ORIGINAL SCIENTIFIC ARTICLE

Addressing Undergraduate Student Misconceptions about Natural Selection with an Interactive Simulated Laboratory

Joel K. Abraham · Eli Meir · Judy Perry ·
Jon C. Herron · Susan Maruca · Derek Stal

Abstract Although evolutionary theory is considered to be a unifying foundation for biological education, misconceptions about basic evolutionary processes such as natural selection inhibit student understanding. Even after instruction, students harbor misconceptions about natural selection, suggesting that traditional teaching methods are insufficient for correcting these confusions. This has spurred an effort to develop new teaching methods and tools that effectively confront student misconceptions. In this study, we designed an interactive computer-based simulated laboratory to teach the principles of evolution through natural selection and to correct common student misconceptions about this process. We quantified undergraduate student misconceptions and understanding of natural selection before and after instruction with multiple-choice and open-response test questions and compared student performance across gender and academic levels. While our lab appeared to be effective at dispelling some common misconceptions about natural selection, we did not find evidence that it was as successful at increasing student mastery of the major principles of natural selection. Student performance varied across student academic level and question type, but students performed equally across gender. Beginner students were more likely to use misconceptions before instruction. Advanced students showed greater improvement than beginners on multiple-choice questions, while beginner students reduced their use of misconceptions in the open-response questions to a greater extent. These results suggest that misconceptions can be effectively addressed through computer-based simulated laboratories. Given the level of misconception use by beginner and advanced undergraduates and the gains in performance recorded after instruction at both academic levels, natural selection should continue to be reviewed through upper-level biology courses.

Keywords Active learning · Evolution · Education · Misconceptions · Natural selection

J. K. Abraham · E. Meir · S. Maruca · D. Stal
SimBiotic Software for Research and Teaching,
Ithaca, NY 14850, USA

J. K. Abraham (✉) · E. Meir · J. Perry
Teacher Education Program,
Massachusetts Institute of Technology,
Cambridge, MA 02139-4307, USA
e-mail: abraham@simbio.com

J. C. Herron
Department of Biology, University of Washington,
Seattle, WA 98195-1800, USA

Introduction

Evolutionary theory, the unifying foundation for most biological concepts, is also one of the most commonly misunderstood areas in biology. It is, therefore, imperative to identify the source of student confusions about evolution and to develop effective teaching tools that can correct them (National Research Council 1996; Novak 2002; Modell et al. 2005; Wescott and Cunningham 2005). Although cultural and political viewpoints affect student understanding (Almquist and Cronin 1988; Sinclair et al. 1997; Hokayem and BouJaoude 2008), even students who claim to accept evolutionary theory often demonstrate little understanding of its basic principles (e.g., Bishop and Anderson 1990; Demastes et al. 1995). These misunderstandings are instead often directly linked to students having their own incorrect conception of the functioning of the world (known as misconceptions), which prevents

them from being able to use scientifically accepted concepts in thinking through some scientific problems (Greene 1990; Ferrari and Chi 1998; but see Hamza and Wickman 2008).

The realization that students and the general public largely misunderstand evolutionary biology principles has spurred educators and researchers to focus on the identification and correction of common misconceptions (Brumby 1984; Bishop and Anderson 1990; Anderson et al. 2002). Much of this work has focused on confusions about the process of natural selection (e.g., Bishop and Anderson 1990; Settlage 1994; Jensen and Finley 1996; Nehm and Reilly 2007). These, and other studies, have identified some widespread misconceptions about principles of natural selection (Table 1). For example, students often ascribe agency to trait shifts in populations, as if a yearning or need of an organism will cause a trait to change, even within that organism′s lifetime. A related misconception is that the origin of genetic variation is nonrandom and is instead the result of external pressures. Many students also struggle with the idea of evolution as changes in the frequency of an allele or trait within a population, thinking instead that populations change their traits gradually as a whole.

Although many of these misconceptions about natural selection are formed early in life (Evans 2000; Samarapungavan and Wiers 1997; Beardsley 2004), students in high school, college, and even graduate school often retain their misconceptions despite receiving formal training in biology (Brumby 1984; Lawson and Thompson 1988; Bishop and Anderson 1990; Ferrari and Chi 1998; Crawford et al. 2005; Nehm and Reilly 2007; Robbins and Roy 2007). These confusions can inhibit understanding of related biological concepts (Greene 1990; Jensen and Finley 1996) and likely limit students′ ability to explain and comprehend other relevant phenomena such as antibiotic resistance in bacteria, impacts of climate change, or biological engineering.

Traditional teaching methods are not well suited for identifying and correcting misconceptions (Morrison and Lederman 2003; Nelson 2008). However, efforts to develop new pedagogical tools for teaching evolutionary biology have achieved some success. Bishop and Anderson (1990) found moderate improvement in undergraduate student understanding of natural selection after intervention with a unit on that subject designed to dispel misconceptions, while Demastes et al. (1995) found minimal improvement using the same instructional methods. However, high school students showed greater gains, perhaps due to differences in the presentation of instructional materials. Robbins and Roy (2007) developed an inquiry-based teaching unit that yielded drastic improvements in college student explanations and overall acceptance of modern evolutionary theory. Recently, interactive computer-based simulations designed to support student and teacher training in biology have also successfully improved learner understanding of biological concepts and dispelled common learner misconceptions (Crawford et al. 2005; Meir et al. 2005a; Kara and Yesilyurt 2007; Perry et al. 2008). This approach may prove to be particularly useful, as it can increase student access to expert knowledge and offer a greater flexibility in teaching situations. Additionally, as a platform through which students can create experiments and modify parameters, interactive simulations offer an opportunity for students to correct their own misconceptions (Crawford et al. 2005; Perry et al. 2008).

In this study, we sought to develop an interactive computer-based simulation that effectively addresses student misconceptions about natural selection. We first used the literature to identify common misconceptions about the process of natural selection held by undergraduate biology students. We used this list to direct the development of a virtual laboratory (henceforward, Darwinian Snails Lab) within the EvoBeaker software package (Meir et al. 2005b)

Table 1 Description, source, and frequency of student use of misconceptions about natural selection in pretest open-response questions

Misconceptions		Description	Reference	# Pretests (N=338)
MC1	Willful Change	Organisms willfully change their traits or the traits of their offspring	e.g. Bishop and Anderson 1990	125 (37%)
MC2	Directed Variation	Variation arises due to environmental or selection pressure	e.g. Bishop and Anderson 1990	152 (45%)
MC3	Intra-generational Change	Trait change occurs within a generation	e.g. Brumby 1984	53 (16%)
MC4	Population Change	Populations change their traits together as a whole	e.g. Bishop and Anderson 1990	68 (20%)
MC5	Inheritable Phenotypic Change	Acquired changes to phenotype of parent are inherited genetically by offspring	e.g. Lawson and Thompson 1988	4 (1.2%)
MC6	Variation through Drift	Genetic drift can create trait variation	e.g. Nehm and Reilly 2007	6 (1.8%)
MC7	Fitness=strength	Fitness is directly related to physical strength	e.g. Bishop and Anderson 1990	1 (0.3%)
MC8	Dominance=Selective Advantage	Dominant alleles are always selectively advantageous	e.g. Nehm and Reilly 2007	1 (0.3%)

to teach natural selection principles and to design the test items on the pre- and post tests constructed for this study. We then conducted a summative assessment of student knowledge before and after using the Darwinian Snails Lab to determine the efficacy of that exercise at correcting misconceptions and teaching the principles of natural selection. In addition to studying the lab itself, we also report results on differences in learning between introductory versus advanced biology students and on differences in information gained from the several types of questions included on our tests.

Methods

Description of Subject Population and Data Collection

We recruited 637 students from 25 two- and four-year colleges and universities for this study. Students were 18 years of age or older and enrolled in introductory and upper-level biology courses. Henceforward, these students are referred to as "beginner" and "advanced," respectively (beginner=170, advanced=426, unreported=43). Course instructors volunteered to use the software in their courses; while some instructors had used previous versions of the software, others were newly recruited contacts. In exchange for their help, we provided the software to the instructors and students at no charge. However, instructors received no form of monetary payment. We directed instructors to give the pretest to their students no earlier than one week before using the Darwinian Snails Lab. Students were assigned the lab during the laboratory section of their biology courses and worked alone or with a partner. Within a week of completing the exercise, students were given the post test. A participant profile sheet was attached to each pretest, in which subjects were asked their gender. The majority of students completed the profile sheet (female=327, male=248).

The Committee on the Use of Humans as Experimental Subjects, the institutional review board at the Massachusetts Institute of Technology in Cambridge, MA, approved this study before data collection.

The Darwinian Snails Lab in EvoBeaker

Each of the EvoBeaker labs includes a series of interactive simulations with which students design experiments and collect data. Students are provided with a workbook for each lab that directs them through different experiments and asks them to organize and interpret data they collect. We (Herron, Maruca, Meir, Perry, Stal) designed the Darwinian Snails Lab to teach the basic principles of natural selection and to correct the most commonly held misconceptions about natural selection. In this lab, students are presented

with a re-creation of a New England rocky shore habitat. The simulated habitat includes populations of the native flat periwinkle snail and their predator, the nonnative European green crab, and is based on the work of Seeley (1986) and, for the final exercise, Trussell (1996). The snails vary in their shell thickness, which affects the efficiency of predation by the European green crab.

Students first read a short section about the simulated system, and then they are introduced to the system by acting as European green crabs feeding on snails. Thicker shells require more effort by the students (i.e., number of mouse clicks needed to feed on the snail). As the students feed on snails, they observe changes in the average shell thickness in the snail population. The population of snails then reproduces, and students are shown how the traits of the remaining snails are inherited by their offspring.

After this initial exercise, students explore three basic requirements for natural selection based on shell thickness: variation in thickness, heritability of thickness, and differential survival of individuals with different shell thicknesses. During this portion of the lab, students sequentially violate each assumption and make predictions about what will then occur when the predator is introduced. Students then quantify changes in the average shell thickness in the snail population.

A third section demonstrates the origin of variation through mutation (but not genetic recombination through sexual reproduction). Students introduce crabs into the system but prevent mutation from occurring in the snail population. They quantify the change in average shell thickness after several generations. Students then allow mutations to occur in the snails, examine individual offspring to see that mutations are random, and compare the average snail shell thickness in the new population after several generations of predation from crabs.

Finally, students design their own experiments to determine if two snail populations differ in average shell thickness due to natural selection. Students can set up common garden experiments and include crabs or crabs with banded claws (crabs unable to feed on snails). Through the experiments, students test their hypotheses about the factors that are driving average shell thickness in the system. Figure 1 shows an example screenshot from this final section of the lab. The full lab takes students 1.5–2 hours to complete, with about 1 hour devoted to the first two sections, 10–15 minutes on the section about mutations, and 30 minutes or more on the final open-ended section

Instrument Design and Validation

The initial design of the test was a series of written open-response questions. We (Herron) asked for feedback on the test from several evolutionary biology instructors and
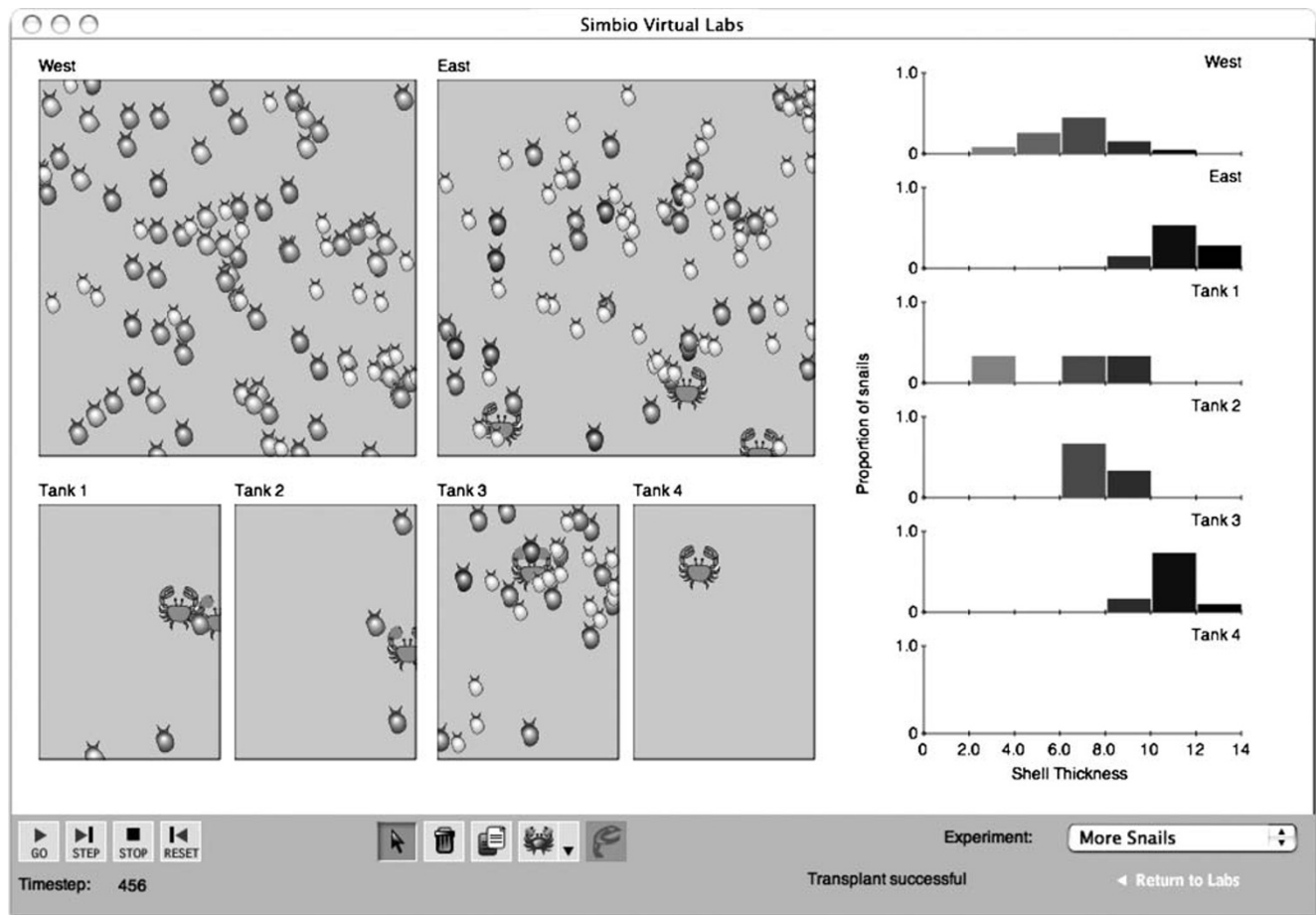
Fig. 1 Screenshot of EvoBeaker Darwinian Snails Lab. In this culminating exercise, students are given two different stretches of coastline (*top two panels*), one "East" which crabs have reached (*right*) and one "West" which is still free of crabs (*left*). They also have four tanks where they can perform experiments. Students must design and carry out experiments to show that the difference in shell average shell thickness between the two areas (*shown in the histograms on the right*) is likely due to natural selection. To do this, they must show that there is variation, it is heritable, and there is selection, the three principles they learned in the more directed earlier parts of the lab

refined the test based on their comments. We (Meir, Perry) then pilot-tested the exams with 20 Boston-area students to further refine problems with test items. We interviewed students after they took the exam, allowed them to further explain their answers in interviews, and used student responses in tests and interviews to develop distracters (incorrect options) for multiple-choice test items.

The final version of the tests designed for this study included nine multiple-choice and seven open-response items about natural selection principles. Four of the multiple-choice items were taken from Settlage and Odum (1995). The remaining multiple-choice and open-response test items presented scenarios from a hypothetical situation and asked students to analyze or predict situations based on the information provided. The majority of the multiple-choice questions included distracters based on student responses and the most common misconceptions about natural selection in the literature. The pre- and post test

were structured identically, but we changed the specific information in each item. We analyzed the internal consistency of the pretest multiple-choice questions using the Kuder–Richardson 20 method, which yielded a reliability coefficient of 0.68. Sample short- and long-response test questions are shown in Appendix A. Full copies of our tests are available by writing to SimBiotic Software® (www.simbio.com)—we avoid posting openly to retain their usefulness for instructors.

To quantify the presence or absence of misconceptions on the open-response test items, we (Abraham, Herron, Meir) first developed a rubric based on the list of misconceptions culled from the literature (Table 1). We then independently coded misconceptions and correct concepts in student responses on a subset of tests. Initial agreement among the authors was 85%. Disagreements about misconceptions in the responses were discussed until all of the authors agreed on 100% of the coding. One author

(Abraham, who did not participate in designing the lab or tests) then coded misconceptions in open responses based on the revised rubric (N=338 students). Coding for the presence of misconceptions was done conservatively; we only assigned a misconception to a response when the student explicitly stated it. Instances where a misconception was indirectly suggested but not clearly stated and instances where an incorrect answer was provided that did not link to a misconception were both coded as unclassifiable.

Data Analysis

We analyzed student performance on the multiple-choice selection and open-response type items separately. We used the full dataset for the multiple-choice responses (N=637) but used a subset of those exams for analysis of the open-responses (N=338). The subset, including tests from twelve institutions, was chosen so that it included a sufficient number of exams from beginner and advanced students and represented the diversity of institution types we included in the study. We first compared the average proportion of correct answers between the pre- and post test multiple-choice items with a one-tailed paired sample Wilcoxon sign-rank nonparametric test. We then used two-tailed Wilcoxon rank sum nonparametric tests to compare the average improvement in student score on the multiple-choice items between two student subgroups: female (n=187) and male (n=145) and beginner (n=128) and advanced (n=210) students. We calculated Cohen′s d effect sizes for each comparison of pre- and post test scores (Cohen′s $d=\frac{\bar{\chi}_1+\bar{\chi}_2}{s}$, χ=mean, s=pooled standard deviation).

In one open-response test item, we asked students to describe what had occurred in the hypothetical situations in the pre- and post tests to elicit descriptions of the process of natural selection (Question 15, Appendix 1). We designated this question as a long-response question because a correct answer necessarily involved more than one sentence. We compared the number of correct concepts ((1) variation in traits; (2) heritability of traits; (3) differential survival to reproduction; (4) change in average trait value in population over generations) provided by students before and after instruction with a one-tailed paired sample Wilcoxon sign-rank test. We then compared average improvement between gender and academic level with two-tailed Wilcoxon rank-sum tests. We used the statistical software package JMP 7.0.2 (SAS Institute 2008) for each of the preceding analyses.

To analyze student performance on the other open-response questions (short response), we scored each misconception as present or absent in the exam. Thus, a student who used a misconception a single time was scored the same as a student who used it multiple times. Students sometimes failed to provide answers for some of the open-response test items. Before analysis, we compared the frequency of incomplete responses to open-response test items between pre- and post tests. We found no difference between tests, so we included incomplete exams in our analysis. We also compared student use of misconceptions between public and private institutions. We found no difference, so this factor was dropped from the analysis.

We compared the prevalence of the four most common misconceptions (MC1=willful change, MC2=directed variation, MC3=intra generational change, MC4=population change; defined in Table 1) between the pre- and post tests with a series of McNemar′s paired sample chi-square tests. Other misconceptions did not occur frequently enough to analyze statistically.

We next compared improvement in students who exhibited misconceptions in either test between male or female and beginner or advanced students with a series of chi-square tests. For these comparisons, we defined improvement as a misconception present in the pretest open responses that was absent in the post test. We defined a lack of improvement as either the presence of a misconception in both the pre- and post test or a misconception present in the post test that was absent in the pretest. Thus, students who did not exhibit a given misconception in either test were excluded from this portion of the analysis.

Results

Student Understanding as Evidenced in Multiple-Choice Questions

Students were readily able to distinguish between correct scientific phrasing and the distracters based on common misconceptions, averaging over 70% correct responses on the pretest. Student performance on the multiple-choice questions significantly improved after instruction (W_+= 9357.5, $p<0.001$; Fig. 2), increasing from 72% to 75% correct. However, while this improvement was highly significant, the effect size of this improvement was small (Cohen′s $d=0.059$).

We did not find any difference in performance on the multiple-choice questions between genders. Both male and female students performed equally well on the pretest ($z=$ 1.83, $p=0.068$) and post test ($z=1.313$, $p=0.189$). However, advanced students significantly outperformed beginner students on both the pretest ($z=-5.42$, $p<0.001$) and post-test ($z=-6.86$, $p<0.001$) multiple-choice questions. Advanced students averaged 10 and 14 percentage points greater than beginner students on the pre- and post tests, respectively (Fig. 3). The pre- to post test improvement we saw across all students was driven almost entirely by improvement among advanced students (Cohen′s $d=0.078$),
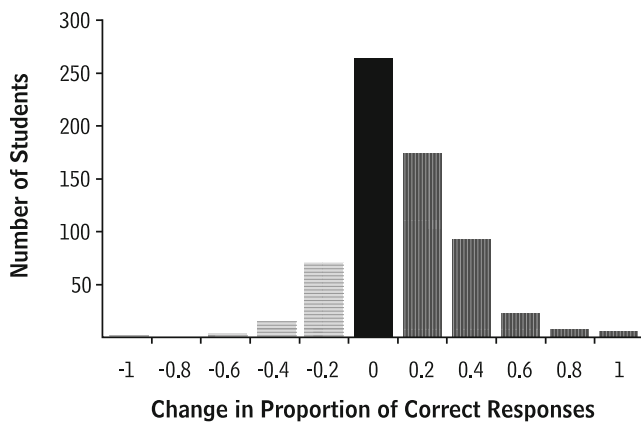
Fig. 2 Histogram of the change in proportion of correct responses by students on the multiple-choice questions, before and after instruction. There was a significant ($p<0.001$), but small, increase in the proportion of correct responses after instruction

while beginner student scores remained largely unchanged (Cohen's $d=0.008$) between tests.

Student Understanding as Evidenced in a Long-Response Question

Although we saw small increases in student performance on the multiple-choice items, we did not see the same pattern in student-generated descriptions of the process of natural selection. A complete, correct response to this test item listed each of the four scientific principles we associated with the process of natural selection. For example (principles are in italics):

> Student: "The packrats came eating to cooler fruits in the valley, allowing only hot fruits to reproduce (*Variation*, *Differential Survival/Reproduction*). Since spiciness is heritable (*Heritability of Trait*), the hot fruits would survive and natural selection would
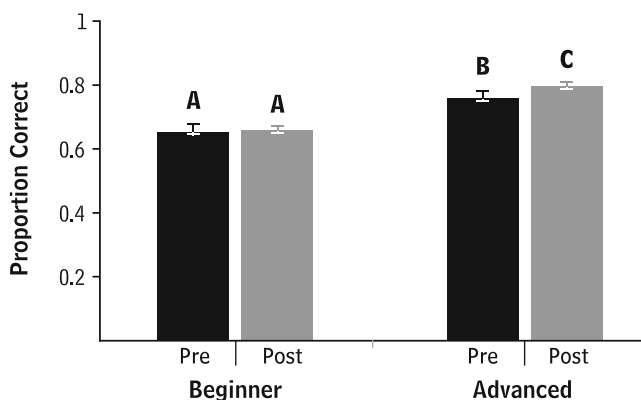
eventually wipe out the cool population and that gene (*Change in Gene/Trait Frequency*)."

> Student: "After the shrikes arrived, they began eating horned lizards. The horned lizards with longer horns were better equipped to defend themselves against the shrikes than the lizards with shorter horns (*Variation*). As a result, the shrikes ended up eating mostly lizards with shorter horns (*Differential Survival/Reproduction*). Since the lizards that survived were mostly long-horned, the offspring inherited the long-horned trait (*Heritability of Trait*) and long-horn lizards eventually dominated the population. The average horn length of the entire population was longer than that of the original population (*Change in Gene/Trait Frequency*)."

Unfortunately, few students provided complete responses to this test item. Typical responses often included just a few of the principles of natural selection or were mixed with misconceptions. For example (principles and misconceptions in italics):

> Student: "The packrats ate the cooler fruits so they didn't reproduce. The hotter plants weren't eaten and thus reproduced (*Variation*, *Differential Survival/ Reproduction*)."

> Student: "When the shrikes arrived they were eating the shorter horn lizards and in order for the lizards to stay alive they had to develop a mutation for long horns so they could pass off to their offspring (*Differential Survival/Reproduction*, *Heritability of Trait*, *Directed Variation (MC2)*, *Need (MC1)*)"

When we compared the number of principles of natural selection supplied by students in this long-response item, we found no evidence that our lab improved overall student understanding ($W_+=82.5$, $p=0.4527$). On average, students

Fig. 3 Proportion of correct responses by beginner and advanced students on the multiple-choice items before and after instruction. *Different letters* indicate significant ($p<0.001$) differences between treatments
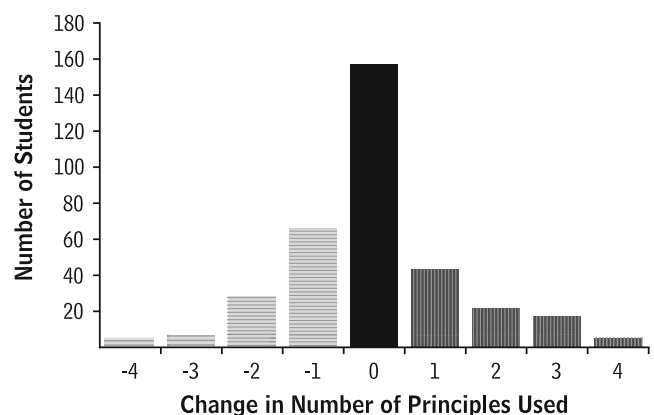

Fig. 4 Histogram of the change in number of principles of natural selection provided by students in a long-response item, before and after instruction. There was no difference in the number of principles used by students before or after instruction

included approximately 2.3 principles in both their pretest and post test explanations. Students improved or worsened almost equally (Fig. 4).

Although, on average, the student mastery of natural selection principles did not change, we saw differences between student academic levels in both the frequency of use and improvement after instruction (Fig. 5). Beginner students used fewer principles than advanced students in the pretest ($z=-4.104$, $p<0.001$), averaging 20% fewer principles than advanced students (Fig. 5). This difference disappeared by the post test ($z=-1.1$, $p=0.272$). Use of the Darwinian Snails lab resulted in significantly greater gains by beginner students (mean delta score=0.31) than advanced students (mean delta score=−0.19; $z=2.897$, $p=0.0038$). We did not see any gender-based differences in the number of natural selection principles used in the pretest ($z=0.49$, $p=0.6243$) or post test ($z=-1.202$, $p=0.23$).

Student Understanding as Evidenced in Short-Response Questions

The open responses on the test were designed primarily to elicit these misconceptions: Willful Change (MC1), Directed Variation (MC2), Intra-Generational Change (MC3), and Population Change (MC4). As expected, these were the most commonly exhibited misconceptions in the pretest (Table 1). Of the 338 sets of student open responses on the pretests, slightly more than half exhibited at least one of the four most common misconceptions. Each of these misconceptions is discussed below.

Willful Change

Approximately 37% of student pretests included the misconception Willful Change (MC1; Table 1). Students
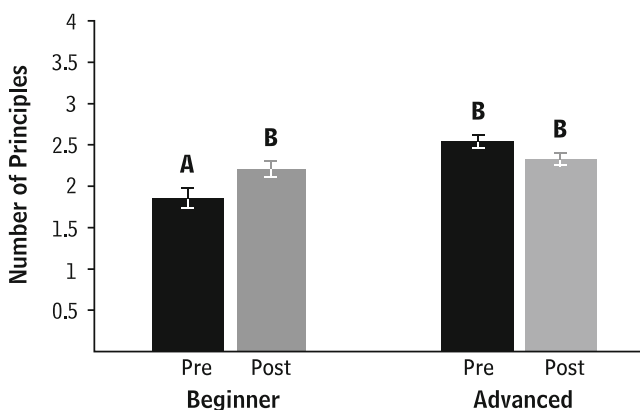

Fig. 5 Difference between advanced and beginner student use of principles of natural selection in a long-response item, before and after instruction. *Different letters* indicate significant ($p<0.001$) differences between treatments

who used this misconception posited a role for the desire or need of an organism in effecting change in that organism's trait expression. Students often used the words "need" or "want" when explaining how or why traits shifted in the population. For example, when asked what would happen to an adult lizard's horn length when a predator was introduced to the system, students answered:

> Student: "It would have needed to grow longer horns out of necessity for survival."

> Student: "The lizards were getting along fine with their short/medium horn lengths, but they suddenly became threatened by a predator, so, in order to adapt to their new environment, they had to develop longer horns"

Directed Variation

The most commonly used misconception in this study, Directed Variation (MC2), appeared in 45% of student pretests (Table 1). This misconception suggests that selection pressure or environmental conditions are the ultimate cause of new trait variation in a population rather than random mutation and genetic recombination through sexual reproduction. Directed Variation (MC2) often appeared in Question 5 (Appendix 1), when students were asked how extreme trait values could appear in a population. Student answers often directly tied trait variation to changes in the environment or the introduction of competitors or predators. For example:

> Student: "If there is no change in the environment there would be no need for change in horn sizes there wouldn't be anything to adapt to out of normal standards"

> Student: "Need directional force (predation, etc.)"

Intra-Generational Change

Approximately 15% of students made statements that evolutionary change could occur within an organism's lifetime (Table 1). When asked about a lizard's individual response to the introduction of predators into the system, these students responded with answers like:

> Student: "This lizard would have grown longer horns because it had to adapt to the environment it was in"

> Student: "The arrival of shrikes will cause the adaptation of longer horns to evolve during the life span of the newly hatched lizards"

Population Change

The last of the most common misconceptions, Population Change (MC4), appeared in 20% of the student pretest exams. Students who exhibited this misconception claimed that populations or species gradually adapt, in concert, from selection pressure rather than through changes in gene or trait frequencies in the population across generations. Of these students, most answered questions with responses similar to the following examples:

> Student: "As generations went on the population developed with longer horns, better adapting them to environment"

> Student: "As the shrikes attack the lizards counter them by trying to grow horns. Short horns are not as effective as long ones, so they grow longer horns, and each generation grows a bit longer"

Overall, student understanding of natural selection, as measured by the presence of misconceptions in the open responses improved after instruction (Fig. 6). We found significantly reduced student use of Willful Change (MC1; $\chi^2=21.01$, $df=1$, $p<0.001$), Intragenerational Change (MC3; $\chi^2=12.5$, $df=1$, $p<0.001$), and Population Level Change (MC4; $\chi^2=15.75$, $df=1$, $p<0.001$). Student use of Willful Change (MC1) dropped 35% after instruction, while student use of Intragenerational Change (MC3) and Population Change (MC4) was reduced by 44% and 49%, respectively (Fig. 6). Although there was a trend towards reduced use of the most common misconception, Directed Variation (MC2; Fig. 6), the difference was nonsignificant ($\chi^2=2.68$, $df=1$, $p>0.05$).

We also compared the prevalence of misconceptions in the short responses on the pretest between subgroups.
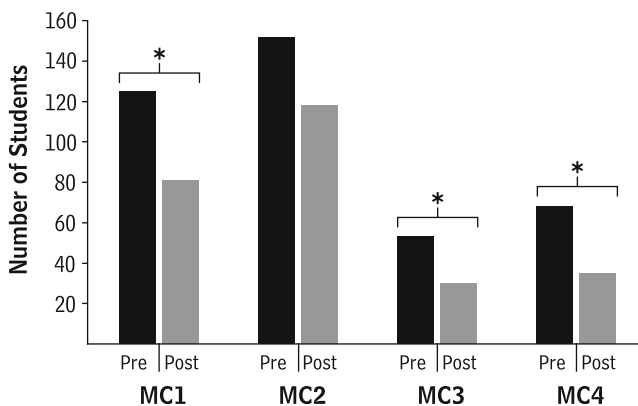
Student gender was not related to the use of any of the four misconceptions (data not shown). We did, however, see differences with student academic level on one of the misconceptions. Beginner students were significantly more likely to explain phenomena with Willful Change (MC1) in the pretest ($\chi^2=11.506$, $df=1$, $p<0.001$) than were advanced students. Nearly half of the beginner students used Willful Change (MC1) in at least one of their responses, as compared to approximately 30% of advanced students (Fig. 7). The prevalence of the other three misconceptions did not differ significantly between academic levels (Fig. 7). When we compared the change in use of misconceptions between the pre- and post tests, we found no evidence that gender or academic level affected the amount of student improvement with instruction.

Discussion

Impact of Darwinian Snails Lab on Student Performance

Overall, our interactive simulation software appears to reduce student misconceptions around natural selection. The prevalence of three of the misconceptions dropped 35–49% after use of the Darwinian Snails Lab. Student performance on the multiple-choice items significantly improved after instruction as well, although the magnitude of change (4%) was quite small (Fig. 2). Together, these results suggest that the Darwinian Snails Lab may be an effective tool for addressing some of the most common student misconceptions about the process of natural selection.

The majority of instructors gave the post test immediately following instruction, but some waited up to a week. Although we asked instructors to refrain from explicit
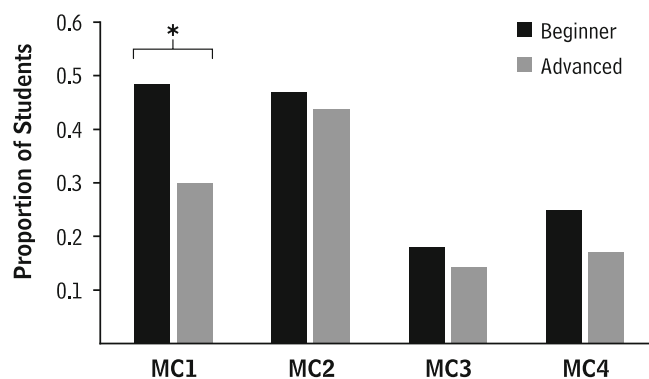

Fig. 6 Change in student use of misconceptions Willful Change (*MC1*), Directed Variation (*MC2*), Intragenerational Change (*MC3*), and Population Change (*MC4*) in short-answer responses after instruction. *Asterisks* indicate a significant ($p<0.001$) reduction in the use of misconceptions between tests

Fig. 7 Relative proportion of beginner and advanced students who used misconceptions Willful Change (*MC1*), Directed Variation (*MC2*), Intragenerational Change (*MC3*), and Population Change (*MC4*) prior to instruction. The asterisk indicates a significant ($p<0.001$) difference between student academic levels

instruction on natural selection between testing phases, it may have been unavoidable in some instances. This information was not consistently reported. Use of the lab may also have prompted students to further explore these topics outside of class. Thus, we cannot rule out the possibility that additional instruction affected student post-test performance or that students who were post tested later performed differently than those tested immediately after instruction. However, given the difficulty of correcting misconceptions through traditional teaching methods or self-teaching (Morrison and Lederman 2003; Nelson 2008), we feel that the gains or stasis in student scores are likely due to the strengths and shortcomings of our software. In future studies, we hope to track post-instruction performance over time to get a better appreciation of student retention of concepts.

Of the misconceptions we tested for, Directed Variation (MC2) was the only one that did not significantly decrease in the open responses after instructional intervention (Fig. 6). This misconception was also the most common in our pretest (Fig. 6). This is not surprising, as we and others have found in previous studies that processes involving randomness are difficult for biology students to understand (Meir et al. 2005a; Garvin-Doxas and Klymkowsky 2008). While we attempted to address this misconception with a separate section in the Darwinian Snails lab, that section was added late in the design process and may have suffered from a lack of clarity. In it, we ask students to attempt to increase average snail shell thickness over time through natural selection, with and without mutations in the system. This may have mistakenly reinforced the misconception rather than identifying mutation as a source of random variation. Alternately, the visuals and student exercises may not have been adequate to facilitate the understanding of the random nature of variation. The section of the lab on the role of mutations comprised only 10–15 minutes of a 1.5–2-hour lab, and the visualization of the mutations was less clear than other aspects of the lab such as the selective advantage of thick shells. We suspect that students′ lack of improvement on this misconception partially results from poor lab design rather than it's being a harder misconception to dispel than the others.

In contrast to the pattern of misconception use by students, we found less convincing evidence that the Darwinian Snails Lab was associated with improvements in overall student use of correct natural selection principles. On the long-response test item, where students wrote a paragraph explaining how natural selection works, the number of principles of natural selection used by students was nearly identical before and after instruction (Fig. 5). While this may be a shortcoming of the software, our study design may have also played a role

in this pattern (see below). But to the extent that this result is real, it indicates that our software does not perform as well in this aspect as we had hoped. Some studies have had greater success at increasing student use of principles of natural selection in their explanations (e.g., Jensen and Finley 1996; Robbins and Roy 2007). These studies used instructional methods that explicitly compared the principles of natural selection to misconceptions to better improve student understanding. This may be something we can improve with a future version of the Darwinian Snails Lab.

Differences between Multiple-Choice, Short-Response, and Long-Response Test Items

Although not directly comparable, in general, students performed better on the multiple-choice questions than they did on open-response questions. While students answered nearly three quarters of the multiple-choice questions correctly, the majority of open-response questions were answered incorrectly or incompletely. This fits with other work that suggests multiple-choice questions are not as powerful at capturing students' underlying conceptual knowledge as other types of questions (e.g., Almquist and Cronin 1988; Nehm and Reilly 2007). However, the use of student misconceptions as distracters in multiple-choice questions allows for a more complete assessment of student conceptual knowledge (Anderson et al. 2002; Tanner and Allen 2005; Settlage and Odom 1995). We successfully assessed conceptual knowledge with multiple-choice questions in a previous study (Perry et al. 2008). In that case, we saw substantial improvements on multiple-choice questions between pre- and post tests and were able to verify these improvements with student interviews.

In contrast, students in this study showed many more misconceptions, and greater improvement, on the open-response items. The open format of these questions may have provided a better opportunity to assess student understanding. Students frequently offered misconceptions in their responses to questions that were designed to elicit other misconceptions and sometimes mixed correct concepts with misconceptions in their responses.

We used a single long-response test item (Question 15, Appendix 1) near the end of the test to directly assess student understanding of the principles of natural selection. In contrast with the above trend, overall, students showed no change between their pre- and post test responses on this item. Given that 100 students used *fewer* correct concepts on the post test, but many of those same students improved on those same concepts in the short-answer questions, we believe that the study design may have caused some fatigue. Most students complete the lab in 1.5–2 hours, not including the time spent on the pretest. By the time they

reached this question, some students may have been tired of the experience and provided less complete answers than they did on the pretest. A slightly higher percentage of advanced students (31%) performed worse on the pretest than did beginner students (27%), which may reinforce the idea that boredom or fatigue played a role in the pattern rather than a change in understanding.

The results of this and previous studies suggest that when possible, multiple-choice questions should be augmented with open-response questions. Although the utility of the test may decrease, the inclusion of multiple question types allows for a more complete assessment of student conceptual knowledge. This is especially true of ongoing assessment during instruction, as the identification of unforeseen student confusions can aid in the development of future instructional priorities.

Gender and Academic Level Effects on Student Performance

In addition to total improvement among all undergraduate biology students, we also gauged the role played by student gender and academic level on student response to the Darwinian Snail Lab. Several studies have found that student gender impacts the use of misconceptions or replacement of those misconceptions after instruction (Almquist and Cronin 1988; Pearsall et al. 1996). However, we found no impact of gender in any of our analyses here or in previous studies (Meir et al. 2005a, 2007; Perry et al. 2008). Male and female biology students did not differ in their performance on multiple-choice questions, use of misconceptions, or improvement after instruction.

In contrast to the similarity between male and female students, we found significant, if varying, effects of student academic level on performance and improvement. In this study, advanced students were responsible for essentially all of the improvement on the multiple-choice questions (Fig. 3). In contrast, beginner students used significantly fewer principles of natural selection in the pretest than did advanced students but showed greater improvement after instruction (Fig. 5). We found that relatively more beginner students used Willful Change (MC1) in their pretest responses than did advanced student but that there was no effect of academic level on the reductions in misconception use that we recorded (Fig. 7).

While the more frequent use of a misconception and the less frequent use of correct principles by beginner students may be intuitive, we were surprised by the greater improvement of advanced students on multiple-choice items after instruction (Fig. 3). Although the reason for this difference is impossible to determine without interviews, we may have recorded a threshold effect from repeated exposure to the topic. Advanced students have

already received instruction on natural selection principles; they are perhaps better able to pick up on the nuances of language differences between the correct answer key versus the distracters presented in the multiple-choice items. Had we seen a similar pattern of greater improvement in misconception use by advanced students after instruction, our results would have provided stronger evidence for this hypothesis. In either case, this study, in agreement with previous research on student misconceptions, provides a strong argument for continued review of natural selection principles through upper-level courses. Student use of misconceptions and inability to explain the process of natural selection is high, even after instruction. However, the significant improvement by advanced and beginner students on either the multiple-choice or open-response questions demonstrates that pedagogical tools designed to challenge misconceptions can be effective at multiple academic levels.

Conclusions

Student use of some of the most common misconceptions about natural selection decreased after exposure to the Darwinian Snails Lab, suggesting that this computer-based interactive simulation is an effective tool. However, the lab was not as successful at improving student mastery of the principles of natural selection. Both beginner and advanced students showed misconceptions about natural selection and were largely unable to fully explain the process of natural selection. These results suggest that basic evolutionary principles should be reviewed in introductory through advanced biology courses. Continued development of teaching practices and tools designed to confront student misconceptions will aid in identifying and confronting student misconceptions and help increase student understanding of the fundamentals of evolutionary theory.

no longer had any financial ties to the company at the time of data analysis or paper submission. Meir, Maruca, and Stal are currently employed by SimBiotic Software®. Herron contracted with SimBiotic Software® on this project and has a financial interest in EvoBeaker.

Appendix

Sample short- and long-response questions from the pretest. Throughout the exam, students answer questions related to an increase in average horn length of a horned lizard population across generations. In this scenario, the lizards initially live in a predator-free environment. Average horn length is of a medium length, and horn length is heritable. The introduction of predatory birds creates a strong selection pressure on horn length. Longer horns help protect the lizards from attack, and over time, the average horn length in the lizard population increases. Additional test items are available upon request (www.simbio.com).

Question 5

Horned lizards have been living in Desert Valley, *free of predators*, for many generations. The diagram at right shows the individual with the shortest horns (0.75 cm) in the Desert Valley population, a typical individual (1.1 cm), and the individual with the longest horns (1.5 cm). *The differences in horn length among lizards are determined by the genes the lizards inherit from their parents.*

Assuming that no new lizards migrate into Desert Valley from elsewhere and that nothing changes in the Desert Valley environment, is there a way that full-grown lizards with very short horns (less than 0.5 cm long) or very long horns (more than 1.75 cm long) could arise in the Desert Valley population?

(A) Yes
(B) No
(C) Don′t know

Briefly explain your answer:

Question 15

Briefly describe the process that caused horn length to change in the lizard population in Desert Valley.

References

Almquist AJ, Cronin JE. Fact, fancy, and myth on human evolution. Curr Anthropol. 1988;29:520–2. doi:10.1086/203672.

Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the conceptual inventory of natural selection. J Res Sci Teach. 2002;39:952–78. doi:10.1002/tea.10053.

Beardsley PM. Middle school student learning in evolution: are current standards achievable? Am Biol Teach. 2004;66(9):604–612.

Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. J Res Sci Teach. 1990;27:415–27. doi:10.1002/tea.3660270503.

Brumby MN. Misconceptions about the concept of natural selection by medical biology students. Sci Educ. 1984;68(4):493–503.

Crawford BA, Zembal-Saul C, Munford D, Friedrichsen P. Confronting prospective teachers' ideas of evolution and scientific inquiry using technology and inquiry-based tasks. J Res Sci Teach. 2005;42(6):6123–637.

Demastes SS, Settlage J, Good R. Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. J Res Sci Teach. 1995;32(5):535–550.

Evans EM. The emergence of beliefs about the origins of species in school-age children. Merrill-Palmer Q. 2000;46:221–54.

Ferrari M, Chi MTH. The nature of naive explanations of natural selection. Int J Sci Educ. 1998;20:1231–56. doi:10.1080/0950069980201005.

Garvin-Doxas K, Klymkowsky MW. Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). CBE Life Sci Educ. 2008;7:227–33. doi:10.1187/cbe.07-08-0063.

Greene ED. The logic of university students' misunderstanding of natural selection. J Res Sci Teach. 1990;27(9):875–885.

Hamza KM, Wickman PO. Describing and analyzing learning in action: an empirical study of the importance of misconceptions. Sci Educ. 2008;92(1):141–164.

Hokayem H, BouJaoude S. College students' perceptions of the theory of evolution. J Res Sci Teach. 2008;54(4):395–419.

Jensen MS, Finley FN. Changes in students' understanding of evolution resulting from different curricular and instructional strategies. J Res Sci Teach. 1996;33(8):879–900.

Kara Y, Yesilyurt S. Assessing the effects of tutorial and edutainment software programs on students' achievements, misconceptions and attitudes towards biology. Asia-Pacific Forum Sci Learn Teach. 2007;8(2):1–22.

Lawson AE, Thompson LD. Formal reasoning ability and misconceptions concerning genetics and natural selection. J Res Sci Teach. 1988;25(9):733–746.

Meir E, Perry J, Stal D, Maruca S, Klopfer E. How effective are simulated molecular-level experiments for teaching diffusion and osmosis? Life Sci Educ. 2005a;4:235–48.

Meir E, Herron JC, Maruca S, Stal D, Kingsolver J. (2005b) EvoBeaker 1.0. Available online at http://www.simbio.com. SimBiotic Software®, Ithaca, NY.

Meir E, Perry J, Herron JC, Kingsolver J. College students' misconceptions about evolutionary trees. Am Biol Teach. 2007;69:71–6. doi:10.1662/0002-7685(2007)69[71:CSMAET]2.0.CO;2.

Modell H, Michael J, Wenderoth MP. Helping the learner to learn: the role of uncovering misconceptions. Am Biol Teach. 2005;67:20–6. doi:10.1662/0002-7685(2005)067[0020:HTLTLT]2.0.CO;2.

Morrison JA, Lederman NG. Science teachers' diagnosis and understanding of students' preconceptions. Sci Educ. 2003;87:849–267. doi:10.1002/sce.10092.

National Research Council. National Science Education Standards. Washington, DC: National Academy Press; 1996.

Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. Bioscience. 2007;57:263–72. doi:10.1641/B570311.

Nelson CE. Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions. Integr Comp Biol. 2008;48(2):213–225.

Novak JD. Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. Sci Educ. 2002;86:548–71. doi:10.1002/sce.10032.

Pearsall RP, Skipper JJ, Mintzes JJ. Knowledge restructuring in the life sciences: a longitudinal study of conceptual change in biology. Sci Educ. 1996;81:193–215.

Perry J, Meir E, Herron JC, Maruca S, Stal D. Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. CBE Life Sci Educ. 2008;7:193. doi:10.1187/cbe.07-01-0007.

Robbins JR, Roy P. The natural selection: identifying & correcting non-science student preconceptions through an inquiry-based, critical approach to evolution. Am Biol Teach. 2007;69:460–6. doi:10.1662/0002-7685(2007)69[460:TNSICN]2.0.CO;2.

Samarapungavan A, Wiers RW. Children's thoughts on the origin of species: a study of explanatory coherence. Cogn Sci. 1997;21:147–77.

SAS Institute. JMP 7.0.2. Cary: SAS Institute; 2008.

Seeley RH. Intense natural selection caused a rapid morphological transition in a living marine snail. Proc Natl Acad Sci USA. 1986;83:9897–6901. doi:10.1073/pnas.83.18.6897.

Settlage J. Conceptions of natural selection—a snapshot of the sense-making process. J Res Sci Teach. 1994;31:449–57.

Settlage J, Odom AL. (1995) Natural selection conceptions assessment: Development of a two-tier test "Understanding Biological Change." Annual meeting of the National Association for Research in Science Teaching, San Francisco, CA

Sinclair A, Pendarvis MP, Baldwin B. The relationship between college zoology students' religious beliefs and their ability to objectively view the scientific evidence supporting evolutionary theory. J Res Dev Educ. 1997;30:118–25.

Tanner K, Allen D. Approaches to biology teaching and learning: understanding the wrong answers - teaching toward conceptual change. Life Sci Educ. 2005;4(2):112–117.

Trussell GC. Phenotypic plasticity in an intertidal snail: the role of a common crab predator. Evolution Int J Org Evolution. 1996;50:448–54. doi:10.2307/2410815.

Wescott DJ, Cunningham DL. Recognizing student misconceptions about science and evolution. MountainRise. 2005;2:1–8.