

# Phylo-Genie: Engaging Students in Collaborative ‘Tree-Thinking’ through Tabletop Techniques

Bertrand Schneider  
Stanford University  
Stanford, CA

Megan Strait  
Tufts University  
Cambridge, MA

Laurence Muller  
Harvard University  
Cambridge, MA

Sarah Elfenbein  
Yale University  
New Haven, CT

Orit Shaer  
Wellesley College  
Wellesley, MA

Chia Shen  
Harvard University  
Cambridge, MA

schneibe@stanford.edu, megan.strait@tufts.edu, lmuller@seas.harvard.edu, sarah.elfenbein@yale.edu, oshaer@wellesley.edu, chia\_shen@harvard.edu

## ABSTRACT

Phylogenetic trees are representations of evolutionary relationships amongst species. Interviews of instructors and students have revealed that novice biologists have difficulty understanding phylogenetics. Moreover, misinterpretations of phylogenetics are common among college-level students. In this paper we present Phylo-Genie, a tabletop interface for fostering collaborative learning of phylogenetics. We conducted an experimental study with 56 participants, comparing students’ conceptual learning and engagement using Phylo-Genie as: 1) a multi-touch tabletop interface and 2) a pen and paper activity. Our findings show that the tabletop implementation fosters collaborative learning by engaging users in the activity. We also shed light on the way in which our design principles facilitated engagement and collaborative learning in a tabletop environment.

## Author Keywords

Education, tabletop; collaborative learning; bioinformatics; multi-touch interaction; tangible user interface.

## ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces:

## General Terms

Human Factors; Design.

## INTRODUCTION

Phylogenetics, the science of constructing and evaluating hypotheses about historical patterns of descent, is at the very core of evolutionary theory. Thus, it has become pervasive inside and outside the field of biology [21]. Phylogenetic trees are the most direct representation of the principle of common ancestry, and as such, hold a prominent place in the public’s understanding of evolution. It is common for biology professors to teach these representations in introductory classes. However, numerous studies indicate that evolutionary trees are difficult to comprehend for college-level students [32], and that many concepts are counter-intuitive and require a more in-depth

understanding of evolution [15]. In addition, the field of phylogenetics depends on computational tools, such as Mesquite [30] and ClustalW [28], for creating and testing evolutionary trees. These tools are not appropriate for novice students since they do not support important aspects of college-level learning, such as collaborative learning or high-level reasoning. Instead, professors often turn to paper-based learning environments (textbooks or worksheets) for teaching phylogenetic principles. This approach does not provide sufficient support for novice students either because representations used in many textbooks have been shown to induce misconceptions [9] and because it lacks support for “active” learning.



**Figure 1. Phylo-Genie tabletop system: tangible and digital exploration of morphological characteristics of Australian snake specimen.**

Recent investigations highlight common misconceptions held by students (e.g. [10]). These misconceptions illustrate fundamental barriers to understanding how evolution operates. Motivated by the need for computational support in college-level phylogenetics and by previous work demonstrating the potential of tabletop interfaces to facilitate learning [1, 3, 4, 7, 8, 13, 16, 17, 20, 23, 26, 37, 46], we developed Phylo-Genie (Figure 1). Phylo-Genie introduces tangible and multi-touch interactions on a horizontal surface. Its design is aimed at fostering collaborative learning and engagement in order to improve students’ understanding of phylogenetics.

In this paper, we first discuss core concepts in phylogenetics and the misconceptions common to college-level students that have been reported in recent studies. We

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI’12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

then review related work in the areas of tabletop and tangible user interfaces for science education, the evaluation of dyadic collaboration, and existing methods and tools for teaching phylogenetics. Next, we describe the Phylo-Genie learning environment and present findings from our comparative study that highlight the strengths and weaknesses of Phylo-Genie compared to traditional pen and paper interactions. Finally, we discuss the implications of our results for designing collaborative tabletop activities.

## RELATED WORK

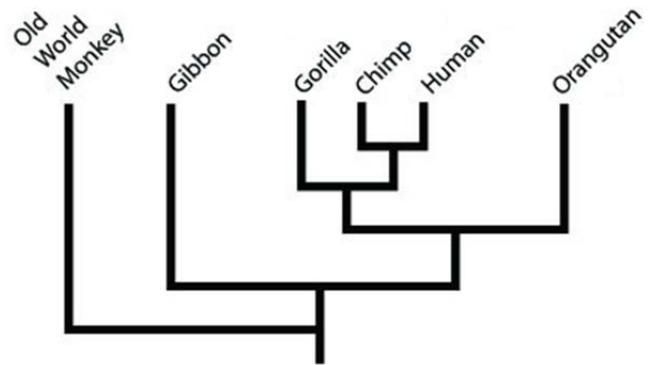
Phylogenetics, the study of evolutionary relatedness among species on earth, is based on morphological and genomic data. A phylogenetic tree shows that some species are more closely related than others because they share a more recent common ancestor. For example, in Figure 2, humans and chimps are more closely related than humans and gorillas. The ability to read and understand a phylogenetic tree is called tree-thinking [21]. More specifically, tree thinking refers to the ability to “(a) interpret and extract information about evolutionary history from phylogenetic trees and (b) use those diagrams to organize knowledge of biodiversity and make inferences to support one’s decisions and actions” [34]. A significant number of studies ([5, 15, 21, 32]) suggest that tree-thinking is a crucial yet difficult skill to acquire, but it is necessary in order to fully grasp evolutionary principles.

### Known Misconceptions

The framework of this paper is a constructivist view of the human mind whereby students hold an everyday understanding of a variety of topics and use their intuition to explain the world. This approach suggests that students use their prior knowledge to build new concepts [36]. The fact that students have pre-conceptions does not mean they are correct; in some cases, they may hold misconceptions that contradict experts’ explanations of a phenomenon. Misconceptions are usually persistent and difficult to correct because they are deeply rooted in existing cognitive structures [15]. In the following section we describe known misconceptions in the field of phylogenetics.

Misconceptions are relatively well-documented in the field of teaching phylogenetics [10, 21, 32]. We review four major misconceptions held by college-level students [32]:

- *Incorrect Mapping of Relative Time:* Many are confused regarding the mapping of time to a phylogenetic tree. For instance, Meir et al. [31] found that some students, when given Figure 2, would draw an arrow from left (past) to right (present) based on the argument that humans descended from gibbons and gorillas. A correct mapping would define the root node of the tree as the oldest ancestor and the tips of the branches as living species.



**Figure 2. Example of a phylogenetic tree.**

- *Proximity Indicates Relationship:* students often believe gorillas are more closely related to gibbons than to humans since they are spatially closer on the diagram. They fail to consider the most recent common ancestor (branching point) as an indicator of relatedness.
- *Node Counting:* a significant number of students also believe that counting internal nodes is the most efficient way to assess evolutionary closeness. Using this strategy would lead to the incorrect interpretation that orangutans are more closely related to gibbons (a distance of 2 nodes) than to humans (3 nodes). Again, a correct observation would consider the most recent ancestor as a sign of evolutionary closeness.
- *Straight Line Equals No Change:* a common misconception of Figure 2 is to interpret the first branching on the left (leading to old world monkeys) as a period where no change happened. This is erroneous since species continue to evolve along the lines of the tree, even though there is no branching.

In summary, misconceptions associated with phylogenetics are widespread and diverse. When designing Phylo-Genie we considered the misconceptions described above and sought solutions to avoid them through an interactive and engaging learning experience.

### Existing Tools for Teaching Phylogenetics

The current approach for teaching phylogenetics relies heavily on textbooks and classroom instruction. Several studies outlined the weaknesses of this method: textbooks are often poorly designed (e.g. using a significant number of ambiguous or incorrectly constructed trees and a large number of cladograms in a ladder format [9]). Cladograms in a ladder format are known to be more difficult to understand and more likely to elicit misconceptions among students [33]. Additional studies have shown that college students [32] and even graduate students [41] hold strong preconceptions that reinforce alternative conceptions of macroevolution. As a consequence, new methods for teaching phylogenetics as well as carefully designed learning environments are necessary to ensure a better understanding of evolutionary principles. These methods should promote collaboration, active learning, and take into account students’ prior knowledge, while using accurate representations of evolutionary trees.

Some prototypes embodying those ideas include the work of Halverson [22], who asked students to build their own phylogenetic trees by connecting pipe cleaners, and Matuk [48], who designed animated trees with multiple external representations to avoid erroneous interpretations. These studies differ from our work in that they have not been thoroughly evaluated, have produced mixed results [22], or do not propose a complete learning activity for teaching phylogenetic concepts [48]. Existing work has extensively studied how students perceive evolutionary trees [33,34], how small interventions can prevent specific misinterpretations [48] and documented students' misconceptions [10,21,32]. However to our knowledge no stand-alone activities that take advantage of recent technological advances have been thoroughly evaluated.

### **Proposition to correct existing misconceptions**

In the design of Phylo-Genie, we adopted Smith, diSessa and Roschelle's [47] view on misconceptions. They propose to "move beyond simple models of knowledge and learning, where novice misconceptions are replaced by expert concepts". Indeed, while most instructional methods try to expose students to the truth hoping it will replace their previous knowledge with appropriate concepts, numerous studies have shown this approach ineffective and that misconceptions are deeply seated and resistant to change (e.g. [11]). As a consequence, replacement directly conflicts with the constructivist premise that learning is the process of adapting prior knowledge [47].

Smith, diSessa and Roschelle suggest a few ways to modify misconceptions in a constructivist fashion. First, they proposed that abstract representations should be removed. Instead, familiar situations where the students held a misconception should be utilized; indeed, novices are often able to exhibit expert-like behavior when describing highly complex systems with which they are familiar. They are more likely to revisit their understanding of concepts based on everyday objects and experience. Secondly, they emphasize the importance of using constructive discussions as a way to reformulate and re-conceptualize students' ideas. It is crucial to remove confrontation from the debate and validate students' pre-conceptions as an evolving structure that can be refined. We draw on Smith, diSessa and Roschelle's framework in the design of the Phylo-Genie learning environment. Misconceptions about phylogenetics need to be addressed in a familiar situation, where students have the opportunity to discuss their experiences in a collaborative and non-confrontational way.

### **Addressing Misconceptions With Collaborative Learning**

In the past decade, research in education has focused much of its effort on social learning and small group cognition. The inspiration for this effort mainly comes from Piaget [36], who postulated that socio-cognitive conflicts cause major cognitive restructuring, and Vygotsky [49] who claimed that learning happens first on a social or cultural level and only then internalized. Those two theories

coalesce under the umbrella of the socio-constructivist theory. This approach emphasizes the importance of collaborating and negotiating knowledge. As a consequence, socio-constructivist researchers have put their effort in analyzing group interactions and successful patterns of collaboration. From an empirical point of view, studies have shown mixed results of this approach [29,40]. As a consequence, we draw on Dillenbourg's view that collaboration works under some conditions, and it is the aim of research to determine the conditions under which collaborative learning is efficient [14]. Thus design guidelines are crucial for building suitable learning environments and activities. In the case of phylogenetic misconceptions, we propose that creating an engaging activity is a prerequisite and a critical factor for fostering positive collaboration.

### **Tabletops and Tangibles for Science Education**

In this section we describe existing tabletops and tangible user interfaces (TUIs) for science education, present the theoretical foundations for our *system design*, and discuss techniques for fostering collaboration and engagement.

Several tangible and tabletop interfaces were designed to augment science education in areas where there is inherent physical structure or where abstract ideas can be clarified with physical representation. Ahmet et al. [1] explored children's learning of concepts in physics through mixed digital and physical interaction techniques. For molecular biology, Gillet et al. [19] created a TUI that employs augmented reality to enhance 3D molecular models. Shaer et al. [45] developed a TUI fostering inquiry-based learning of genomics (G-nome Surfer). Fjeld et al. [16] and Glaeser et al. [20] constructed TUIs for chemistry education. Schkolne, Ishii and Schroeder [41] developed a tangible immersive interface for the design of DNA molecules. Similarly, Involv [26] is a multi-touch tabletop interface for exploring biodiversity and the relationships among organisms. These systems have the potential to contribute to science education in areas related to phylogenetics. Some of those projects focus on the representation and manipulation of objects that have inherent physical structure, while others propose a purely digital output (G-nome Sufer, Involv).

### **Collaboration and Engagement**

Most closely related to our work are two studies conducted by Shaer et al. [45] and Piper and Hollan [37]. Shaer et al. compared the affordances of multi-touch versus multiple and single mice on dyadic collaborative learning for undergraduate genomics. Piper and Hollan [37] studied pairs of undergraduate students, comparing the affordances of tabletop displays and traditional paper handouts. In both cases, they found that using a multi-touch surface increased the quality of collaborative processes. In a study comparing multi-touch and tangible interfaces, Schneider et al. [43] found that the tangible interface fostered collaboration, playful learning and supported a larger exploration of the

problem space. Our study draws largely upon this body of work in the context of enhancing learning of phylogenetics through interaction techniques that support collaborative processes. More specifically, we propose to blend multi-touch and tangible inputs to create a learning environment that engages users in collaborative problem-solving.

Other studies have focused on the linkages between engagement and learning [6, 8, 13, 39, 47]. Studies that focused on tangible interfaces found that physical objects fostered engagement and exploration among children [4] and apprentices in logistics [43]. Based on those studies, we propose that tangibles support users' engagement due to their aesthetic and affordances as physical objects.

### PHYLO-GENIE

Phylo-Genie is a tabletop interface for supporting collaborative learning of phylogenetics. As a learning tool, Phylo-Genie targets the understanding of phylogenetic principles through engaging content and tabletop interaction techniques. Gregory [21] holds that lessons at the undergraduate level should de-emphasize the technical aspects of phylogeny reconstruction in favor of a focus on the concepts underlying 'tree-thinking'. For this reason, we designed the scenario of Phylo-Genie as a hands-on activity that focuses on exploratory and active learning.

### Design Goals and Guidelines

To identify design elements for supporting engagement in learning of college-level phylogenetics, we collected data from a series of recurring individual 1.5-hour meetings with four instructors and two doctoral candidates in the field of evolutionary biology. These domain experts participated in brainstorming sessions and provided specific feedback on a series of prototypes. This iterative and participatory process allowed us to increase the fidelity and ecological relevance of our learning environment. Moreover, a review of related literature and teaching materials provided us with additional design requirements for the learning activity.

From the collected data we found that a learning environment focusing on phylogenetics needed to address the following questions: What is a phylogenetic tree? How do biologists build it (e.g. by combining morphological and genomic data)? How can it be interpreted (e.g. how closely related are two species)? It is also necessary to introduce a vocabulary for fostering discussion with a peer.

In addition to those requirements, we propose the following design principles to foster collaborative learning:

- **Engagement:** Providing an engaging learning experience supported by an interesting story, a visually appealing interface, and natural interactions techniques, so that students focus on discovering concepts.
- **Territoriality:** several studies [27, 44] suggest that territoriality is a crucial factor for supporting collaboration. A learning environment should provide ways to reinforce users' ownership of both physical objects and areas on top of the interactive surface.

- **Opportunities for reflection:** Clements [12] argues that the danger of concrete objects is to trap the students in an "action mode". Transitioning from physical action to abstract knowledge can be challenging for students. As a consequence our system needs to provide various "reflection stages" where users can step back and articulate their discoveries as formal concepts.
- **Autonomy:** traditional classroom exercises need a constant supervision from teachers. A more efficient system should provide automatic feedback and give users a strong sense of control and autonomy.

The next section describes how we addressed these requirements into the design of Phylo-Genie.

### Scenario

The Phylo-Genie environment presents users with a scenario that motivates the learning activity. Participants 'travel' to Australia as researchers to assist in data collection and analysis. The trip is executed as a series of nine stages. During the trip, a user is bitten by a venomous snake. The users then have to choose between 4 equidistant hospitals; each hospital has only one type of anti-venom. They have time to reach only one hospital before the venom irreversibly affects the bitten user. Since the species of the snake is unknown, participants must learn tree-building techniques to assess the common ancestry of the Australian snakes. Closely related snakes share the same venom (and thus anti-venom). Successful tree construction and interpretation allows users to select a treatment. This scenario, anchored in a real life setting, provides participants with motivation for learning phylogenetics.

We developed a scenario that progresses in several stages to introduce and reinforce fundamental phylogenetic concepts in an engaging manner. First, participants receive and review several vocabulary notes and information in preparation for their trip to Australia. In Figure 3, users are introduced to the fundamental concepts and snake organisms by exploring and discussing 'sticky note' items. Phylo-Genie utilizes the metaphor of sticky notes to avoid overwhelming users with textual information: notes can easily be removed from the display with a swipe gesture, and retrieved by touching the corresponding area of the surface.



Figure 3. Exploration of basic concepts in phylogenetics, organized by virtual 'sticky notes'.

### Organizing Heterogeneous Information

When constructing phylogenetic trees, scientists often involve more than a dozen variables. In this system, Phylo-Genie allows users to directly manipulate and explore the information relevant to each snake specimen of question. Snakes are embodied by small physical tokens and can be dropped on the table to access snakes' characteristics. Tokens are randomly distributed around the table at the beginning of the activity; this implicitly supports a sense of territoriality among participants, since they 'own' the tokens on their side of the surface. In addition, those tangibles facilitate sharing of data between partners because of their physical nature.

Figure 4 shows an image of simultaneous exploration of the information for three different snakes. Users fill in boolean values in a character table (right) to aggregate information for all snakes. Organizing the table allows users to create a matrix in a row echelon form; this, in turn, facilitates the tree building process. Snakes sharing a higher no. of characteristics are most closely related, while the species sharing only one or no feature with others represents the outgroup (Note that this is a simplified learning example.).

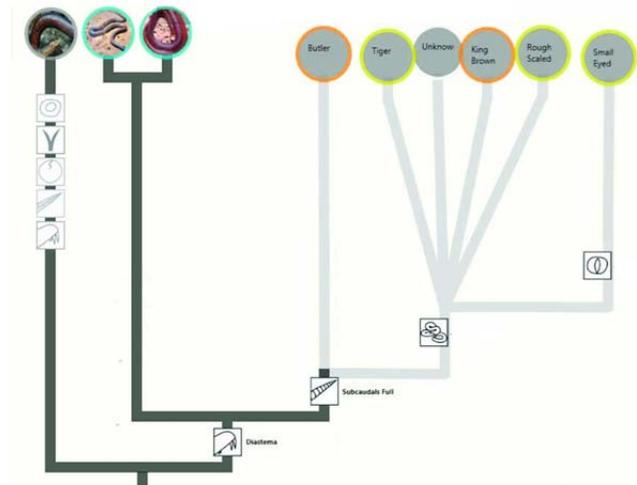
It should be noted that computational tools usually calculate these values for users; however, we propose that the physical completion of the task facilitates a connection between collecting real-world data and building a phylogenetic tree.

### Constructing and Interpreting Scientific Representations

Phylo-Genie takes the users through two stages of construction, each employing one of two common methodologies: morphology-based and genomics-based. To build a tree, which shows the evolutionary relationships of 10 Australian snakes, participants place tangible tokens that represent species and character tokens that represent common traits.



**Figure 4. Collection and organization of morphological information of the Australian snake specimen.**



**Figure 5. Construction of a phylogenetic tree based on the collected morphological information.**

In the first step, the morphology-based method is used to build a basic phylogenetic tree. Since this tree does not provide enough information to resolve the evolutionary relationship amongst all snakes, users are motivated to learn genomic-based methods to choose the appropriate treatment for the bite.

Figure 5 shows a constructed tree based on morphological characteristics. The outgroup, the species most ancestral in the constructed tree, is shown (left) with all ancestral versions of the traits. In the Phylo-Genie scenario, traits refer to various characteristics of the snakes: being ovipara or vivipara (do they lay eggs or give live birth), having round or ellipse-shaped eyes, etc. Divergence is marked by the trait icons. The parts of the tree in dark gray are presented to the user prior to construction, and the items in light gray are those added by the participants.

Following each tree-building stage, participants reflect on the tree previously built to select an anti-venom for the bitten partner. This requires users to interpret the structure of the tree to derive and understand the historical relationships of the corresponding species. These steps are crucial for knowledge-building since they encourage users to reflect on their experience and formulate formal concepts. Reflection is encouraged by having users “step back” from their previous task (organizing information and building a tree): a new screen is presented, where users access the previously built tree in a non-editable format. This forces users to leave the “concrete mode” as defined by Clements [12], where manipulating objects is the focus of the activity and move to an “abstract mode” where they are required to think in abstract terms in order to correctly interpret the phylogenetic tree.

### Interacting with Physical Tokens

In Phylo-Genie, students interact with physical tokens that represent various species of snakes. Users manipulate those tokens to accomplish two key actions: organizing

heterogeneous information and building phylogenetic trees. In the first case, positioning a token on the surface will display the snake's characteristics and its outward appearance. Thus, the token becomes an embodiment of the snake and a self-contained object: information is easily retrievable by dropping a token on the table. Moreover, tokens are used to build the tree in the next step. By sliding a snake from a branch of the tree users can extend the given template and use the collected data to build their own phylogenetic tree. We designed those tokens as a way to foster collaboration: users implicitly "own" the tangibles situated on their portion of the table. Ownership, in turn, helps users to coordinate tabletop interactions [44].

### Implementation

Phylo-Genie is written in C# using the Microsoft Surface SDK V1. The morphological and genomic information is adapted from Wallach [50] and material and exams from a course on Introductory Biology. Tangible objects are tagged with the Microsoft Surface fiducial tags.

### EVALUATION

To evaluate Phylo-Genie's strengths and limitations in supporting collaborative learning of phylogenetics we conducted a between-subjects experiment with 56 undergraduate and graduate students. We compared the system implemented on a multi-touch tabletop to a traditional pen and paper implementation of the Phylo-Genie scenario. The rationale for this choice was to conduct an ecologically valid comparison of our system as pen and paper is the premier media used for teaching in college settings. In this study, we examined the similarities and differences of the two implementations in terms of quantitative performance and qualitative behavior.

### Participants

56 undergraduate and graduate students (28 female, 28 male, average age = 21.28, SD = 3.70) volunteered to participate in the study, composing 28 dyads. None of the participants had received college-level instruction in evolutionary biology before the study. We also pre-screened students' knowledge in biology prior to the session with a basic 5-question quiz on phylogenetics; no participant correctly answered more than 2 questions.

### Procedure

A between-subjects study design was used, in which 28 dyads were randomly assigned to one of the two experimental conditions. Sessions were held in two private laboratory spaces. Both spaces and conditions maintained the same spatial setup. Subjects received a 5-minute introduction on phylogenetics before the study task. Upon completion of the Phylo-Genie activity, participants individually filled out three post-task questionnaires: the NASA-TLX (a measure of the users' workload [24]), standard engagement questionnaire [35] and a test on phylogenetics (written by the same professor whose teaching materials were used to construct our learning activity). The session concluded with a short debriefing and

the participants received \$10 for their participation. Each session lasted approximately 60 minutes. All sessions were videotaped and observed by a researcher.

### Paper Activity

The paper implementation of our learning scenario is identical to the tabletop implementation and contains similar stages and tasks. Instead of interactive screens users proceed through the pages of a booklet. Table 1 shows the differences between experimental conditions. The workspace was the same in both conditions. The display size in the paper condition was constrained to 8" x 11" pieces of paper; however, relative image and font sizes were consistent across conditions. Users accessed information by flipping pages of the booklet. In addition they built phylogenetic trees via traditional material (pen and paper).

|                             | Paper                    | Tabletop    |
|-----------------------------|--------------------------|-------------|
| <b>Workspace</b>            | 42"                      | 42"         |
| <b>Display size</b>         | 8"                       | 30"         |
| <b>Display orientation</b>  | Horizontal               | Horizontal  |
| <b>Interaction style</b>    | Bimanual, multiple users | Multi-touch |
| <b>Input</b>                | Direct                   | Direct      |
| <b>Tangible Interaction</b> | No                       | Yes         |
| <b>Tutorials</b>            | Paper                    | Video       |

Table 1. Differences across experimental conditions.

### Results

#### Learning Performance

We assigned a score on a scale of 0 to 100 to each subject's test based on the correctness of their answers (see Figure 6). We did not use a pre-test since none of our participants had received college-level instruction in phylogenetics. A between-subjects ANOVA of the learning performances from the two conditions showed that the subjects in the tabletop condition (mean = 90.3, SD = 11.3) obtained statistically significantly higher scores on the test than those in paper condition (Mean = 78.6 and SD = 15.8):  $F(1,54) = 10.11, p = 0.002$ .

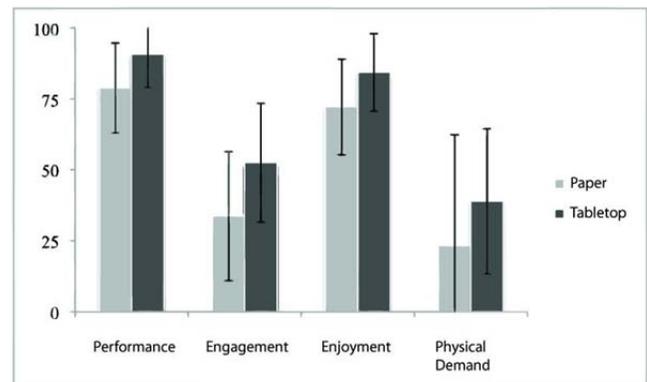


Figure 6. Average ratings of overall attitude measures and performance

### Engagement

Using a five-step bipolar scale, participants gave a score from -2 to 2 on 56 questions adapted from O'Brien's measures of engagement [35]. This provides an estimate of users' engagement by aggregating answers on the following dimensions: focused attention, perceived usability, aesthetics, endurance, novelty and involvement. Ratings were normalized to a percentage out of 100 (Figure 6).

Overall, participants rated the task as being more engaging in the tabletop condition compared to the paper condition ( $F(1,54) = 10.31, p = 0.002$ ). More specifically, participants gave higher ratings for the following factors: *Felt Involvement* - "how much fun users' were having during the interaction and how drawn in they were able to become" ( $F(1,54) = 11.84, p = 0.001$ ), *Endurability* - "assessment of users' perception of success with a task, and their willingness to use an application in future or recommend it to others" ( $F(1,54) = 5.14, p = 0.027$ ) and *Aesthetics* ( $F(1,54) = 45.65, p < 0.001$ ). Finally, it should be noted that the difference in engagement was not influenced by a novelty effect ( $F(1,54) = 2.08, p = 0.16$ ). Ensuring that subjects did not learn only because they were intrigued by the novelty of the interface is important.

### Task Load

Task load was measured using a questionnaire adapted from Hart et al.'s NASA-TLX [24]. Using a twenty-step bipolar scale, participants gave ratings from 0 to 100 for each item of the questionnaire. A between subjects ANOVA of the task load's sub-dimensions indicate statistically significantly higher ratings of Engagement and Enjoyment on the tabletop condition ( $F(1,54) = 10.31, p = 0.0022$  and  $F(1,54) = 8.77, p = 0.0045$  respectively). Physical demand was also found to be significantly higher in the tabletop condition ( $F(1,54) = 6.38, p = 0.015$ ). This is probably because questions of this dimension were formulated in terms of computer interface and not pen and paper activity (e.g. "How much physical activity was required (e.g. pointing, typing, pressing, turning, dragging, rotating)?"). It should be noted that questions from other dimensions were formulated in more general terms.

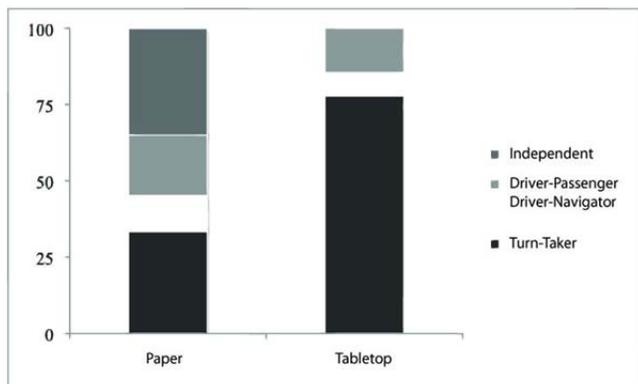


Figure 7. Average collaboration profile relative to total task stages.

### Collaboration

Based on the collaboration profiles developed by Shaer et al. [45] we coded each dyad at each stage of the task, for nine total ratings per pair. A second judge double-coded 20% of the data; inter-reliability index using Krippendorff's alpha was 91%. An alpha higher than 80% is considered as a reliable agreement between judges [25]. Categories are defined as follows:

- Independent: users are absorbed in their own activity; there is a minimal verbal communication.
- Driver-Passenger: the driver is fully engaged; the passenger is not focused on the task.
- Driver-Navigator: both users are engaged. The navigator contributes with suggestions and observations.
- Turn-Taker: both users make and accept suggestions and observations.

The results of the average collaboration profile relative to the total number of task stages are shown in Figure 7.

Analysis of the profiles by total stages reveals the dyads in the paper condition were classified as Turn-Taker in only 33% of the stages in comparison to 78% for those assigned to the tabletop. Furthermore, an average 37% of the stages were classified as Independent in the paper condition in comparison to 0% of stages in the tabletop condition.

We further analyzed dyadic collaboration using the rating scheme developed by Meier et al. [31]. Each subject pair was assigned post-task a rating from -2 to 2 in nine process dimensions. Inter-reliability index using Krippendorff's alpha was 83% [25]. Figure 8 shows the results of the five aspects (aggregated from the nine process dimensions) of overall collaboration.

A between subject ANOVA of the overall measures revealed that the collaboration aspects of Dialogue management, Information Pooling, Technical Coordination, and Individual Task Orientation were all statistically significantly higher in the tabletop condition ( $F(1,26) = 3.91, p = 0.0254$ ;  $F(1,26) = 10.46, p = 0.0033$ ;  $F(1,26) = 12.61, p = 0.0015$ ;  $F(1,26) = 7.93, p = 0.0091$  respectively).

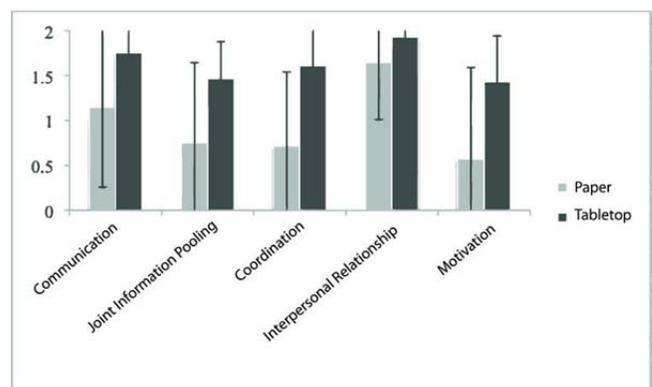
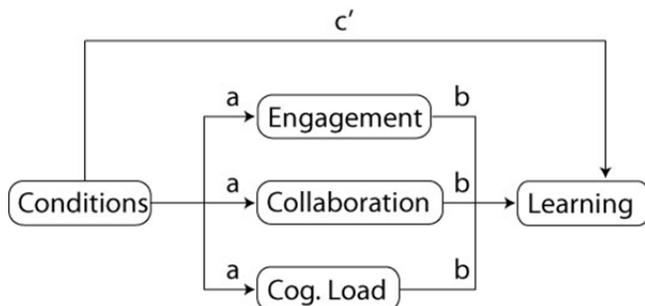


Figure 8. Average overall collaboration ratings by type.

It means that users in the tabletop condition had a smoother flow of communication (e.g. where turn-taking was facilitated by small questions like ‘what do you think?’), they tried more intensively to collect and share pieces of information and better coordinated themselves. For instance, where participants in the paper activity tended to silently read the instructions, users in the tabletop condition actively shared information and spent more time discussing the best way to solve the problem at hand.

#### *Mediatory Effect of Collaboration and Engagement on Learning*

We conducted a mediation analysis [38] to isolate the factors associated with positive learning gains (see Figure 9). A mediation model assumes the existence of one or several variables (called “mediators”) between the dependent (DV) and independent measure (IV). A mediational model assumes that the IV causes the mediator variable, which in turns causes the DV. Confidence intervals containing zero are interpreted as not significant; in other words, non-significant results mean that there are no causal relationships between the IV, the mediator and the DV. Our results show that a strong collaboration acts as a mediator between the two conditions and learning (CI:[0.074,5.95]), while engagement (CI:[-3.85,6.31]) and cognitive load (CI:[-0.66,5.54]) do not. If we consider collaboration as a dependent variable, our analysis shows that engagement acts as a positive mediator for an efficient collaboration (CI: [-0.75,-0.067]).



**Figure 9. Model for the mediatory effects. Our results show that in our study only collaboration acts as a mediator for learning.**

## DISCUSSION

In this study we examined the similarities and differences of the two implementations in terms of quantitative performance and qualitative behavior. Our results suggest that our proposed design guidelines promoted a deeper understanding of phylogenetics.

More specifically, four guidelines shaped the design of our system: firstly, we focused on developing an activity that engaged and motivated students. Interactive technology helped reach this goal. We created rich and viscerally compelling contents that were not available in paper format (e.g. immersive videos, interactive contents, animated

tutorials). The engagement questionnaire suggests that students felt more involved, viewed the learning experience more successful, and rated the task as being more aesthetic. Secondly, we supported collaboration through physical tokens and tabletop implementation embodying the main actors of this scenario. The goal was to reinforce users’ ownership of both physical objects and areas on the interactive surface. We interpret the more balanced division of labor between users as an indirect evidence for this hypothesis: in the tabletop implementation, participants took more turns and spent more time collaborating (e.g. pooling information, coordinating activities) rather than working independently. The results are supported by other studies showing that territoriality plays an important role in tangible learning environments [27,44]. Thirdly, we dedicated a significant part of the learning activity for students to reflect on the content. Even though the paper activity also offered this opportunity, we argue that switching from a very active role (interacting with physical tokens on a multi-touch surface) to a more reflective stage (articulating discoveries as formal concepts) provided better opportunities for understanding and refining phylogenetic concepts. Fourthly, we tried to provide a sense of control and autonomy to our participants by building automatic feedback and scaffolding in our activity. Although we did not devise any measures for this dimension, we believe that this principle played a major role in the learning process; in traditional classrooms, students often have a passive role and depend on the teachers for advancing in the domain taught. In Phylo-Genie, users control the pace of the activity and decide when to move to the next step without any social pressure. Future studies should develop metrics to measure this construct and investigate how it helped students' learning.

In summary, the four guidelines helped develop a strong learning activity. Additionally, we observed that the learning process took an indirect path. The mediation analyses revealed that collaboration was the strongest factor in supporting knowledge-building and that engagement acted as mediator for a productive collaboration. Thus, Phylo-Genie succeeded in supporting learning by increasing users’ collaboration, which in turn was improved by a high engagement. It suggests that interactive environments can and should support learning in very diverse ways.

## Limitations

There is one concern in using new technologies for education: a novelty effect may be responsible for the learning gain observed (e.g. students are more involved because they are intrigued by the interface). The main issue is that a novelty effect fades over time and brings the learning gain to the same level as with regular activities. Our results suggest that this is not the case for our system. More specifically, the “Novelty” dimension of the engagement questionnaire was not found to be significantly different across conditions. However, future studies should investigate this issue by using longitudinal measures.

This study has limitations: first, we did not take into account users' preconceptions; we assumed that the diversity of their pre-existing knowledge would be randomly distributed across both conditions. Future studies should better control for this dimension. Moreover, our evaluation of participants' knowledge only considered a few phylogenetic concepts and did not address the breadth and richness of evolutionary theory. In addition to assess students' knowledge more completely, future studies should investigate whether discovery learning activities on tabletops improves students' transfer of knowledge (e.g. in situations that seem different from our activity but share the deep structure of the base concept [18]).

### CONCLUSIONS AND FUTURE WORK

While the domain of phylogenetics provides the frame for this work, this research makes four contributions to HCI in general and to tabletop interaction in particular: 1) we identify requirements and design principles for fostering collaborative learning through tabletop interaction; 2) we designed and implemented an interactive system, Phylo-Genie, which employs multi-touch and tangible interaction techniques to support the learning of college-level phylogeny; 3) we present findings from an experimental study that show that Phylo-Genie promotes engagement, collaboration, and learning to a greater extent than traditional learning tools (i.e. pen and paper); 4) we discuss implications for the design of tabletop user interfaces for promoting collaborative learning.

Future studies should explore the benefits of tabletops combining multi-touch and tangible interaction for longitudinal and less structured tasks and observe how students' thinking is impacted over longer periods of time.. Moreover, we plan to expand and explore Phylo-Genie's functionality to further support tree building and engagement in science learning through continued collaborations with domain scientists. We also intend to integrate Phylo-Genie into undergraduate biology courses for longitudinal measures of learning.

### ACKNOWLEDGMENTS

Special thanks go to Professor Jonathan Losos (Organismic and Evolutionary Biology, Harvard University) whose teaching and exam materials were used as part of our learning activity. We thank our science advisors Hopi Hoekstra from the Department of Organismic and Evolutionary Biology at Harvard University, Andrea Sequiera and Emily Buchholtz from the Department of Biological Sciences at Wellesley College. We also thank students Courtney Stepan from the Committee on Evolutionary Biology at University of Chicago, and Harriet Alexander from the WHOI program at Massachusetts Institute of Technology. This work was partially support by a grant from the Encyclopedia of Life Learning + Education group (<http://education.eol.org/>) and the Museum of Comparative Zoology at Harvard University. Finally, we

thank the instructors and students that participated in our study.

### REFERENCES

1. Ado, L. Ahmet, Z., Jonsson, M., Sumon, S.I., and Holmquist, L.E. Supporting embodied exploration of physical concepts in mixed digital and physical interactive settings. In *Proc. TEI '11*.
2. Anderson, M. L. Embodied cognition: A field guide. *Artificial Intelligence 149, 1* (2003), 91–130.
3. Antle, A.N., Bevans, A., Tanenbaum, J., Seaborn, K., and Wang, S. Futura: design for collaborative learning and game play on a multi-touch digital tabletop. In *Proc. TEI '11*.
4. Antle, A.N., Droumeva, M., and Ha, D. Hands on what?: comparing children's mouse-based and tangible-based interaction. In *Proc. IDC '09*.
5. Baum, D.A., Smith, S.D., and Donovan, S.S. The Tree-Thinking challenge. *Science 310, 5750* (2005), 979–980.
6. Beeland, W.D. Student engagement, visual learning and technology: Can interactive whiteboards help. In *Annu Conf of the Assoc of IT for Teaching Edu '02*.
7. Bolchini, D., Finkelstein, A., Perrone, V., and Nagl, S. Better bioinformatics through usability analysis. *Bioinformatics 25, 3* (2009), 406–412.
8. Carini, R.M., Kuh, G.D., and Klein, S.P. Student engagement and student learning: Testing the linkages\*. *Research in Higher Education 47, 1* (2006), 1–32.
9. Catley, K.M., and Novick, L.R. Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience 58, 10* (2008), 976.
10. Catley, K.M., and Novick, L.R. Digging deep: Exploring college students' knowledge of macroevolutionary time. *J of Research in Science Teaching 46, 3* (2009), 311–332.
11. Clement, J. The use of analogies and anchoring intuitions to remediate misconceptions in mechanics. (1987).
12. Clements, D.H. 'Concrete' Manipulatives, concrete ideas. *Contemporary Issues in Early Childhood 1, 1* (2000), 4560.
13. Corno, L., and Mandinach, E. The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist 18, 2* (1983), 88–108.
14. Dillenbourg, P., Baker, M., Blaye, A., and O'Malley, C. The evolution of research on collaborative learning. In *E. Spada & P. Reiman (Eds) Learning in Humans and Machine: Towards an interdisciplinary learning science*. (1995), 189–211.
15. Evans, E.M., Spiegel, A.N., Gram, W., Frazier, B.N., Tare, M., Thompson, S., and Diamond, J. A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching 47, 3* (2010), 326–353.
16. Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Boetschi, K., Voegtli, B., and Juchli, P. Tangible user interface for chemistry education: comparative evaluation and re-design. In *Proc. CHI '07*.
17. Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., and Bonnett, V. Actions speak loudly with words: unpacking collaboration around the table. In *Proc. ITS '09*.
18. Gick, M.L., and Holyoak, K.J. Schema induction and analogical transfer. *Cognitive Psychology 15, 1* (1983), 1–38.

19. Gillet, A., Sanner, M., Stoffler, D., Goodsell, D., and Olson, A. Augmented reality with tangible Auto-Fabricated models for molecular biology applications. In *Proc. IEEE Visualization Conference '04*.
20. Glaeser, T., Franke, J., Wintergerst, G., and Jagodzinski, R. Chemieraum: tangible chemistry in exhibition space. In *Proc. TEI '09*.
21. Gregory, T. R. Understanding evolutionary trees. *Evolution: Education and Outreach* 1, 2 (2008), 121–137.
22. Halverson, K. L. Improving Tree-Thinking one learnable skill at a time. *Evolution: Educ and Outreach* 4, 1 (2011), 95–106.
23. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., and Rogers, Y. Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? *Proc. CSCL '09*.
24. Hart, S., and Staveland, L. Development of NASA-TLX (Task load index): Results of empirical and theoretical research. In *Advances in Psychology*, 52 (1988), 139–183.
25. Hayes, A.F. and Krippendorff, K. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1, (2007), 77-89.
26. Horn, M. S., Tobiasz, M., and Shen, C. Visualizing biodiversity with voronoi treemaps. In *Proc. IEEE Symposium on Voronoi Diagrams in Science and Engineering '09*, 265–270.
27. Jermann, P., Zufferey, G., Schneider, B., Lucci, A., Lepine, S., and Dillenbourg, P. Physical space and division of labor around a tabletop tangible simulation. In *Proc. CSCL '09*.
28. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.(2007). Clustal W and Clustal X 2.0. *Bioinformatics*, 23, 2947-2948.
29. Lou, Y., Abrami, P. C., and d'Apollonia, S. Small group and individual learning with technology: A Meta-Analysis. *Review of Educational Research* 71, 3 (2001), 449–521.
30. Maddison, W.P., Maddison, D.R. Mesquite: A modular system for evolutionary analysis. *Evolution* 62, 5 (2005), 1103-1118.
31. Meier, A., Spada, H., and Rummel, N. A rating scheme for assessing the quality of computer-supported collaboration processes. *J of CSCL* 2, 1 (2007), 63–86.
32. Meir, E., Perry, J., Herron, J. C., and Kingsolver, J. College students' misconceptions about evolutionary trees. *The American Biology Teacher* 69, 7 (2007), e71–e76.
33. Novick, L. R., and Catley, K. M. Understanding phylogenies in biology: The influence of a gestalt perceptual principle. *Journal of Experimental Psychology: Applied* 13, 4 (2007), 197–223.
34. Novick, L. R., Shade, C. K., and Catley, K. M. Linear versus branching depictions of evolutionary history: Implications for diagram design. *Topics in Cog. Sci.* 3, 3 (2011), 536–559.
35. O'Brien, H. L., Toms, E. G., Kelloway, E. K., and Kelley, E. Developing and evaluating a reliable measure of user engagement. *Proc. American Soc. for IS&T* 45, 1 (2008), 1–10.
36. Piaget, J. *The language and thought of the child*. Psychology Press, 1998.
37. Piper, A.M., and Hollan, J.D. Tabletop displays for small group study: affordances of paper and digital materials. In *Proc. CHI '09*.
38. Preacher, K. J., and Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 40, 3 (2008), 879.
39. Roschelle, J., Tatar, D., Chaudbury, S. R., Dimitriadis, Y., Patton, C., and DiGiano, C. Ink, improvisation, and interactive engagement: Learning with tablets. *Computer* 40, 9 (2007), 42–48.
40. Salomon, G., and Globerson, T. When teams do not function the way they ought to. *J of Educ Research* 13, 1 (1989), 89–99.
41. Sandvik, H. Tree thinking cannot be taken for granted: challenges for teaching phylogenetics. *Theory in Biosciences* 127, 1 (2008), 45–51.
42. Schkolne, S., Ishii, H., and Schroeder, P. Immersive design of DNA molecules with a tangible interface. In *Proc. IEEE Visualization Conference '04*.
43. Schneider, B., Jermann, P., Zufferey, G., and Dillenbourg, P. Benefits of a tangible interface for collaborative learning and interaction. *IEEE Trans.on Learning Technologies PP*, 99, 1–1.
44. Scott, S.D., Sheelagh, M., Carpendale, T., and Inkpen, K.M. Territoriality in collaborative tabletop workspaces. *Proc. CSCW '04*.
45. Shaer, O., Kol, G., Strait, M., Fan, C., Grevet, C., and Elfenbein, S. G-nome surfer: a tabletop interface for collaborative exploration of genomic data. In *Proc. CHI '10*.
46. Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., and Wang, H. Enhancing genomic learning through tabletop interaction. In *Proc. CHI '11*.
47. Smith, J., diSessa, A., and Roschelle, J. Misconceptions reconceived: A constructivist analysis of knowledge in transition. *J. of the Learning Sciences* 3, 2 (1994), 115–163.
48. Matuk, C.F. *Animated cladograms: Interpreting evolution from diagrams*. In: G. Stapleton, J. Howse, and J. Lee, (Eds.) *Diagrams 2008 - Diagrammatic Representation and Inference - 5th International Conference 2008*, 395-397.
49. Vygotskij, L. S. *Mind in society: the development of higher psychological processes*. Harvard University Press, 1978.
50. Wallach V. A cladistic analysis of the terrestrial Australian elapidae. In: Grigg G, Shine R, Ehmann H, eds. *Biology of the Australasian Frogs and Reptiles*. Sydney: Surrey Beatty and Sons, 223–253, 1985