Using Project-Based Learning to Teach Phylogenetic Reconstruction for Advanced Undergraduate Biology Students: Molluscan Evolution as a Case Study

Andrew A. David

Abstract
Phylogenetics plays a central role in understanding the evolution of life on Earth, and as a consequence, several active teaching strategies have been employed to aid students in grasping basic phylogenetic principles. Although many of these strategies have been designed to actively engage undergraduate biology students at the freshman level, less attention is given to designing challenges for advanced students. Here, I present a project-based learning (PBL) activity that was developed to teach phylogenetics for junior and senior-level biology students. This approach reinforces the theories and concepts that students have learned in their freshman courses along with incorporating Bioinformatics, which is essential for teaching zoology in the 21st century.

Key Words: systematics; morphology; discussion; pedagogy; active; learning.

Introduction
Phylogenetics is the study of the evolutionary relationships of individual species and groups (Nei & Kumar, 2000). The discipline plays a central role in the understanding of modern evolutionary theory but also extends to other fields such as DNA barcoding, phylogeography, and conservation biology (Avise, 1989; Hajibabaei et al., 2007). Undergraduates in most biology programs across the United States are usually introduced to phylogenetics early in their introductory coursework with technical terms such as “monophyly,” “parsimony,” and “synapomorphy/symplesiomorphy” becoming staple terms in their vocabulary by the end of their freshman year. Students are also expected to have covered the “morphological versus molecular tree” debate in which themes such as convergent evolution and homology are reinforced. In upper-level biology courses, where advanced students are expected to delve into the literature in detail, many instructors often assume that their class has at least a basic understanding of phylogenetics and is able to interpret the topology of a phylogenetic tree if one is included in an assigned research paper. However, previous studies have shown that even advanced biology students often harbor a variety of misconceptions about building and interpreting phylogenetic trees (Lents et al., 2010).

The obvious solution would be to implement more effective pedagogical techniques in introductory biology classes for teaching phylogenetics. Multiple case studies have attempted this, and the use of tree-building exercises have been shown to improve students’ abilities to read and interpret tree topologies (Eddy et al., 2013). In this paper, I describe a project-based approach to teaching phylogenetic reconstruction. Project-based learning (PBL) consists of a broad range of pedagogical strategies that use projects as a central component. These projects extend over a period of time, vary in complexity based on the students’ aptitude, and actively engage them in design, problem-solving, investigative activities and decision making (Blumenfeld et al., 1991; Bell, 2010). This strategy facilitates autonomous work, stimulates critical thinking, and often culminates in a final realistic product such as a presentation or report. A major component of PBL is instructor feedback at crucial intervals during the course of the project. It is during these intervals that the important processes of reflection and recalibration occurs, which in turn, facilitate deep learning.

Significance of the Mollusca as the Focal Phylum for Investigation
In this project we use the phylum Mollusca as a case study. Molluscs are the second most speciose group of animals on our planet, consisting of more than 85,000 extant species with many more represented in the fossil record. Even more impressive is their ecological and physiological diversity and the fact that the diverse body plans observed across the different families is believed to have evolved from a basic ground pattern. Despite a series of systematic and phylogenomic studies in the late 1990s and 2000s, the monophyly of the group remains a contentious issue. Furthermore, some important questions on molluscan evolution remain unanswered. For example, does the shell-less condition of the Aplacophora represent an ancestral condition? This question is critical for molluscan evolution as it could shed light on the origin of the molluscan shell and
whether it is truly an ancestral trait for this group. Also, how can one explain the extreme divergence of many cephalopods such as squids and octopuses, whose complex sensory structures like as eyes and advanced cognitive capabilities place them in stark contrast to their gastropod (snails) and bivalve (clams, oysters, scallops) cousins? The Mollusca therefore represents an ideal phylum for zoology students to investigate through PBL as it allows them to address broad evolutionary questions through phylogenetic reconstruction of highly diverse taxa.

Methods

This project was developed as part of a student-centered introductory zoology course geared toward advanced biology students. The class size was 16, comprised of juniors and seniors. Pre-requisite coursework included introductory-level molecular biology along with ecology and evolution. It is advisable that instructors complete their lectures on molluscs prior to assigning this project. The project was carried out over four class periods of 75 minutes each: Day 1, prep stage; Day 2, morphological analysis; Day 3, molecular analyses; Day 4, presentation and wrap up (Figure 1). On Day 1 students were organized into groups of four, briefed on the project (~20 minutes) and watched a 15-minute video on Mollusca (https://www.youtube.com/watch?v=xKjeJlDBCQ). The debriefing should consist of an objective for the entire class—e.g., “The objective of this project is to elucidate the evolutionary relationships among selected molluscan taxa using morphological and molecular data”—a timeline for completion (Figure 1), and a grading rubric, along with expectations and a list of deliverables for the project. Students are then instructed to organize into groups (five to six students per group worked well in my class) and choose their group leader. Leaders designate tasks and must be willing to accept responsibility for completion of the project. After assigning this project in 2015 and again in 2017, I found that the most productive group configuration involved having two students complete the morphology tree (one student builds the character matrix, the other constructs the tree), two students on the molecular tree (one student mines and compiles sequences, the second edits, aligns, and builds the tree), and a fifth student who mines the literature and helps the group leader with the progress report and presentations. The progress report will include a summary of each group member’s contribution for each time the group meets (either outside or during class), problems that arose, and possible solutions. This configuration controls for “free-riding,” ensuring that all students in a group contribute to the final product, however it does incur “transaction costs” as groups must also meet outside the classroom (Yamane, 1996).

On Day 1, groups are also given eight molluscan taxa representing six classes and an outgroup taxon to root their trees (Table 1). In this project, students used the annelid Chaetopterus caudus as an outgroup because phylogencic analyses have placed annelids as the sister group to Mollusca (Kocot et al., 2011). It should be noted that in Table 1 some classes (e.g., Bivalvia) are represented by more than one species. The purpose of this is to reinforce the concepts of “sister taxa” and monophyly by having students observe the bifurcating branches that they will recover from their analyses.

Morphological Tree Construction

On Day 2, each group must choose at least five phylogenetically informative traits to be used in creating their character matrix for their morphology tree (Table 2). The simplest way to execute this task is to have students review lecture notes and their textbook and search for traits that distinguish the higher molluscan groups (e.g., gastropods vs. polyplacophorans, cephalopods vs. aplacophorans, etc.). As these molluscan groups are very distinct, there are a variety of traits from the basic molluscan body plan that can be used for the morphological analysis. Once the character matrix is completed, student groups are then tasked with manually creating the most parsimonious morphological tree. Based on the level of difficulty desired, instructors can also provide a “trait bank” to help students or give a list of mandatory traits for use to ensure that all groups are using the same characters.

There are two options for constructing morphology trees: the first involves using software such as PAUP* (Swoford, 1993) or Mesquite (Maddison & Maddison, 2003), and the second is by hand. The problem with using PAUP* and Mesquite is that traits will need to be coded, which is not as straightforward as working with molecular data (Pleijel, 1995). However, if the instructor is familiar with coding morphological traits, then they may opt for this option, keeping in mind that the project duration will be extended to accommodate the time needed for students to familiarize themselves with additional software. For the sake of simplicity, groups in my course constructed their morphology trees by hand using the following steps:

Step 1: Choose at least five phylogenetically informative traits. Instructors should provide either a trait bank or have students choose their own traits, which they can discuss with the instructor prior to attempting tree construction. Students should be aware that increasing the number of traits increases the number of possible trees that can be produced.

Step 2: Create a character matrix using both the ingroup and outgroup taxa. The outgroup taxon will be least similar to all of the ingroup taxa and as such will be used to “root” the tree. Table 2 shows an example of a character matrix developed by one of the groups in the class (hereafter referred to as Group A).

Step 3: Construct the tree in such a way that you can see when each of the traits either develops or is lost, and ensure that organisms are grouped by shared traits. At this point instructors should

---

**Figure 1.** Summary of upper-level phylogenetics project for zoology course.
reiterate the philosophy of parsimony in constructing phylogenetic trees, i.e., the most preferred tree is the one with the least number of mutational steps/character state changes.

**Step 4:** Groups should be prepared to give a five-minute presentation of their morphology tree to which instructors can provide feedback. These presentations are not graded (although that can be left to the preference of the instructor) but are used to help students produce an improved phylogeny in the final presentation.

**Molecular Tree Construction**

On Day 3, each group is given GenBank accession numbers corresponding to each taxon (Table 1). For this project, I use archived sequences from the cytochrome c oxidase I (CO1) gene due to its ubiquity in phylogenetic analyses and the short sequence lengths (400–700 bp), which is convenient for teaching purposes. Two Bioinformatics software were used, BioEdit (Hall, 1999) and MEGA (Tamura et al., 2007), both of which are available for free online and offer a graphical user interface (GUI) that allows for a user-friendly working environment. BioEdit is a very simple Windows-based program that possesses extensive and easy-to-use sequence editing capabilities, and MEGA is the preferred program for tree building in many undergraduate classrooms and compatible with both Windows and Macintosh operating systems (Newman et al., 2016). I created a simple one-page manual for using both programs but explicitly omitted troubleshooting, as having advanced students struggle with the computational glitches and coding issues that may occur is a crucial part of the learning process (especially considering that these struggles are also common among professional biologists!). Each group leader delegates the following tasks to specific group members:

1. Mine and compile the sequence data from GenBank using the accession numbers provided. Choose the FASTA format in GenBank, and copy and paste all the sequences into a text-editing file (Figure 2).

2. Use the ExPASY online translational tool to ensure gene functionality, i.e., ensure that each sequence can be translated into a functional protein (http://web.expasy.org/translate/). This is important as it ensures that the archived sequences being used are not pseudogenes and do not contain sequencing errors.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Molluscan Groups</th>
<th>Accession Number**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parachiton politus</td>
<td>Polyplacophora</td>
<td>HQ907881</td>
</tr>
<tr>
<td>Falcidens halanychi</td>
<td>Aplacophora</td>
<td>DQ371484</td>
</tr>
<tr>
<td>Crepidula fornicata</td>
<td>Gastropoda</td>
<td>KJ566746</td>
</tr>
<tr>
<td>Halotis midae</td>
<td>Gastropoda</td>
<td>AB236717</td>
</tr>
<tr>
<td>Crassostrea gigas</td>
<td>Bivalvia</td>
<td>AF280608</td>
</tr>
<tr>
<td>Mytilus edulis</td>
<td>Bivalvia</td>
<td>EU915580</td>
</tr>
<tr>
<td>Pusellum salishorum</td>
<td>Scaphopoda</td>
<td>AY260832</td>
</tr>
<tr>
<td>Loligo reynaudi</td>
<td>Cephalopoda</td>
<td>AF075406</td>
</tr>
</tbody>
</table>

*Outgroup taxon: Chaetopterus cautus (Annelida), GenBank Accession number KX896507.
**Students are given accession numbers on Day 3.

**Table 2. Example of Group A’s character matrix for in-group and outgroup taxa.**

<table>
<thead>
<tr>
<th></th>
<th>Trochophore larvae</th>
<th>CaCO₃ shell</th>
<th>Plated Shell</th>
<th>Radula</th>
<th>Torsion</th>
<th>Byssus</th>
<th>Gladius (internal shell)</th>
<th>Captacula</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. cautus (outgroup)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P. politus</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F. halanychi</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. fornicata</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H. midae</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. gigas</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M. edulis</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P. salishorum</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>L. reynaudi</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
introductory biology courses, students learn to translate the nucleotide sequence of a protein-coding gene into an amino acid sequence using the mRNA codon chart. In ExpASY, students copy and paste their edited sequences into a translation window, which executes the same process in much a shorter time. Students can confirm that the gene is functional if an open reading frame (ORF) is available.

3. The compiled text file with each sequence in FASTA format can then be imported into BioEdit for alignment and editing (Figure 3). Once imported, the Clustal W alignment tool should be selected to align the sequences. For simplicity, editing consists of eliminating gaps in the flanking regions of the sequences so a uniform size is recovered.

4. The edited sequence file can then be exported into a separate file, and then imported into MEGA to build a neighbor-joining (NJ) tree. In this exercise, I used a step-by-step instructional video provided by the National Institute of Health (https://www.youtube.com/watch?v=d_-NTsJDvn8) for building the NJ tree in MEGA (parameters are also specified in the video). The NJ method was used as opposed to the maximum-likelihood (ML) method because it is computationally faster and more accurate than the latter when dealing with smaller datasets, such as the one being used in this exercise (Tamura et al., 2004).

I have found it particularly useful to have one teaching assistant in the classroom to aid with sequence editing and technical glitches that may arise during the class activity. Alternatively, the instructor may secure a computer lab on campus for this activity, as it would allow for consistent hardware and software performance across all groups and thus less frustration on the part of both instructor and students.

Out-of-Classroom Activities

To facilitate Day 3 activities, I have found it useful to incorporate a flipped classroom approach. Prior to Day 3, I uploaded taped lectures along with PowerPoints, which students were required to watch and read. These materials explained the general practice of aligning and editing sequences, and background information on the different types of tree building such as the NJ and ML methods. In addition, the instructional video for using the MEGA software can be assigned as an out-of-classroom activity, which allows students to toggle with the settings and become familiarized with the program before attempting to build the molecular tree on Day 3. Students were encouraged to start forum threads on the course webpage if they had any questions, and either I or another student who understood the material would post responses to begin discussing the material in more detail. Extra-credit points were used as incentives for these out-of-classroom activities. Finally, all groups were required to meet outside the classroom to compare and reconcile their morphology and molecular trees and to provide an explanation for their results. Once these were accomplished, each group

---

Figure 2. Partial compilation of CO1 sequences of selected taxa in FASTA format using the notepad text editor.
was required to give a 15-minute presentation on the final day of the project (Day 4). Groups should also be prepared to meet outside class to prepare their presentations for Day 4. I have found it useful to arrange time slots to meet with groups to discuss any issues or problems they may have encountered during their final analyses.

**Oral Presentation**

If the instructor is not a trained zoologist or does not carry out research in phylogenetics, it may be helpful to students if a faculty member in this research area can be persuaded to sit in on the presentations. Evaluating student work at this point is completely up to the instructor, but for this zoology course, each group was required to present three deliverables in their oral presentations:

1. **Morphological and molecular trees**—Groups should present both trees along with the methodology used to construct them and the revisions that were made after instructor feedback and peer-review from their classmates.

2. **Interpret and compare both trees**—Groups should be prepared to discuss similarities and differences between their morphology and molecular trees. In addition, students should also be able to compare their own results against published studies, providing proper explanation for any differences observed.

3. **Caveats identified during the study**—This is perhaps the most critical aspect of the project since groups should be able to explain the limitations of their methodology that could have affected their results. For example, would altering one or two traits in their character matrix significantly alter the morphology tree? How different would the molecular tree be if a nuclear gene or multiple genes were used in the analysis instead of a single mitochondrial gene fragment? Here, the instructor may explain the difference between gene trees and species trees, and why that difference is important for interpreting molecular phylogenies.

**Results & Discussion**

Examples of Group A’s morphology and molecular trees are shown in Figures 4 and 5. In the morphology tree, the shell-less condition of the Aplacophora is ancestral, and as a consequence, it is likely that the ancestor to all molluscs did not have a shell. Their tree also placed cephalopods and gastropods as sister taxa along with the bivalves and scaphopods, but interestingly, they did not identify a synapomorphic trait that would justify the two clades. The molecular tree complemented the morphology tree by recovering a distinct clade of gastropods and bivalves; however, the Cephalopoda was more closely related to the bivalves than the gastropods in the

---

**Figure 3.** Compiled CO1 sequences in BioEdit before and after alignment (top and bottom panels, respectively).
molecular tree. In addition, a third clade consisting of Polyplacophoran and Aplacophoran specimens was recovered using the molecular data. Group A concluded that their molecular tree was more accurate as it "conformed more closely to the trees presented in the latest phylogenomic study on molluscs.

In terms of timing, the project was assigned after two days of lectures and labs on molluscan diversity. During the project, it is important to reinforce to students that each phylogenetic tree is a hypothesis because the dataset used to generate it is limited. This is obvious in the morphological part of the study, as students must manually draw different trees and choose the most parsimonious evolutionary scenario. In contrast, for the molecular part of the study, the tree-building algorithm built into MEGA automatically chooses the best possible tree. In this case, I often direct students to review the metadata generated during the heuristic search so they can observe the number of tree arrangements that have occurred before the program produces the "best possible" tree.

This project emphasizes the importance of using PBL to actively engage students in the scientific process by having them complete the procedures that were used to generate the results and theories that they learn about during lecture (Kolb & Kolb, 2005). In addition, we incorporated certain aspects of Bioinformatics into the project, such as DNA sequence mining and editing, along with multiple alignments. This is crucial training for undergraduates at all levels, as the future of the Biosciences is strongly associated with the field of Bioinformatics (David, 2017). One caveat of this project is that it was designed for a zoology course in which phylogenetics played a central role in the course curriculum. As many professors know, zoology is arguably the broadest discipline of all the biological sciences and can be taught in a variety of styles, with some instructors opting for alternative approaches to the phylogenetic-based framework. Furthermore, for this project to be worthwhile for students, it is imperative that they have a firm understanding of phylogenetic theory, which can be accomplished by reinforcing basic concepts in the early lectures and assigning key phylogenetic papers during the semester. Papers that were assigned that I considered to be most critical to zoology include a landmark review piece by Halanych (2004), who addressed the influence of molecular data on the tree of life, along with papers by Struck et al. (2007) on Annelid evolution, Kocot et al. (2011) on molluscan evolution, and more recently, a controversial but informative commentary by Halanych (2015) and Whelan et al. (2015), both of whom argue that the basal position on the tree of life should be occupied by ctenophores, not sponges—a hypothesis that is still met with considerable backlash by some zoologists.

Aside from organismal-centered courses, this exercise is also appropriate for any upper-level biology course where phylogenetic trees play a central role, e.g., Systematics, Bioinformatics, Population Genetics, and Evolution. In addition, this project could also be given as a challenge to high-performing students in introductory biology courses, but will need to be repackaged and executed in a different manner than is outlined in this paper. I would suggest that at the freshman or sophomore level, instructors should edit and align the sequences so that students would only be required to execute the NJ analysis in MEGA. The instructor should also omit the dense background information on NJ and ML methods of building phylogenetic trees, and should provide a one-page info sheet with biological information (physiological and ecological) on the ingroup taxa along with pictures of each taxon. Although we used molluscan evolution as a case study, the instructor could in theory choose any group providing that (a) enough taxonomic information is available to clearly distinguish the in-group taxa morphologically, and (b) DNA sequences for a specific genetic marker are available for each taxon on public online repositories that students can access.

○ Acknowledgments

Fruitful discussions at the National Academy of Sciences’ (NAS) Summer Institute workshop in June 2016 were instrumental in the development of this project. The financial support of the Biology Department at Clarkson University was also extremely helpful in designing a zoology course consistent with the NSF’s Vision and Change manifesto. Finally, this paper is dedicated to Clarkson’s Biology undergraduates, whose critical feedback over the past two years has played the biggest role in changing my view of how zoology should be taught in the 21st century.
References


ANDREW DAVID (adavid@clarkson.edu) is Assistant Professor and the Director of Freshman Biology in the Biology Department at Clarkson University (Box 5805), Potsdam NY 13699.