CBE—Life Sciences Education
Vol. 7, 193–201, Summer 2008

Article

Evaluating Two Approaches to Helping College Students Understand Evolutionary Trees through Diagramming Tasks

Judy Perry,* Eli Meir,*† Jon C. Herron,‡ Susan Maruca,† and Derek Stal†

*Teacher Education Program, Massachusetts Institute of Technology, Cambridge, MA 02139-4307; †SimBiotic Software, Ithaca, NY 14850; and ‡Department of Biology, University of Washington, Seattle, WA 98195-1800

Submitted February 1, 2007; Revised December 23, 2007; Accepted January 8, 2008
Monitoring Editor: W. Bradley Kincaid

To understand evolutionary theory, students must be able to understand and use evolutionary trees and their underlying concepts. Active, hands-on curricula relevant to macroevolution can be challenging to implement across large college-level classes where textbook learning is the norm. We evaluated two approaches to helping students learn macroevolutionary topics. Treatment 1 is a laboratory for the software program EvoBeaker designed to teach students about evolutionary trees. We tested Treatment 1 among nine college-level biology classes and administered pre/posttests to assess learning gains. We then sought to determine whether the learning gains from Treatment 1 were comparable to those derived from an alternate hands-on treatment, specifically the combination of a prerecorded lecture on DVD and paper-based activity based on Goldsmith's Great Clade Race (Treatment 2). Comparisons of pre- and posttests among participants using either Treatment 1 or 2 show large learning gains on some misconceptions and skills beyond knowledge gained from reading standard textbook entries. Both treatments performed equivalently in overall learning gains, though both had areas where they performed better or worse. Furthermore, gains among students who used Treatment 1 representing a wide range of universities suggest that outcomes are potentially applicable to a variety of "real-world" biology classes.

INTRODUCTION

Evolutionary thinking is widely regarded as a lynchpin of modern biological science (Dobzhansky, 1973), but it can be an unintuitive and challenging subject to master. Much research has gone into understanding students' conceptual knowledge of evolutionary biology, with a particular emphasis on the mechanisms of natural selection (Bishop and Anderson, 1990, Ferrari and Chi, 1998, Anderson *et al.*, 2002). Additional researchers, such as Soderberg and Price (2003), have sought to develop and assess pedagogical materials to help secondary and postsecondary students better grasp these core concepts. Little research, however, has focused on students' conceptual understanding of macroevolution (Marco, 1993, Brewer, 1997a, Brewer, 1997b, Baum *et al.*, 2005). We set out to design and assess techniques for improving college students' understanding of macroevolutionary concepts, specifically construction and interpretation of evolutionary tree diagrams.

In a previous study (Meir *et al.*, 2007), the authors sought to identify common misconceptions and important skills in the area of "tree-thinking," the ability of students to read, interpret, construct, and use tree-diagrams showing the relationship between different species or populations. This paper identified four misconceptions and three key skills with which many college students have difficulty (Tables 1 and 2). The four misconceptions included:

DOI: 10.1187/cbe.07–01–0007
Address correspondence to: Judy Perry (jperry@mit.edu).
Conflict of Interest: Perry was paid by SimBiotic Software for her work on this project, but no longer had any connection with SimBiotic at the time of submission of this paper and no financial stake in the commercial success of the product. Meir, Maruca, and Stal are all currently employed by SimBiotic. Herron contracted with SimBiotic on this project and has a financial interest in EvoBeaker, as well as the textbook used for comparison within the "local" study.

Table 1. Descriptions and prevalence of skills (based on subset of pretest questions)

Skill	Description	Percent of population *lacking* this skill (n = 410)
S1: Reading Traits from Tree	Given a tree with trait transitions marked, the student is able to correctly deduce the traits of a particular species on the tree.	65%
S2: Deducing Ancestral Traits	Given the terminal taxa (species) on an evolutionary tree with sets of traits given for each extant species, the student is able to deduce the most likely traits shared by the common ancestor of those species.	84%
S3: Reconstructing Trees	Given a small set of extant species with no convergence or loss of derived traits, and the common ancestor of these species, the student is able to draw an evolutionary tree showing the most likely phylogeny.	70%

(*M1*) *Confusion with the direction time flows on the evolutionary tree (or "Incorrect Mapping of Time")*. Many students thought, for instance, that the oldest species was shown at the left-hand leaf of a vertically drawn tree (base at the bottom, extant species on top), and progressively more recent species were shown from left to right across the tips of the branches. They did not recognize that all species shown at the top of the tree were currently living species and equally recent.

(*M2*) *Thinking that species that were drawn closer together at the tips of the tree are more closely related (or "Tip Proximity Indicates Relationship")*. Students did not recognize that branches can be rotated around branch points.

(*M3*) *Using the number of branch points between two species (which we refer to as "nodes" herein) to determine how closely related the species were (or "Node Counting")*. For instance, if there were many speciation events in a short time on one side of the tree, creating many nodes in the tree, students would say that those species were more distantly related to each other than species in another area of the tree where speciation rates had been lower.

(*M4*) *Confusing straight lines on the tree with ancestry (or "Straight Line Equals No Change")*. Many students would start at the bottom ancestral node of the tree and move up the left hand branch until they reached the tip, thinking that the species at that tip was the ancestor of all the others.

The three "tree-thinking" related skills the authors found missing among students were (S1) the ability to deduce the traits of an extant species from a tree with trait changes marked (or "Reading Traits from Tree"), (S2) the ability to deduce the traits of an ancestral species from the traits of all the extant species (or "Deducing Ancestral Traits"), and (S3) the ability to reconstruct the most likely tree for a small set of species and their traits (or "Reconstructing Trees").

The prevalence of each of these missing skills and misconceptions among students is shown in Tables 1 and 2 respectively, and is discussed in more detail elsewhere (Meir *et al.*, 2007). As the data show, at least one-fourth, and in many cases more than half, of the college students whom the authors surveyed had trouble with each misconception or skill (Tables 1 and 2). These students all had exposure to lectures in evolutionary biology, and many were in upper-level evolution classes (see below), clearly showing the need for better tools for teaching these topics.

Dispelling misconceptions and building critical skills are tasks best accomplished through active learning. The didac-

Table 2. Descriptions and prevalence of misconceptions (based on subset of pretest questions)

Misconception	Description	Percent of population exhibiting this misconception (n = 410)
M1: Incorrect Mapping of Time	Confusion about spatial representation of the flow of time on an evolutionary tree.	31%
M2: Tip Proximity Indicates Relationship	Assumption that species that were drawn closer together at the tips of the tree were more closely related to each other than those drawn farther apart.	24%
M3: Node Counting	Assumption that the number of nodes crossed in tracing a path between two species on a tree indicated how closely related they were.	38%
M4: Straight Line Equals No Change	Interpretation that species at the top of a ″straight line″ (i.e., one lineage as straight line extending from common ancestor deep into the tree to one of the tips, with another species branching off) is the common ancestor.	40%

tic transmission model, wherein students passively receive information, is largely inadequate when seeking to dispel deeply harbored misconceptions (McDermott, 2001, Roy, 2003). A more effective teaching strategy involves actively presenting students with opportunities to critically observe where their current ideas can (or cannot) explain phenomena, allowing them to modify and test new ideas, and providing opportunities to self-assess whether their new conceptions now appear sufficient. This is likely best done in a small class with a skilled instructor using hands-on tasks such as comparing and classifying skeletons, or tools like The Great Clade Race (Goldsmith, 2003). However, small classes with instructors skilled in evolutionary biology are not always available to students, especially in introductory biology classes.

As an alternative, the authors wrote EvoBeaker (Meir *et al*., 2005), a computer-based interactive software package, to help teach evolutionary biology. EvoBeaker includes a laboratory called "Flowers and Trees," which addresses basic tree-thinking skills. Here we assess the ability of this laboratory to teach students about evolutionary trees and overcome their misconceptions. We also compare EvoBeaker's "Flowers and Trees" lab's efficacy relative to a combination of prerecorded lecture and hands-on interactive activities guided by a skilled instructor.

Description of EvoBeaker's "Flowers and Trees" Lab

EvoBeaker is an interactive, desktop-based software package designed to model concepts in micro- and macroevolution, from the assumptions behind natural selection to how to interpret an evolutionary tree to reconstructing phylogenies from sequence data. EvoBeaker teaches by asking students to do realistic experiments on simulated biological systems. As a companion to the EvoBeaker software, students receive a printed lab workbook that provides background narrative, instructions for each activity, and questions for the students to answer as they work. More information about EvoBeaker, and samples of all materials discussed in this paper, are available from SimBiotic Software (www.simbio.com). An example of EvoBeaker software is shown in Figure 1.

Of the six available labs for EvoBeaker, the "Flowers and Trees" lab primarily focuses on evolutionary trees and reconstructing phylogenies, topics typically covered in both college-level introductory biology and introductory evolution classes. In "Flowers and Trees," students are introduced to evolutionary trees through the diversification of Columbine flowers (*Aquilegia*). The lab begins by showing populations of plants splitting and diverging from each other, with an evolutionary tree growing alongside. The next several sections ask students to manually "mutate" and "split" populations of plants, and predict what the evolutionary tree will look like after each event. They also manipulate the evolutionary tree itself by, for instance, flipping branches, to see how the way the tree is drawn does (or does not) relate to its interpretation. The final sections have the mutations and population splits happening invisibly to the students, and then ask the students to sample each ending population and reconstruct the phylogeny by comparing traits. In all exercises, students are asked to predict the results of their actions, and then can see whether they predicted correctly.

DO EVOBEAKER'S "FLOWERS AND TREES" LAB TASKS HELP STUDENTS UNDERSTAND "TREE-THINKING"?

Materials and Methods

Sample and Procedures. Professors at nine U.S. colleges or universities volunteered to use the EvoBeaker "Flowers and

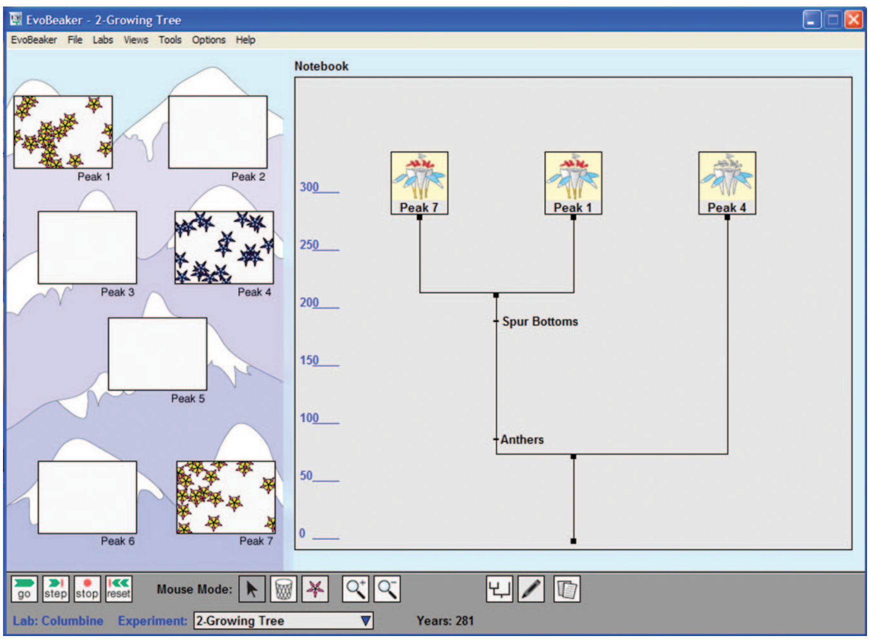

Figure 1. Screenshot from "Flowers and Trees" lab in EvoBeaker. The simulation initially contained a population of flowers on one of the seven mountain peaks, shown on the left. As the model runs, there is occasionally a colonization event from one of the populated peaks to an empty peak. There are also occasionally mutation events that spread through the population on a peak, which are shown by a change in the coloring and shape of those flowers. The colonizations and mutations are shown on the tree at right, which grows dynamically upwards as time progresses. The relations between the populations are shown by the branching, and the current phenotype in each population is shown in the picture at the tip of the branch. In this case, seeds from Peak 4 colonized either Peak 7 or Peak 1, which then colonized the other of those. Before the second colonization event, the Peak 7/Peak 1 population had a mutation that changed the anther color to red, and a second mutation making the spur bottoms longer.

Trees" lab as part of their general biology or evolutionary biology classes, with EvoBeaker software and labs made available to them free of charge for the duration of the study. Their students (n = 262) (henceforth referred to as "remote" students) volunteered to participate in the study, which included taking the pretest before the lab, completing EvoBeaker's "Flowers and Trees" lab, and finally taking the posttest. The majority of remote students conducted the lab as a pair with a classmate (although all testing was done individually). Students were not financially compensated for their time.

It was our intent that this sample would represent a varied and realistic usage of EvoBeaker's "Flowers and Trees" lab as it might be integrated into typical college-level courses covering relevant macroevolutionary content. As such, the amount of in-class time covering curricular materials varied from class to class. Specifically, three of the nine participating professors covered some material on evolutionary trees before the pretest, and three professors (two included above and one additional professor) covered some tree-thinking material (other than the lab) before the posttest. Two other professors did not cover any tree-thinking material during lectures, with EvoBeaker's "Flowers and Trees" lab being the only exposure these students had to these concepts before the posttest. We did not receive this information from the remaining three professors.

To preserve anonymity, no participating colleges or universities are identified in this paper by name. Schools ranged in size from three small schools (with approximately 1100 to 1400 undergraduates), to three midsized (with approximately 5000 to 7000 undergraduates), to three large school (with more than 15,000 undergraduates). Class sizes ranged from small seminars to large lectures. Three classes were introductory level (200s), and the remaining six classes were upper level (300+). All classes had "evolution" in the title of the course.

Data Collection. Professors were instructed to give the pretest no more than one week before covering EvoBeaker's "Flowers and Trees" lab materials, and to give the posttest no more than one week after the lab/material. Pre- and posttest instruments were identical in content, length, and difficulty, but used different examples for each question asked (i.e., different sets of species on pre- and posttests). Both the pre- and posttests contained 21 questions, several of which had multiple parts, allowing a maximum score of 33 points. The majority of students also voluntarily self-reported their gender. Pre- and posttests were given in proctored, controlled environments. However, tests were not scored by professors and did not count toward students' final course grades.

Several test questions used in this study are reprinted in Meir *et al*. (2007). Samples of all other materials used in this study, including pre- and posttests, are available by request from SimBiotic Software (info@simbio.com).

Results

All nine classes showed improvement between pre- and posttest scores. Overall, we saw a 35% improvement from pre- to posttests across all students (Figure 2), which was highly significant (paired two-tailed *t* test assuming equal variance, $p < .0001$). The average changes in pre/posttest scores for individual classes ranged from a minimum delta of + 5% to a maximum delta of + 62%.

Among these nine classes, there were no significant differences (two sample two-tailed *t* test assuming equal variance, corrected for multiple comparisons) between the gains among students in lower-level (200 level) classes (n = 171, 40.1% gain) and when compared with learning gains among students in upper-level (300 and 400 level) classes (n = 91, 26.7% gain).

A subset of classes were known to have received some relevant prior instruction, and as such we compared pre/post learning gains between students known to have had prior instruction with those who did not. (Information about prior instruction was not available for all students, and those for whom this information was not available (n = 53) were excluded from both the "prior instruction" and "no prior instruction" groups.) We found learning gains between these two groups (those with and without prior instruction) significantly different from one another (two sample two-tailed *t* test assuming equal variance, corrected for multiple comparisons, $p < .05$). Specifically, students with relevant prior instruction (n = 161) demonstrated large (36.9%) and highly significant (paired two-tailed *t* test, $p < .0001$) average learning gains between pre- and posttest scores after using EvoBeaker's "Flowers and Trees" lab (Figure 3). Among the subset of students known *not* to have had relevant prior instruction (n = 48), learning gains were also substantial (30.6%) and significant (paired two-tailed *t* test, $p < .0001$), but less than those among students known to have had prior instruction, suggesting that professors assigning EvoBeaker "Flowers and Trees" lab may want to consider delaying use of the lab until they have introduced some relevant material to their students.

The majority of students self-reported gender. When segmented by gender, there were no statistically significant

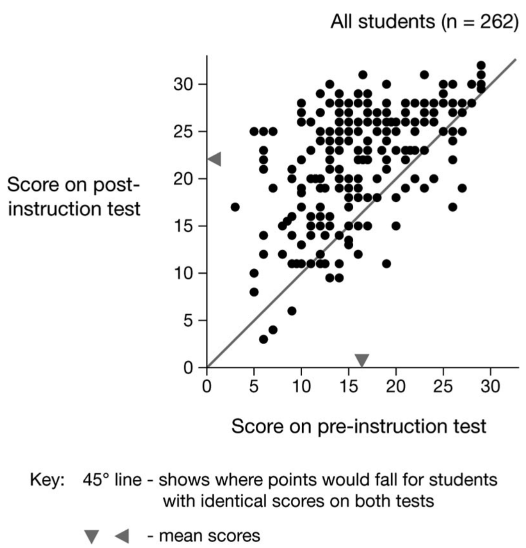

Figure 2. Scatterplot showing pre- and posttest scores among remote students who used EvoBeaker's "Flowers and Trees" lab. Arrows along *x* and *y* axes indicate average pre- and posttest scores.

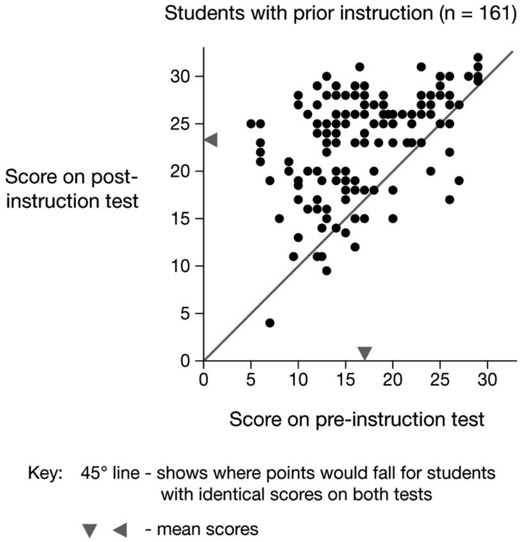

Figure 3. Scatterplot showing pre- and posttest scores among subset of remote students who used EvoBeaker's "Flowers and Trees" lab, and were known to have received relevant in-class instruction before pretest. Arrows along *x* and *y* axes indicate average pre- and posttest scores.

differences in average test score delta between males (n = 89) and females (n = 147).

In addition to examining overall performance, we sought to examine the lab's efficacy at teaching the missing key skills and overcoming common misconceptions found in our previous study. Subsets of pre- and posttest questions were aligned with each of the skills and misconceptions listed above (see Meir *et al*., 2007, for details). Students' scores on these question subsets were then averaged to determine students' improvement or decline with respect to these particular areas. Figures 4 and 5 summarize pre- and posttest scores for each of these areas.

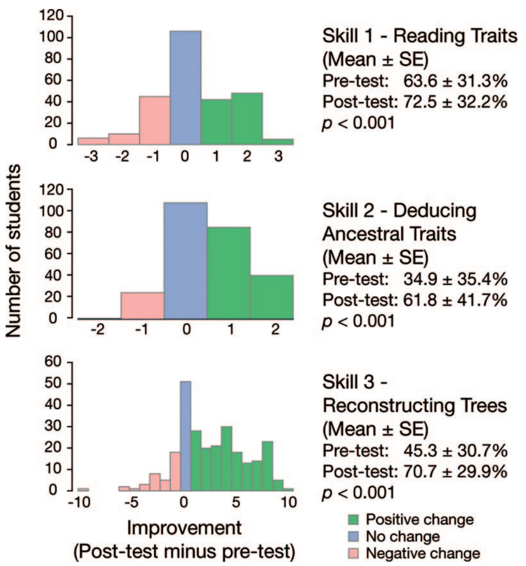

Figure 4. Histograms of the delta (posttest minus pretest scores) for subset of test items relevant to each skill among remote students who used EvoBeaker's "Flowers and Trees" lab. The average pre- and posttest scores for these subsets, along with SE, are also included. Differences between pre- and posttests were significant at the $p < .001$ level (paired, two-tailed *t* tests).

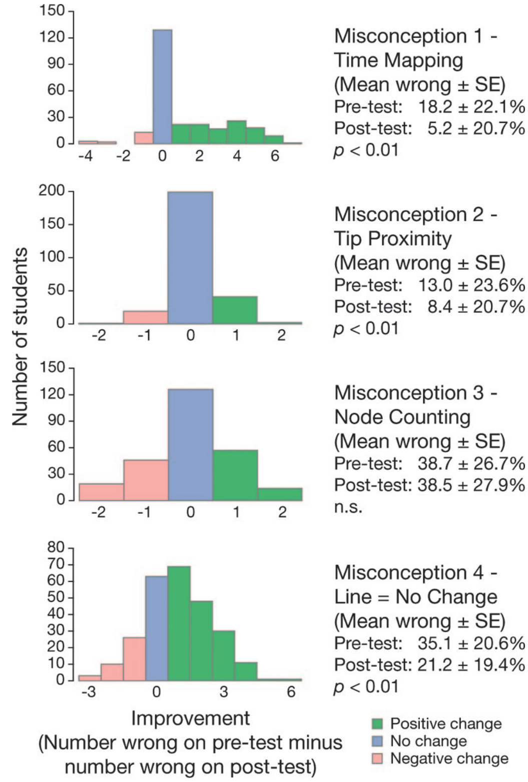

Figure 5. Histograms of the delta (pretest minus posttest scores) for subset of test items relevant to each misconception among remote students who used EvoBeaker's "Flowers and Trees" lab. Note that a positive delta indicates the *reduction* of that misconception. The average pre- and posttest scores for these subsets, along with SE, are also included. Again, note that a lower average indicates a reduction in that misconception. Differences between pre- and posttests were significant at the $p < .01$ level (paired, two-tailed *t* tests), with the exception of M3: Node Counting, where there was no change.

Students demonstrated statistically significant improvements (paired two-tailed *t* test, $p < .001$) on all three skills (Figure 4), with the most dramatic improvement on (S2) Deducing Ancestral Traits. Students also showed significant improvements (paired two-tailed *t* test, $p < .01$) on three of the four misconceptions (Figure 5). However, students showed no improvement on (M3) Node Counting.

HOW DOES EVOBEAKER'S "FLOWERS AND TREES" LAB COMPARE TO LECTURE/PAPER-BASED PHYLOGENETICS CURRICULA?

Having seen improvements among students who used EvoBeaker's "Flowers and Trees" lab, we wanted to know how the computer-based lab compared with an alternate

Table 3. Average overall pre/post scores of local students comparing EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups*

	Average overall pretest scores (SD)	Average overall posttest scores (SD)	Delta (% ±)
EvoBeaker's "Flowers and Trees" lab (n = 27)	0.491 (0.191)	0.619 (0.200)	0.128 (+26%)
Clade-Race+Lecture (n = 22)	0.477 (0.121)	0.680 (0.173)	0.203 (+43%)

* Average scores have been scaled from 0 to 1 to facilitate comparisons.

hands-on treatment, specifically a prerecorded DVD of a lecture delivered by an expert teacher coupled with a hands-on paper-and-pencil activity. We also wanted to know to what extent both hands-on activities demonstrated learning gains beyond knowledge gained by solely reading relevant textbook excerpts. To do this, we conducted an additional round of testing in which we randomly assigned students to do either Treatment 1 (EvoBeaker's "Flowers and Trees" lab) or Treatment 2 (viewing the DVD and completing the Clade-Race activities). Both treatments covered similar content and process skills.

Materials and Methods

Sample and Procedures. College students (n = 57) were recruited as volunteers from various Boston-area 4-yr colleges and were paid for their time. All students were undergraduates who had taken a college-level introductory biology course within the past 3 yr which covered evolutionary biology for at least one lecture.

Students in this study, whom we will refer to as "local students," participated in two separate sessions. During the 1-h Part I session, local students (n = 57) read an assigned textbook excerpt (Freeman and Herron, 2004, pp. 47–49) covering the basic aspects of tree-thinking, followed by the pretest, which they completed individually under controlled testing conditions. Based on participants' pretest scores, students were categorized into lower, middle, and upper levels. One local student who scored perfectly on his pretest was disqualified from the study, because his potential for improvement was minimal. Among the 56 remaining qualified students, we were able to schedule a Part II session with 49 students, which lasted approximately 2 h, during which random stratified sampling was used to assign roughly even numbers of students from low, middle, and high levels to either the Clade-Race+Lecture group (n = 22) or EvoBeaker's "Flowers and Trees" lab group (n = 27). All students completed EvoBeaker's "Flowers and Trees" lab or Clade-Race+Lecture activity individually in a proctored setting. We have only included pre/post scores among students who completed both Part I and Part II in this article.

EvoBeaker's "Flowers and Trees" Lab Group. Local students assigned to the EvoBeaker group completed the "Flowers and Trees" lab described previously.

Clade-Race+Lecture Group. Local students assigned to the Clade-Race+Lecture group individually viewed a prerecorded lecture on DVD, with the ability to pause or review material as necessary. Jon Herron (an author on this paper, lecturer in biology, and author of a leading textbook on evolutionary biology) developed and delivered the lecture, which included subject matter he has taught in evolutionary biology courses for many years. The lecture was videotaped in front of a live student audience (who were not themselves study participants, but who had taken college-level introductory biology and were paid for their time, hereafter referred to as the "audience"). At various points throughout the lecture, the DVD prompted students with instructions to pause and complete pencil-and-paper exercises. Some of these exercises were drawn from Goldsmith's Great Clade Race activity (2003) and others were designed by Herron.

The lecture was book-ended with examples of real-world puzzles that have been solved by reconstructing phylogenies, such as determining the origins of HIV. The lesson itself began with The Great Clade Race. This pencil-and-paper exercise engages students in cladistic reasoning in an easily understandable context that is free of evolutionary biology and the jargon that comes with it. (The 'ninth card' activity, designed by Goldsmith for advanced classes, was omitted.) After the students had completed the Clade Race, the lecture gave an illustrated presentation on how evolutionary trees represent the history of divergence and change in groups of populations or species, on how descent with modification from common ancestors produces nested sets of taxa that can be defined by shared derived traits, and on how shared derived traits can be used to reconstruct evolutionary history. In the final pencil-and-paper activities, designed by Herron, students reconstructed the evolutionary history of seven fictitious lizard species. The trees they drew during this exercise served as the basis for a discussion of how the same evolutionary history can be described with superficially different phylogenies, and of what it means to say that species are, or are not, closely related. At the end of the lecture, members of the audience were given a chance to ask questions. These questions and the instructor's responses were included at the end of the DVD recording.

Data Collection. Local students took pretests no more than one month before using either the Clade-Race+Lecture or EvoBeaker's "Flowers and Trees" lab treatment. Posttests were given immediately after the treatment. Pre- and posttest instruments were identical to those used in the first study and were given in a controlled environment.

Results

Table 3 shows the average pretest score versus average posttest score for both the EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups. Average overall test scores improved for both EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups. However, there was no statistically significant difference found between the EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups at the $p < .05$ level. Additionally, there were no statistically significant differences found between male and female populations.

As in the remote students, we sought to examine both interventions' efficacy at teaching the missing key skills and overcoming the common misconceptions found in our previous study. Overall, both treatments (Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab) were effective in improving skills (Table 4). The Clade-Race+Lecture activity was significantly more effective at teaching the (S2) Deducing Ancestral Traits skill than EvoBeaker's "Flowers and Trees" lab activity, but there were no significant differences between the groups with respect to the other two skills. Likewise, both interventions were shown to reduce or virtually eliminate the misconceptions (Table 5), with no statistically significant differences between Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab groups' changes in misconceptions at the $p < .05$ level when corrected for multiple comparisons. For example, (M1) Incorrect Mapping of Time was significantly reduced among both Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab populations. Similarly, whereas (M2) Tip Proximity Indicates Relationship was not widespread among the local students (as demonstrated by their pretest scores), both interventions were effective at virtually eliminating this misconception. Both EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture interventions were effective at reducing, but not eliminating, the other two misconceptions. Figure 6 shows histograms displaying the delta for each skill and misconception among Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab groups. Although the histograms show improvement among many students, a minority (in both Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab groups) did the same or worse after the treatment, suggesting that the interventions were not universally effective. Overall, however, the computer laboratory and the paper-based activity/lecture appear to be roughly equivalent in their ability to teach tree-thinking skills and overcome common tree-thinking misconceptions.

Discussion

As evidenced by their pretest scores, many students initially lacked critical skills necessary for sound understanding of phylogenetic trees and related evolutionary biology con-

Table 4. Average pre/post scores for key skills questions among local students comparing EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups[a]

Skill	Group	Average pretest scores (SD)	Average posttest scores (SD)	Delta (% ±)
S1: Reading Traits from Tree	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.654 (0.313)	0.691 (0.345)	0.037 (+6%)
	Clade-Race+Lecture (n = 22)	0.515 (0.321)	0.667 (0.291)	0.152 (+29%)
S2: Deducing Ancestral Traits[b]	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.259 (0.321)	0.500 (0.439)	0.241 (+93%)
	Clade-Race+Lecture (n = 22)	0.227 (0.255)	0.773 (0.369)	0.545 (+240%)
S3: Reconstructing Trees	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.489 (0.285)	0.626 (0.341)	0.137 (+28%)
	Clade-Race+Lecture	0.427 (0.273)	0.732 (0.301)	0.305 (+71%)

[a]Scores have been scaled between 0 and 1 to facilitate comparisons. Note that a higher score indicates a higher measure of that skill.

[b]Denotes statistical significance at the $p < .05$ level when corrected for multiple comparisons between EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups.

Table 5. Average pre/post scores for key misconceptions questions among local students comparing EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups[a,b]

Misconception	Group	Average pretest scores (SD)	Average posttest scores (SD)	Delta (% ±)
M1: Incorrect Mapping of Time	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.226 (0.236)	0.056 (0.115)	-0.170 (-75%)
	Clade-Race+Lecture (n = 22)	0.123 (0.169)	0.055 (0.15)	-0.068 (-56%)
M2: Tip Proximity Indicates Relationship	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.130 (0.263)	0.093 (0.242)	-0.037 (-29%)
	Clade-Race+Lecture (n = 22)	0.091 (0.197)	0 (0)	-0.091 (-100%)
M3: Node Counting	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.420 (0.286)	0.259 (0.267)	-0.160 (-38%)
	Clade-Race+Lecture (n = 22)	0.424 (0.234)	0.348 (0.281)	-0.076 (-18%)
M4: Straight Line Equals No Change	EvoBeaker's "Flowers and Trees" lab (n = 27)	0.365 (0.199)	0.291 (0.204)	-0.074 (-20%)
	Clade-Race+Lecture (n = 22)	0.357 (0.175)	0.273 (0.192)	-0.084 (-24%)

[a]Scores have been scaled between 0 and 1 to facilitate comparisons. Note that a lower score indicates a smaller measure of that misconception.

[b]There were no statistically significant differences between EvoBeaker's "Flowers and Trees" lab and Clade-Race+Lecture groups at the $p < .05$ level when corrected for multiple comparisons.

cepts (Meir *et al.*, 2007). Given this prevalence of misconceptions and lack of skills, there were statistically significant gains among students using the "Flowers and Trees" lab, which both improved skills and reduced or virtually eliminated misconceptions in nearly all areas.

EvoBeaker's "Flowers and Trees" lab and the Clade-Race+Lecture treatment were roughly equivalent in their efficacy. There were few statistically significant differences among local students who used EvoBeaker's "Flowers and Trees" lab versus those who used the Clade-Race+Lecture in terms of improved skills and reduced misconceptions. Given both treatments' emphases on active learning, and specifically on practice at reconstructing and interpreting evolutionary trees, we feel that this demonstrates the efficacy of this pedagogical technique. This is supported by the fact that students read subject-specific textbook materials just before taking the pretest. Although students did spend a relatively shorter amount of time reading (approximately 15 min) than they did doing the hands-on software or paper-and-pencil tasks (approximately 1 h), the prevalence of misconceptions and lack of skills on the pretests (after reading the chapter) suggests that merely reading about evolutionary trees does not sufficiently prompt students to consider and rethink their conceptual models; active, hands-on tasks did.

Among the three skills that we specifically measured, it appears that the most difficult skill to impart to students was (S1) Reading Traits from Tree. Both remote and local students who used EvoBeaker's "Flowers and Trees" lab or the Clade-Race+Lecture demonstrated relatively little improvement in this area. The other two skills saw large improvements among both groups. Although the differences between Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab groups were not statistically significant, it appears that skills may have improved slightly more among students who used the Clade-Race+Lecture activity than

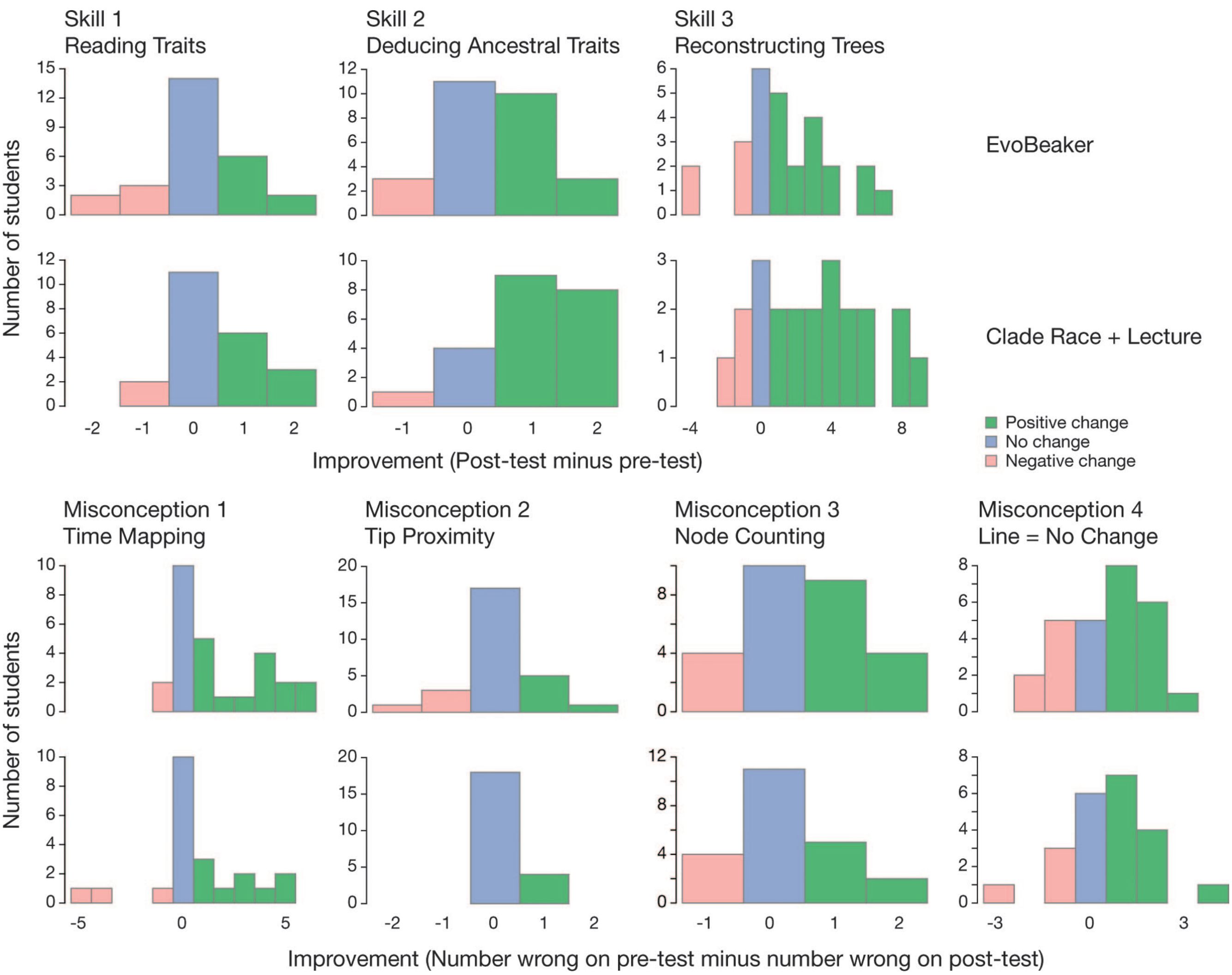

Figure 6. Histograms of the delta for skills (posttest minus pretest scores) and misconceptions (pretest minus posttest scores) for subsets of questions relevant to each skill and misconception for students who used EvoBeaker's "Flowers and Trees" lab and those who used Clade-Race+Lecture treatment.

among students who used EvoBeaker's "Flowers and Trees" lab. This suggests that EvoBeaker's "Flowers and Trees" lab simulation could benefit from improvements in these areas, perhaps by adding more exercises where students are tasked with labeling traits on trees. The Clade-Race+Lecture activity placed greater emphasis on developing these particular skills, suggesting that more time devoted to those areas in the lab will lead to improved learning.

One of the four misconceptions, (M3) Node Counting, appeared to be particularly strongly held, with little or no improvement among remote students who used EvoBeaker's "Flowers and Trees" lab. We noticed this while designing the "Flowers and Trees" lab as well and piloted several different versions of the lab to ameliorate this, but with little success. This is a second area of the lab that would likely benefit by expanding the exercises devoted to it. One potential exercise would be to have one branch of a tree undergo very rapid speciation, while a second branch from the same ancestor goes through little or no speciation, and ask questions about relatedness on this tree. Although the current lab has something like this, it is perhaps not dramatic enough for students to reconsider their misconceptions.

Like EvoBeaker's "Flowers and Trees" lab, the Clade-Race+Lecture activity was generally successful in improving students' understanding of evolutionary trees. We attribute this to two main factors: First, as with EvoBeaker's "Flowers and Trees" lab, the content covered by Clade-Race+Lecture was specifically designed to target the skills and misconceptions identified in our previous study (Meir *et al.*, 2007). Second, the Clade-Race+Lecture was designed to synthesize the experiences of a small class led by a skilled instructor with that of a larger class that included a well-designed, hands-on lab activity. We strove to make the lecture portion of the video feel somewhat interactive, by encouraging the videotaped audience to ask the instructor questions (which were included in the edited version viewed by participants in the Clade-Race+Lecture group). However, because subjects in the Clade-Race+Lecture group did not have real-time access to the lecturer, they could not ask their own additional questions or seek clarification. We expect that a skilled instructor helping students in real time would improve students' scores and, more importantly, their conceptual understanding of evolutionary trees, in both the Clade-Race+Lecture and EvoBeaker's "Flowers and Trees" lab treatments.

In conclusion, active, hands-on tasks appear to help students gain a better understanding of "tree-thinking" beyond knowledge gained merely by reading traditional textbooks. The combination of diagramming tasks, predictions, and reconstruction of trees, whether computer based or paper based, helped students partially overcome many of the common misconceptions relating to reading evolutionary trees and taught them how to reconstruct trees themselves. Results from this study demonstrate that computer-based simulations or active-learning curriculum such as The Great Clade Race can improve student learning, suggesting the potential for widespread dissemination where more individualized instruction from subject matter experts is not available.

ACKNOWLEDGMENTS

Thanks to Eric Klopfer, Ilona Holland, Mark Hartman, and Emily Franklin, and to the group of biologists at the University of North Carolina, Chapel Hill, who reviewed our tests. Eric Klopfer generously donated lab space to conduct student interviews. Thanks also to professors and students who volunteered to assist in our study (names withheld to preserve anonymity). This work was partially funded by grant #0341202 from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. EvoBeaker is a trademark of SimBiotic Software for Teaching and Research, Inc.

REFERENCES

Andersen, D. L., Fisher, K. M., and Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. J. Res. Sci. Teach. *39*(10), 952–978.

Baum, D. A., Smith, S. D., and Donovan, S.S.S. (2005). The tree-thinking challenge. Science *310*, 979–980.

Bishop, B. A., and Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. J. Res. Sci. Teach. *27*(5), 415–427.

Brewer, S. (1997a). Constructing student problems in phylogenetic tree construction. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. March 24–28, 1997.

Brewer, S. (1997b). A model of desired performance in phylogenetic tree construction for teaching evolution. Paper presented at the 70th Annual Meeting of the National Association for Research in Science Teaching, Oak Brook, IL, March 21–24, 1997.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. Am. Biol. Teach. *35*, 125–129.

Ferrari, M., and Chi, M.T.H. (1998). The nature of naïve explanations of natural selection. Int. J. Sci. Educ. *20*(10), 1231–1256.

Freeman, S., and Herron, J. C. (2004). Evolutionary Analysis, 3rd Edition, Upper Saddle River, NJ: Prentice Hall.

Goldsmith, D. (2003). The Great Clade Race: presenting cladistic thinking to both biology majors & general science students. Am. Biol. Teach. *25*(9), 679–682.

Marco, O. B. (1993). A simple model to think about the evolutionary rate in macroevolution. Am. Biol. Teach. *55*(7), 424–29.

McDermott, L. C. (2001). Oersted Medal Lecture 2001, Physics education research—the key to student learning. Am. J. Phys. *69*(11), 1127–1137.

Meir, E., Herron, J. C., Maruca, S., Stal, D., and Kingsolver, J. (2005). EvoBeaker 1.0, Ithaca, NY: SimBiotic Software. www.simbio.com.

Meir, E., Perry, J., Herron, J. C., and Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. Am. Biol. Teach. *69*(7), 71–76.

Roy, H. (2003). Studio vs interactive lecture demonstration – effects on student learning. Bioscene *29*(1), 3–6.

Soderberg, P., and Price, F. (2003). An examination of problem-based teaching and learning in population genetics and evolution using EVOLVE, a computer simulation. Int. J. Sci. Educ. *25*(1), 35–55.