetic systematics. We included this course in the present study to assess misconceptions that persist until the end of a student's undergraduate career. Subjects were offered extra credit as an incentive to participate in the study and were recruited with a classroom announcement. 75% of students in the EVO course participated in the survey.

Measuring evolutionary tree understanding in students from these two courses allowed us to see how students entering the university compared with those who were near the end of their studies. Other studies have used a similar design to compare differences in student thinking (Atman, Cardella, Turns, & Adams, 2005; Azizi-Fini, Hajibagheri, & Adib-Hajbaghery, 2015; Genco, Hölttä-Otto, & Seepersad, 2012; Kögce & Yıldız, 2011; Meir et al., 2007).

Study Design

To measure misconception prevalence, we used an assessment that contained at least two items to elicit each of the misconceptions described earlier (Reading the Tips, Node Counting, Ladder Thinking, and Similarity Equals Relatedness). To create a valid assessment, we used two independent researchers who study tree-thinking with extensive experience in teaching these concepts to undergraduate students. Each researcher chose items from the previously published *Tree-thinking Quizzes I and II* that corresponded to misconceptions identified above (Baum et al., 2005). Eight items were selected. The authors wrote two additional items (17/18 and 19/20) that were included in the assessment. These items were based on previous student responses and interactions where misconceptions were demonstrated. Each question had the potential to elicit multiple misconceptions depending on the answer choice chosen and on the reasoning described (see Figure 5 and Table 2). Reliability statistics are described below. We used a multiple-choice format with the goal of producing an easily scored objective assessment. One issue with using a multiple-choice format in determining the prevalence of misconceptions among students is that the same wrong answer can be selected due to several different misconceptions. For example, in Figure 1, students may incorrectly choose answer choice A using the Reading the Tips misconception or Similarity Equals Relatedness. Similarly, students can choose the correct answer (answer choice B) using the wrong reasoning (Node Counting or Ladder Thinking). To overcome this issue students answered the multiple-choice content-based question and then answered a follow-up free-response question explaining the reasoning behind their choice. Doing this allowed us to more accurately determine any misconception the subject held. This approach is similar to the pattern used on other assessments such as Lawson's Classroom Test of Scientific Reasoning (Halverson et al., 2011; Lawson, 1978).

Table 2. The misconception most commonly associated with selected answer for each item pair and the correct answer for each item pair.

Question Pair	Answer Option	Most Commonly Categorized Misconception		Correct Answer
		INTRO	EVO	
1/2	A	Reading the Tips	Ladder Thinking	B
	B	Similarity Equals Relatedness	Branch Length	
	C	Reading the Tips	Node Counting	
3/4	A	Reading the Tips	Reading the Tips	B
	B	Ladder Thinking	Branch Length	
	C	Node Counting	Node Counting	
5/6	A, B	Ladder Thinking	Ladder Thinking	E
7/8	A, B, C, D, E	Ladder Thinking	N/A	E
9/10	A, B, D, E	Ladder Thinking	Ladder Thinking	C
11/12	A	Reading the Tips	Node Counting	C
	B	Similarity Equals Relatedness	Ladder Thinking	
	C	Ladder Thinking	Branch Length	
13/14	A, B, E	Ladder Thinking	Ladder Thinking	C
15/16	A, B	Ladder Thinking	Ladder Thinking	C
	D	Similarity Equals Relatedness	N/A	
17/18	A, B, D	Similarity Equals Relatedness	Similarity Equals Relatedness	C
	C	Similarity Equals Relatedness	Ladder Thinking	
19/20	A	Ladder Thinking	Ladder Thinking	D
	D	Similarity Equals Relatedness	Ladder Thinking	

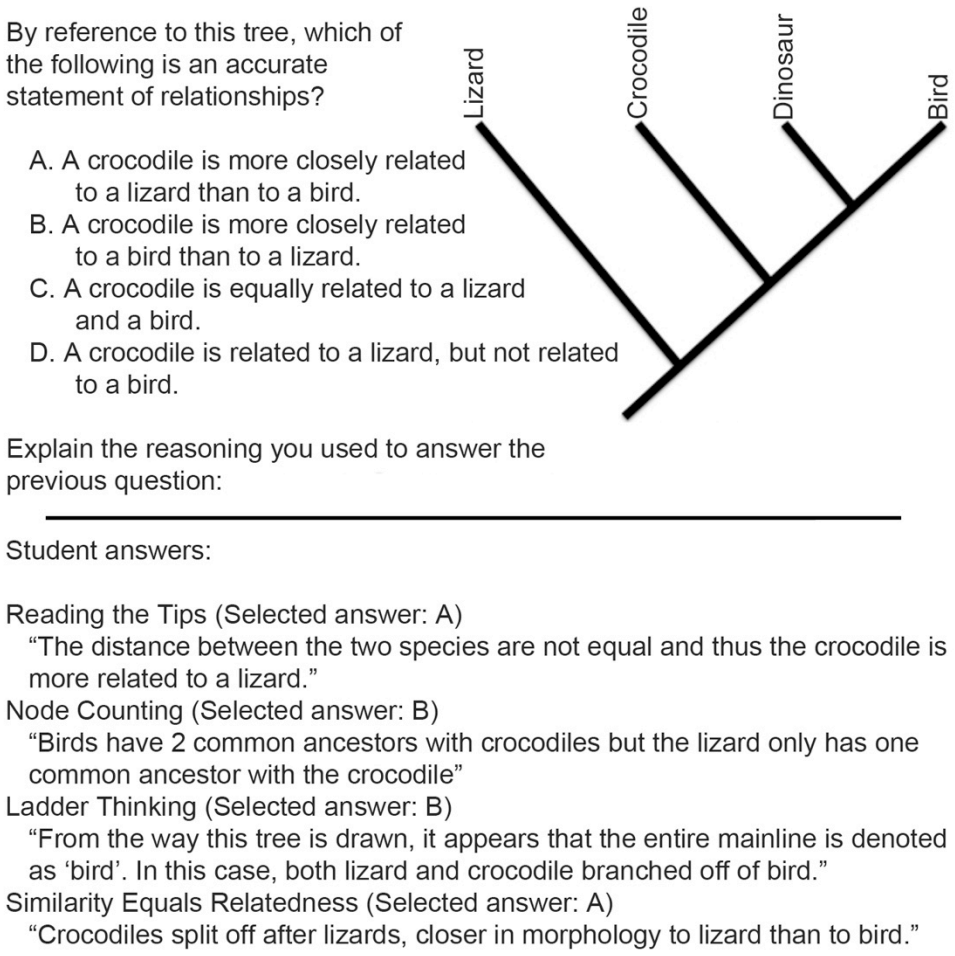


Figure 5. A sample item set from the assessment used to measure student misconception and examples of student responses to the item set.

The assessment consisted of 20-paired items: 10 multiple-choice content-based items and 10 open response follow-up reasoning items (see appendix). We administered the assessment to students using an online survey system. Students had the opportunity to take the assessment during a one-week period after being recruited with an in-class announcement and an email containing the link. Not all courses were surveyed at the same time. We administered the assessment to students in the INTRO course prior to the students receiving any formal instruction on tree-thinking related topics. This was done to assess what level of misconception the students had at the beginning of their undergraduate study of life sciences. We administered the survey to

the EVO course near the end of the course to assess the misconception levels of students near the completion of the capstone course.

We calculated a student's score on the exam using only the 10 multiple-choice content-based items. We used the follow-up reasoning items in conjunction with the multiple-choice items to diagnose the presence of common misconceptions held by students. To interpret the results of the reasoning questions two raters individually evaluated each pair of items with an emphasis on the written response. Raters were science education researchers and instructors of introductory biology. They were both familiar with the tree-thinking assessment items and had experience with student responses to these items. Data was anonymous so that raters were unaware of student identities; however, they were aware of the populations (INTRO and EVO) from which the data came. Raters evaluated subject responses item by item. Raters classified each response as correct, correct with one of the described misconceptions, incorrect with one of the described misconceptions, or incorrect with no clear misconception. Misconceptions were identified by first looking at the answer selected, for many of the questions a selected answer indicated a misconception was likely held. Raters would then evaluate phrasing in the written response to either confirm or identify a misconception not indicated by the selected answer. Example responses that would indicate each of the given misconceptions are shown in Figure 1; additional examples are provided in the Supplementary Materials.

Raters evaluated written responses with no clear misconceptions for any commonalities between them in an effort to identify misconceptions not previously described. If a subject was deemed to have demonstrated a misconception on any one item, we classified them as holding that misconception. This approach allowed us to effectively measure the prevalence of each misconception but it did not give us an indication of how strongly each subject held a

misconception. After individually evaluating each response the raters met together and discussed differences in evaluation in an attempt to reach agreement on the evaluation.

Statistical Analysis

We analyzed the persistence of each misconception using Mann-Whitney U tests using SPSS software v. 21. These analyses allowed us to evaluate the significance of any differences found between the groups. Mann-Whitney U tests were used because the data failed to meet the assumption of normality.

We ran a Spearman's correlation of the two raters' evaluations to measure the degree of correlation between them. We found that reviewer categorizations were significantly correlated with each other with an inter-rater reliability of 0.992.

We also used Cronbach's alpha to measure the internal reliability of the assessment for only the multiple-choice item responses. The analysis produced a Cronbach's alpha coefficient of 0.638. A Cronbach's alpha of 0.638 is within the acceptable range for an assessment of this type (Kline, 2000). Cronbach's alpha functions as an equivalent of KR-20 when used with dichotomous data. Because multiple misconceptions could manifest from the same item we did not evaluate the internal reliability of subsets that we intended to measure specific misconceptions.

Results

Overall Prevalence of Misconceptions

The prevalence and persistence of tree-thinking misconceptions varied among the four misconceptions we measured in this study. We compared student responses from the EVO course to student responses in the INTRO course. The results of the Mann-Whitney U tests showed that significantly fewer students from the EVO course gave answers that were based on the Reading the Tips misconception. However, we found that students from the EVO course demonstrated significantly higher levels of the Node Counting misconception than those in the INTRO course. The most prevalent and persistent misconception we measured in this study was the Ladder Thinking misconception; students at both levels demonstrated a high level of this misconception. The Similarity Equals Relatedness misconception was equally prevalent in both courses.

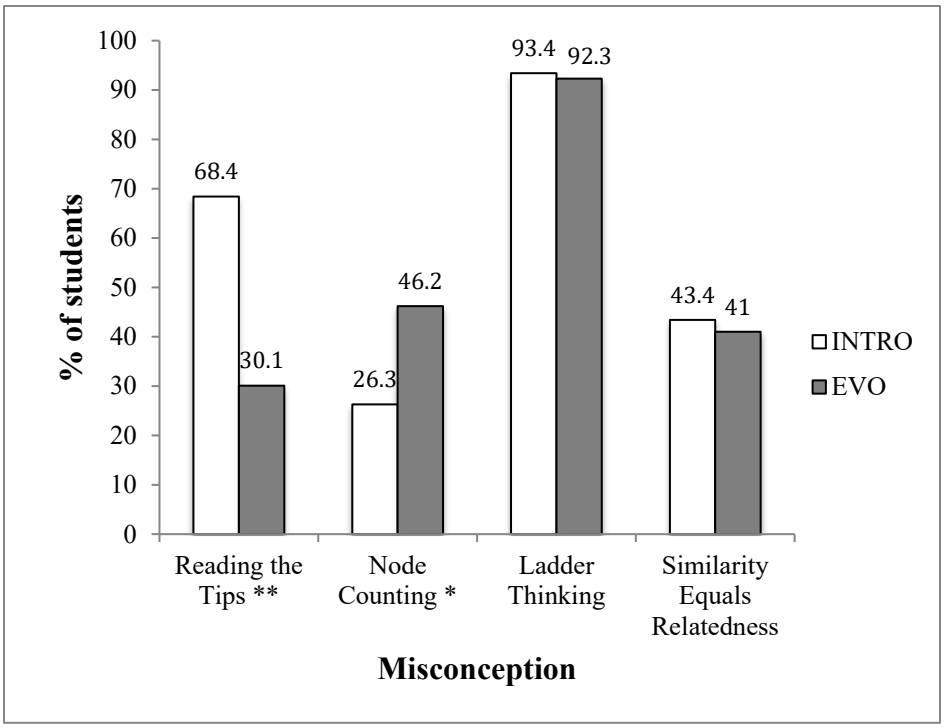


Figure 6. The proportion of students who gave answers indicating they held each of the misconceptions assessed in this study for the INTRO course and the EVO course. ** represents a p-value < .01 * represents a p-value < .05.

Table 3. The results from the Independent-samples Mann-Whitney U Test.

Misconception	Subjects (n)	Mann-Whitney U	*p*-value
Reading the Tips	115	924.00	0.000
Node Counting	115	1776.0	0.033
Ladder Thinking	115	1465.5	0.825
Similarity Equals Relatedness	115	1446.5	0.807

Detailed Student Response Rates

To look for further evidence as to how tree-thinking differs between students in the EVO course and those in the INTRO course, we compared the percentage of each misconception used to justify student answers on each of the questions in the survey. Several interesting patterns were found in responses. Percentages pertaining to each question are displayed in Supplementary Table l. We will highlight the main findings below.

Reading the Tips

Looking at question pairs 1/2, and 3/4, we see that explanations using Reading the Tips reasoning were much more frequent in the INTRO course than in the EVO course. For example, in question 1, distractor 'C' was designed to elicit the Reading the Tips misconception and was chosen for this reason 42% of the time among the INTRO students while only 15% of EVO students did this. Likewise, in question pair 3/4, 55% of INTRO students chose distractor 'A' and used Reading the Tips reasoning compared to only 18% in the EVO course. The same trend is seen in item pair 11/12.

Unexpectedly, we saw a large proportion of students in the INTRO course using Reading the Tips when answering item pair 15/16, a question not specifically designed to elicit this misconception. In fact, 16% of students chose distractor 'D' citing reasoning such as, "Student D described how the closer the species are on the tree, the more closely related they are" or "All of the branches are related to the branches next to them." Interestingly, distractor 'D' was

designed to elicit Similarity Equals Relatedness. Students in the EVO course almost exclusively used Similarity Equals Relatedness when choosing this distractor (18%), using reasoning such as, "The alga are most closely related because the only thing differing between them is their color, and moss and pine are most closely related because their common ancestor developed into a multi-cellular organism."

Node Counting

Opposite of Reading the Tips, the Node Counting misconception appears to be more common in EVO students than INTRO students. On item pair 1/2, we see that instead of choosing distractor 'C' using Reading the Tips reasoning like INTRO students, the majority of students in the EVO course that chose "C" did so using Node Counting reasoning, such as, "There is just one node between a trout and a coelacanth and there is just one node from a trout to a stingray." We see a similar shift on item pair 3/4. In INTRO students, only 3% of students chose distractor 'A' and 11% chose distractor 'C' using Node Counting reasoning; whereas, 15% of EVO students chose distractor 'A' and 21% chose distractor 'C' using Node Counting reasoning.

Item pair 11/12 resulted in a similar pattern. 66% of students in the INTRO course and 62% of the students in the EVO course selected distractor 'A'. However, 45% of students selected "A" using Reading the Tips in the INTRO course while only 7% used Node Counting. In the EVO course we see 38% of students selected "A" using Node Counting compared to 21% using Reading the Tips.

Ladder Thinking

The prevalence of Ladder Thinking was found to be equal and at high levels in both groups. Looking at responses in finer detail shows that there were differences between the two

groups even though the overall prevalence was similar. In item pair 7/8, Ladder Thinking is the only misconception that manifests in student explanations. However, students in the INTRO course were nearly equally distracted by answers 'B' (12%) and 'D' (16%) and only slightly more by 'C' (28%); whereas, EVO students overwhelmingly favored distractor 'C' (54%) while not entertaining 'D' at all. It appears that students in the EVO course were less distracted by words such as 'intermediate' or 'advanced' but still maintained 'ancient species' as a probable explanation. Interestingly when 'ancient' is not used as a distractor, as in item pair 15/16, students are more likely to consider the possibility that a species could be 'most advanced', i.e., 28% of students chose answer choice 'A' in the EVO course as opposed to only15% in the INTRO course. Explanations like the following were used to justify this answer: "The pine is the latest to diverge and builds on the changes made to each previous species."

Similarity Equals Relatedness

On item pair 1/2, the use of Similarity Equals Relatedness in student explanations is somewhat common among INTRO students (18%). In the EVO course we see that Similarity Equals Relatedness is never used. We see a similar pattern on item pairs 3/4 and 11/12. Despite this small but convincing shift, the overall prevalence of Similarity Equals Relatedness was not significantly different. Two item pairs can help to explain this phenomenon. On item pair 15/16, distractor 'D' suggested similarity as the bases for relatedness. In the EVO group 28% of students selected D with 18% indicating Similarity Equals Relatedness as the reasoning. In comparison, 19% of the students in the INTRO class used this reasoning. We also see that on item pair 17/18, when students are asked to place a species on a tree, EVO students routinely (36%) cited similarity, regardless of the selected answer, as the basis for determining where in

the tree it should fit (e.g., “I know that dolphins are fish-like, but they are also mammals so I put them in between fish and mouse”).

Branch Length

On the items that asked students to compare relatedness (item pairs 1/2 and 3/4) we noticed a common reasoning pattern among EVO students that seemed to be different than our four defined misconceptions. We labeled this misconception Branch Length because students were explaining their reasoning by describing the length of the branches connecting the two species. Interestingly, this misconception manifested despite the student getting the correct answer. For example, in item pair 1/2, 21% students chose the correct answer, ‘B’, but did so by comparing branch lengths as did this student: “The branch length is longer between stingray and trout than it is between trout and coelacanth, so the trout is more closely related to a coelacanth than to a stingray.” The same pattern was seen on item pair 3/4, with 10% of students choosing the correct answer, ‘B’, by using reasoning like the following: “The line drawn from Crocodile to Bird is shorter than the line drawn from Crocodile to Lizard.” A similar pattern was seen on item pair 11/12. We compared the prevalence of these branch length-based responses between the courses. We used a Mann-Whitney U test and found that branch length-based responses were significantly higher in the EVO course (n = 115, U = 1,842.5, p-value = .000; see Figure 7).

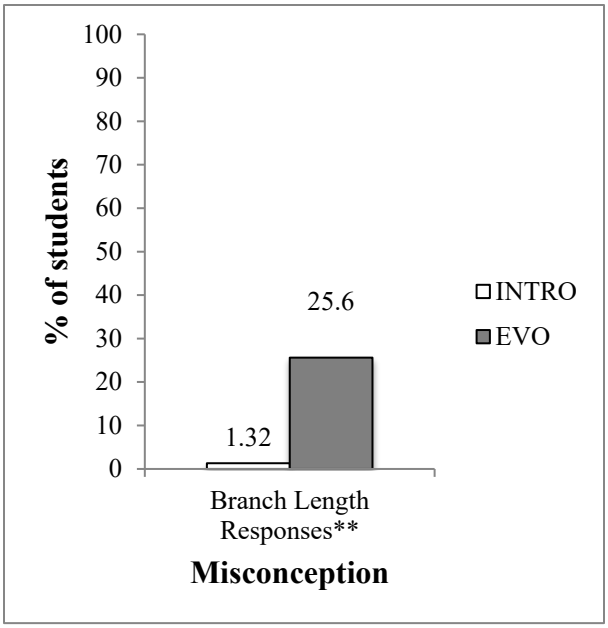


Figure 7. The proportion of students who gave response based on branch length in the INTRO course and EVO course. ** represents a *p*-value < .01.

Student Performance

We compared student overall performance on the content component of the assessment between courses. The average score for both groups was below 50% on the assessment. Using a Mann-Whitney U test we found no significant difference between the groups in assessment performance (n = 115, $U = 1{,}635.5$, p-value = .240). Despite changes in some misconception levels, student performance on the assessment overall remained low for both groups.

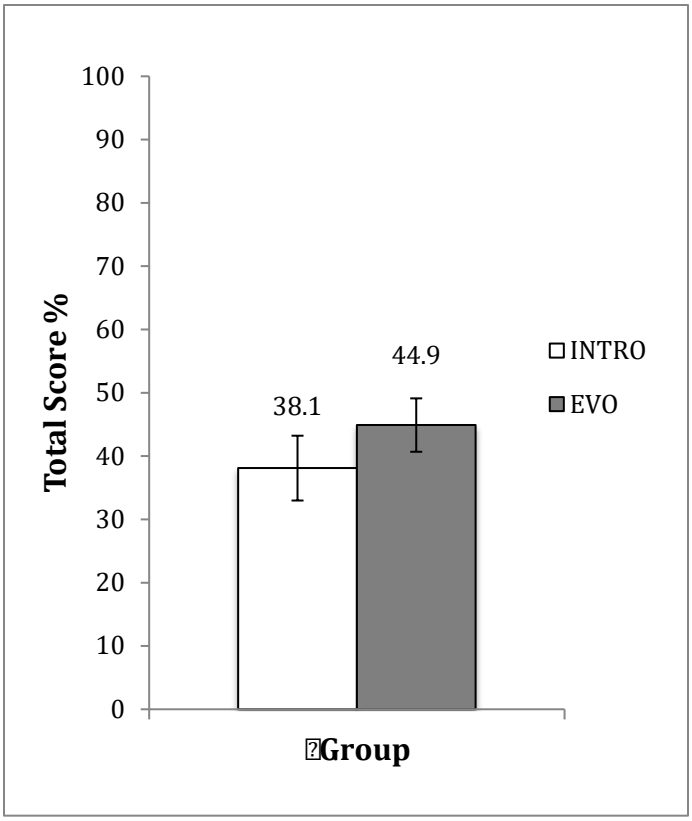


Figure 8. A comparison of overall performance on the assessment for each group in the study. Error bars represent one standard error.

Discussion

The prevalence of the misconceptions in our study varied in interesting and informative ways. Based on our study populations being at the beginning of their education and at the end, we expected to see a decrease in all misconceptions due to specific instruction on this topic presumably throughout their biology degree, including a capstone course on evolution. This was indeed the case for some misconceptions but not for others. Reading the Tips and Node Counting potentially demonstrate an inverse relationship. Reading the Tips went from highly prevalent in the INTRO course to extremely low in the EVO course. This indicates that students are, for the most part, rejecting this misconception as they progress in their education, presumably due to course exposure. Node Counting, however, appears to have the opposite

trend. Students in the EVO course demonstrated significantly higher levels of Node Counting when compared to the INTRO course students, indicating that instruction is not addressing this misconception and perhaps is introducing it.

One potential explanation for these two findings is that many students are abandoning Reading the Tips and replacing it with Node Counting. Students may be experiencing disequilibrium as they attempt to interpret relatedness using the closeness of the tips but are instructed that the nodes are what matter (Bransford, Brown, & Cocking, 1999). Thus, they may be abandoning Reading the Tips as a method for interpreting evolutionary trees and instead adopting a Node Counting misconception. This potential explanation is supported by the lack of a significant difference in overall performance on the assessment. If the Reading the Tips misconception had been replaced by a correct understanding, we would have likely seen a significant difference between the overall scores of the groups. Evidence from the question detail also supports this potential explanation. In all three question pairs that focused on having students interpret evolutionary relationships we saw dramatic differences between the INTRO group and the EVO group when it came to Reading the Tips and Node Counting. EVO students in all three pairs used Node Counting at higher rates and used Reading the Tips at lower rates than INTRO students. However, certainly there may be other explanations that fit this data. Poor performance on interpreting relationships in this study is supported by another study that found no significant difference in student ability to interpret relationships from an evolutionary tree before and after instruction on evolutionary trees (Dees et al., 2014). The authors suspected that students were using the number of steps separating taxa to determine relatedness. The high levels of Node Counting held by EVO students found in this study appears to support their suspicion.

In a 2007 study that also includes Reading the Tips (Tip Proximity) and Node Counting, researchers saw a different pattern (Meir et al., 2007). They saw no significant differences between lower level course students and upper level course students for prevalence of either of these misconceptions. One potential explanation for this difference is in the study design. For their study students from both groups received at a minimum one lesson of instruction on evolutionary trees. In contrast, our study compared students (INTRO) who had received no instruction on evolutionary trees to students who had received multiple lessons on the topic. It is possible that both groups in the 2007 study more closely represent those in the EVO course than they do those in the INTRO course even if they are enrolled in a lower level course. This would explain why there was no significant difference between the upper and lower level courses. We do see one important commonality between their upper level students and our EVO students, both groups have higher levels of Node Counting than of Reading the Tips. In other words, for these two misconceptions we see a similar pattern in both studies for the advanced courses.

While the inverse relationship between Reading the Tips and Node Counting is potentially explained as described above we also see a similar pattern with Reading the Tips and Branch Length. Indeed, on the same three questions pairs (1/2, 3/4, 11/12) we see Branch Length used at a higher percentage and Reading the Tips used at a lower percentage by EVO students. A number of possible explanations may describe how the thinking of an INTRO student changes to that of an EVO student in light of the patterns we see with Branch Length. Not only might they adopt Node Counting as previously described but they may also go from Reading the Tips to Branch Length or perhaps from Reading the Tips to Node Counting to Branch Length. A study consisting of student interviews, pre and post assessment, and tree-thinking interventions that include phylogram introduction is needed to test these potential

explanations. This would allow researchers to identify students who had Reading the Tips as a misconception and then after instruction demonstrated that they held Node Counting or Branch Length. After identifying these students, a series of interviews could be conducted to determine why students abandoned the Reading the Tips misconception and why they adopted Node Counting or Branch Length. Our current study can only provide us with patterns that allow us to generate hypotheses for future tests.

Perhaps the most troubling finding in this study is the high levels of the Ladder Thinking misconception found in both groups. When looking at the question detail we do see some slight differences between the groups though the overall prevalence remains unchanged. Students still held Ladder Thinking ideas but the interpretation of the trees may have been different. For example, a student in the EVO course appears to be more likely to favor explanations declaring species to be advanced or primitive over explanations that one species on the tree was an intermediate on a path of transformation. Ladder Thinking is still indicated by both types of explanation but the view of what the tree is depicting is different. As mentioned previously this misconception indicates a fundamental misunderstanding of evolution based in teleological reasoning. Teleological reasoning is a pervasive idea that contributes to misconceptions in many fields of science (Kampourakis, 2007). The persistence of the Ladder Thinking misconception is perhaps understandable when we consider how entrenched the underlying reasoning is among students. Teleology is pervasive because it is fundamental to how most people see and interpret the world around them. Fundamental presuppositions, like teleology, are very difficult to overcome (Vosniadou, 1994). If the fundamental presupposition is not addressed, a misconception will be persistent. We believe lessons and activities that seek to address Ladder Thinking must also address teleological reasoning to be successful.

Meir et al. also looked at a subset of Ladder Thinking, that a straight line equals no change (2007). This means that if no branching is occurring on the line to a particular species than there must be no evolution occurring. Given that this is just one subset of responses that we would categorize as Ladder Thinking it is expected that they saw lower levels of Straight Line Equals No Change (40%) than we saw of Ladder Thinking (our percentage). Gregory and Ellis conducted a study to look at conceptions of evolution among science graduate students (2009). They found that the most commonly held misconception among graduate students was teleological thinking, and viewing evolution as a progressive process as the second most commonly held misconception. Given the ties of both of those concepts to Ladder Thinking we believe our study, along with Gregory and Ellis, show that Ladder Thinking and teleology relating to evolution are misconceptions that still need to be addressed even with advanced students.

Similarity Equals Relatedness was shown to be persistent across both courses. This misconception can be difficult because of the role similarity plays in the building of evolutionary trees. Many instructors focus on tree reconstruction methods using character matrices when they teach students about evolutionary trees. It is possible that students are mistaking similarity as the basis for building the tree rather than parsimony. When a dominant characteristic (morphological or ecological) is due to an ancestral state (plesiomorphy) or convergence rather than synapomorphy, it could lead to incorrect interpretations of evolutionary relationships by students who mistake basic similarity as the basis for determining relatedness. Novick, Catley, and Funk found one potential explanation for why similarity can be such a challenge for students when looking at evolutionary relatedness (2011). This study asked students to compare evolutionary relationships between three taxa. Researchers found that when familiar taxa were

used in the comparison students who could correctly determine relationships with unfamiliar taxa often gave incorrect responses. Students' familiarity with the physical characteristics of the taxa caused them to override their reading of the tree. Unfortunately, our study used taxa that were likely familiar to the students in each of the questions asking the students to compare relationships so we were not able to see if a similar pattern of obstruction was seen in our students when answering.

The question detail provided some interesting patterns that were not apparent from comparing the total prevalence from both groups. In the INTRO group we saw students using Similarity to determine relatedness on several questions including those that required students to compare evolutionary relatedness and those that required students to place a new taxon in a tree. Alternatively, in the EVO group we saw that students on almost every evolutionary relatedness question did not use Similarity as an explanation. One exception was on question pair 15/16. Answer D specifically appeals to similarity as the basis for determining relationships. In the EVO group we saw Similarity Equals Relatedness used to justify a determination of evolutionary relationships but only on answer D. The overall trend from all the questions seems to indicate that EVO students do not rely on similarity when being asked to determine relatedness but some may be open to using it if it is suggested by the answer choice, as was the case in pair 15/16.

Despite this difference in using similarity for analyzing an existing tree, we did not see a difference in the overall prevalence of this misconception primarily due to question pair 17/18. This question asked students to place a new taxon in an existing tree. When placing the taxon in the tree EVO students were just as likely to place it in the tree based on similarity as the INTRO students were.

In addition to the four misconceptions discussed above, we found that many students in the EVO course referred to the overall length of the branches (e.g., in millimeters) between the taxa in their written explanations as something that had meaning, indicating a new misconception. The reasoning was that the longer the branch lengths were between two taxa, the less related these taxa were. We believe this may be a potential misconception that forms when students are introduced to phylograms. Branch length in a phylogram is informative, but the branch length provides information on the rate of change not on the relationships between taxa. An alternative explanation is that this pattern is a special case of the Similarity Equals Relatedness misconception. Branch length in a phylogram represents some sort of evolutionary change (DNA mutations) that occurred after the most recent branching. The higher the number of evolutionary changes, the longer the branch. It is possible that students are interpreting longer branch lengths between two taxa as dissimilarity and using that to infer less relatedness. Student responses were not extensive enough to distinguish the underlying reasoning. Further study and characterization of students using this explanation is needed.

Although our findings are compelling and informative, certain limitations should be considered. Our research subjects came from a highly selective institution for academics. Thus, the prevalence of these misconceptions may not necessarily reflect what would be seen at a less selective institution. In addition, the institution is a private religious institution with neutral views toward evolution but cultural biases may exist that could potentially influence results (Ludlow & Evenson, 1992). To account for these biases, we examined student responses to items 19/20, which discusses human evolution to determine whether this item was more frequently missed than others. The data showed that this item had the highest number of accurate responses in comparison to all other items suggesting that students were capable of

responding scientifically to potentially controversial evolution statements. In addition, in a recently published study using subjects from the same institution, researchers found that student acceptance of evolution, as measured by the Measure of Acceptance of the Theory of Evolution (MATE), was over 80% and knowledge of evolution, as measured by the Knowledge of Evolution Exam (KEE), was nearly 80% (Manwaring, Jensen, Gill, & Bybee, 2015). This is in contrast to a study of students at the University of Minnesota who scored an average of 54.2% on the KEE and a study of life science faculty at a large Midwestern university who scored 74.3% on the KEE and 87.6% on the MATE (Moore & Cotner, 2008; Rice, Clough, Olson, Adams, & Colbert, 2015). Thus, our student body does not appear to have different acceptance or knowledge of evolution than other student bodies.

Another potential limitation is in regards to our assessment. Although we took most items from a previously published assessment it had not been subjected to rigorous statistical analyses for reliability and was not necessarily intended to be a concept inventory (Baum et al., 2005). We added an additional two items and included an overall reliability analysis. This instrument appeared sufficient to solicit student misconceptions enough to quantify them. Since a published tree-thinking concept inventory does not yet exist, we are limited to what experts have produced.

In addition, the assessment was only 10 questions. It is possible that students who have a strongly held misconception may also have other misconceptions that are not elicited by the limited number of questions. Given the choice of two answers on a given question, it may be that the strongly held misconception influenced the choice even though the other misconception was held. We attempted to overcome this issue by having targeted questions for each

misconception and including free-response opportunities to explain reasoning, but it does not preclude this possibility.

Conclusion

From this study, we conclude that tree-thinking is a difficult skill for undergraduate students. Despite our students performing exceptionally well on the ETS Biology Field Exam (in the 96th percentile), they have a fundamental misunderstanding of and inability to correctly interpret phylogenetic trees. In addition, we see that each of the misconceptions, which we measured, have different trajectories across a student's educational career. Misconceptions related to reading the graphic (Reading the Tips and Node Counting) proved to be tricky to address in that we may be able to correct one practice (i.e., Reading the Tips) but we may simply be replacing it with an alternative erroneous practice (i.e., Node Counting, Branch Length). This indicates that instructors should be meticulous in the way that they teach students to read a phylogenetic tree and ensure that students have indeed understood what each component of the tree represents (e.g., a branch, a node). Using formative assessment often is a way to ensure that students understood the concept in the way in which the instructor intended. It would also allow instructors to detect any new misconceptions that may have been introduced (e.g., measuring branch lengths). Using a tree-thinking assessment would be useful to this end.

Alternatively, misconceptions related to the fundamental underpinnings of evolutionary theory (Ladder Thinking and Similarity Equals Relatedness) proved nearly completely resistant to change, remaining equally prevalent in freshman and seniors. This indicates the need for instructors covering evolutionary topics at all stages to attempt to directly address these fundamental misconceptions. Ladder Thinking is by far the most prevalent and persistent of all the misconceptions addressed in this study, indicating a significant need for the development and

testing of educational interventions that address this problematic misconception. Ultimately, by understanding the prevalence and persistence of misconceptions of students in our classroom, we can make pedagogical decisions targeted for our course.

DEVELOPMENT OF A TREE-THINKING CONCEPT INVENTORY

Introduction

Assessing student understanding is a critical tool for educators as they try to help students learn the important content of a given class or course (Miller, Linn, & Grunland, 2013). What and how an assessment measures student understanding is also important for researchers who are trying to understand the effectiveness of various interventions targeted at improving or understanding student learning. Educators and researchers are often required to develop their own assessments due to a myriad of reasons (e.g., unique subject area or appropriate level of rigor). Unique assessments also pose a problem for researchers when we attempt to evaluate the effectiveness of an intervention and put it in context with other studies. Without a widely used and accepted concept inventory, it becomes difficult to appropriately compare the results of one study with another.

To address these problems educational communities have developed concept inventories. Concept inventories are typically multiple-choice assessments that focus on important concepts relating to a subject area. This focus on conceptual understanding allows the assessment items to focus on assessing understanding and applying concepts rather than knowledge level information that can be memorized (Garvin-Doxas, Klymkowsky, & Elrod, 2007; Krathwohl, 2002). A concept focused assessment that has been properly developed provides educators with confidence in what the assessment reveals about their students' understanding and provides researchers with a more objective and meaningful form of evidence as we study student learning of the given subject.

Evolutionary trees are crucial in modern biology (Thanukos, 2010). Biologists utilize evolutionary trees to study biological phenomena ranging from genes to biogeography (Baum et

al., 2005). Evolutionary tree concepts have been dubbed tree-thinking by researchers who advocate the importance of these concepts (O'Hara, 1997). In the last decade, significant research has been conducted on the learning of tree-thinking. Researchers have provided great insight, into how students think about evolutionary trees, common misconceptions students exhibit, and how best to teach evolutionary trees to students (e.g., Eddy et al., 2013; Gregory, 2008; Halverson et al., 2011; Perry et al., 2008)

While important and interesting insights have resulted from research on tree-thinking the research into this area is inhibited by the lack of a concept inventory for tree-thinking that has been published and made available to researchers (Walter et al., 2013). Without a widely available concept inventory researchers are left to create their own or modify an existing assessment that was not published as a concept inventory. While a few published assessments do exist, they do not meet the characteristics and standards of a concept inventory because they lack proper development/evidence to be concept inventory or they cover only one component of tree-thinking (Evolutionary relatedness) (Baum et al., 2005; Blacquiere & Hoese, 2016; Smith et al., 2013). The Tree-thinking Concept Inventory is the most promising assessment that has been published but it has since had some of its evidence of validity called into question (Walter et al., 2013). Due to the clear need for a concept inventory on Tree-thinking we have sought to develop a new inventory targeted to undergraduate students and to provide the appropriate evidence of validity and reliability that would give researchers and educators confidence in the measurement it provides.

Methods

Subjects

We recruited a total of 581 subjects from a variety of life science courses (both majors and non-majors) to participate in this study. We used three subsets of subjects from the total of 581 to conduct different analyses. The Final Group consists of all 531 subjects who completed the final version of the assessment. The Convergent-Discrimination Group consists of 124 subjects who completed the final version of the assessment as well as two other assessments. The Test-Retest Group consists of 120 subjects who completed the final version of the assessment twice in a six-week period. Subsets used are identified in Table 4. The host institution IRB approved this study. Subject consent was obtained for the use of their scores in this project.

Table 4. Groups of subjects used to provide evidence of validity and reliability.

Group	Subjects (n)
Final Group	531
Convergent-Discriminant Group	124
Test-Retest Group	120

Content Validity

Student Understanding

We began the process by trying to learn more about how students think and reason with evolutionary trees. In order to do this, we administered a set of multiple-choice items and free response items to subjects in several biology courses. We reviewed student responses and coded them as being correct or as demonstrating one of the common misconceptions described in the literature (Baum & Smith, 2013; Gregory, 2008; Halverson et al., 2011; Meisel, 2010; Omland et al., 2008). This process helped us become familiar with patterns of student thinking including concepts they understood well and concepts they did not. We also met with eight students to hold discussions about evolutionary trees and the items they had responded to. Evaluating the

responses and discussing the items aided us in creating items and distractors that would both target key concepts and were worded in ways that were compatible with student thinking.

Learning Outcomes

We developed an initial list of learning outcomes with the focus on what would be appropriate for undergraduate students of biology to understand about evolutionary tree concepts. Determining a set of learning outcomes is critical to the development of an assessment (Wiggins & McTighe, 2005). It allows us to develop items that directly address key concepts of tree-thinking and it ensures that each outcome is given coverage in the assessment. A 4-person panel of experts (three evolutionary biologists and a science education researcher) reviewed the initial list. Each expert provided feedback on the learning outcomes and helped as well as proposed additional outcomes. Tree-thinking learning outcomes defined by others are comparable to the learning outcomes we developed for this concept inventory (Eddy et al., 2013; Novick & Catley, 2012), see Table 5.

Table 5. Identified learning outcomes and the hypothesized constructs.

Learning Outcomes
1- Accurately interpret information depicted in an evolutionary tree using an understanding of common ancestry
a. Distinguish monophyletic, paraphyletic and polyphyletic groups b. Compare evolutionary relationships between taxa c. Identify what various components of an evolutionary tree represent d. Distinguish between evolutionary trees with differing topologies and evolutionary trees with depicting differing evolutionary relationships
2 - Demonstrate an understanding of how characters are inherited from common ancestors by accurately interpreting an evolutionary tree
a. Identify cases of homology and analogy when interpreting an evolutionary tree b. Analyze character information and evolutionary trees using parsimony c. Identify synapomorphies for a group on a given evolutionary tree d. Identify character states as derived or ancestral on a given evolutionary tree e. Use an evolutionary tree to identify characters a given taxon would exhibit
3 - Demonstrate an understanding of evolution as a continuing and non-teleological process
a. Identify why using simplicity and complexity to categorize organisms as primitive and advanced species is inappropriate from an evolutionary perspective b. Demonstrate an understanding that all extant populations continue to evolve and have evolved their entire existence

Item Development

We used student responses described previously and the learning outcomes to develop multiple-choice items. We developed at least two items to address each learning outcome in the hopes of providing robust coverage of the outcomes. Each item consisted of a questions pair. The first question was directly related to the content the item was designed to assess and the second question addressed what reasoning was used to answer the content question. This method is patterned after Lawson's Classroom Test of Scientific Reasoning (LCTSR) (Lawson, 1978). Using paired questions benefits us in two ways. First, it reduces the impact of guessing by requiring the subject to answer both questions correctly to receive credit for a correct

response to the item. We found in our discussion and reading responses from students that they could often answer a content question correctly but for an incorrect reason. The paired questions allow us to account for this and more accurately differentiate students with accurate understanding from those with incorrect understanding. Second, it helps us better identify misconceptions that a student might be using when they answer an item.

Student responses and discussions were valuable in developing questions that were appropriately worded for students while still targeting our learning outcomes. Student responses also served as the primary source for the wording of distractors that were appropriate for the question but also represented common misconceptions (Garvin-Doxas et al., 2007).

Research has shown that when some students, who are able to accurately interpret evolutionary trees using abstract or unknown taxa, are given an evolutionary tree with known taxa they are unable to interpret the phylogeny correctly (Novick et al., 2011). This is likely due to a common misconception, which was defined previously as Similarity Equals Relatedness. This misconception results when students rely on a similarity of features to determine relatedness rather than what is depicted in the evolutionary tree. When the taxa are abstract or unknown to the student they cannot rely on similarity to interpret the evolutionary tree. We developed items that used both abstract taxa and well-known taxa with this finding in mind. We believe this will allow the concept inventory to distinguish between students with no understanding of how to interpret evolutionary trees, those who can only do so with abstract or unknown taxa, and those who can regardless of the taxa used.

We used multiple rounds of revision to refine our items. Both major and non-major students reviewed our initial items. These students were asked to review the questions and then comment on anything that seemed out of place or confusing about the question. We selected 20

students to interview and asked them to describe their thinking about the items and give us feedback. We conducted a group discussion with six students after this revision asking them to discuss the items and provide feedback on any aspect that may have been confusing. After this final round of student comment and revision, the items were administered to a set of non-major introductory biology students. Following the piloting of the inventory, we submitted 26 two-part multiple-choice items to the expert panel. The panel reviewed each item, suggested additional revisions, and gave approval.

Table 6. The items intended to address each learning outcome. R1 and R2 were items removed from the final version of the concept inventory as a result of item analysis.

Learning Objective	1a	1b	1c	1d	2a	2b	2c	2d	2e	3a	3b
Items	1	6	2	21	3	5	8	7	9	11	12
	20	15	10	22	4	R1	R2	13	14	19	17
		16				23					
		18				24					

Item Analysis: Subjects from the Final Group were assessed using the Evolutionary Tree Assessment (ETA). We analyzed student responses to each item to determine if the items were effective at measuring the construct. We used item difficulty and item discrimination to evaluate the effectiveness of each item. Item Difficulty was determined by calculating the proportion of students who correctly answered each item correctly. Item Discrimination was evaluated in two ways. First, we calculated discrimination by taking the top scoring 27% of subjects and comparing the number of correct responses with the number of correct responses in the lowest scoring 27% of subjects (Doran, 1980). Next, we calculated a point-biserial correlation for each item to the total score. We used these three values to evaluate the effectiveness of each item and to decide on whether to include them on the ETA. Based on poor item performance with difficulty and discrimination we removed two items (R1 and R2) from the ETA (high difficulty and/or poor discrimination).

Convergent and Discriminant Validity

Convergence is the degree to which scores on two assessments that purport to measure the same construct correlate with one another. Discrimination, in contrast, is the degree to which scores on two assessments, which claim to measure two differing constructs, correlate to one another. We used two assessments developed for a non-majors introductory biology course to provide evidence of convergence and discrimination by comparing the student scores from the Convergent-Discriminant Group. The convergent assessment was designed to assess tree-thinking (TT) and consisted of 11 multiple-choice items. The discriminant assessment was used to assess the nature of science (NOS) and consisted of 12 multiple-choice items. We then used a Pearson and Filon's *z* test to compare the correlations.

Factor Analyses

We hypothesized that the learning outcomes created fit into three distinct categories as outlined in Table 6: accurately interpret information depicted in an evolutionary tree using an understanding of common ancestry (LO 1a-1d), demonstrate an understanding of how characters are inherited from common ancestors by accurately interpreting an evolutionary tree (LO 2a-2e), and demonstrate an understanding of evolution as a continuing and non-teleological process (LO 3a-3b). To test this initial hypothesis, we used student responses from the Final Group to conduct an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). The EFA allowed us to generate a statistically supported hypothesis about the relationships of the items and the underlying structure of evolutionary tree concepts while the CFA was used to test how well the hypothesis fits the data (Matsunaga, 2010). The CFA included generating a modification index to see if any theoretically sound improvements to the model were justified.

We also calculated multiple fit indices that are robust to differing data patterns to better evaluate the fit of the model.

Shortened Version

In our process of evaluating the ETA we recognized its length would be a potential issue that reduces its utility to instructors. Its length allowed us to have multiple items for all but one learning outcome; however, instructors may not wish to have an exam entirely devoted to tree-thinking. To compensate for the length, we selected 10 items that covered most of the learning objectives and had excellent item characteristics, see Table 3. We then ran a Pearson product-moment correlation to examine the relationship of the total score for the 10 items to the ETA as a whole.

Table 7. The learning outcomes associated with the 10 items selected for the shortened version of the ETA.

Learning Objective	**1a**	**1b**	**1c**	**1d**	**2a**	**2b**	**2c**	**2d**	**2e**	**3a**	**3b**
Items	1	15	2	21	4	24	8	7	-	19	12

Reliability

We used responses from the Final Group and Test-Retest Group to gather two forms of evidence of reliability. We measured the internal consistency of student responses by calculating a Cronbach's alpha coefficient for the Final Group. The second method we used was test-retest. Test-retest allows us to estimate the stability of the scores over time. We assessed 120 subjects using the ETA and then six weeks later we assessed them with the ETA a second time. We then calculated a Pearson's product-moment correlation between the two scores for each student. Using these two methods allowed us to produce multiple forms of reliability evidence for the results reported in this study.

Results

Validity

Item Analysis

We calculated item difficulty, item discrimination, and a point-biserial correlation for each item with the total score; results for each item are shown in Table 8. We used widely accepted standards to evaluate the values produced in the item analysis (Doran, 1980). Average item difficulty was 0.61 with a low of 0.91 (Item 9) and a high of 0.22 (Item 5). Average item discrimination was 0.52 with a low of 0.17 (Item 9) and a high of 0.68 (Item 21). The average point-biserial correlation was 0.45 with a low of 0.25 (Item 9) and a high of 0.56 (Item 19). Item 9 and Item 14 also had low discrimination but this could be explained by the low difficulty of the items. While traditional discrimination values were low for these two items, the point-biserial correlations (another common means of evaluating discrimination) for both were in the acceptable range (> .2). We decided to keep these two items as part of the ETA because they were the only items that targeted learning outcome 3e.

Table 8. The difficulty (p), discrimination (D), and point-biserial correlation (r_{pb}) for each item on the final version of the ETA. Boxed cells indicate values of concern.

Item	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
p	.65	.40	.47	.45	**.22**	.69	.56	.69
D	.47	.65	.60	.44	.40	.64	.62	.34
r_{pb}	.39	.50	.44	.32	.38	.53	.47	.29
Item	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
p	**.91**	.69	.71	.87	.64	.89	.50	.55
D	**.17**	.59	.60	.34	.47	**.24**	.71	.66
r_{pb}	.25	.50	.53	.41	.39	.34	.55	.51
Item	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
p	.65	.77	.72	.67	.64	.72	.38	**.22**
D	.59	.59	.64	.54	.68	.57	.55	.41
r_{pb}	.46	.54	.56	.47	.55	.52	.43	.41

Convergent and Discriminant Validity

We used a Pearson product-moment correlation to compare the relationship between scores from the Convergent-Discriminant Group on the TT assessment and the ETA as a measure of convergent validity. We found a large positive correlation between the scores on the two, $r(124) = .616$, $p = .000$.

We also used a Pearson product-momentum correlation to compare the relationship between scores from the Convergent-Discriminant Group on the NOS assessment and the ETA. A smaller positive correlation was found, $r(124) = .362$, $p = .000$.

Correlation coefficients relating scores of two assessments between .4 and .7 are considered to be strongly correlated while values between .2 and .4 are considered to be weakly correlated (Doran, 1980). We compared the two correlations with the ETA to determine if the TT-ETA correlation was significantly different from the NOS-ETA correlation. A Pearson and Filon's z test conducted using the cocor package for R showed that the TT-ETA correlation was significantly larger than the NOS-ETA correlation, $p = .002$ (Diedenhofen & Musch, 2015).

Factor Analyses

We conducted a principal axis factor analysis (PAF) of the 24-item ETA on the responses of 265 randomly sampled subjects from the Final Group. This random sampling allowed for a confirmatory factor analysis on responses of the 266 remaining subjects. The Kaiser-Meyer-Olkin (KMO) value was 0.84 indicating that, overall, the data was likely to be factorable. In addition, all individual item KMO measures were greater than 0.76 indicating that each item was suitable to be included. A Bartlett's Test of Sphericity was statistically significant ($p < .000$), which also indicates the data is likely to be factorable.

We used a scree plot and the total variance explained to help evaluate the number of factors that were appropriate to extract. Visual inspection of the scree plot indicated two potential inflection points at either three or five factors. Total variance explained exceeded 5% in five factors. These five factors cumulatively explained 45.3% of the total variance as opposed to only 35.0% in the three-factor solution. Based on these initial results we decided a five-factor solution was more appropriate than the three-factor solution, which may have matched our initial hypothesis. To provide further evidence we used a PAF parallel analysis on the 265-subject data set with 1000 randomly generated and normally distributed data sets. A parallel analysis uses randomly generated data to determine if factors produced from actual data are larger than would be expected by chance. The results of the parallel analysis support our decision to extract five factors.

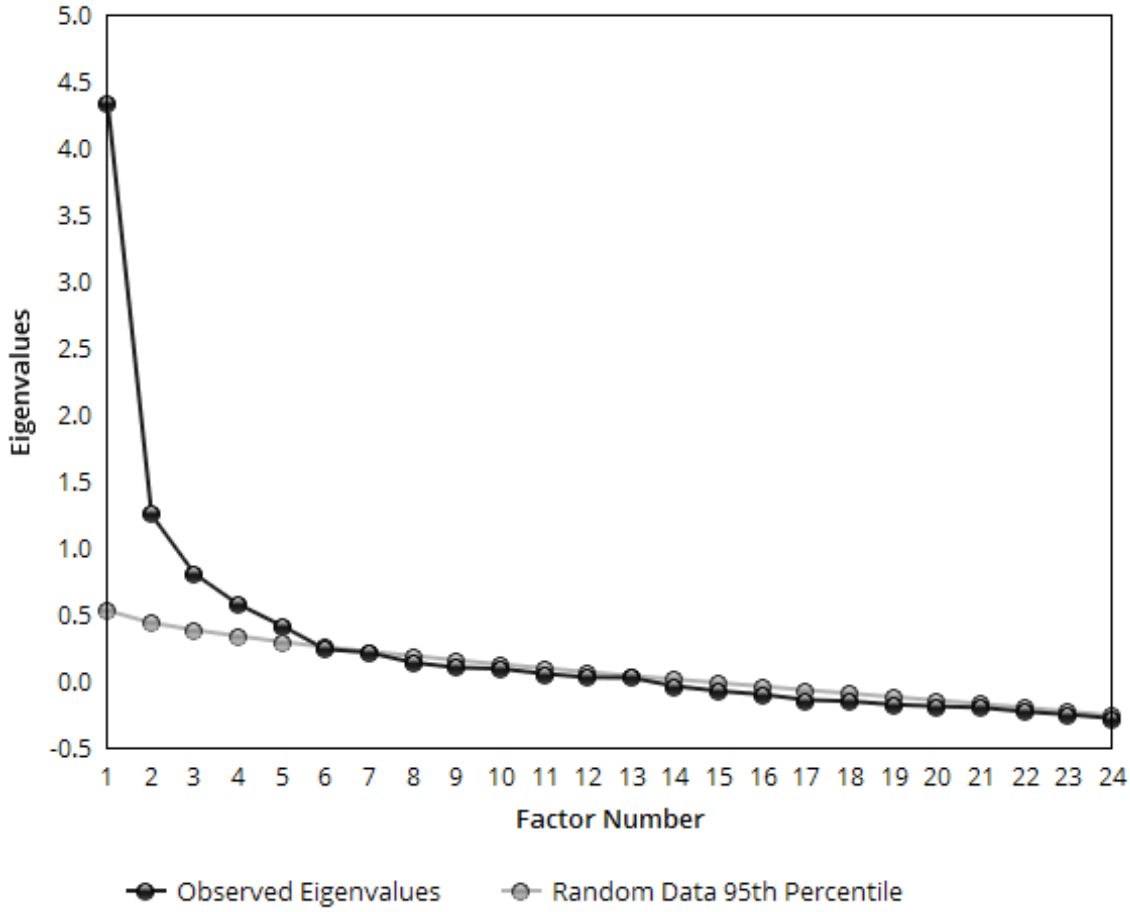


Figure 9. Scree plot of the observed eigenvalues and the 95th percentile of eigenvalues generated by random data.

We used a Promax rotation on the data and the Item loadings on the five factors are shown in Table 9. The Item factor relationships differ in a number of ways from our originally proposed three-factor grouping. Item 1 most strongly loaded on the same factor as items that asked students to compare evolutionary relatedness when it theoretically should have loaded on Factor 5 with Item 20, which measured the same construct (clade type). Item 10 also loaded most strongly on the factor with evolutionary relatedness items rather than with Item 2 on Factor 5. While the factor loading of Items 1 and 10 differed from what we expected, their loading was relatively weak on either factor meaning the evidence that they measure the same factor as the

other items on Factor 1 questionable. In addition to Items 1 and 10 Items 8, 9, 20, and to a lesser extent 4 all had low factor loadings (< .3).

Table 9. Largest factor loadings on the five extracted factors for each item and their corresponding learning outcome. Highlighted rows indicate items that share the same learning outcome loading on separate factors. * Indicates weak factor loadings.

		Factors				
Item	LO	1	2	3	4	5
1	1a	.256*				.021*
20	1a	.053*				.205*
6	1b	.713				
15	1b	.594				
16	1b	.803				
18	1b	.639				
2	1c	.011*				.393
10	1c	.229*				.016*
21	1d				.664	
22	1d				.755	
3	2a					.415
4	2a					.292*
5	2b					.471
23	2b					.356
24	2b					.429
7	2c			.485		
13	2c			.415		
8	2d			.173*		
9	2e			.227*		
14	2e			.552		
11	3a		.917			
19	3a		.758			
12	3b		.382			
17	3b		.382			

We conducted a CFA based on the five factors and item loading patterns seen in the EFA using the lavaan package in R (Rosseel, 2012). The CFA allowed us to compare the fit of the 5-factor model to the second half of the data we excluded from the EFA. After performing the CFA, we also computed a modification index that showed how the model might be altered to improve the fit. Based on our theoretical reasoning we adopted two of these suggestions that allowed for two covariance terms in the model: one between Items 6 and 16 and the other

between Items 11 and 19. We accepted these suggestions because they significantly improved the model and each set of items loaded on the same factor, respectively. We used three fit indices to evaluate the fit of our model. Two indices indicated that our model was an acceptable fit to the data (RMSEA = .034 and SRMR = .043). The third index we used was the incremental fit index (IFI), which had a reported value of .94. This falls just below the conservative threshold of .95 and above the threshold of .90 that some recommend (Hu & Bentler, 1999).

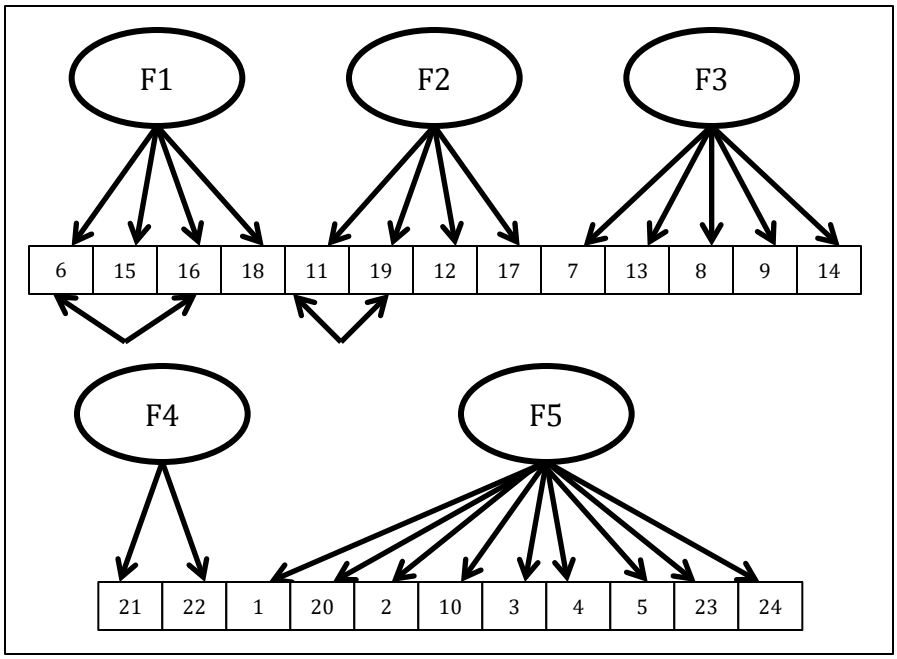


Figure 10. Five-factor model analyzed for fit in the CFA. Lines between factors and items indicate loading. Lines between two items indicate covariance.

Shortened Version

To evaluate the shortened 10-item version of the ETA as a predictor of the full version we used a Pearson product-moment correlation and treated the 10 items as a separate assessment and compared it to the ETA as a whole. We found the scores of these 10 items to be strongly correlated with the scores of the ETA, $r(531) = .918$ $p = .000$

Reliability:

We calculated a Cronbach's alpha coefficient as a measure of internal reliability for the 531subject group who took the ETA. The ETA was shown to have a Cronbach's alpha of 0.845.

We used a Pearson's product-moment correlation to assess the relationship between subject scores on the first test attempt and second attempt of our test-retest group of subjects. We found a large positive correlation between the two scores which is to be expected, $r(120) = .828$, $p = .000$.

Discussion

We developed items to directly address learning objectives that were developed for undergraduate students. Experts reviewed, revised, and approved both the learning objectives and corresponding items. Student interviews, student open answer responses, and literature defining common misconceptions were used to guide the development of each item. We believe the characteristics reported for each item in Table 8 demonstrate
that the process of item development produced appropriate items that measure what we intended them to measure and distinguish between students of differing ability.

The EFA results showed that our three-factor model proposed based on our theoretical grouping of the learning objectives was not justified based on the pattern of student responses. We used the results of the EFA to propose a new five-factor model as seen in Figure 10. This new five-factor solution differed from our proposed three-factor model in a number of ways but most importantly in the grouping of items targeting LO 1a and 1c (clade type and evolutionary tree components) with items targeting LO 2a and 2b (homology/analogy and using parsimony). All other factors consisted of a subset of items that fell in the same theoretical category they were initially placed in (e.g., Factor three consists of items targeting 2c, 2d, and 2e but none from

groups one or three). The results of the CFA show this new model is an acceptable fit to the data. We created a new classification of our learning outcomes based on the results of the factor analyses (Table 10). We believe this new classification reflects sound theoretical groupings and is consistent with the underlying construct structure supported by the factor analyses.

While this new classification and the model used to create it was a good fit to our data we did have a number of items that only weakly loaded on a factor. This means that our model could likely be improved with further research**.**

Table 10. Learning outcomes aligned to the five-factor solution.

Learning Outcomes
1 - Compare evolutionary relationships between taxa
2 - Distinguish between evolutionary trees with differing topologies and evolutionary trees with depicting differing relationships
3 - Use an understanding of the theoretical aspects of evolutionary trees to evaluate group and character evolution based on common ancestry and parsimony
a Identify cases of homology and analogy when interpreting an evolutionary tree b Analyze character information and evolutionary trees using parsimony c Distinguish monophyletic, paraphyletic and polyphyletic groups d Identify what various components of an evolutionary tree represent
4 - Demonstrate an understanding of how characters are inherited from common ancestors by accurately interpreting an evolutionary tree with characters
a Identify synapomorphies for a group on a given evolutionary tree b Identify character states as derived or ancestral on a given evolutionary tree c Use an evolutionary tree to identify characters a given taxon would exhibit
5 - demonstrate an understanding of evolution as a continuing and non-teleological process
a Identify why using simplicity and complexity to categorize organisms as primitive and advanced species is inappropriate from an evolutionary perspective b Demonstrate an understanding that all extant populations continue to evolve and have evolved their entire existence

The significant difference found between the correlation coefficients (Tree-thinking-ETA vs. Nature of Science-ETA) and the differing classification of the correlation coefficients (strong and weak) serve as evidence that the ETA measured the constructs we intended. While the

correlation between the TT-ETA was significantly higher, the correlation between NOS-ETA was still significant. We believe that the higher-order cognitive skills required to answer the majority of items on both assessments can explain this significance. A published assessment focused on evolutionary relatedness found a significant correlation with scientific reasoning (Blacquiere & Hoese, 2016). Scientific reasoning has been found to be highly correlated with performance on assessment items that require higher-order cognitive skill (Lawson, Alkhoury, Benford, Clark, & Falconer, 2000). We believe the significant correlation found between the NOS and the ETA is likely due to both requiring higher-order cognitive skills. Students with higher scientific reasoning ability performed better on both assessments leading to a significant correlation that was not due to similar constructs being assessed.

We calculated a Cronbach's alpha coefficient and conducted a test retest to gather evidence of reliability. Results from our reliability analyses provide solid evidence that student responses to the EFA were reliable. Our estimate of internal reliability produced a Cronbach's alpha coefficient well within the range of values expected for a concept inventory (Kline, 2000). The strong correlation found during the test-retest analysis gives us a second estimate of reliability this time to provide evidence of stability over time. Given the amount of time between the two attempts, our correlation falls well above the acceptable cutoff ($r > .7$) (Kline, 2000). Reliability is a required condition for the validity of results to be claimed. We believe that the results of the analyses performed provide strong evidence for the reliability of responses to the ETA.

One use of evolutionary trees that was uncovered by our concept inventory and learning outcomes was their use in depicting the evolution of genes (Omland et al., 2008). While scientists commonly use gene trees in their research, these types of trees are rarely included in an

introductory study of biology. We believe it would have been beyond the scope of this assessment to include gene trees and related concepts in our learning outcomes and concept inventory.

While we believe we have demonstrated that the ETA is an adequate measure of tree-thinking, we recognize that its focus is on conceptual understanding and does not ask students to necessarily complete tasks that would be more authentic to how practicing scientists use evolutionary trees. We believe we appropriately focused on conceptual understanding given the goals and intended use of the ETA, but it does not represent the entirety of ways in which an instructor may want to assess evolutionary tree concepts.

As we previously mentioned, the length of the ETA is likely to be of concern for those who wish to use it in academic settings. The large correlation found between the shortened version and the full version indicates that the 10 items selected serve as a good predictor of student scores on the full version. The correlation between the shortened version and full version is higher than a similarly shortened version of the Meiosis Concept Inventory to the full version (Kalas, O'Neill, Pollock, & Birol, 2013). The evidence of reliability and validity of student responses to the ETA outlined in this research only apply to the full version. Due to this, we would not recommend using the 10-item version for research but it may be useful to instructors as a pre-assessment, quiz, or as part of a unit assessment.

The ETA has significant potential to help researchers and instructors as a concept inventory. Researchers can use the ETA in multiple ways. First, it can be used to better understand how tree-thinking concepts are related to each other. As we have shown our own theoretical understanding differed from the pattern shown in our results. As we better understand the relationship tree-thinking concepts have to one another we can design instruction

to account for these patterns. The ETA can also be used by researchers to measure student learning of tree-thinking. Doing this would allow researchers to make better comparisons between their own research and the research of others. Instructors, of course, can also use the ETA as an assessment to determine how effective their tree-thinking related instruction has been in teaching tree-thinking concepts.

THE PERILS OF BUILDING TREE-THINKING

Introduction

Evolutionary trees are crucial tools used in nearly every biological field ranging from molecular biology to medicine to biogeography (Wiley, 2010). Because evolutionary trees are so broadly used all biologists regardless of field need to be able to read and interpret trees and understand the basic concepts that underlie them. Undergraduate education in the life sciences must include a significant amount of instruction on evolutionary trees to produce biologists who are capable of understanding or using this tool in their future.

In 2005, a call was made to spread the teaching of evolutionary tree concepts or tree-thinking and further our understanding of how students learn about and understand evolutionary trees (Baum et al.). The number of researchers studying student learning and understanding of evolutionary trees and associated publications since this call has grown significantly. Important insights have been gained as to what inhibits learning and to the common misconceptions students hold in relation to evolutionary trees (Baum & Smith, 2013; Dees et al., 2014; Gregory, 2008; Halverson, 2011; Halverson et al., 2011). In addition to this important conceptual work, educators have sought to develop lessons that incorporate best practices in teaching evolutionary tree concepts.

How best to teach evolutionary trees is an important question. Two primary methods of teaching tree-thinking to students are found in the published literature. The first and most commonly used method, referred to as 'tree building', is to teach students the concepts and principles that underlie how an evolutionary tree is built (e.g., Davenport et al., 2015a; Kumala, 2010; Singer, Hagen, & Sheehy, 2001; Goldsmith, 2003). Students are generally asked to use morphological characters as the evidence to produce an evolutionary tree. It is thought that if

students can understand the evidence used to make a tree and put that evidence to use in building a tree they will be able to understand and use the information portrayed by evolutionary trees appropriately.

An alternative method to a tree building approach is to instead focus on how to read and interpret evolutionary trees, i.e., 'tree analysis' (e.g., Davenport et al., 2015b; Offner, 2016). Instead of focusing on how and why one tree is favored over another students are expected to be able to analyze information shown in a pre-built tree such as the evolutionary relationships depicted or the evolution of the characters. Those who support this method argue that it is the skill of interpreting the tree that is most important and should be the focus of instructional time. It has even been argued that knowing how systematists infer an evolutionary tree is unnecessary for knowing how to interpret the information the trees convey (Baum et al., 2005).

Halverson postulated, based on a study of how students understand evolutionary trees, that student learning would be benefitted more by analyzing and interpreting evolutionary trees than by building evolutionary trees (2011). This recommendation was based on findings that students were able to interpret evolutionary trees before they were able to accurately build evolutionary trees from a data set. The recommendation and findings indicate that interpreting evolutionary trees and building evolutionary trees are separate but related constructs rather than a single interconnected construct (a latent characteristic of an individual such as a mental ability or psychological trait) (Miller et al., 2013). If the goal of instruction is to have students read and interpret evolutionary trees and they are separate constructs it would not be necessary to teach both unless building an evolutionary tree was also a goal of instruction.

Teaching both tree analysis and tree building may actually introduce misconceptions and unnecessary complexity because students inappropriately apply how they build an evolutionary

tree to what information is conveyed by the tree. When building an evolutionary tree one seeks to build it based on having the fewest number of changes required to produce the pattern of characteristics present in the extant taxa (Farris, 1970). This is done primarily by finding clade-defining characters that were derived after a given clade separated from its sister taxa but before any of the taxa in the clade diverged. These special characters, or synapomorphies, reduce the number of evolutionary changes required to produce the pattern of biodiversity that exists. Different evolutionary trees are evaluated based on the number of changes each would require and the one that requires the fewest number of changes is selected as the most likely to have occurred. The logic behind this is known as parsimony. Introducing students to evolutionary trees via tree building requires students be taught how to infer the tree from a dataset of characters.

Unfortunately, many students view these synapomorphies as the cause of divergence between two taxa or clades. In other words, the node in the tree is when the synapomorphy evolved causing the population/lineage to split into two. So instead of correctly identifying the node as the most recent common ancestor the node is viewed as a character even though it is more likely that the character evolved well before the most recent common ancestor. Students mistakenly conclude that the number of nodes (which they are interpreting as shared characters) determines the degree of evolutionary relatedness, a misconception we will refer to as Nodes Equal Character Change.

Eddy et. al. (2013) conducted a study that put these ideas to the test. Different sections of the same course were taught using each approach (tree building vs. tree analysis). Their results showed that students who were taught using tree building instead of tree analysis performed better on a common tree-thinking assessment. This is counter to what should have

occurred if Halverson's idea was correct. It also indicates that tree building may not be a separate construct but actually the same construct as tree analysis because when tree building understanding increased, students showed a complementary increase in tree analysis understanding. If they are a single construct, as this indicates, teaching both potentially will increase student understanding by complimenting each other and addressing weaknesses that might be inherent to each individually.

While the study described previously is very well conducted and does address the hypothesis appropriately, it does not address how the two approaches might work together. Perhaps an instructor does not have to choose an either-or approach but could use both strategies to help students understand evolutionary trees and confront the potential misconceptions that each strategy might introduce. To test this idea, we taught tree-thinking to lab sections of introductory biology for majors using each strategy individually and together and then used a common assessment to measure the learning for each group. We predicted that students taught using both strategies would perform better on a summative assessment than those taught only using one strategy because the strategies would work to complement one another.

Methods

Subjects

Subjects were students in an introductory biology course for majors at a large private university in the United States. Subjects participated as part of the normal course work but had the option of removing their data from the study, to satisfy IRB requirements. Over the course of two academic years we had 474 participants from 28 sections. Sections were randomly assigned to a treatment as shown in Table 11.

Table 11. The four treatment groups and their corresponding instructional strategy and number of subjects.

Treatment	Instructional Strategy	Subjects (n)
SKULL	Tree building	135
GCR	Tree building	115
CAR	Tree analysis	109
GCR+CAR	Tree building and Tree analysis	115

Treatments

Three lessons were used in this experiment: The Great Clade Race (GCR), a Skull Analysis Activity (SKULL), and an activity that utilizes car manufacturers as an analogy (CAR). Two used tree building methods (GCR and SKULL) and one focused on tree analysis along with what an evolutionary tree theoretically represents (CAR). The SKULL activity is a variation of a commonly used method that requires students to build a dataset and then use parsimony to build an evolutionary tree. This lesson asks students to examine skulls of various animals, build a dataset using a required number of characters from the skulls, and then build an evolutionary tree using the dataset as evidence based on parsimony. This particular lesson is the method that had been used in the course since its inception.

The second lesson used is also a tree building activity called The Great Clade Race (GCR) (Goldsmith, 2003). This activity is commonly used to introduce students to evolutionary trees at both the college and secondary levels (Perry et al. 2008). It has also been used into compare the effectiveness of evolutionary tree lessons. This lesson requires students to recreate a map of a hypothetical race. In the race, runners all started on the same path and the path splits repeatedly. The runners could have taken either path. In addition, the race also had checkpoints scattered throughout. When a runner happened upon a checkpoint the runner received a stamp

on a card that was unique to that checkpoint. Students recreate the paths the runners took by comparing the stamps the runners received.

This serves as an analogy for what occurred in the evolution of organisms and how we can use current characters to infer how organisms diverged from one another in the past. Students are then given a set of animals and their characteristics and asked to fill out a matrix. After filling out the matrix with the animals and their characters students are then asked to create an evolutionary tree from the matrix.

The final lesson was one created specifically for this study to serve as the tree analysis lesson (CAR). The primary learning objective for this lesson is to introduce students to what an evolutionary tree is representing in theory and how organizing and analyzing organisms and their evolution based on their history creates a meaningful classification. Students start this lesson by classifying a set of twelve cars (four trucks, four minivans, four sedans) based on any characteristics they would like. After a discussion about meaningful vs. arbitrary classifications students then read a short history of automobile manufacturing during which they learn about the relationships between four automobile manufacturers. Students are then asked to reclassify the vehicles in a way that reflects their history (manufacturer) and to make a diagram that depicts the history and classification. Finally, students are asked to make an analogy between the activity and the evolution of organisms.

We had four treatment groups in this study. Three treatments consisted of using each lesson individually GCR, CAR, and SKULL. The fourth treatment group consisted of using both GCR and CAR together. GCR+CAR started with the CAR lesson that was immediately followed by the GCR lesson. Research Assistants led the activities and discussions for these lab sessions in place of the regular Teaching Assistants.

Assessments

Measure of Group Equivalence

Students were given two assessments as part of this study. The first served as a measure of group equivalence. We used the Biology Concept Inventory (BCI; Garvin-Doxas & Klymkowsky, 2008) to determine the general biology knowledge of students before any treatment.

Tree-thinking Assessment

The second assessment was administered one week after the treatment for each section. It served as a summative assessment to measure the content knowledge of students after the treatment. This assessment consisted of ten pairs of questions. Each pair had a multiple-choice question focused on the content and a follow-up, open-answer question that asked students to explain the reasoning they used in answering the first question of the pair. We chose this format because it provides confidence that the student did in fact understand the question and did not just select the correct answer by chance or for the wrong reason. In order to receive a point for the pair students had to answer both the multiple-choice question and the reasoning question correctly. The content questions were primarily taken from an assessment published by Baum in 2005 though two questions were developed internally. Graders were given example explanations to evaluate student responses for correctness. If graders had a questions relating to the correctness of a response they discussed it with other graders until an agreement on how to categorize the response was reached. Scores on the summative assessment were compared between each treatment group.

As we evaluated written responses a pattern on the items asking subjects to evaluate the evolutionary relatedness of organisms, a pattern emerged. It became apparent that many students

were appealing to the Node Equals Character Change misconception when explaining their answer. We wanted to see if the treatment type had an impact on the number of students demonstrating the Node Equals Character Change misconception. We predicted that the tree building treatments would result in a higher proportion of students demonstrating this particular misconception based on the thought process described in the introduction. Subject responses were categorized as demonstrating this misconception if, when comparing evolutionary relatedness, they claimed that the relationship was based on the number of similar characteristics the two organisms had in common. We then compared the proportion of subjects who demonstrated this misconception between each treatment group.

We also compared the responses from this study to responses from a pre-assessment (using the same assessment and scored in the same way) given to a different set of students in the same course. This was done to allow insight as to whether the misconception was being introduced by the treatments or was present and left unaddressed by the treatments.

Table 12. Representative responses of subjects who were classified as demonstrating the Node Equals Character Change Misconceptions.

Sample Quotes From Item Pair 11/12
"The horse is only 2 traits off from the seal while the whale is 4 traits off."
"A whale develops more traits along the phylogenetic tree so the horse is more closely related to the seal."
"A horse has all of the seal's traits plus one extra. The whale has way more extra traits than the horse."
"Since the whale has evolved so many more traits, the seal is more closely related to the horse than the whale."
"The horse differs from the seal by two trait junctions whereas the whale is separated from the horse by 4. The horse and seal have more common traits."
"The seal and horse have less differences of traits than a seal and whale according to this cladogram. Which doesn't really make sense."

Statistical Analysis

All analyses were conducted using SPSS V.24. Data from both assessments were examined to determine if they met the assumptions of normality and equal variance. Data from the BCI was found to be normally distributed and had equal variance between the four treatment groups. A one-way ANOVA was used to look for a significant difference between the groups. Data from the summative assessment was not normally distributed but it was found to have an equal variance. A log transformation was used to normalize the data. A one-way ANOVA followed by a post-hoc Tukey-Kramer test was used to compare the summative assessment from the four treatment groups.

To compare the proportion of students in each group who demonstrated the Similarity Equals Relatedness misconception on the questions described previously we used a Chi-square test of homogeneity. We followed this post hoc analysis to compare each treatment with a Bonferroni correction for multiple comparisons.

We used Cronbach's alpha to gather evidence of reliability for the summative assessment. Scores for all groups were combined for this analysis since it is looking for a characteristic of the assessment itself. The internal consistency value reported from the Cronbach's alpha was 0.711, which falls in the acceptable range for an assessment of this type (Kline, 2000).

Results

Measure of Group Equivalence

The results from the one-way ANOVA comparing BCI scores indicated no significant difference between the treatment groups, $F(3,470) = .182$, $p = .909$. This test suggests that the

treatment groups had no significant difference in their biology content knowledge prior to the treatment. As a result, BCI score was not used as a covariate in other analyses.

Tree-thinking Assessment

The results from a one-way ANOVA comparing the log transformed summative assessment scores (Tree Score) indicated that there were significant differences between the treatment groups $F(3, 470) = 22.53$, $p = .000$. A Tukey-Kramer post-hoc analysis was conducted to compare the means between each pair of treatment groups, statistics are shown in Table 13. The analysis revealed that all treatment groups scored significantly higher than the SKULL group. Both the CAR group and the GCR+CAR group were found to have performed significantly higher when compared to the GCR group. No significant difference between the means was found between CAR group and the GCR+CAR group.

Table 13. Results from one-way ANOVA of log transformed scores on the tree assessment.

Comparison	Mean Difference (95% CI)	*p*-value
GCR-SKULL	0.116 ± 0.069	.000
CAR-SKULL	0.188 ± 0.07	.000
GCR+CAR-SKULL	0.189 ± 0.069	.000
GCR-CAR	0.072 ± 0.072	.050
GCR-GCR+CAR	0.073 ± 0.072	.045
CAR-GCR+CAR	0.001 ± 0.073	1.00

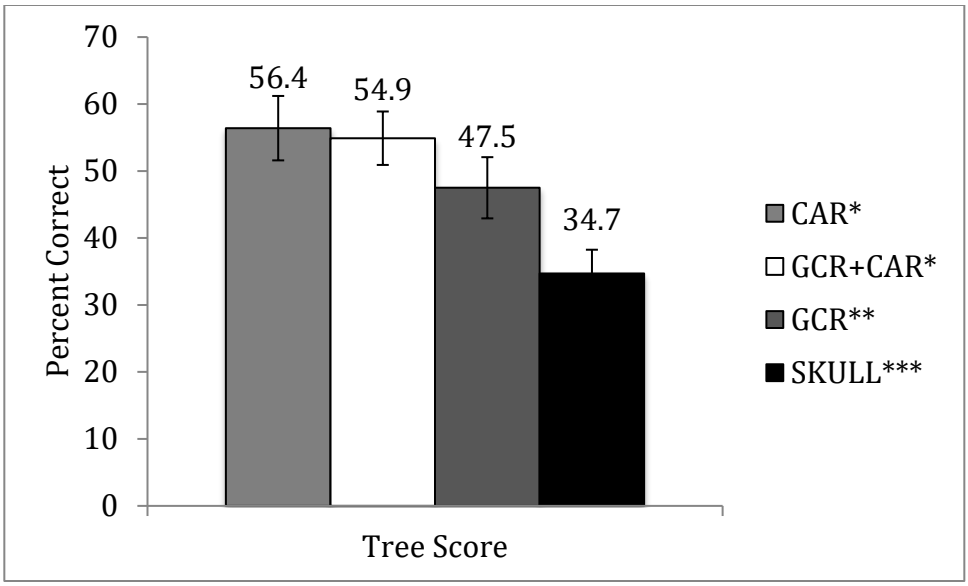


Figure 11. Mean percentage and 95% confidence intervals for the Tree Score of each treatment group. Each * denotes a subset of group categories that do not significantly differ from each other at the ≤ .05 level.

Prevalence of Node Equals Character Change

We examined student responses categorizing them as Node Equals Character Change being present or absent as described in the methods section. We ran a Chi-square test of homogeneity to compare the proportion of students who demonstrated this misconception in each group. The test showed statistically significant differences in proportions, $p = .000$. A post hoc analysis was used to make pairwise comparisons with a z-test and a Bonferroni correction. The proportion of students demonstrating the misconception in the CAR group was significantly lower than every other group except the pre-assessment group, $p < .05$. The proportion of student's demonstrating the Node Equals Character Change misconception in the GCR+CAR was also significantly different from every other group except for the pre-assessment group, $p < .05$. The proportion of students demonstrating this misconception was not significantly different between the GCR and SKULL groups.

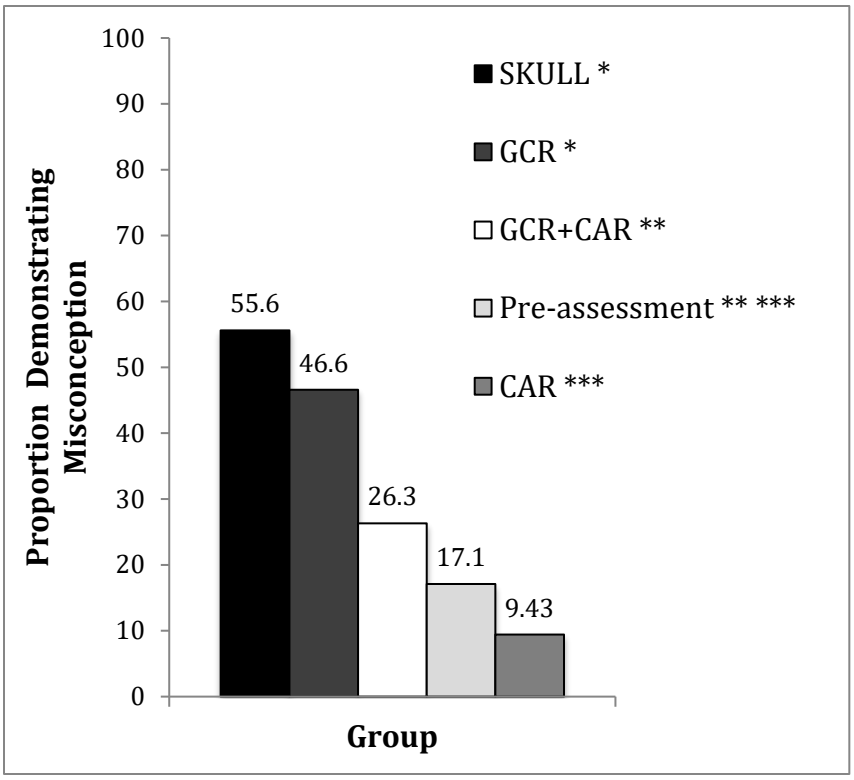


Figure 12. Proportion of subjects in each treatment group that gave a response indicating the Node Equals Character Change misconception. Each * denotes a subset of group categories that do not significantly differ from each other at ≤ .05 level.

Discussion

The goal of this study was to compare the impact of using both tree building and tree analysis as instructional strategies for teaching evolutionary tree concepts. We predicted that using both strategies would result in improved student learning compared to using either strategy individually. We instead saw that students in the CAR group (a tree analysis activity) performed equally to those in the GCR+CAR group (GCR being a tree building activity). This result suggests that teaching tree building does not improve student understanding of tree-thinking if tree analysis has already been taught.

Students in these two groups also significantly outperformed both 'tree building only' treatments we used in this study. This is perhaps the most surprising result of the study. Eddy et

al. (2013) found that students taught using the tree building strategy outperformed students taught using the tree analysis strategy. Our results run counter to this finding and instead support the claim that teaching tree analysis results in better student understanding because it does not introduce the unnecessary concepts related to tree building (Halverson, 2011).

To further understand this surprising result we analyzed student responses. As described earlier, we believe tree building strategies either introduce or do not address a misconception that nodes in an evolutionary tree represent or are caused by character changes. Many students seem to view the tree as a flow chart representing a dichotomous key. This prevents students from being able to accurately interpret the relationships depicted.

The results show that students taught using a tree building strategy alone demonstrated significantly higher levels of this misconception when compared to the students taught using tree analysis alone and students taught using both strategies. When we compared the teaching tree analysis alone to teaching both strategies we also found that including tree building led to a significantly higher proportion of students indicating this misconception in their responses (26.3% to 9.43%). This indicates that tree building strategies are likely introducing this misconception. If this misconception were held prior to instruction, we would have expected to see the CAR treatment mitigate this misconception equally in both treatments (CAR and GCR+CAR). Instead, it appears that the CAR treatment helped students more when taught alone than when the tree building activity was taught in conjunction. This conclusion is also supported by the comparison of student responses from this study to responses from the pre-assessment group. If the misconception were present before the treatments, we would have expected the proportion of responses demonstrating the misconception on the pre-assessment to be similar to

the tree building groups and not similar to the CAR and GCR+CAR groups as was seen in the results.

The quotes in Table 12 are representative of how most students used similarity to interpret the evolutionary trees. These quotes provide a very interesting look at how students who hold this misconception interpret an evolutionary tree when comparing the relatedness of a seal to a horse and a whale. The correct answer according to the tree is that a seal, horse, and whale are equally related but students overwhelmingly said that the seal was more closely related to the horse. It has even been shown that this prior knowledge can even overcome a correct understanding of interpreting evolutionary trees (Catley et al., 2010). In our study students were not relying on prior knowledge of similarity but instead they were inappropriately inferring the tree as showing the number of shared characters. This leads them to conclude against their prior knowledge that a seal and horse have more similar characters than do a seal and whale. We believe this further illustrates that tree building is introducing the misconception rather than it being a prior conception held by the students before receiving the treatments.

It is possible that differences between our results and those found by Eddy et al. (2013) can be explained by differences in the assessments. Three of the 10 pairs of questions we had on the summative assessment for our study required students to accurately interpret evolutionary relationships given a tree. The assessment used by Eddy et al only had a single question requiring students to do this. In addition, we required that student not only answer the question correctly but that they do so using the appropriate reason in their written response. This was not the case in the other study; students who selected the correct answer for the wrong reasoning were given credit for correct understanding. These two aspects likely explain the difference in

our findings. Other explanations may include differences in the overall lessons used in each treatment and placement of the lessons within the course structure. Further study is warranted.

It might be suggested that knowing how evolutionary trees are built is a necessary aspect of learning evolutionary tree concepts because it is what experts do in their practice and if students are to gain expert ways of thinking they need to think about the same aspects that experts do. We would argue however, that it is actually somewhat rare for experts to focus on tree reconstruction methods in their practice and even more rare for them to use parsimony in that process.

In practice, most evolutionary biologists focus on tree reconstruction as a tool to analyze an evolutionary question relating to genes, populations, and species (Yang & Rannala, 2012). These experts do not actually score and compare trees nor do they actually create the programs that do. While most should and do understand the fundamentals behind these tree reconstruction programs it is not the focus of their research. In fact, it is no longer the norm to even use the fundamental principle taught in most tree building lessons, parsimony, in tree reconstruction. Maximum Likelihood and Bayesian methods are more widely used and use a different rationale in evaluating which evolutionary tree is most consistent with the evidence (Kolaczkowski & Thornton, 2004). While parsimony may be an appropriate starting point for learning about these methods it is not what most experts use or focus on when conducting their research. To argue that tree building is necessary for students to learn based on that claim seems to ignore what most researchers actually do in this field.

Teaching tree building instead appears to be inhibitory to learning. We offer a potential explanation for this phenomenon utilizing Cognitive Load Theory (Chandler & Sweller, 1991). If it is true that the actual process of tree building is a unique construct to analyzing and

interpreting the information conveyed in evolutionary trees, then teaching tree building concepts is extraneous and unnecessary to learning how to analyze evolutionary trees. The concepts unique to tree building become extraneous cognitive load and, while it is seemingly useful to understand these concepts in addition to understanding how to analyze evolutionary trees, adding them to the learning process appears to actually inhibit learning. This of course is not a problem unique to learning evolutionary tree concepts; it has been demonstrated that adding nonessential material inhibits learning by adding extraneous cognitive load in many subjects (Sweller, Merrienboer, & Paas, 1998). Further research is needed using assessment and factor analysis to determine if these constructs are indeed unique.

The instructional strategy by which we teach evolutionary trees is an important decision we face as educators. Our results support the hypothesis that tree analysis alone supports student learning of evolutionary trees by removing extraneous concepts. Based on this and the fact that tree building appears to introduce misconceptions as to what the evolutionary tree represents, we recommend that tree analysis, not tree building, be used as the instructional strategy to introduce students to concepts relating to evolutionary trees. Our results also show that using both strategies together does not significantly differ from using tree analysis alone. Given the increased amount of time and potential introduction of misconception we would not recommend using both strategies as an introduction unless the course learning objectives require that students be able to build an evolutionary tree. If tree building is part of the course learning objectives, special attention should be given to addressing the relationship between characters and what an evolutionary tree represents. Recognizing that our study stands in direct contrast to a previous study we believe further research on this important question is needed.

EVOLUTIONARY TREES AND LEARNING PLANT DIVERSITY

Introduction

Organismal diversity is a pillar in biological education. Knowing and understanding the natural history of organisms is fundamental to an understanding of biology as a whole. Organisms are the embodiment of (genes, behavior, etc.) or are the basic units of (populations, ecosystems, etc.) every field of biological study (Greene, 2005). Studying the diversity of organisms sparks many of the questions scientists seek to answer and provides the means of testing hypotheses as we seek to understand the biological world around us (Schwenk, Padilla, Bakken, & Full, 2009). Understandably, this importance is reflected in the number of courses offered on organismal diversity at universities and their requirement in many programs of study.

Organismal diversity is often taught as a walk through the phyla moving from one taxonomic group to the next, usually beginning with taxa that exhibit dominant ancestral features (e.g., crocodiles) and finishing with taxa that exhibit dominant derived features (e.g., birds). This pattern is not in and of itself flawed but it does have weaknesses among which is the lack of connectivity and lack of evolutionary context. Due to these weaknesses, it is common for instructors to utilize evolutionary trees as a component of their instruction. The ways in which instructors use evolutionary trees in their instruction varies widely from using them primarily as a means of introducing the next taxa of study to fully integrating them in the course (assessments, labs, and lectures) (Smith & Cheruvelil, 2009; Staub, Pauw, & Pauw, 2006; White, 2009).

Evolutionary trees are an obvious and natural tool to aide in the learning of organismal diversity because trees provide a connective and explanatory thread between the taxa. Using evolutionary trees in this way allows students to understand not just the characters taxa hold but

the evolution that occurred to produce the patterns they see. Students can be asked to not only memorize but to make predictions and hypotheses about character evolution or the placement of a new taxon based on their knowledge of the evolutionary relationships of the organisms. In doing this students are learning the material at a higher cognitive level and developing important skills related to scientific reasoning (Krathwohl, 2002; Julius & Schoenfuss, 2005). Learning diversity in the context of evolution should also improve student understanding of evolution itself. For example, the relationship between homology, analogy, common ancestry, and convergent evolution is confronted time and time again as students study the patterns present in the taxa of study. This allows students to understand and think about organismal diversity in ways that experts think about it, which is in alignment with goals put forward by Vision and Change (AAAS, 2011). Rather than a course solely teaching organismal diversity by marching through the phyla, a course that fully integrates evolutionary trees becomes a course that teaches organismal diversity, scientific reasoning, *and* evolution.

Evolutionary trees also have the potential to reduce cognitive load when learning organismal diversity by grouping the diversity in meaningful ways and reducing the total amount of information to memorize. Baum and Smith (2013) claimed that using evolutionary trees to organize information is a far more efficient way to organize character information than any other method. This efficiency reduces the raw amount of information students have to learn by making the students learn only a single evolutionary point on the tree for a character change rather than individually learning the character state for each taxon (Baum & Offner, 2008). Having less to learn reduces the overall cognitive load required to learn the material.

We illustrate this point by showing the information conveyed in a recreated summary table from the end of a chapter on Gymnosperm diversity in a plant diversity textbook (Table

14). The table organizes the taxa and key characters in a single location allowing students to compare and contrast the groups (Raven, Evert, & Eichhorn, 2005). This can be compared to an evolutionary tree conveying the exact same information but in the context of the evolution of Gymnosperm diversity (Figure 13). To illustrate how the evolutionary tree reduces cognitive load we can look at the column representing the pollen tube. In the table, we have four taxa with a 'yes' or 'no' answer for each. In contrast, the evolutionary tree conveys the same information in a single mark on the tree, indicating that it evolved prior to the most recent common ancestor of the Coniferophyta and Gnetophyta clades.

Table 14. Recreated summary table from the end of a chapter on Gymnosperm diversity.

Phylum	Representative Genus or Genera	Type of Tracheary Element(s)	Produce Motile Sperm?	Pollen Tube a True Sperm Conveyor?	Type of leaves produced	Miscellaneous Features
Cycadophyta	*Cycas and Zamia*	Tracheids	Yes	No	Palmlike	Ovulate and microsporangiate cones simple on separate plants
Ginkgophyta	*Ginkgo*	Tracheids	Yes	No	Fan-shaped	Ovulate and microsporangiate cones on separate plants; fleshy-coated seeds
Coniferophyta	*Abies, Picea, Pinus, and Tsuga*	Tracheids	No	Yes	Most needlelike or scalelike	Ovulate and microsporangiate cones on same plants; ovulate cones compound; pine needles in fascicles
Gnetophyta	*Ephedra, Gnetum, and Welwitschia*	Tracheids and vessel elements	No	Yes	*Ephedra*: small scalelike leaves; *Gnetum*: relatively broad, leathery leaves; *Welwitschia*: two enormous, strap-shaped leaves	Ovulate and microsporangiate cones compound' borne on separate plants, have conifer and angiosperm-like features, leaves borne in opposite pairs

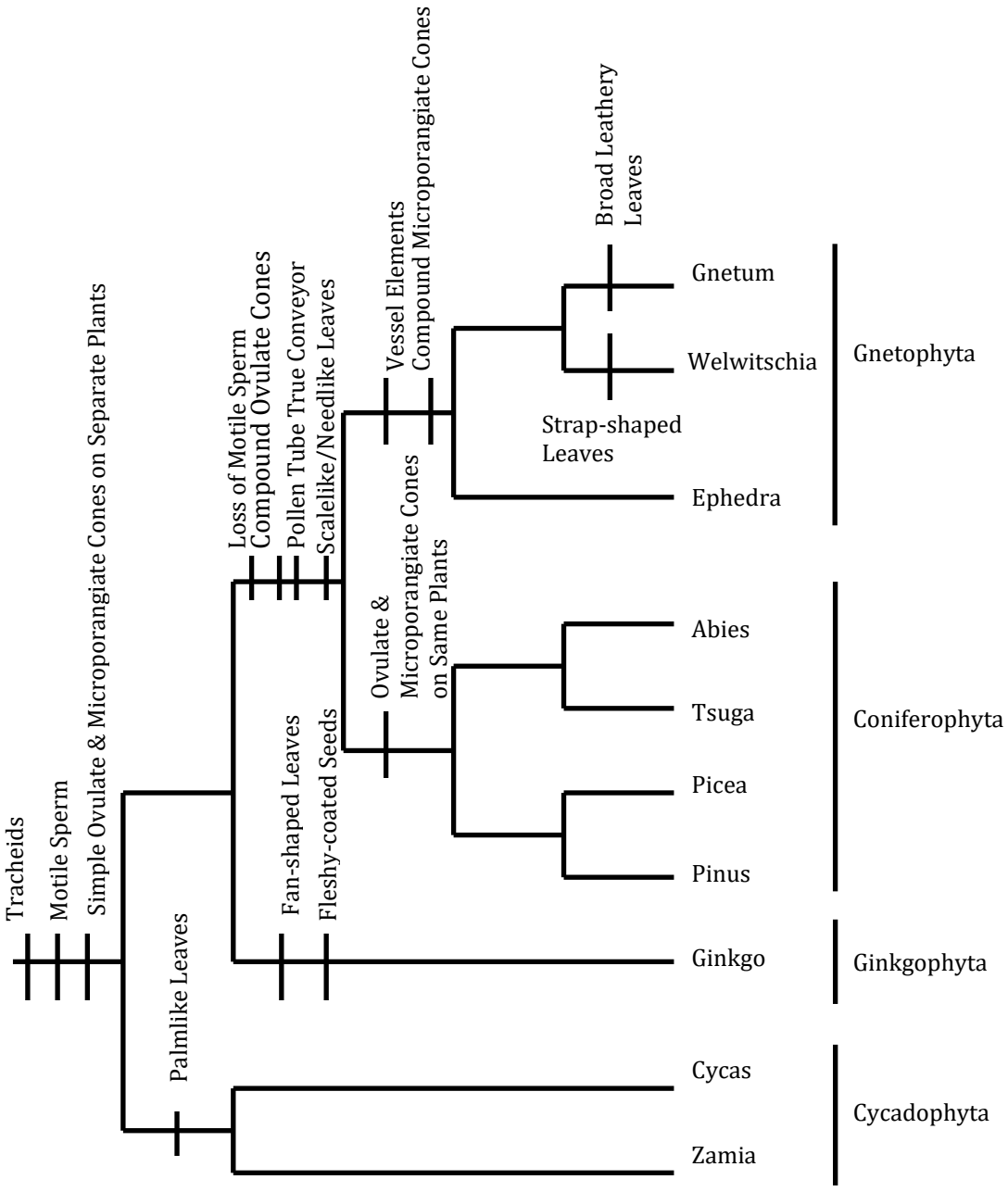


Figure 13. Evolutionary tree depicting how a student might map the same character information from Table 14.

While the use of evolutionary trees to teach organismal diversity is common, the effectiveness of using them in this way has had little study. What study has been done has focused on student learning of evolutionary tree concepts rather than the learning of organismal

diversity content (Smith et al., 2013). We have addressed two goals in the current research using three independent experiments. First, we tested the hypothesis that evolutionary trees exhibit less cognitive load than character matrices and predicted that this efficiency of using evolutionary trees to store information about organismal diversity would translate to better learning of organismal diversity. Second, we sought to implement this hypothesis into a real-world learning scenario to test whether the effect is transferable to the actual classroom during a one-week diversity unit. In the third experiment, we sought to test our hypothesis on a whole course scale by testing it in a full semester diversity course.

Experiment 1: Experimental Comparison

Methods

This experiment was conducted to determine if studying organismal diversity via an evolutionary tree reduces cognitive load over memorizing a list of characters in a strictly experimental setting with students randomly assigned to each treatment and a controlled amount of study time. If cognitive load is reduced by studying organismal diversity via evolutionary trees rather than a list of characters, we would predict that students in the treatment using evolutionary trees would score significantly higher on an assessment testing their knowledge of which organisms have each character. This experiment allows us to see if there is a difference in the initial acquisition of the content by the learner by removing factors that might explain differences in authentic learning situations.

Subjects

Students were recruited from three sections of an introductory biology course for non-majors. 146 students participated in the first stage of the experiment with 120 students

completing both stages. Student consent was obtained for use of their data as part of the study but participation was part of the course. Because the study took place over 2 class periods those who did not participate the first day were not included in the study. Students who participated the first day but did not participate the second day were included in the analysis of the first stage of the study but not the second stage. The Institutional Review Board at the host institution reviewed and approved this study.

Experimental Design

Two treatments were used in this study. In both treatment groups, students were asked to take 20 minutes to study 12 characters possessed by 7 taxa. To prevent familiarity with characters from confounding the study the characters attributed to the taxa were abstract and represented only by numbers. In the first treatment group, referred to as the 'Tree group', students were given a phylogenetic tree depicting the evolutionary relationships of the 7 taxa. The tree also had the 12 characters mapped on it indicating the hypothesized timing of character evolution. In the second treatment, referred to as the 'List group', students were given a table that listed the characters of each taxon. This allowed the students to learn the same information but removed the connectivity and grouping that the evolutionary tree depicted, see appendix for materials used.

All students had completed a unit on what an evolutionary tree represents and how they are interpreted. This allowed us to be confident that no matter which treatment a student was assigned to they were capable of accurately learning the information conveyed to them.

After the 20-minute study period, all students were assessed on what they learned using a 12 question multiple-choice assessment. The assessment had one question for each character the students were to learn. The question would ask the students to identify each taxon that possessed

the character in question. This meant that in many cases students had to select multiple taxa. To get credit for a correct response students needed to select ALL taxa that possessed that character. If they missed a taxon or included one that should not have been included the answer was marked as incorrect.

At the beginning of the next class period (two days later) students were given the same assessment described previously. Students were not informed of this second attempt on the assessment beforehand. We did not inform them of this so they would not spend additional time studying for the second attempt. This allowed us to compare retention of the information learned over a longer period of time rather than just immediately after study.

Statistical Analysis

Analyses were conducted using SPSS V.24. We examined the student scores on the character assessment for both attempts and found that the scores were not normally distributed and that they had unequal variance. Given the characteristics of the data the Mann-Whitney U Test was selected to compare student scores between the groups on both assessment attempts.

Results

We calculated the mean score and 95% confidence interval for the 20-minute administration of the assessment. We compared the two groups' 20-Minute assessment scores with a Mann-Whitney U Test which showed that the assessment scores of the Tree group (Mdn = 100.0) were significantly higher than the List group (Mdn = 83.3), U = 1674.5, z = -3.894, p = .000

The mean scores for the 2-day post administration of the assessment were calculated along with 95% confidence intervals for the means. We used a Mann-Whitney U test to compare

the assessment scores for the Tree group and List group. The assessment scores were significantly higher for the Tree group (Mdn = 83.3) than for the List group (Mdn = 41.7), U = 1096.5, z = -3.616, p = .000.

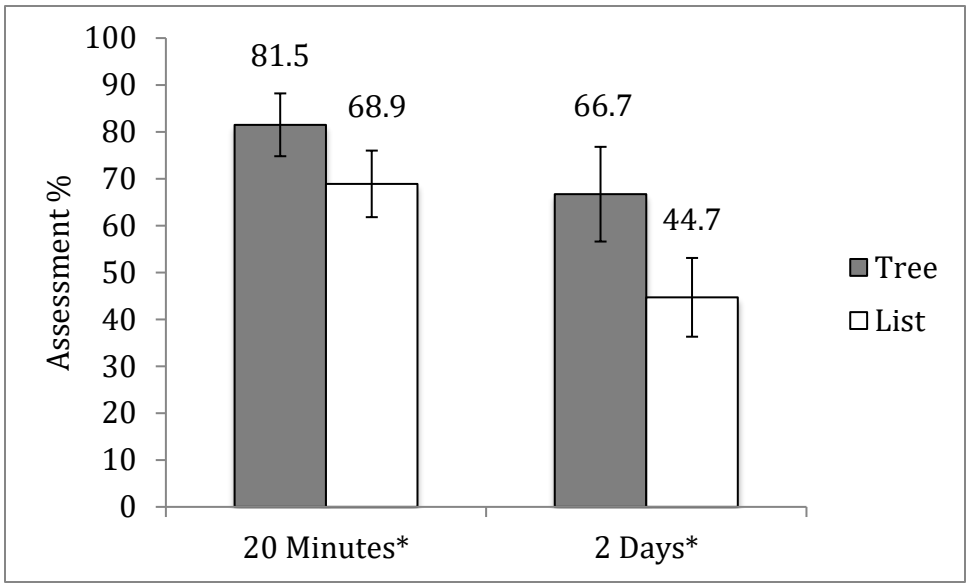


Figure 14. Means and 95% confidence intervals for the scores of each treatment group on the assessment. * Indicates p-value ≤ .01.

Experiment 2: Unit Comparison

For the second experiment, we test the effect of using evolutionary trees on student learning in an authentic learning environment. Students were randomly assigned to one of two treatment groups for the duration of a one-week unit of instruction. Each treatment group was taught a unit of organismal diversity with the same learning objectives and concept attainment lessons (part one of the lesson). The treatments differed in the application lessons (part two of the lesson) with one treatment using evolutionary trees as an integral part of the lessons (Tree group) and the other not utilizing trees but rather just discussing the taxa independently (Traditional group). At the end of the unit we gave an assessment to compare differences in the organismal diversity content knowledge of the students. If studying organismal diversity via

evolutionary trees reduces cognitive load we would expect students in the Tree group to score higher on the assessment than those in the Traditional group.

Methods

Subjects

The students for this study were recruited from an introductory biology course for non-majors. Student consent was required for the use of data but participation in the learning activities was required as part of the course. 442 students participated with 221 students randomly assigned to each treatment group. Only students who completed the assessment were included in the study limiting it to 219 students in the Tree group and 217 students in the Traditional group. The Institutional Review Board at the host institution reviewed and approved this study.

Experimental Design

The courses used in this study were using a flipped classroom model of instruction throughout the entire course. Students were acquiring their content knowledge before class through interactive online instruction (part one) and then completing activities that required students to think about this information on a deeper level and apply it in different scenarios during class (part two). This model of instruction allowed for random assignment into the two treatment groups. Half of the students were asked to come to class and complete the apply activities (part two) while the other half of the students were asked to complete the apply activities online.

The two treatments used in this study differed in how students applied what they learned about organismal diversity (part two). Both groups learned about the basic characters and life cycles of the taxa with the same instruction and activities (part one, completed online prior to

class). After this introduction, the treatments differed by having the Tree group attend class and use evolutionary trees by mapping characters on a tree to look at major evolutionary transitions while the Traditional group was asked to stay home and complete an online activity focused on comparing and contrasting the characters and on major transitions without the use of an evolutionary tree. Using this approach, we could be sure that each group received the instruction needed to meet the course learning objectives while simultaneously using evolutionary trees differently.

The assessment (Diversity Assessment) was used to evaluate students' content knowledge that was gained during the unit. The Diversity Assessment consisted of 16 multiple-choice questions about the organismal diversity covered during the unit. The assessment required no use of evolutionary trees to prevent understanding of evolutionary trees (which potentially differed between treatment groups due to practice) from confounding the results.

Data analyses

Analyses were conducted using SPSS V.24. We used Cronbach's alpha to determine the reliability of the Diversity Assessment (0.693). The reliability was found to be in the acceptable range for an assessment of this type. The student scores on the Diversity Assessment were found to be normally distributed. This meets the assumptions of the Independent-samples t-test, which we used to analyze the data for differences between the two groups.

Results

We calculated the means and 95% confidence intervals for the Diversity Assessment scores of both groups. The results of the Independent-samples t-test showed there was a

significant difference between the Diversity Assessment scores with the Tree group being higher than the Traditional group M = 5.43, 95% CI [1.73, 9.13], $t(434) = -2.884$, $p = .004$.

We used Cohen's d to calculate an effect size for the difference between the 2 groups. The effect size was found to be d = .275. An effect size of this size falls just above the .2 that is considered a small effect size.

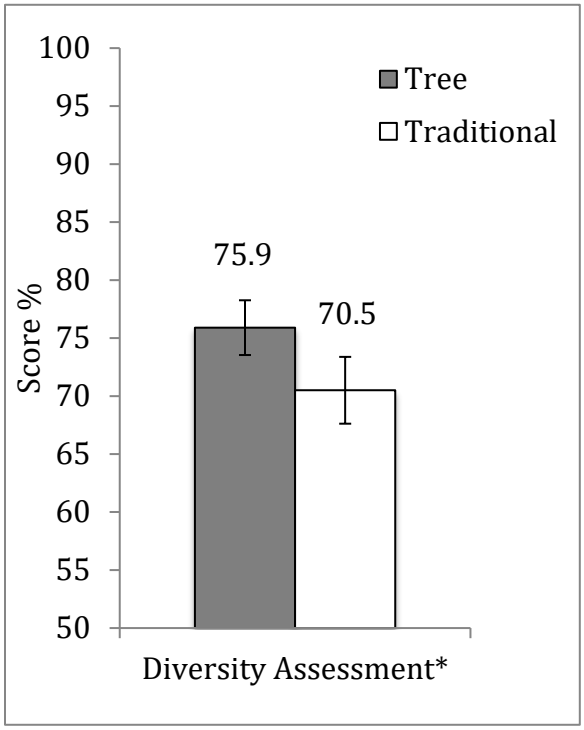


Figure 15. Means and 95% confidence intervals for both treatment groups. * Indicates p-value ≤ .01.

Experiment 3: Course Comparison

Methods

The final experiment used in this study was to implement our approach at the course level. This provides us with another authentic learning situation and one that is probably the most common way that organismal diversity is taught, meaning a course devoted to the topic. We used two sections of a plant diversity course with each serving as a separate treatment group for this experiment. We taught one section with an extensive focus on evolutionary trees in the

curriculum (Tree section) while the other section was taught using evolutionary trees but in a comparatively limited way (Traditional section). We used multiple assessments to measure learning in these sections including an assessment to measure plant diversity content knowledge at the end of the course. If studying plant diversity via evolutionary trees reduces cognitive load we would expect to see students in the Tree section score higher on the Plant Diversity Content Assessment (PDCA) than students in the Traditional section.

Subjects

Students for this experiment were students enrolled in two plant diversity sections. Students in these sections should have already been introduced to evolutionary tree concepts in introductory courses. Student permission was obtained to include data in the study. If students did not complete each assessment they were not included in the study. 44 total students participated in the study with 27 in the Tree section and 17 in the Traditional section.

Experimental Design

Different instructors taught each section covering the same content but in different ways: one with evolutionary trees being fully integrated and the other using them in a traditional manner (e.g., introducing new taxa, depicting character states). Using the traditional approach was important in order to compare this new integrated approach to an approach that authentically represents how organismal diversity is most commonly taught. Evolutionary trees are not excluded in most diversity courses but many instructors perhaps, underutilize them. Ideally, we would have used the same instructor to teach both sections but due to logistical limitations, this comparison was not possible.

The Tree section used evolutionary trees in multiple ways that differ from the Traditional section, yet commonalities between the sections exist. First, both sections had a lab associated

with the course during which students learned principles related to evolutionary trees (e.g., a modified form of The Great Clade Race) (Goldsmith, 2003). Second, both sections used evolutionary trees as part of lectures though to differing degrees.

The Tree section differed from the Traditional section in the structure of labs, lectures, and assessments. The first lab in the Tree section had an extra lab activity on evolutionary tree concepts. The first activity, completed by both courses, used tree building to introduce evolutionary tree concepts while the second activity, completed by the Tree section only, focused on what an evolutionary tree theoretically represents and why evolutionary trees are meaningful.

The labs in the Tree section also differed in the lab quizzes. The labs were held once a week and the majority of the labs focused on the characters of various taxa that were being discussed in lecture that week. Students used various tools in the lab to examine and compare characters. In the Traditional section course lab, quizzes consisted of a series of questions related to the introductory reading students were to complete prior to the lab. These quizzes were altered in the Tree section so that students only answered two questions on the readings and the remaining questions required students to know the evolutionary relationships of the taxa to be covered in the lab that week. Students were given an evolutionary tree with a bank of taxa that they would have to correctly place on the evolutionary tree.

The lectures in the two sections also differed in multiple ways. The Tree section had three lectures devoted explicitly to evolutionary tree concepts. This consisted of readings from a textbook, lectures on key concepts, and working through example problems as a class. The three lectures along with the two lab activities gave us confidence that students would be able to understand and use information conveyed to them in lectures using evolutionary trees. Lectures in the Tree section also differed in the integration of evolutionary trees into lectures and

discussions throughout the semester. The integration was at a level far beyond what was done in the Traditional section it included activities such as identifying monophyly and paraphyly among traditional taxonomic groups and determining the evolutionary change in characters implied by a given evolutionary tree. This meant that the students did not just know taxonomic groups and characters but that they could evaluate the evolutionary history of these taxonomic groups.

The final difference between the Tree and Traditional sections was in the summative assessments. Each weekly quiz and unit exam used in the Tree section required that students understood the plant diversity content in the context of the taxa's evolutionary history. It was hoped that emphasizing this understanding in the assessments would change how and what the students studied in preparation for their assessments and as a consequence what they learned in the course.

We used three assessments and a survey to compare students between our two treatment groups. The Biology Concept Inventory (BCI) is a commonly used assessment that measures general biology content knowledge at the university level. The BCI was given to the students near the beginning of the course. We used the BCI in this study as our measure of group equivalence. Without this, we could not be confident that differences between our groups were due to the treatment rather than a pre-existing difference in knowledge or ability.

The final two assessments and survey were given at the end of the course and were used to measure changes in response to the treatment. One of these assessments was the summative assessment, which we will refer to as the Plant Diversity Content Assessment (PDCA) mentioned previously. The PDCA consisted of 15 questions on plant diversity. Each question on this assessment was given to both instructors for evaluation. Both instructors agreed that these items represented content that students should know at the end of the course. We did not

include any questions on the PDCA that would require students to use evolutionary trees to ensure that any difference in performance was limited to content knowledge only.

An evolutionary tree assessment was also used to determine if the increased emphasis on trees also made a difference on student understanding of evolutionary tree concepts. The Tree-thinking Assessment (TTA) consists of 10 pairs of questions that included a multiple-choice component and a free response portion that asks students to explain the reasoning they used to answer the multiple-choice component. The TTA has previously been used to measure student understanding of evolutionary trees and to identify commonly held misconceptions (Kummer, Whipple, & Jensen, 2016). We used it for the same purposes in this study.

Table 15. Items on the four-point Likert scale.

Student Attitudes and Study Behavior
I regularly used evolutionary trees to study and learn in this course overall
I regularly used evolutionary trees to prepare for exams in this course
I found evolutionary trees to be helpful in learning about plant diversity
I feel confident in my ability to understand and interpret evolutionary trees

We used a four-question survey to measure student behavior and attitudes at the end of the course. The questions covered student study habits and their confidence level using a four-point Likert scale, see Table 15. This allowed us to see how the differences in the treatments might have impacted student attitudes and behaviors.

Data Analyses: Analyses were conducted using SPSS V.24. To determine the reliability of the assessments used as outcomes we used Cronbach's alpha. The reliability of both the TTA (0.724) and the PDCA (0.683) fell within the acceptable range for an assessment of this type though the content assessment was on the low end of that range (Kline, 2000).

We found that the scores on the TTA and PDCA were not normally distributed and had unequal variance. Given these characteristics, we used a Mann-Whitney U Test to compare the scores of each group on these assessments. The BCI scores were found to meet the assumptions of normality making it appropriate to use an Independent-samples t-test.

Results

To determine if either section had a significantly higher level of biology content knowledge prior to taking the plant diversity course we compared their scores on the BCI. Using an Independent-samples t-test we found that there was no significant difference between the mean scores of the sections, M = .806, $t(44) = -.225$, $p = .823$.

We calculated the mean scores and 95% confidence intervals for each group on the TTA. The mean score for the Tree section was 85.9 ± 6.76 while the mean score for the Traditional section was 67.7 ± 12.13. A Mann-Whitney U Test showed that the difference between the Tree section (90.0) and Traditional section (Mdn = 60.0) was significant, U = 121.0, $z = -2.675$, $p = .007$. We used Cohen's d to calculate effect size. Effect size was found to be d = .885 which is considered to be a large effect size.

The PDCA was used to measure differences in student learning of diversity content. We calculated the mean score and 95% confidence intervals for both groups. We used a Mann-Whitney U Test to compare student scores on the PDCA and the results showed that the scores of the Tree section (Mdn = 74.2) were significantly different from the Traditional section (Mdn = 43.6), U = 26.0, $z = -4.92$, $p = .000$. Cohen's d was used to calculate the effect size and it was found to be d = 2.31 which is a large effect size.

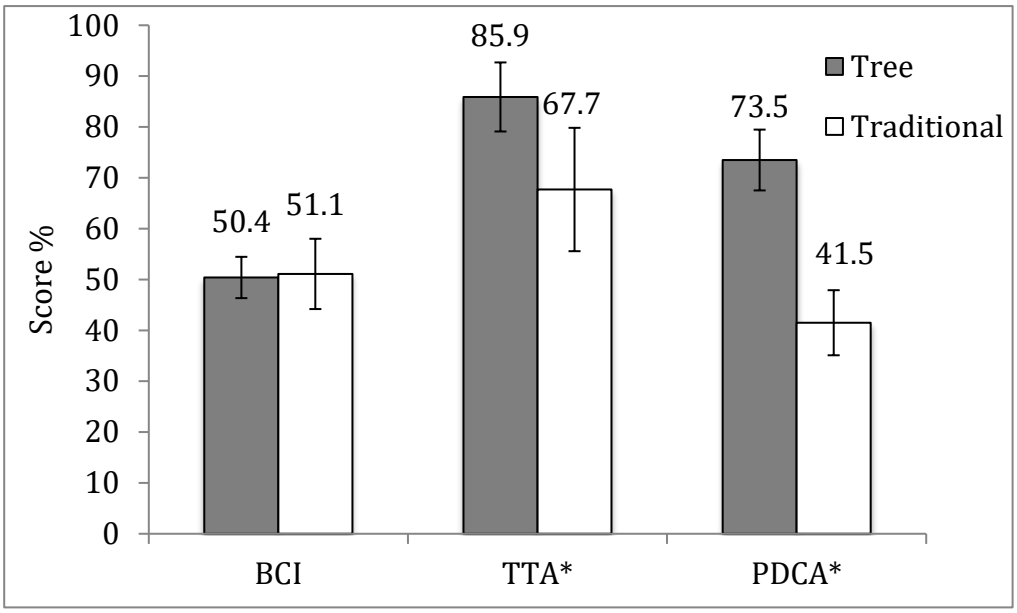


Figure 16. Means and 95% confidence intervals for each treatment group on the BCI, TTA, and PDCA. * Indicates *p*-value ≤ .01.

We asked the students to mark their level of agreement with four statements to assess differences in study habits and attitudes. The first statement related to the overall use of evolutionary trees when studying for the course. 33% of students in the Tree section reported that they strongly agreed that they regularly studied with evolutionary trees during the course compared to 24% in the Traditional section. The second statement specifically referred to using evolutionary trees to prepare for exams. On this statement, 70% of students in the Tree section strongly agreed while only 23.5% of students in the Traditional section strongly agreed.

The second pair of statements related to how students felt about the usefulness of evolutionary trees in studying plant diversity and how confident students felt about their ability to interpret evolutionary trees. 70% of students in the Tree section strongly agreed that evolutionary trees were helpful when studying plant diversity compared to just 47% in the Traditional section. Finally, only 33% in the Tree section strongly agreed that they were confident in their ability while 23% strongly agreed in the Traditional section.

Discussion

The experimental comparison provides evidence in support of our hypothesis that evolutionary trees improve learning of organismal diversity by reducing the cognitive load of the learner. We also see that the impact of using evolutionary trees appears to become even greater as the time from the period of study increases. This implies that in addition to improving learning the use of evolutionary trees may also improve retention. If retention is also improved this gives us two potential mechanisms by which student learning may benefit from using trees in authentic learning environments. The first mechanism involves reducing cognitive load described previously. The second mechanism is that trees allow for better retention due to the picture superiority effect (Defeyter, Russo, & McPartlin, 2009). The evolutionary tree with the characters mapped can be considered an image that is potentially easier for students to recall than the text of the table the other group had to memorize.

While the results from the experimental comparison provide strong support in an experimental setting, this is not how students necessarily study for an actual course. Students are not limited to an equal amount of study nor are they limited to assigned study materials. The initial acquisition might be better with evolutionary trees but in authentic situations, that advantage might be overcome with more time studying as students recognize they do not know the material yet. By comparing learning organismal diversity using evolutionary trees to traditional methods of instruction in an authentic learning situation we can see if these differences have practical implications.

The unit level comparison provides evidence that using evolutionary trees to learn diversity helps in authentic situations, as well. While we did see significant differences between

the groups the effect size was relatively small. This indicates that the benefits seen in the experimental comparison are reduced when study time is not constrained. While the results from the unit comparison support the findings of the experimental comparison it is possible that the benefits of using evolutionary trees to study diversity might be seen because the assessment consisted of only low-level items that require rote memorization. It is possible that using evolutionary trees would not impact student learning with higher-level assessment items that would be more common in courses devoted to organismal diversity.

The course comparison is the most authentic learning situation for the majority of students who are studying organismal diversity. When comparing the two treatment groups, we found that students who were in the tree infused course outperformed those in the traditional course when it came to learning organismal diversity content.

We also found that students in the Tree course understandably scored higher than those in the Traditional course when it came to their understanding of evolutionary tree concepts. This finding is consistent with other research and is to be expected when greater time and emphasis is given to a topic (Smith et al., 2013).

While the differences on the content assessment are drastic, we feel that those findings should be viewed with caution. A different instructor taught each treatment condition. This instructor effect might have a greater influence on the results than did the difference in treatment, although every effort to make covered learning outcomes consistent was taken. In particular, the content assessment score between treatments had a difference with a magnitude much larger than we saw in the previous experiments, and thus should be considered with caution.

We do not believe that we can reliably conclude that the course comparison provides unequivocal evidence that organismal diversity learning will be improved by fully integrating

evolutionary trees into diversity courses. However, we do feel confident in concluding that learning of content is not hindered by the integration of evolutionary trees even though it took substantial course time away from teaching strictly the plant diversity content. We believe that we also have shown that organismal diversity courses can be used as effective avenues to not only teach the traditional content but also help students better understand evolutionary tree concepts, which are critical to biological literacy (Baum & Offner, 2008; Baum et al., 2005).

While we express caution with regard to the content findings of experiment 3 we do feel that given the results of the first two comparisons we have provided evidence that evolutionary trees can positively influence student learning of organismal diversity. This has significant implications for instructors of organismal diversity courses. We recommend that instructors revise their curriculum to include evolutionary trees as the unifying theme of the course with substantial use in lectures, activities, and assessments.

Assessment, in particular, is a key indicator to students of what you want them to learn. Assessments are how students learn what you, as the instructor, value and in turn, it becomes what they will study in order to prepare for future assessments (Ramsden, 2003). If assessments (both formative and summative) only address organismal diversity content without introducing elements of evolutionary trees students will not learn the content in the desired context. Results from the attitudinal survey support the importance of assessment in altering student-learning behaviors. Students in the Tree section reported that far more use of evolutionary trees in preparing for exams than did students in the Traditional section.

Fully integrating evolutionary trees into a diversity course is not an easy undertaking. A number of publications provide possible paths to bring evolutionary trees into a course. One method that was utilized in this study is the mapping of characters on a given evolutionary tree

(Smith & Cheruvelil, 2009). Another commonly used practice consists of using evolutionary trees to provide context when introducing new taxa (Staub et al., 2006). A different approach involves having students either build an evolutionary tree or evaluate the evidence supporting competing evolutionary trees based on the content learned throughout the course (Singer, Hagen, & Sheehy, 2001; Smith et al., 2013; White, 2009). These methods cited here are all useful and important contributions to the literature but they have very little evidence as to their effectiveness at helping students learn organismal diversity. The lack of evidence makes it difficult to recommend one approach over another. We recommend considering each approach along with the learning goals of the course as instructors determine how they might integrate evolutionary trees into organismal diversity courses.

We recommend instructors utilize pre-assessments to assess student understanding of evolutionary tree concepts near the beginning of the course (Baum et al., 2005). This would allow the instructor to determine who is in need of extra assistance in this area so that students who lack understanding of these concepts are not lost in the course.

We suggest two areas that we feel warrant further study. First, we would like to better understand the mechanisms that appear to result in improved learning in the evolutionary tree conditions. We believe this might be addressed through an experiment similar to experiment 1. Because the first mechanism appeals to cognitive load and a more efficient acquisition of the content knowledge, we believe we can test its influence by allowing different durations of study. If we increase the duration of study for the List group we would expect the scores on both administrations of the assessment to increase and be closer to the scores seen in the Tree group. Using images as part of the table and comparing results with the images to the results of this study might address the impact of the evolutionary tree as an image.

The second area in which we recommend further study is using a single instructor to teach each treatment in an organismal diversity course. Along with potentially providing more reliable evidence to the findings of this study, this could also be an opportunity to compare the effectiveness of the various methods described previously at enhancing student learning of organismal diversity and in enhancing instructor satisfaction.

REFERENCES

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: a call to action*. Washington, D.C.

Atman, C. J., Cardella, M. E., Turns, J., & Adams, R. (2005). Comparing freshman and senior engineering design processes: an in-depth follow-up study. *Design Studies*, *26*(4), 325–357.

Azizi-Fini, I., Hajibagheri, A., & Adib-Hajbaghery, M. (2015). Critical Thinking Skills in Nursing Students: a Comparison Between Freshmen and Senior Students. *Nursing and Midwifery Studies*, *4*(1).

Baum, D. A., & Offner, S. (2008). Phylogenics & Tree-Thinking. *The American Biology Teacher*, *70*(4), 222–229.

Baum, D. A., & Smith, S. D. (2013). *Tree-thinking: An introduction to phylogenetic biology* (Vol. 1). Greenwood Village, Colorado: Roberts and Company Publishers.

Baum, D. A., Smith, S. D., & Donovan, S. S. S. (2005). The Tree-Thinking Challenge. *Science*, *310*(5750), 979–980.

Blacquiere, L. D., & Hoese, W. J. (2016a). A valid assessment of students' skill in determining relationships on evolutionary trees. *Evolution: Education and Outreach*, *9*(1), 5.

Bransford, J., Brown, A., & Cocking, R. (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press.

Brooks, D. R. (2010). Sagas of the Children of Time: The Importance of Phylogenetic Teaching in Biology. *Evolution: Education and Outreach*, *3*(4), 495–498.

Catley, K. M. (2006). Darwin's missing link—a novel paradigm for evolution education. *Science Education*, *90*(5), 767–783.

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

Catley, K. M., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. *Journal of Research in Science Teaching*, *47*(7), 861–882.

Catley, K. M., Phillips, B. C., & Novick, L. R. (2013). Snakes and Eels and Dogs! Oh, My! Evaluating High School Students' Tree-Thinking Skills: An Entry Point to Understanding Evolution. *Research in Science Education*, *43*(6), 2327–2348.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* (1st ed.). London: John Murray.

Davenport, K., Milks, K. J., & Tassell, R. V. (2015a). Investigating Tree-thinking & Ancestry with Cladograms. *The American Biology Teacher*, *77*(3), 198–204.

Davenport, K., Milks, K. J., & Tassell, R. V. (2015b). Using Evolutionary Data in Developing Phylogenetic Trees: A Scaffolded Approach with Authentic Data. *The American Biology Teacher*, *77*(4), 274–283.

Dees, J., Momsen, J. L., Niemi, J., & Montplaisir, L. (2014). Student Interpretations of Phylogenetic Trees in an Introductory Biology Course. *CBE-Life Sciences Education*, *13*(4), 666–676.

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development*, *24*(3), 265–273.

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE*, *10*(4).

Dominici, S., & Eldredge, N. (2010). Brocchi, Darwin, and Transmutation: Phylogenetics and Paleontology at the Dawn of Evolutionary Biology. *Evolution: Education and Outreach*, *3*(4), 576–584.

Doran, R. L. (1980). *Basic Measurement and Evaluation of Science Instruction.* National Science Teachers Association, 1742 Connecticut Ave., N.W., Washington, DC.

Eddy, S. L., Crowe, A. J., Wenderoth, M. P., & Freeman, S. (2013). How should we teach tree-thinking? An experimental test of two hypotheses. *Evolution: Education and Outreach*, *6*(1), 13.

Eldredge, N. (2010). How Systematics Became "Phylogenetic." *Evolution: Education and Outreach*, *3*(4), 491–494.

Farris, J. S. (1970). Methods for Computing Wagner Trees. *Systematic Zoology*, *19*(1), 83–92.

Garvin-Doxas, K., Klymkowsky, M., & Elrod, S. (2007). Building, Using, and Maximizing the Impact of Concept Inventories in the Biological Sciences: Report on a National Science Foundation–sponsored Conference on the Construction of Concept Inventories in the Biological Sciences. *CBE Life Sciences Education*, *6*(4), 277–282.

Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding Randomness and its Impact on Student Learning: Lessons Learned from Building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education*, *7*(2), 227–233.

Gee, H. (2002). Progressive evolution: Aspirational thinking. *Nature*, *420*(6916), 611–611. https://doi.org/10.1038/420611a

Genco, N., Hölttä-Otto, K., & Seepersad, C. C. (2012). An Experimental Investigation of the Innovation Capabilities of Undergraduate Engineering Students. *Journal of Engineering Education*, *101*(1), 60–81.

Goldsmith, D. W. (2003). The Great Clade Race. *The American Biology Teacher*, *65*(9), 679–682.

Goldstein, A. M. (2010). Exploring Phylogeny at the Tree of Life Web Project. *Evolution: Education and Outreach*, *3*(4), 668–674.

Greene, H. W. (2005). Organisms in nature as a central focus for biology. *Trends in Ecology & Evolution*, *20*(1), 23–27.

Gregory, T. R. (2008). Understanding Evolutionary Trees. *Evolution: Education and Outreach*, *1*(2), 121–137.

Gregory, T. R., & Ellis, C. A. J. (2009). Conceptions of Evolution among Science Graduate Students. *BioScience*, *59*(9), 792–799.

Halverson, K. L. (2010). Using Pipe Cleaners to Bring the Tree of Life to Life. *The American Biology Teacher*, *72*(4), 223–224.

Halverson, K. L. (2011). Improving Tree-Thinking One Learnable Skill at a Time. *Evolution: Education and Outreach*, *4*(1), 95–106.

Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree-thinking expertise in an undergraduate systematics course. *Science Education*,

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Julius, M. L., & Schoenfuss, H. L. (2005). Phylogenetic Reconstruction as a Broadly Applicable Teaching Tool in the Biology Classroom: The Value of Data in Estimating Likely Answers. *Journal of College Science Teaching*.

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a Meiosis Concept Inventory. *CBE-Life Sciences Education*, *12*(4), 655–664.

Kampourakis, K. (2007). Teleology in biology, chemistry and physics education: what primary teachers should know. *Review of Science, Mathematics and ICT Education*, *1*(2), 81–96.

Kline, P. (2000). *The Handbook of Psychological Testing*. Psychology Press.

Kögce, D., & Yıldız, C. (2011). A comparision of freshman and senior mathematics student teachers' views of proof concept. *Procedia - Social and Behavioral Sciences*, *15*, 1266–1270.

Kolaczkowski, B., & Thornton, J. W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, *431*(7011), 980–984.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218.

Kumala, M. (2010a). A Natural History of You. *Evolution: Education and Outreach*, *3*(4), 532–538.

Kumala, M. (2010b). The Gummy Tree Challenge—Building Connections One Treat at a Time. *Evolution: Education and Outreach*, *3*(4), 520–525.

Kummer, T. A., Whipple, C. J., & Jensen, J. L. (2016). Prevalence and Persistence of Misconceptions in Tree-thinking †. *Journal of Microbiology & Biology Education*, *17*(3), 389–398.

Kuzoff, R. K., Kemmeter, S. B., McKinnon, J. S., & Thompson, C. P. (2009). Phylogenetic Analysis: How Old are the Parts of Your Body? *Evolution: Education and Outreach*, *2*(3), 405–414.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11–24.

Lawson, A. E., Alkhoury, S., Benford, R., Clark, B. R., & Falconer, K. A. (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. *Journal of Research in Science Teaching*, *37*(9), 996–1018.

Lents, N. H., Cifuentes, O. E., & Carpi, A. (2010). Teaching the Process of Molecular Phylogeny and Systematics: A Multi-Part Inquiry-Based Exercise. *CBE-Life Sciences Education*, *9*(4), 513–523.

Ludlow, D., & Evenson, W. (1992). *Encyclopedia of Mormonism*. New York: Macmillan Publishing.

Manwaring, K. F., Jensen, J. L., Gill, R. A., & Bybee, S. M. (2015). Influencing highly religious undergraduate perceptions of evolution: Mormons as a case study. *Evolution: Education and Outreach*, *8*(1), 1–12.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97–110.

McLennan, D. A. (2010). How to Read a Phylogenetic Tree. *Evolution: Education and Outreach*, *3*(4), 506–519.

McLennan, D. A. (2010). Sociobiology and the Comparative Approach: One Way to Study Ourselves. *Evolution: Education and Outreach*, *3*(4), 548–557.

Meikle, W. E., & Scott, E. C. (2010). Why Are There Still Monkeys? *Evolution: Education and Outreach*, *3*(4), 573–575.

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

Meisel, R. P. (2010). Teaching Tree-Thinking to Undergraduate Biology Students. *Evolution*, *3*(4), 621–628.

Miller, M. D., Linn, R. L., & Grunland, N. E. (2013). *Measurement and Assessment in Teaching* (11th ed.). Boston: Pearson.

Moore, R., & Cotner, S. (2008). Educational Malpractice: The Impact of Including Creationism in High School Biology Courses. *Evolution: Education and Outreach*, *2*(1), 95–100.

Neagle, E. (2009, April). *Patterns of thinking about phylogenetic trees: A study of student learning and the potential of tree-thinking to improve comprehension of biological concepts.* Idaho State university.

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, *45*(10), 1131–1160.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: the influence of a Gestalt Perceptual Principle. *Journal of Experimental Psychology. Applied*, *13*(4), 197–223.

Novick, L. R., & Catley, K. M. (2012). Reasoning About Evolution's Grand Patterns: College Students' Understanding of the Tree of Life. *American Educational Research Journal*, 2831212448209.

Novick, L. R., Catley, K. M., & Funk, D. J. (2010). Characters Are Key: The Effect of Synapomorphies on Cladogram Comprehension. *Evolution: Education and Outreach*, *3*(4), 539–547.

Novick, L. R., Catley, K. M., & Funk, D. J. (2011). Inference Is Bliss: Using Evolutionary Relationship to Guide Categorical Inferences. *Cognitive Science*, *35*(4), 712–743.

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

Offner, S. (2016). Using Great Ape Phylogeny to Teach Evolutionary Thinking. *The American Biology Teacher*, *78*(3), 263–265.

O'Hara, R. J. (1997). Population thinking and tree-thinking in systematics. *Zoologica Scripta*, *26*(4), 323–329.

Omland, K. E., Cook, L. G., & Crisp, M. D. (2008). Tree-thinking for all biology: the problem with reading phylogenies as ladders of progress. *BioEssays*, *30*(9), 854–867.

Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D. (2008). Evaluating Two Approaches to Helping College Students Understand Evolutionary Trees through Diagramming Tasks. *CBE Life Sciences Education*, *7*(2), 193–201.

Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2012). Teaching Tree-thinking to College Students: It's Not as Easy as You Think. *Evolution: Education and Outreach*, *5*(4), 595–602.

Ramsden, P. (2003). *Learning to Teach in Higher Education*. Routledge.

Raven, P., Evert, R., & Eichhorn, S. (2005). *Biology of Plants* (7th ed.). W.H. Freeman and Company.

Rice, J. W., Clough, M. P., Olson, J. K., Adams, D. C., & Colbert, J. T. (2015). University faculty and their knowledge & acceptance of biological evolution. *Evolution: Education and Outreach*, *8*(1), 1–15.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling (PDF Download Available). *Journal of Statistical Software*, *48*(2).

Sandvik, H. (2008). Tree-thinking cannot taken for granted: challenges for teaching phylogenetics. *Theory in Biosciences*, *127*(1), 45–51.

Schwenk, K., Padilla, D. K., Bakken, G. S., & Full, R. J. (2009). Grand challenges in organismal biology. *Integrative and Comparative Biology*, *49*(1), 7–14.

Singer, F., Hagen, J. B., & Sheehy, R. R. (2001). The Comparative Method, Hypothesis Testing & Phylogenetic Analysis – An Introductory Laboratory. *The American Biology Teacher*, *63*(7), 518–523.

Smith, J. J., & Cheruvelil, K. S. (2009). Using Inquiry and Tree-Thinking to "March Through the Animal Phyla": Teaching Introductory Comparative Biology in an Evolutionary Context. *Evolution: Education and Outreach*, *2*(3), 429–444.

Smith, J. J., Cheruvelil, K. S., & Auvenshine, S. (2013). Assessment of Student Learning Associated with Tree-thinking in an Undergraduate Introductory Organismal Biology Course. *CBE Life Sciences Education*, *12*(3), 542–552.

Staub, N. L., Pauw, P. G., & Pauw, D. (2006). Seeing the Forest Through the Trees: Helping Students Appreciate Life's Diversity by Building the Tree of Life. *The American Biology Teacher*, *68*(3), 149–151.

Sweller, J., Merrienboer, J. J. G. van, & Paas, F.G.W.C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296.

Thanukos, A. (2010). Evolutionary Trees from the Tabloids and Beyond. *Evolution: Education and Outreach*, *3*(4), 563–572.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69.

Walter, E. M., Halverson, K. M., & Boyce, C. J. (2013). Investigating the relationship between college students' acceptance of evolution and tree-thinking understanding. *Evolution: Education and Outreach*, *6*(1), 26.

White, B. T. (2009). Exploring the Diversity of Life with the Phylogenetic Collection Lab. *The American Biology Teacher*, *71*(3), 157–161.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by Design*. ASCD.

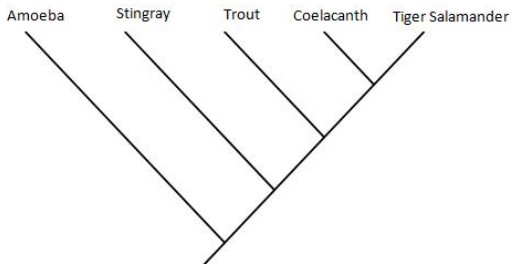
Wiley, E. O. (2010). Why Trees Are Important. *Evolution: Education and Outreach*, *3*(4), 499–505.

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, *13*(5), 303–314.

Young, A. K., White, B. T., & Skurtu, T. (2013). Teaching undergraduate students to draw phylogenetic trees: performance measures and partial successes. *Evolution: Education and Outreach*, *6*(1), 16.

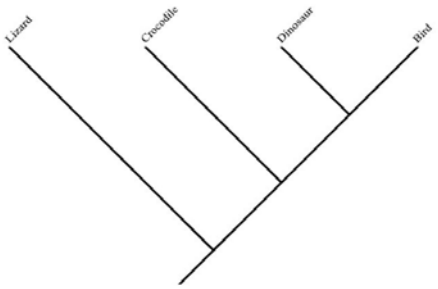
APPENDIX

Tree-thinking Assessment



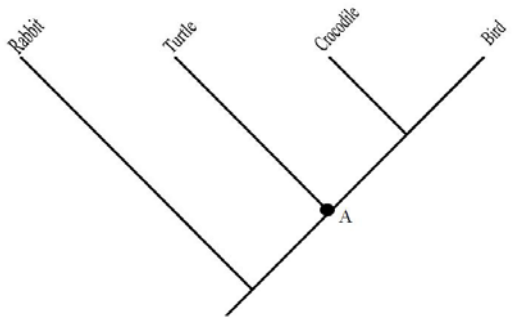
1. By reference to the tree above, which of the following is an accurate statement of relationships?
    a. A trout is more closely related to a stingray than to a coelacanth
    b. A trout is more closely related to a coelacanth than to a stingray
    c. A trout is equally related to a stingray and a coelacanth
    d. A trout is related to a stingray, but is not related to a coelacanth

2. Explain the reasoning you used to answer the previous question



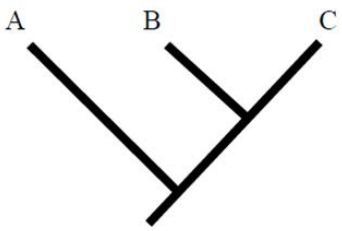
3. By reference to the tree above, which of the following is an accurate statement of relationships?
    a. A crocodile is more closely related to a lizard than to a bird
    b. A crocodile is more closely related to a bird than to a lizard
    c. A crocodile is equally related to a lizard and a bird
    d. A crocodile is related to a lizard, but is not related to a bird

4. Explain the reasoning you used to answer the previous question



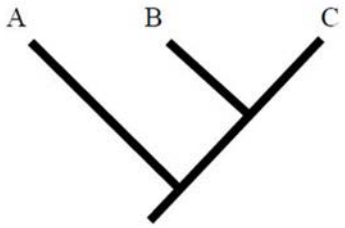
5. Given the tree above, what would you expect the common ancestor marked ‘A’ to look like
    a. Most like a rabbit because it is the only species that is an ancestor of A
    b. Most like a turtle because it is the most direct descendant of A
    c. Most like a crocodile because a crocodile is known to be a “living fossil”
    d. An equal mix of rabbit, turtle, crocodile, and bird features, because it is an ancestor of all of them
    e. One cannot say without a model of how traits evolved along the branches of this tree

6. Explain the reasoning you used to answer the previous question



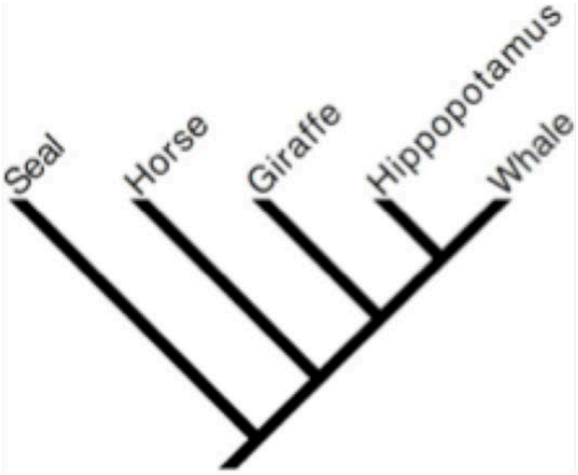
7. Which of the following is a correct interpretation of the tree shown above?
    a. "C" is descended from "B", which is descended from "A"
    b. "C" is the most advanced species
    c. "A" is the most ancient species
    d. "B" is an intermediate between "A" and "C"
    e. None of the above

8. Explain the reasoning you used to answer the previous question



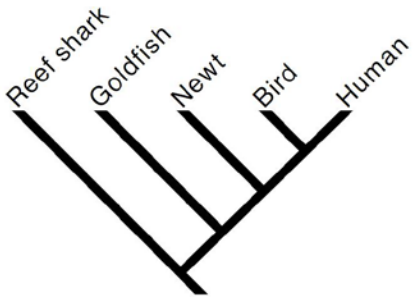
9. Referring to the above tree, which statement about common ancestry hold?
    a. A is the common ancestor of B and C
    b. The common ancestor of A and B lived after the common ancestor of A and C
    c. B and C share a more recent common ancestor than B and A
    d. Any common ancestor of C and B is also an ancestor of A
    e. None of the above

10. Explain the reasoning you used to answer the previous question



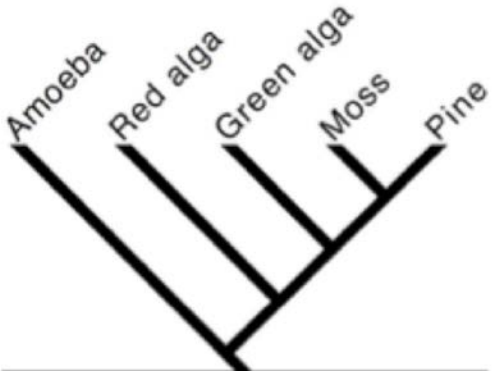
11. By reference to the tree above, which of the following is an accurate statement of relationships?
    a. A seal is more closely related to a horse than to a whale
    b. A seal is more closely related to a whale than to a horse
    c. A seal is equally related to a horse and a whale
    d. A seal is related to a whale, but is not related to a horse

12. Explain the reasoning you used to answer the previous question



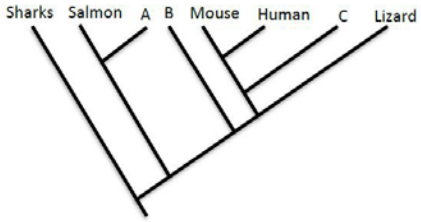
13. Assume that the tree above is correct. Which of the following is true?
    a. Reef sharks are older than newts
    b. Reef sharks gave rise to goldfish
    c. The common ancestor of goldfish and humans lived before the common ancestor of birds and humans
    d. Reef sharks and goldfish have no common ancestor
    e. Birds came before humans

14. Explain the reasoning you used to answer the previous question



15. Looking at the tree above, five students have different interpretations. Student A says that pine is the most advanced species because it is the most recent. Student B says that the pine is the least advanced species because all the others branch off it. Student C says that all the species are equally advanced because they have all evolved the same amount of time from their common ancestor. Student D says that based on similarities, Red and Green Alga are most closely related to each other and Moss and Pines are most closely related to each other. Student E says that based on similarity, only red and green alga are closely related. Which student is correct?
    a. Student A
    b. Student B
    c. Student C
    d. Student D
    e. Student E

16. Explain the reasoning you used to answer the previous question



17. Looking at the tree above, where would you place a dolphin?
    a. Position A
    b. Position B
    c. Position C
    d. None

18. Explain the reasoning you used to answer the previous question

19. *Homo sapiens* evolved from *Pan troglodytes* (Chimp)
    a. The above statement accurately reflects current scientific thought
    b. The above statement is false because *Homo sapiens* did not evolve
    c. The above statement is false because *Homo sapiens* evolved from *Homo neanderthalensis*
    d. The above statement is false because *Homo sapiens* and *Pan troglodytes* each evolved from a common ancestor

20. Explain the reasoning you used to answer the previous question

Supplementary Table 1

Report of the percentage of students classified with a misconception for each question and response. M1 is Reading The Tips, M2 is Node Counting, M3 is Ladder Thinking, M4 is Similarity Equals Relatedness, M5 is Branch Length, and Other is no clear misconception.

	Intro	Evo	Intro	Evo	Intro	Evo	Intro	Evo	Intro	Evo
Question	1&2		3&4		5&6		7&8		9&10	
A (Total)	13.5	5.1	66.2	43.6	5.4	10.3	5.4		13.5	15.4
M1	9.5		55.4	18.0						
M2			2.7	15.4						
M3	2.7	2.6		2.6	5.4	10.3	5.4		12.2	12.8
M4			5.4	2.6					1.4	
M5				5.1						
Other	1.4	2.6	2.7							
B (Total)	27.0	43.6	20.3	33.3	64.9	46.2	12.2	2.6	2.7	2.6
M2	5.4		1.4							
M3	1.4		4.1	5.1	63.5	46.2	10.8	2.6	2.7	
M4	6.8		4.1							
M5		20.5		10.3						
Correct	12.2	20.5	9.5	15.4						
Other	1.4	2.6	1.4	2.6	1.4		1.4			2.6
C (Total)	59.5	51.3	13.5	23.1			28.4	53.9	44.6	71.8
M1	41.9	15.4		2.6						
M2	12.2	28.2	10.8	20.5						
M3							25.7	51.3	1.4	
M4	1.4								1.4	
M5	1.4									
Correct									39.2	69.2
Other	4.0	7.7	2.70				2.7	2.6	2.7	2.6
D(Total)					1.4		16.2		23.0	2.6
M3					1.4		14.8		10.8	
M4									2.7	
Other							1.4		9.5	2.6
E (Total)					28.4	43.6	37.8	43.6	16.2	7.7
M3							4.1	2.6	6.8	
M4									5.4	
Correct					25.7	35.9	29.7	33.3		
Other					2.7	7.7	4.1	7.7	4.0	7.7
	Intro	Evo	Intro	Evo	Intro	Evo	Intro	Evo	Intro	Evo
Question	11&12		13&14		15&16		17&18		19&20	
A (Total)	66.2	61.5	27.0	23.1	14.9	28.2	2.7	5.1	35.1	48.7
M1	44.6	20.5								

M2	6.8	38.5								
M3			25.7	23.1	12.2	28.2			27.0	46.2
M4	9.5						2.7	5.1	1.4	
M5	1.4	2.5								
Other	4.0		1.4		2.7				6.8	2.6
B (Total)	13.5	10.2	10.8	5.1	5.4	7.7	35.1	43.6		
M1					1.4					
M2	1.4	5.1								
M3	5.4	5.1	10.8	5.1	2.7	7.7	4.1	12.8		
M4	6.8				1.4		28.4	28.2		
Other							2.6	2.6		
C (Total)	16.2	28.2	47.3	69.2	37.8	30.8	58.1	48.7	2.7	2.6
M3	1.4			2.6				2.6		2.6
M4	1.4		1.4				6.8			
M5		5.1								
Correct	10.8	18.0	43.2	64.1	35.1	28.2	48.7	46.1		
Other	4.0	5.1	2.7	2.6	2.7	2.6	2.6		2.7	
D(Total)	4.1		6.8		40.5		4.1	2.6	62.2	48.7
M1					16.2					
M2					2.7					
M3	1.4		1.4							5.1
M4					18.9		2.7	2.6	1.4	
Correct									55.4	38.5
Other	2.7		5.4		2.7		1.4		5.4	5.1
E (Total)			8.1		1.4					
M3			6.8							
M4					1.4					
Other			1.4							

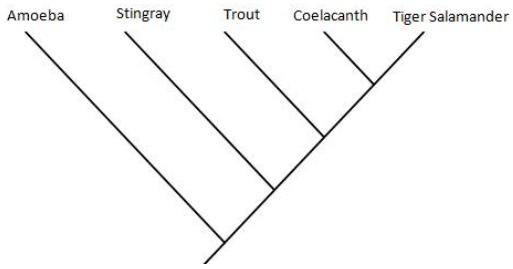
Supplementary Table 2

Representative quotes for each misconception. These quotes are each indicative of the types of responses categorized under each misconception during the coding process.

Reading the Tips
"A trout is one line away from both a stingray and a coelacanth so they are a similar difference away in relation."
"They are equal distances away. If the stingray were next to the coelacanth, then my statement would be false."
"The branches for a seal and a horse are closer"
"The seal and horse are closer together than the seal and whale (which are at opposite ends)."
"Horse is closer to a seal on the tree than a whale is."
Node Counting
"There are more evolutionary steps between the seal and the whale than between the seal and the horse."
"There is only one node between the seal and the whale, and there are two (if you count the bottom node, the first split) before you can get to the horse from the seal."
"The seal is only one break away from the horse while 3 from the whale."
"The bird is two steps away from crocodile while the lizard is only one step away."
"The same number of evolutionary divergences separate the two and they have a common ancestor."
Ladder Thinking
"I reasoned that since the trout was in between the coelacanth and the stingray, it seems to have evolved from the stingray while the coelacanth evolved from the trout."
"A seal is more closely related to a whale than to a horse because it branches off from the whale line, while the horse does also, it is not directly connected to the seal."
"The seal is more closely related to the whale because its species came from the whale species."
"The seal would be a descendent of the whale and so would the horse so they're related."
"The most recent should have evolved the most."
Similarity = Relatedness
"The trout has similar characteristics with the stingray but not the coelacanth."
"The trout has all the same traits as the stingray and does the coelacanth, but the coelacanth has traits that the trout does not."
"It has very few common traits with both of them."
"They share more common traits than the seal and the whale."
"The seal is a marine mammal as with the whale. Horses are land mammals."
Branch Length

"It is difficult to measure the edge lengths in this tree. It seems to me that there is equal distance for horse-seal and whale-seal paths."
"The branch length is longer between stingray and trout than it is between trout and coelacanth, so the trout is more closely related to a coelacanth than to a stingray."
"The branch is longer to the stingray. Usually the length of the lines determines the amount of relatedness one species holds to another."
"The length of the branches equates to amount of relatedness, shorter branches are more closely related. So the trout is more closely related to the Coelcanth (*sic*) because the lines are shorter, meaning there is less difference."
"The horse and the whale have the same branch lengths."

Initial Student Interview Protocol



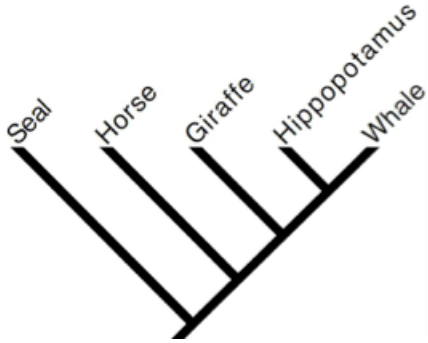
By reference to the tree above, which of the following is an accurate statement of relationships?

A trout is more closely related to a stingray than to a coelacanth
A trout is more closely related to a coelacanth than to a stingray
A trout is equally related to a stingray and a coelacanth
A trout is related to a stingray, but is not related to a coelacanth

Explain the reasoning you used to answer the previous question

What does an evolutionary tree represent?

What evidence do scientists use to build/support one evolutionary tree over another?



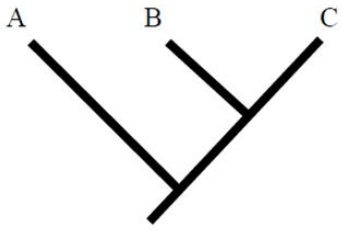
By reference to the tree above, which of the following is an accurate statement of relationships?

A seal is more closely related to a horse than to a whale
A seal is more closely related to a whale than to a horse
A seal is equally related to a horse and a whale
A seal is related to a whale, but is not related to a horse

Explain the reasoning you used to answer the previous question

What does a node represent?

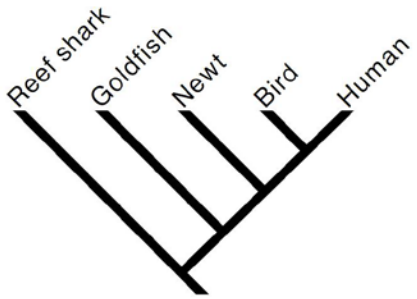
What does the line after the node represent?



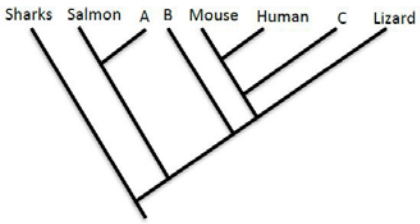
Which of the following is a correct interpretation of the tree shown above? (select all that apply)

"C" is descended from "B", which is descended from "A"
"C" is the most advanced species
"A" is the most ancient species
"B" is an intermediate between "A" and "C"
None of the above

Explain the reasoning you used to answer the previous question



Where on the tree did Goldfish stop evolving?



Looking at the tree above, where would you place a dolphin?

Position A
Position B
Position C
None

Explain the reasoning you used to answer the previous question

Item Interview Questions

Take a moment to review this item

What do you believe are the correct answers?

Was there any wording in the questions or answers that was confusing?

How might you change this question to make it easier to understand?

Tree vs. List Assessment

Which of the following animals exhibited trait 1?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 2?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 3?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 4?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 5?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 6?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 7?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 8?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 9?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 10?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

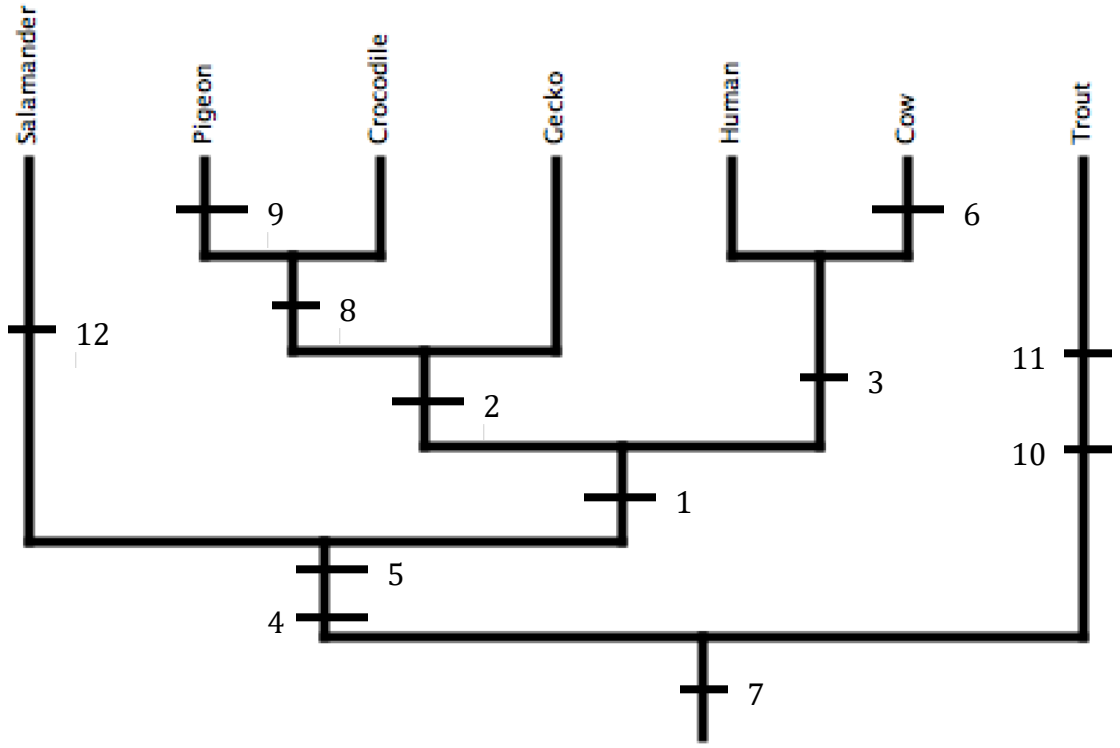
Which of the following animals exhibited trait 11?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Which of the following animals exhibited trait 12?

a. Cow
b. Trout
c. Pigeon
d. Crocodile
e. Salamander
f. Human
g. Gecko

Tree vs. List Study Materials



Pigeon	Trout	Cow	Human	Gecko	Salamander	Crocodile
2	11	3	3	2	12	2
1	10	1	1	1	7	1
9	7	4	4	4	4	4
4		5	5	5	5	5
5		6	7	7		7
7		7				8
8						