

Teaching DNA Barcoding for the Identification of Algae

Amber Tateno-Bisel^{*1} and Kauahi Perez²

¹ Department of Plant and Environmental Protection Sciences, University of Hawai‘i at Mānoa, Honolulu, HI, USA

² Department of Tropical Plant and Soil Sciences, University of Hawai‘i at Mānoa, Honolulu, HI, USA

* Corresponding author, ambaht@gmail.com

Abstract

Here we discuss the design and implementation of an introductory DNA Barcoding module that we developed for the University of Hawai‘i at Mānoa’s Science in Action Program, a two-week summer program that teaches high school students about Hawai‘i’s biodiversity. Students used the concept of characterization to explain the relationships among organisms using morphological, ecological, and molecular data. Additionally, students gained experience in the scientific practice of generating explanations by gathering multiple lines of evidence to support or refute a claim, linking claims with evidence, and presenting such claims in written and oral formats to identify unknown algae samples. During this activity, students also gained real-world research experience in the field of biodiversity research. We also discuss potential modifications for future iterations of this module.

Keywords: activity design, argumentation & explanation, characterization, collaboration, DNA

1. Introduction

The Institute for Scientist & Engineer Educators (ISEE) Professional Development Program (PDP) is a program for scientists and engineers at the early stages of their careers, with a primary focus on graduate students and postdocs (ISEE, 2020). Participants receive training on teaching inquiry and designing lesson plans through multi-day workshops, working collaboratively in a team to design an inquiry activity, and implementing their new teaching skills into practice in ISEE-affiliated educational programs or courses. One focus of the PDP is to teach participants how to become better facilitators, or instructors that guide the learning of stu-

dents. Facilitators provide opportunities and guidance that help students come to their own understandings of core concepts. The key is not telling students the answer, but to guide them to reaching the answer and learning for themselves. Through funding from the National Science Foundation and Thirty Meter Telescope, our team developed an inquiry activity that involved characterizing an unknown algae specimen primarily using DNA barcoding methodology and secondarily by using morphological or ecological data.

2. Venue and audience

Every year, the University of Hawai‘i at Mānoa Outreach College hosts the Science in Action Program, a two-week summer course that runs for three hours daily and is led by a course instructor. This program was designed to give high school students the opportunity to explore various fields of life sciences through hands-on activities and field trips. In 2015, our activity on DNA Barcoding (Table 1) was incorporated as a portion of a course that focused on the biodiversity of Hawai‘i's wetlands. A total of six students were enrolled, with three being local students from Hawai‘i and the others from the US mainland.

Table 1: Timeline of our DNA Barcoding activity.

Day	Activity	Timeline
1	Introduction to Barcoding Starter Activity Thinking Tools	10 min 30 min 1 h 30 min
2	Research Scenario Group Formation Focused Investigation Scaffolding Tool	1 h 30 min
3	Jigsaw Activity Synthesis Culminating Assessment	15 min 30 min

3. Goals for learners

We felt that the concept of characterization (content goal) and the practice of generating explanations (practice goal) go hand-in-hand and were fundamental concepts in the realm of science, so we chose to focus primarily on these two learning goals. Characterization is an extremely broad concept and is interpreted distinctly in different fields of Science, Technology, Engineering, and Math (STEM). For our activity, our content goal was for students to use the concept of characterization to explain relationships among organisms using molecular data and supplementing their claims with morphological data. In this case, identifying algae

using morphology was taught in an earlier portion of the course prior to our teaching activity. We wanted our students to understand that closely related species have similar DNA sequences and sometimes share similar morphological features, and that the characterization of a species is based on multiple lines of evidence.

Constructing explanations is an authentic research practice that enables students to justify their claims with appropriate lines of evidence and reasoning (McNeill & Krajcik, 2009). Furthermore, generating explanations is a practice that cuts across all disciplines in science, and is used to support or refute an explanatory account of a phenomenon (National Research Council, 2012). Thus, our practice goal was to engage students in constructing or generating explanations through a Claim, Evidence, and Reasoning (CER) framework (Metevier et al., 2022), and thereby support or refute their claims for identifying a sample of algae. By the end of our activity, we wanted students to be able to answer the question: “How do you know that this sample of algae is (species X)?”

As an additional goal, we wanted our students to get hands-on experience in doing real-world research. A part of this hands-on experience involved receiving training on how to use different molecular software programs or bioinformatics software (i.e., Chromas Lite, CAP3, Genbank, and Molecular Evolutionary Genetics Analysis or MEGA) to generate different kinds of data, running different phylogenetic analyses, and collaborating in teams to answer a research question. We also generated DNA barcode data directly from algae samples that students collected because we felt that students would feel more ownership if we used samples that they collected themselves, as opposed to simply downloading sequences from GenBank. We also believed this approach would be the best way to pique students' interest in the subject while simultaneously gaining a better appreciation for the over-all procedures involved in molecular research.

4. Description of our activity

Our activity spanned a total of three days as part of a two-week summer course for high school students. The general timeline is listed above in Table 1, whereas the activities for each class meeting are detailed below.

4.1 Prelude to our teaching portion

Prior to our DNA barcoding section of the course, students collected live samples of algae from Diamond Head Beach and learned how to identify various algae species based on morphological taxonomy. They also learned a basis on ecology and how habitat can create a favorable environment for the growth of certain algae species. Students were then taught by the course instructor how to conduct polymerase chain reaction (PCR) by extracting DNA from their collected algae samples and running gel electrophoresis. Then students selected their six samples to be genetically sequenced for use in our DNA barcoding activity.

4.2 Introduction (starter activity & thinking tools)

At the start of our activity on Day 1, we gave the students context about who we were, what fields of science we came from, our affiliation with the PDP, and what our teaching unit was going to entail so they would know their purpose and what was expected of them as learners. Formal instruction of our activity began with an explanation of the traditional classification system of taxonomy and how morphology is used to assemble organisms into different taxonomic ranks.

We designed a starter activity to get students thinking about how algae can be characterized using morphological characteristics. Morphological taxonomy includes the use of unique structures of an organism to characterize them into different families, genera, and species. In the field of Botany, some examples of this include similarities in shape, color, and size of different plant structures. How-

ever, nowadays it is almost commonplace to include molecular data in the characterizations of species.

Pictures of various algae species displaying different colors and unique structures were shown through PowerPoint. This presentation brought up the issue that characterization using morphology can sometimes be deceiving because certain traits, such as similar colors and shapes, are not suitable characteristics that distinguish all algae. We asked students for ideas about alternative ways to classify algae without using morphology to get them to think about how characterization is conducted in science nowadays, especially since molecular characterization is becoming commonplace. Also, students introduced another line of evidence that was unplanned in our lesson plan — ecological data. We did not know that the course instructor would be teaching this concept, but this worked out well by providing students with another means of identification. This encouraged students to understand our practice goal, and more so take ownership of gathering their own evidence since they came up with ecological data as a line of evidence on their own.

The remainder of the period was reserved for introducing the concept of DNA barcoding. The main purpose of DNA barcoding is to identify various species of organisms ranging from flora and fauna to microorganisms, by comparing relatively short segments of DNA from a gene or set of genes. DNA barcoding makes use of PCR to amplify many copies of a specific segment from DNA that is extracted from a sample. The DNA is then verified by gel electrophoresis and sent to a sequencing facility to obtain the DNA sequence (the “DNA barcode”) of a given sample. Once the sequence is obtained, the use of bioinformatics software to assess the properties of DNA sequences and comparing sequences with others in molecular databases is used to identify an organism. This is a method that students would be using to characterize an unknown sample of DNA. However, as with real-world research in species identification, DNA barcoding is simply a

means to an end and findings must be corroborated with other evidence. Thus, we wanted our students to understand that an assessment of multiple properties (i.e., incorporating morphological and/or ecological data) are required to truly characterize a given sample.

In order for our students to gain a better understanding of the processes involved with DNA barcoding, we showed two- to three-minute videos on the concepts of how PCR is used to amplify DNA segments and how DNA is sequenced. We also introduced our thinking tools, which were essentially step-by-step instructions on how to use different molecular software programs to evaluate DNA sequences. This was done as a guided in-class activity. Through guided tutorials, students learned how to edit DNA sequences from specimens they sequenced, search and retrieve sequences from a database, understand results from a database query, create alignments, generate phylogenetic trees using different algorithms, and interpret phylogenetic trees. Students would be using these thinking tools to gather evidence and generate explanations during the focused investigations the following day.

4.3 Focused investigations with scaffolding

Day 2 began with a brief review of the previous day's topic and then segued into their focused investigations. Student groups were composed of two students and each group was randomly assigned an unknown sequence of DNA. They were given a research scenario in which a colleague forgot to label the sample prior to sequencing, so they were tasked with characterizing an unknown sample of DNA based primarily on molecular data and supplemented with morphological and/or ecological data. The "unknown" samples belonged to one of the six algae samples that students selected for sequencing. Due to time constraints, we also provided them with a scaffolding tool (Appendix A) of questions that helped them organize their thoughts. We believed this scaffolding tool would aid in their practice of generating explanations using Claim, Evidence,

and Reasoning because it built upon the skills they acquired from our thinking tools and allowed them to direct their attention to the information that they would need to know for their focused investigations.

During the focused investigation, students used the thinking tools from the previous day to answer questions from the scaffolding tool. Once students felt that they successfully identified their unknown sample down to the species level, they were directed to an online algae database to retrieve morphological and/or ecological data of their presumed sample. The intention here was to aid students in gathering multiple lines of evidence to supplement their findings, which in turn would lead to proper characterization of their sample. We also encouraged them to refer back to the previous week's notes on how they identified algae samples based on morphology in hopes that they would link their molecular data to that of their morphological findings from the week prior.

4.4 Jigsaw activity (used for culminating assessment)

As a culminating assessment task for our learners, we had students present their findings based on the scaffolding tool we provided for their focused investigations. We elected to do a jigsaw style presentation since students could present in a comfortable group setting, versus a solo oral presentation in front of the class. Each student was paired with a non-group member to share their findings. We also developed simple content and practice rubrics to formally assess our learners' understanding of the concept and practice of our unit (Appendices B & C).

The content for these rubrics was made in a simple three-point design that was modeled after other rubrics we reviewed. We developed our particular scoring criteria based on scenarios that we expected our students would struggle with (i.e., using DNA sequence similarity as evidence of identity while ignoring morphological data), which mirrors scenarios that real scientists struggle with today. We

wanted our learners to partake in real-world research methodology and practices, and we used these rubrics to evaluate their understanding in hopes that they would achieve a deeper understanding that closely related organisms share similar DNA sequences and morphological features, and we wanted them to use Claim, Evidence, and Reasoning to show this.

Each team was assigned a facilitator, who would conduct assessments of each student's understanding of the concept of characterization and their ability to generate explanations using our content and practice rubrics, respectively. Each student's presentation began by a student making a claim about what genus and species they thought their "unknown" specimen belonged to based on molecular analyses. They supported their claim with evidence, which included showing where their specimen was placed in a phylogenetic tree they generated (Figure 1), and discussing bootstrap values and percent sequence similarity. To the extent that they could find morphological and ecological information on their putative specimens, students provided these data as further lines of evidence. Finally, students used reasoning to link their claims with their evidence. At the end of each student's

presentation, there was a Q&A session for the group's facilitator and students to ask questions.

4.5 Synthesis

The synthesis portion of our activity was designed to present a summary of our teaching activity and what core concepts were taught while linking the activity to an authentic research experience. Regarding the concept of characterization, we explained to the students that we wanted them to understand that: 1) closely related organisms share similar molecular traits, 2) closely related organisms share morphological/ecological traits, and 3) a combination of molecular and morphological/ecological data is needed to adequately characterize species. We also discussed how the practice of generating explanations involved making a claim, providing scientific data to support a claim, and providing a justification that links the claim with evidence. Additionally, the synthesis included a reflection on what the students did, what they had learned, and how these activities closely aligned with the way real researchers carry out characterizations of species. Our unit concluded with a post-teaching survey for students to evaluate our teaching activity.



Figure 1. Phylogenetic trees generated by student groups (A. Team Maverick; B. Team Panda) showing the placement of their "unknown" samples within clusters of closely related species. This is how groups determined the species of their unknown algae samples via molecular characterization.

5. Developing an identity as a person in STEM

Equity and Inclusion is a theme that is incorporated as a component of the PDP, and according to Seagroves et al. (2022) is further organized into four foci: 1) Multiple ways to productively participate, 2) Learners' goals, interests, and values, 3) Beliefs and biases about learning, achievement, and teaching, and 4) Developing an identity as a person in STEM. Within our activity, we chose to incorporate the focus area of developing an identity as a person in STEM.

In our activity, we tried to give our learners as much ownership of their learning as time would allow. As such, we did allow the students to choose which samples to send for sequencing, to provide this ownership. We also felt that if students were able to choose the samples to be sent in for sequencing they would have a stronger vested interest in the DNA analysis component of our activity, thus addressing the element of relevancy. In addition, students were able to modify certain parameters of models in the phylogenetics program they used to construct phylogenetic trees, what morphological/ecological data to include in their analyses, and how they chose to organize their findings to present in the jigsaw activity.

We also felt that we could better develop their STEM identity through guided practice within our activity (showing them how to conduct DNA Barcoding research) and having our students apply this practice to answer a real-world research question. In addition, having students work in teams allowed them to practice collaborating with others, an authentic soft skill that students must hone to be successful in STEM fields. Additionally, when students engage in explanation, they increase their understanding of the content through metacognition, make their thought process and reasoning visible for assessment, and take ownership of their learning.

6. Considerations for the future

There are a number of ways that our activity could be modified in the future. For our teaching activity, we incorporated four different bioinformatics tools (Chromas Lite, CAP3, Genbank, and MEGA). We recommend allowing more time for students to learn more about the molecular concepts involved within this activity, and perhaps spreading it over the course of four to five full days of instruction. Daily instruction time could also be extended for more hands-on learning and in-depth discussion of DNA barcoding methodology, such as PCR, gel electrophoresis, DNA sequencing and alignment, and evaluating the utility of several DNA barcodes for a given plant group. We would also recommend that instructors spend more time discussing the difference between percent similarity and bootstrap values, as students in our class indicated confusion and sometimes referred to these terms interchangeably. For clarification, these terms are both expressed as percentages, but are calculated very differently. Furthermore, these values address different aspects of DNA analyses — percent similarity describes how similar two sequences are to one another, while bootstrap values indicate the reliability of a given relationship between and among organisms on a phylogenetic tree. Thus, we recommend spending more time explaining to students how these values are calculated and perhaps a guided activity that exemplifies how to properly interpret these values prior to a focused investigation.

If replicating this activity in the same time frame allotted for our teaching activity, we would suggest simply focusing on a guided tutorial on how to use GenBank and the Basic Local Alignment Search Tool (BLAST) within this program, and visualizing relationships among species to show students how to interpret phylogenetic trees using evidence. Students would be able to focus on distinguishing similar algae sequences and familiarizing themselves more with using and mastering this program. If working with this time frame, we also recommend

simply mentioning the morphology activity in brief during the synthesis portion as an additional method that plant scientists use to accept or reject the accuracy of a DNA barcode. Or, the use of the morphology/ecology activity could be omitted and the activity could solely focus on molecular characterization and the bioinformatics tools used for DNA barcoding.

Learners demonstrated mastery of an additional piece of content—that ecological evidence can also be used for characterizing organisms. During the design phase of our teaching activity, the only additional line of evidence to support molecular characterization using DNA barcoding that we expected students to use was morphological data. While conducting our teaching activity, students suggested ecological data as supporting evidence to identify their unknown sample, which was sparked from the course instructor's teaching prior to our activity. Nonetheless this additional piece of content complemented the (unplanned) practice of arguing using evidence and showed that students had an elevated understanding of the practice goal and gathering other lines of evidence to support their claim.

In light of our content goal, assessments focused on whether learners understood the ideas that related species have similar genetic sequences and morphological features. Students did not have as strong an understanding of morphological characterization as we had hoped, which was due to a combination of factors: 1) the classes taught prior to ours did not delve into teaching morphological taxonomy as expected; 2) we didn't have time to add more of this topic into our lesson plan. Further, students needed to demonstrate that characterizations must be based on both molecular and morphological data. Students displayed their understanding of the content during a culminating assessment in a jigsaw format, which was from a combination of our starter activity and different tools.

Considering our practice goal, we planned for students to make qualified claims based on different bioinformatics and morphological data, and that

these data would be combined in a logical way to back up their claim. Students did an excellent job of providing claims and evidence but didn't always understand why certain evidence was used or what that evidence meant. In the jigsaw activity the thought processes and reasoning of learners were not always clear. This could be improved by facilitation and design that focused more on asking why certain kinds of evidence were used and having the students explain why these lines of evidence were chosen to support their claim. This limitation may have been due in part to our schedule, which was more constricted than we had planned. And in turn, we had less time to allow students to fully immerse themselves in the content, engage in the practice, and draw their own conclusions.

In conclusion, if replicating this activity, having more time would be a big advantage. Allowing more time for the students to familiarize themselves with the bioinformatics tools will help with the inquiry portion of the teaching activity and to comprehend these programs to focus on the content and practice goals of future teaching activities. Also, once it has been determined how much time will be allotted for this activity, the amount of activities, bioinformatics tools, and content to cover can be scaled up or down based on the timeline. In this way, an appropriate amount of time and teaching can be included into this activity to obtain the best retention of subject matter for learners.

Acknowledgements

The authors would like to acknowledge Matthew Bond (University of Hawai'i at Mānoa, Department of Botany) for his significant contributions to the design, implementation, and teaching of our DNA barcoding module. We also thank the University of Hawai'i at Mānoa Outreach College and Emily Johnston (University of Hawai'i at Mānoa, Department of Botany) for allowing us to pilot our activity within her course of the Science in Action Program - Biodiversity of Hawai'i's Wetlands. We also thank Lisa Hunter, Anne Metevier, and Austin Barnes for

advising our team during the design phase of our module. We acknowledge the Institute for Scientist and Engineer Educators and the Center for Adaptive Optics for the opportunity to participate in the Professional Development Program, and funding support from the Thirty Meter Telescope. The PDP was a national program led by the UC Santa Cruz Institute for Scientist & Engineer Educators. The PDP was originally developed by the Center for Adaptive Optics with funding from the National Science Foundation (NSF) (PI: J. Nelson: AST#9876783), and was further developed with funding from the NSF (PI: L. Hunter: AST#0836053, DUE#0816754, DUE#1226140, AST#1347767, AST#1643390, AST#1743117) and University of California, Santa Cruz through funding to ISEE.

Seagroves, S., Palomino, R., McConnell, N., Metevier, A. J., Barnes, A., Quan, T., & Hunter, L. (2022). *ISEE's equity & inclusion theme*. UC Santa Cruz: Institute for Scientist & Engineer Educators.
<https://escholarship.org/uc/item/8cz4r718>

References

- Institute for Scientist & Engineer Educators, ISEE. (2020). Professional Development Program.
<https://isee.ucsc.edu/programs/pdp/index.html>
- Metevier, A.J., Hunter, L., Seagroves, S., Kluger-Bell, B., McConnell, N., & Palomino, R. (2022). *ISEE's inquiry framework*. UC Santa Cruz: Institute for Scientist & Engineer Educators.
<https://escholarship.org/uc/item/9q09z7j5>
- McNeill, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *Journal of the Learning Sciences, 18*(3), 416-460.
<https://doi.org/10.1080/10508400903013488>
- National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington DC: *The National Academies Press*, 67-69. <https://doi.org/10.17226/13165>

Appendix A – Scaffolding Tool for DNA Barcoding

Please use this tool to help you conduct your analyses in MEGA. You have the option of either starting all over again (like what we did yesterday) or continuing

OPTION 1: To Start From Scratch

Open up MEGA

ALIGN→Edit/Build Alignment→Retrieve Sequences From File

- Select the **.fas** file you created and saved from seqdump file
- Drag or double-click that **.fas** file
- Add your “unknown” to the alignment
- Edit-->Insert sequence from file
- Double-check all sequences are in right orientation (reverse complement)

Align→MUSCLE (non-coding, Plant Plastid DNA sequences)

- Trim the beginning of your alignment
- Export as MEGA File (.meg)→Name your file (group name_sample number) to desktop

Close Alignment Explorer window (not the Main MEGA Window)

OPTION 2: To continue using your sequences you already edited from yesterday

Open up MEGA

File→Open a File/Session→select your **.meg** file

Phylogeny→Construct/Test [*UPGMA tree / Neighbor-Joining tree (students focus on this)*]

→Select Bootstrap Method→No. of Replicates→Select other parameters if you like→COMPUTE!

Save your tree session to the desktop.

Print out your tree to answer questions on the reverse side of this sheet.

Also, you may want to select the “**Caption**” option in the tree-viewing window to review the options you selected to create this respective tree.

Please use this tool to help you and your team generate explanations.

1. [CLAIM] State what your unknown sample most likely is (down to *Genus*, at least).

2. [EVIDENCE] What lines of molecular evidence support your claim? [EVIDENCE]

a. Which analysis did you run--UPGMA and Neighbor-Joining

b. Have you tried running analyses using different replications?

c. Use your printout to show your overall tree topology--circle/highlight where your “unknown” sample is in the tree of relationships.

Appendix B - Content Rubric

Content Goal: Students will use the concept of characterization to explain relationships among organisms.

Dimensions: Components or “knowledge statements”	0 Evidence of misunderstanding or incomplete understanding	1 Evidence of partial understanding	2 Evidence of deep understanding	Score
Related species have similar DNA sequences	<p>Students refer to “gene” similarity, but do not indicate in their dataset that DNA sequences are similar.</p> <p>Students refer to bootstrap value as an indication of similarity (which it is not)</p>	<p>Students understand that related species have similar DNA sequences, but are unable to explain it.</p> <p>Only indicate that they confirmed that they identified their “unknown,” but lack evidence.</p>	<p>Students recognize that nucleotide sequence similarity is correlated with relatedness by saying something along the lines of: ‘Closely related individuals share more similar sequences.’</p>	
Related species tend to have similar morphological features	<p>Students characterize algae on only one characteristic-- but is not a defining characteristic for the taxa/group under study. Not taking into account other morphological features.</p>	<p>Students characterize algae on several morphological features, but do not use robust ones that are unique to the group to which the organism belongs.</p>	<p>Students characterize algae based on robust (multiple) morphological features.</p>	
Species characterizations are drawn from molecular data and strengthened with the use of morphological data	<p>Students base characterizations on either morphological or molecular data--not both.</p>	<p>Students acknowledge the use of morphology in identifying a species, but do not provide actual evidence with morphological data to identify their unknown sample.</p>	<p>Students recognize that morphological features are often superficial, and use both molecular and morphological data as evidence to characterize their unknown sample.</p>	

Appendix C – Practice Rubric

Practice Goal: Students will engage in the practice of generating explanations by using claim, evidence, and reasoning or the CER framework (Metevier et al., 2022).

Aspects of core practice:	0 Lack of evidence did not observe learners enough to decide between A and B	1 Evidence of difficulty what it looks like when a learner needs to work more on the practice	2 Evidence of proficiency what it looks like when a learner is proficient with the practice	Score
CLAIM -“We think our unknown sample most likely is...”		Student claims not qualified- possibly exaggerate or overstep boundaries. {say that their unknown IS species X}	Students use qualified claims. {suggest that their unknown is most likely species X: Say that phylogenetic trees/data do not guarantee relatedness-struggle of real data.}	
EVIDENCE -bootstrap values -percent similarity between “unknown” sequence and closely related taxa that it clustered with. -physical traits -ecological traits		Students use only one line of evidence (i.e. sequence similarity). {morphological or molecular, no integration of both. Lacks comprehension of molecular evidence..“And so, here’s the tree that we generated.” Fail to discuss other lines of evidence (i.e. branch support, percent sequence similarity, morphology, & ecology) that they used to conclude that their unknown sample is most likely species X.}	Students include, understand, and integrate multiple lines of evidence. {sequence and morphological/ecological data- percent sequence similarity and bootstrapping. Show & explain phylogenetic tree, refer to the branch support values on the tree, refer to percent similarity, indicate whether molecular data supports the morphological features that associates their unknown sample with other taxa that cluster with it.}	
REASONING - linking a claim with evidence) -how they reason with the evidence		Students do not use logical chains of reasoning and are unable to reconcile conflicting evidence. {Unable to explain with evidence what their unknown sample is- Reasoning with either morphological, ecological and/or molecular evidence, but not both}	Students explain results that logically link to their claim. {molecular, morphological & ecological data: Interpretation of their phylogenetic tree, percent similarity, bootstrapping, morphological/ecological data}	