

One Fish, Two Fish, Real Fish, Fake Fish: Using Free Online Tools to Teach Tree Thinking

Hans D. Lemke¹ and Anna M. Davis²

¹College of Computer, Mathematics and Natural Science, University of Maryland, College Park MD 20742 USA

²Marine-Estuarine Environmental Science, University of Maryland, College Park MD 20742 USA
(hlemke@umd.edu; amdavis@umd.edu)

Tree thinking is a fundamental skill in biology. This workshop highlights two exercises that are designed to provide students with experience interpreting trees and using them to test a hypothesis. Both use a freely available set of web tools (www.phylogeny.fr) to analyze sequences downloaded from GenBank. The first exercise uses molecular data to simulate testing the identity of fish species served in a sushi restaurant. The second is a guided inquiry exercise that allows students to use phylogenetic techniques to test the claims of a simulated advocacy group. In this workshop, we will work through these exercises and then discuss how the techniques can be adapted for other scenarios.

Keywords: phylogenetics, tree thinking, evolution, inquiry-based learning, molecular phylogeny

Link to Supplemental Materials: <https://doi.org/10.37590/able.v41.sup11>

Introduction

Tree thinking has become an essential part of modern biological thinking (Baum et al. 2005), but it is often a difficult concept for students to understand. The exercises presented here are used to reinforce concepts of tree thinking taught in an introductory biology sequence course focusing on ecology, evolution

and biodiversity. These exercises represent the second of a two-week sequence of three-hour labs covering tree thinking. They provide additional practice developing and reading trees as well as an introduction to molecular tools for creating and interpreting trees. The first week, students learn the basics of constructing and interpreting trees using models of imaginary organisms (similar to Lemke and Jensen, 2012).

Student Outline

Exercise 1 – Food Fraud?

In 2013, several instances of fraud in the global food industry were revealed. In the United States and Europe, horsemeat was eaten by consumers who thought they purchased beef. Honey from China was labeled as coming from a half-dozen other countries to get around U.S. laws. Lastly, cheap or even toxic fish species were deceptively sold under the guise of being expensive species that humans prize for their culinary value. This last example includes a phylogenetic study by the advocacy group Oceana (Warner et al., 2013) that showed that 113 out of 120 samples of "red snapper" in the U.S. were actually another species. When you see a whole cow, you can tell it is a cow, and you can see that a horse is a horse. What about when it is ground up in your burrito or lasagna? A slice of expensive fish on rice doesn't have the distinctive fins, color, or shape to help you tell it apart from other fish, allowing some wholesalers to be dishonest. With DNA testing, we can identify animals or plants even if they are in small parts (sushi), ground up (herbal supplements), or mixed with other species (ground horse and beef mixture).

Your best friend is visiting from out of town and you want to take them to the best sushi restaurant in town. You know that they are very fussy about their sushi, so you decide to do a reconnaissance mission to see if the restaurant is serving the fish that they claim. Before you go, you need to use molecular tools to create a tree based on similarities in their genetic sequence to create a tree of known species. There are many different genetic markers that can be used for this. For our purposes, we're going to focus on the gene for *cytochrome oxidase subunit 1* (denoted "COI").

1. Open the file *sushispecies_COI.docx*. This file contains genetic sequences for a set of species that you would commonly find in sushi restaurants as well as a few unknowns that you will be testing later.
2. Open <http://www.phylogeny.fr>
3. Scroll down the page until you find a link that says 'One Click.' Click on this.
4. You should now be on the 'One Click Mode' page. In the middle of the page, you'll see a large text box. Copy the **known** sequences from *sushispecies_COI.docx* and paste into this box. One person from your group should enter their email address in the box below so the results will be sent to you (this is just so you have a backup file. The final tree will also appear on the page after you've run the program). After entering both of these things, click 'Submit'.

"One Click" Mode

Alignment MUSCLE → Curation Gblocks → Phylogeny PhyML → Tree Rendering TreeDyn

1. Overview 2. Data & Settings

Name of the analysis (optional):

Upload your set of sequences in FASTA, EMBL or NEXUS format from a file:

no file selected

Or paste it here [\(load example of sequences\)](#)

Maximum number of sequences is 200 for proteins and 200 for nucleic acids.
Maximum length of sequences is 2000 for proteins and 6000 for nucleic acids.

Use the Gblocks program to eliminate poorly aligned positions and divergent regions

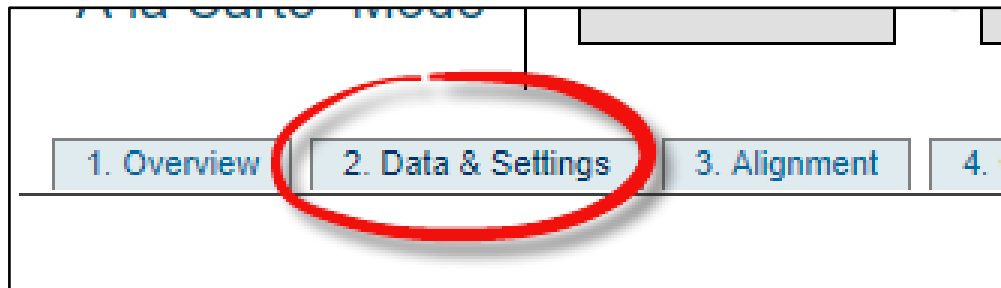
To receive the results by e-mail, enter your address(es):

5. It will take a couple of minutes for the software to produce a tree for you.
6. Copy the tree into the space below.

Now that you have a tree of known species, it is time to see if the restaurant is on the up-and-up. While at the restaurant, you discreetly collected samples of the fish that you ordered and brought them back to the lab for testing. After generating sequences, you can compare them to what you found when you created your earlier tree.

The first sample that you want to check is the yellowtail (*Seriola quinqueradiata*).

7. Add **Unknown 1** to your tree
 - a. Click on the **2. Data & Settings** tab



- b. Copy the sequence labeled Unknown 1 from *sushispecies_COI.docx*
 - c. Paste the sequence to end of your previous data file in the Input Data box
8. Repeat steps 4-5.
9. Do you think that you were served genuine yellowtail? Why or why not? Draw any relevant branches of the tree.

- The next sample that you want to process is the bluefin tuna (*Thunnus thynnus*), labeled as **Unknown 2**.
10. Repeat Steps 7-8, being sure to replace Unknown 1 with Unknown 2.
 11. Did you get bluefin tuna from the restaurant? Why or why not? Draw any relevant branches of the tree.
 12. Repeat this procedure for your last sample, which is from the chef's special spicy snapper roll, labeled **Unknown 3**.
 13. Is the chef serving snapper, or something else? What makes you think so? Draw any relevant branches of the tree.

Exercise 2 – Testing A Claim

Who Ate All the Crabs?

Striped bass (*Morone saxatilis*) (known regionally as 'rockfish') are an important fish in the Chesapeake Bay, both ecologically and economically. They are a particularly popular fish with recreational anglers. Their populations in the Chesapeake Bay declined throughout the 20th century, but more stringent fishing regulations initiated in the 1980s helped their populations to rebound. Today, they are an abundant fish in the Chesapeake Bay.

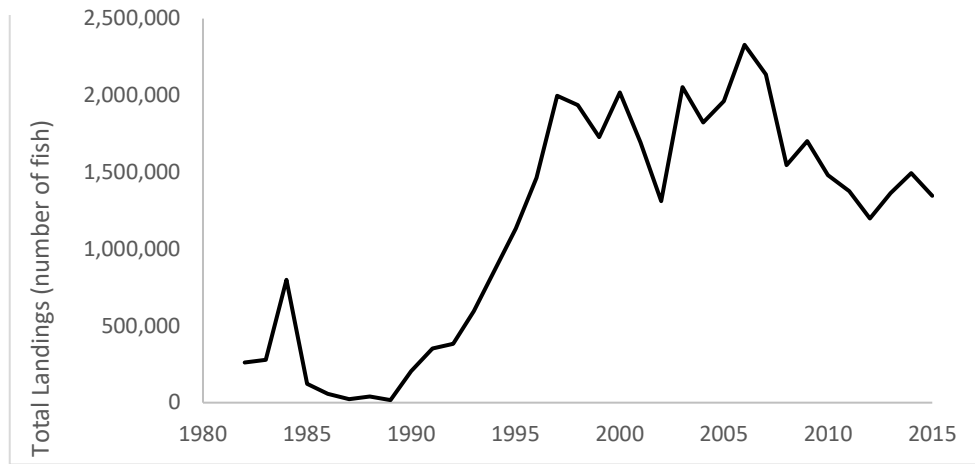


Figure 1. Striped bass landings (number of fish) in the Chesapeake Bay 1982-2015. Data from Atlantic States Marine Fisheries Commission 2016 Atlantic Striped Bass Stock Assessment (ASMFC, 2016).

You're reading the local paper one morning, and you come across an op-ed from a local recreational angler's organization. The op-ed reads:

Rockfish and blue crabs make up two of the Bay's most prized fisheries, both to recreational and commercial fishermen. While historically rockfish populations had low abundance, today their populations are experiencing a significant boom. Based off of observations made by members of the Maryland Angler's Association, there is reason to believe that blue crabs make up a significant portion of the diet of striped bass. With the large and growing abundance of striped bass, there is concern that this could drive down the abundance of blue crabs in the Chesapeake Bay. Currently, fishing regulations limit recreational anglers to a maximum of two fish per person per day during brief spring and fall fishing seasons. We believe these regulations should be reconsidered in light of our observations, and an increase in allowable catch should be implemented to protect blue crab populations.

-- Maryland Angler's Association

You question the assertions of this group and decide to run an experiment to test whether they are true. You and the rest of the members of the class make up a team of fisheries biologists who are going to examine the gut contents from a sample of striped bass caught in the Chesapeake Bay to see if blue crabs indeed make up a significant proportion of the striped bass's diet. Sixty striped bass are collected, their stomach contents removed, and a genetic analysis conducted on the primary contents. Each group will randomly select a number of samples from the list (remember our discussion on random

sampling?). You will determine how many fish your group will sample and how best to randomly sample them. It's important to note that results from the class will be pooled at the end of the activity so that we can get a sense of how many fish are consuming blue crabs.

List the fish your group will be analyzing:

Part 1- Creating A Tree of Knowns

Before we get started, you will want to do some background research on what striped bass eat. Visit a few websites (keep in mind what you know about reliable internet sources) and come up with a list of the top five food items that make up the diet of the striped bass. Make sure you include fish as well as invertebrates in your list. **NOTE - Because we're interested in whether the stomach contents contain remnants of blue crab (*Callinectes sapidus*), you'll also want to make sure you include them in your analysis.**

List the common and scientific name for each species in Table 1.

Note: If the scientific name isn't immediately apparent on the website(s) you use, some additional searching should help identify the scientific name of the species.

Table 1. Common prey species of striped bass.

Common Name	Scientific Name	Number of Base Pairs (from Step 4)
Blue Crab	<i>Callinectes sapidus</i>	

We'll be using molecular tools to create a phylogenetic tree with the unknown samples, as well as some known samples, to see if we can identify what is present in the striped bass stomach. There are different genetic markers that can be used for this. For our purposes, we're going to focus on the gene for *cytochrome oxidase subunit 1* (denoted "COI").

Creating Your Tree:

1. Visit <https://ncbi.nlm.nih.gov/genbank>
2. Using the search bar at the top of the page, type in the name of the first organism from your list, followed by 'coi' (to indicate the marker of interest). Make sure that the drop down to the left of the search bar is set to 'Nucleotide'. We're interested in the sequence of nucleotides for this particular gene. Click **Search**.



3. After searching the database, there should be a number of matches. Verify that the species and marker correctly match your search term (see *Note* following if not). If your results look correct, you'll first want to sort the results by the length of the sequence to make sure you're getting the longest sequence available. Using the 'Sort by' pull down

menu at the top of the page, select 'Sequence Length'. The results should re-sort, and the longest available sequence should now be at the top.

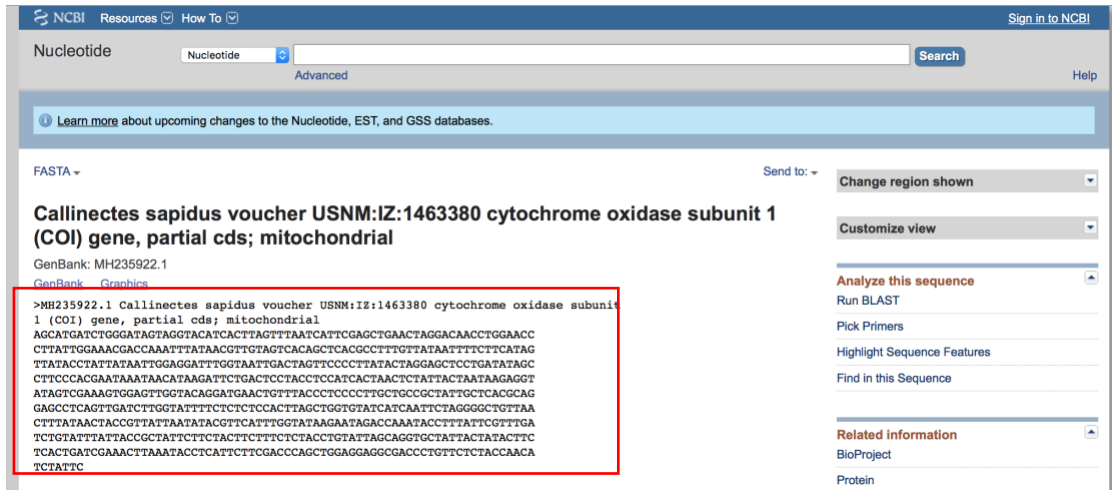
Note: Sometimes a species hasn't yet been added to GenBank. In this case, you won't get any results or, occasionally, the results will be for a different species. If this is the case, you'll want to add a new species to your list above and do a new GenBank search for this organism.

The screenshot shows the NCBI Nucleotide search interface. The search term is 'callinectes sapidus coi'. The 'Sort by' dropdown menu is open, with 'Sequence Length' selected. The search results are sorted by sequence length. The first result is 'Callinectes sapidus from Connecticut cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial', which is 1,534 bp linear DNA. The second result is 'Callinectes sapidus from Maryland cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial', also 1,534 bp linear DNA. The page shows 'Items: 1 to 20 of 697'.

- Now that you've sorted your results, the first match on the list should be the most appropriate. Ignore results that include the entire mitochondrial genome. Choose one of the sequences that is approximately 655 base pairs. Record the length of the sequence in Table 1.

The screenshot shows the NCBI Nucleotide search interface. The search term is 'callinectes sapidus coi'. The search results are sorted by default order. The first result is 'Callinectes sapidus voucher (USNM-17-1463380) cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial', which is 637 bp linear DNA. A red arrow points to this result. The second result is 'Callinectes sapidus voucher (USNM-17-1463380) cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial', which is 658 bp linear DNA. The page shows 'Items: 1 to 20 of 243'.

- Below the correct sequence, you'll see a link entitled 'FASTA'. FASTA files are how molecular scientists store genetic information. They consist of an identifier line that gives some information about the sequence (species, the accession number on GenBank, the marker, etc.) and the actual nucleotide sequence. Click on this link.
- You'll now be looking at a page with your nucleotide sequence. You'll notice the identifier line, denoted with a greater than symbol (>), followed by a sequence of As, Ts, Gs, and Cs. This is the sequence of nucleotides for your marker. Copy the entire FASTA file (including the identifier line) and paste this into a new Word document.



7. Repeat steps 2-5 with the other organisms from your list. Make sure that you're using the same 'coi' marker each time. Copy the FASTA files for each of these into the same Word document. Don't forget to include the identifier line (including the >).
8. To make this a little easier to read once we have our final tree, we'll want to edit the identifier line to be shorter for the sequences you pulled off GenBank. For each sequence, remove all text in the identifier line except for the name of the species. It should look something like this:

>Callinectes sapidus

ATGCGTCACCTTGAA...

>Fundulus heteroclitus

ATGCCGAATGACAGGT...

9. Okay, now we're ready to build the tree! Visit <http://phylogeny.fr> to get started.
10. Scroll down the page until you find a link that says: 'One Click.' Click on this.
11. You should now be on the 'One Click Mode' page. In the middle of the page, you'll see a large text box. Copy the entire list of sequences from your Word file and paste into this box. One person from your group should enter their email address in the box below so the results will be sent to you (this is just so you have a backup file. The final tree will also appear on the page after you've run the program). After entering both of these things, click 'Submit'.

"One Click" Mode

Alignment MUSCLE → Curation Gblocks → Phylogeny PhyML → Tree Rendering TreeDyn

1. Overview 2. Data & Settings

Name of the analysis (optional):

Upload your set of sequences in FASTA, EMBL or NEXUS format from a file:

no file selected

Or paste it here [\(load example of sequences\)](#)

Maximum number of sequences is 200 for proteins and 200 for nucleic acids.
Maximum length of sequences is 2000 for proteins and 6000 for nucleic acids.

Use the Gblocks program to eliminate poorly aligned positions and divergent regions

To receive the results by e-mail, enter your address(es):

12. The program will take a little time to run. When finished, a page will appear with a tree on it. This is the tree showing the relatedness between the sequences you entered. Note: We won't be focusing on the red numbers on the nodes of the tree for our interpretation today, but you may wonder what they indicate. They represent our confidence in the relationships depicted on the tree and range between 0 and 100, with numbers closer to 100 indicating a high level of confidence.

Draw your tree of known species below:

Now you want to add the unknown samples onto your tree. On the same Word document as before, paste the unknown sequences (identifier line and all) below the known sequences. Repeat steps 9-11 from above.

Draw your full tree including all known and unknown species below:

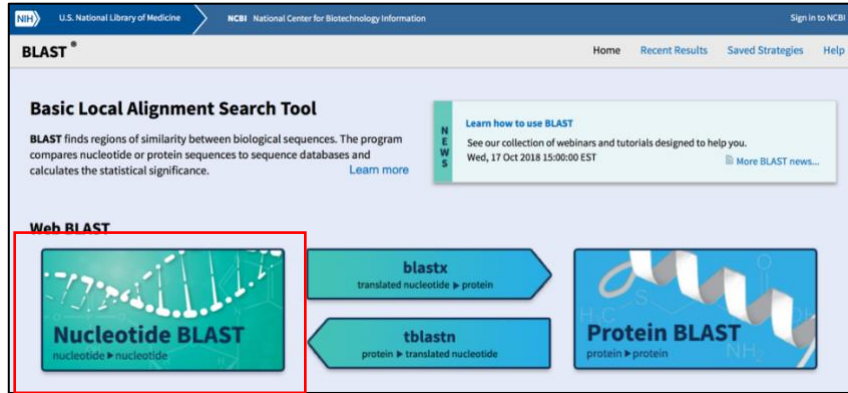
Questions:

1. Based on your tree, are there any species you can identify? What do you suspect they are? Justify your answer.
2. Does your sample include blue crab? How do you know?

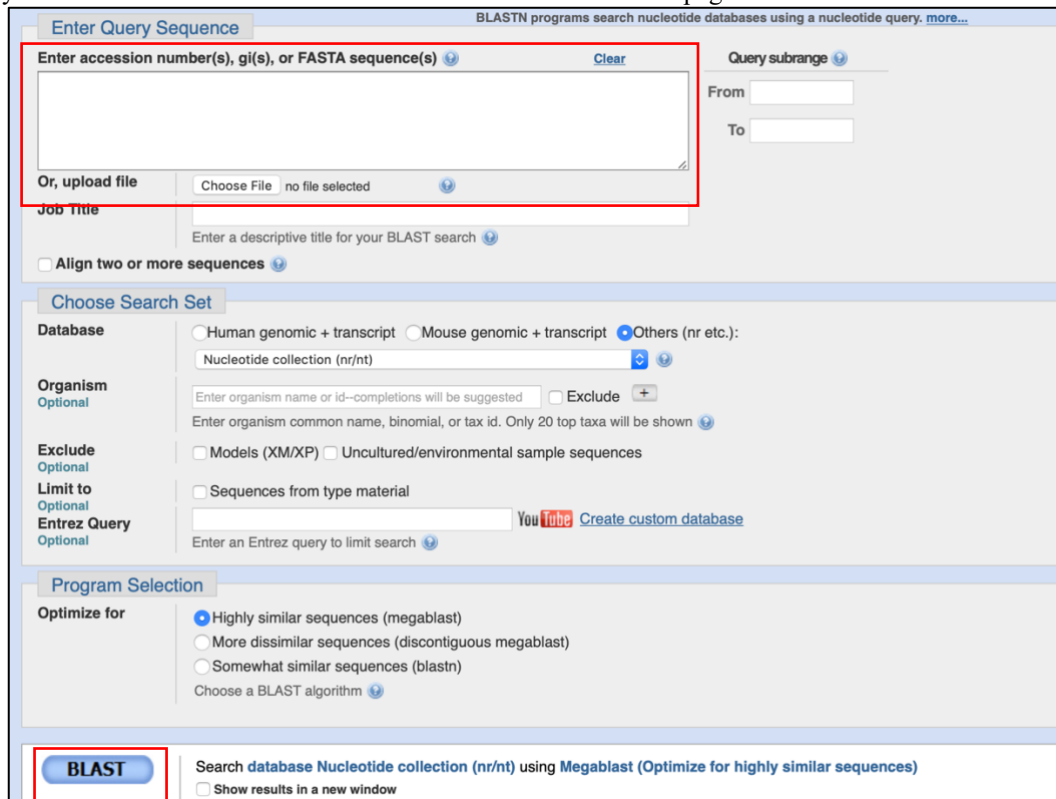
Part 2 - Identifying Unmatched Unknowns

It's possible that you have some stomach content samples that didn't match anything on your tree of known species. We can determine what they are by using a BLAST search. If you didn't have any unknowns, you can verify the identification of one of your matched species.

1. Visit <https://blast.ncbi.nlm.nih.gov/>
2. On the front page, click on the box labeled 'Nucleotide Blast'.



- On the next page, copy and paste one unknown sequence into the text box under 'Enter Query Sequence'. It isn't necessary to include the identifier line. Scroll down to the bottom of the page and click on the 'BLAST' button.



- On the next page, scroll past the 'Graphic Summary' to the list under 'Descriptions'. The first line is the best match for your sequence. Write down the species name.

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments [Download] [GenBank] [Graphics] [Distance tree of results]

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Callinectes sapidus from Ocean Springs, MS cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2833	2833	100%	0.0	100%	AY882077.1
<input type="checkbox"/>	Callinectes sapidus from Maryland cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2811	2811	100%	0.0	99%	AY882073.1
<input type="checkbox"/>	Callinectes sapidus from Connecticut cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2800	2800	100%	0.0	99%	AY882072.1
<input type="checkbox"/>	Callinectes sapidus from Brownsville, TX cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2789	2789	100%	0.0	99%	AY882079.1
<input type="checkbox"/>	Callinectes sapidus from Apalachicola, FL cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2789	2789	100%	0.0	99%	AY882076.1
<input type="checkbox"/>	Callinectes sapidus from Naples, FL cytochrome c oxidase subunit I (COI) gene, complete cds; mitochondrial	2789	2789	100%	0.0	99%	AY882074.1

5. Repeat steps 1-4 for any remaining unknowns you have.

Now that you have identified your gut content samples, we're going to pool the results from the whole class. Your TA will facilitate this. After the unknowns have been pooled, answer the following questions.

1. How many fish had blue crab in their stomach? What proportion of the fish does this make up?

2. Do you agree with the assertions of the op-ed? Does blue crab make up a significant proportion of striped bass diet?

Assessment

Write a response to the Maryland Angler's Association op-ed indicating whether you agree or disagree with their assertions. Provide your tree and the results of the pooled class data to support your response.

Materials

Websites

A web-based set of tools for conducting phylogenetic analyses is available at www.phylogeny.fr. Users can choose between a simple one-click interface that makes important decisions for you, or a more flexible a-la-carte version that allows you to choose between multitudes of options. It is freely available and has a ton of options for exporting the trees.

GenBank (www.ncbi.nlm.nih.gov/genbank/) is an NIH sponsored database of annotated genetic sequences. From here, you can download sequences and perform BLAST searches.

Notes for the Instructor

Scenarios

By varying the species used with the basic procedures included here, you can easily adapt this lab to virtually any scenario you would like. There are countless examples of food fraud that lend themselves to relatable scenarios. For example, for our course we adjusted the first scenario to match Oceana's report on the substitution of other crab species for Maryland blue crab (Warner et al., 2015).

Exercise 2 includes an op-ed from the Maryland Anglers Association. This is a fictitious organization, but the ideas are based on common feelings among fishermen in the Chesapeake Bay region.

Cited References

- ASMFC. 2016. 2016 Atlantic striped bass stock assessment update. http://www.asmfc.org/uploads/file/581ba8f5AtlStripedBassTC_Report2016AssmtUpdate_Oct2016.pdf
- Baum DA, Smith SD, Donovan SS. 2005. The tree thinking challenge. *Science*. 310: 979–980.

Lemke HD, Jensen J. 2012. Exploring systematics and phylogenetic reconstruction using biological models. *Tested Studies for Laboratory Teaching. Proceedings of the Association for Biology Laboratory Education*.33: 123-144. <http://www.ableweb.org/volumes/vol-33/?art=9>

Warne K, Lowell B, Disla C, Ortenzi, K, Savitz, J, Hirshfield M. 2016. Oceana reveals mislabeling of iconic Chesapeake blue crab. *Oceana* 1-15. https://oceana.org/sites/default/files/crab_testing_report_final_3.27.15.pdf

Acknowledgments

We would like to thank the countless teaching assistants who gave us feedback as they used these exercises in the classroom. We would also like to express our appreciation for the support received from the Biological Sciences Program for the development of new lab exercises.

About the Authors

Hans Lemke is the Lab Coordinator for the Principles of Ecology and Evolution Lab (BSCI161) at the University of Maryland. He holds a B.A. in Biology from St. Mary's College of Maryland, an M.S. in Entomology from the University of Maryland, and an M.D.E. in Distance Education from University of Maryland, University College. His current research focuses on educational outcomes in laboratory and online settings and a survey of tiny Miocene shark and ray fossils found along the Chesapeake Bay.

Anna Davis is a doctoral student in the Marine-Estuarine Environmental Science program at the University of Maryland. She holds a B.S. in Biology from George Washington University and an M.S. in Marine Biology from College of Charleston. Her doctoral research focuses on oyster ecology and physiology.

Appendix A: Sample Results

Exercise 1

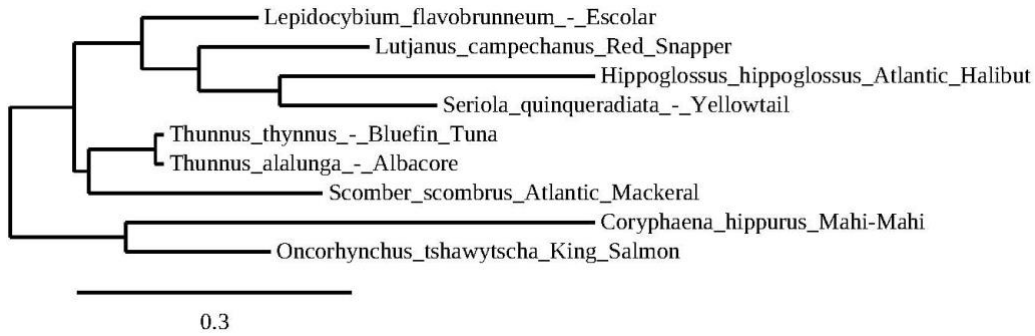


Figure 2. Tree created with known fish species.

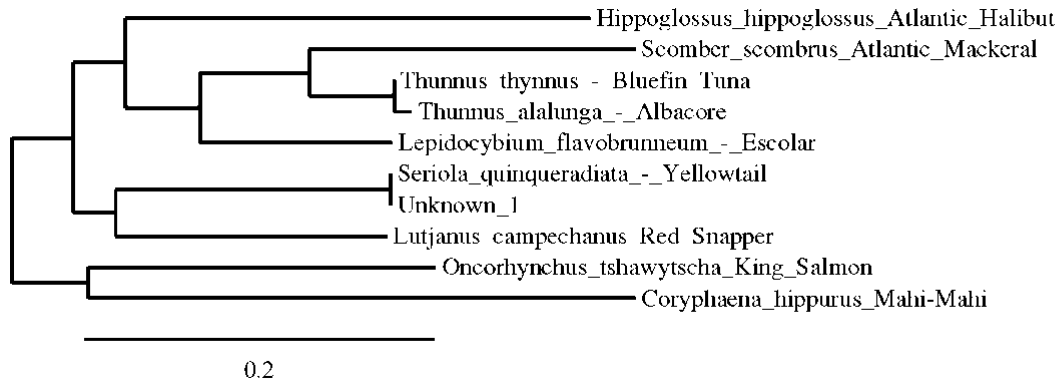


Figure 3. Tree with Unknown 1 included.

Exercise 2

Part 1

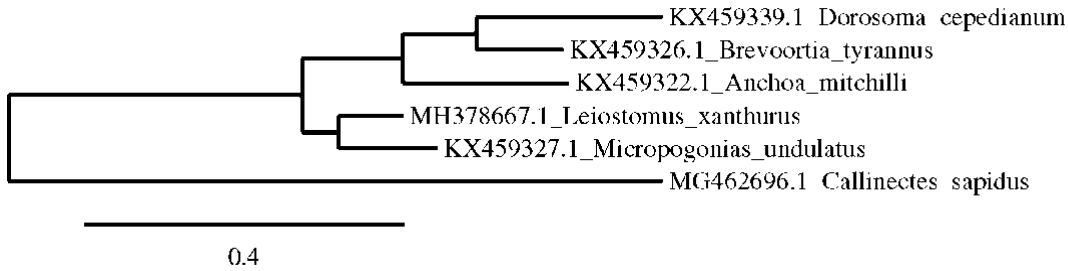


Figure 4. Tree constructed with six of the most common prey species that are found in *Morone saxatilis*

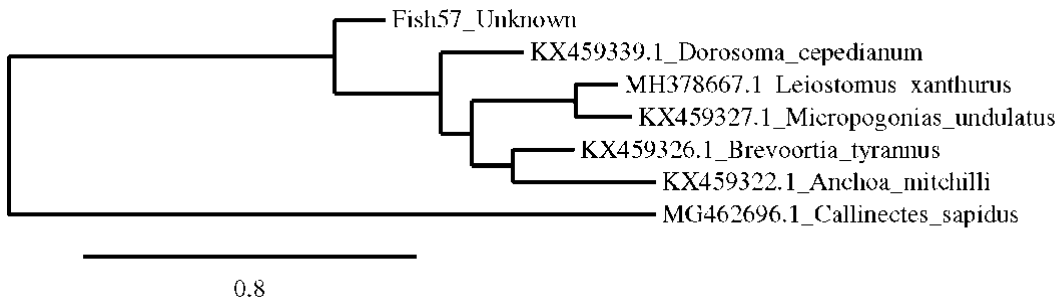


Figure 5. Tree with known prey species including unknown gut sample #57.

Part 2

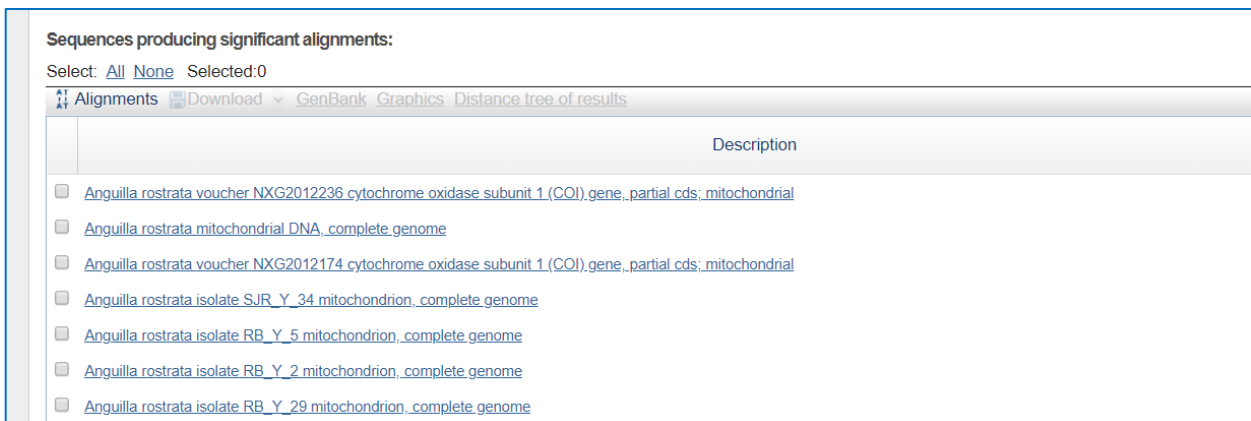


Figure 6. Results of BLAST search on Unknown 57.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit <http://www.ableweb.org/>.

Advances in Biology Laboratory Education is the peer-reviewed publication of the conference of the Association for Biology Laboratory Education. Published articles and extended abstracts are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board. Published abstracts are evaluated and selected by a committee prior to presentation at the conference.

Citing This Article

Lemke HL, Davis AM. 2020. One fish, two fish, real fish, fake fish: using free online tools to teach tree thinking. Article 11 In: McMahon K, editor. *Advances in biology laboratory education*. Volume 41. Publication of the 41st Conference of the Association for Biology Laboratory Education (ABLE). <https://doi.org/10.37590/able.v41.art11>

Compilation © 2020 by the Association for Biology Laboratory Education, ISBN 1-890444-17-0. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.