Publication of the Association for Biology Laboratory Education
Volume 44, Article 33 2024

ABLE Association for Biology Laboratory Education | ADVANCES IN BIOLOGY LABORATORY EDUCATION

Using DNA Barcoding to Examine Mislabeling in Tuna

Michael P Martin, Madison Sobonya, and Christopher A Sheil

John Carroll University, Department of Biology, 1 John Carroll Boulevard, University Heights, OH, 44118

Abstract

This two-part lab exercise allows students to evaluate the accuracy of food labeling by methods of DNA barcoding. Students extract, isolate, and amplify the DNA sequence for the cytochrome oxidase c subunit 1 protein, from samples of sushi-grade tuna acquired from restaurants or grocery stores. DNA sequences from these tuna samples are analyzed with the Basic Local Alignment Search Tool (BLAST) to identify similar sequences that exist in the GenBank database. BLAST results are used to infer the identity and evaluate the accuracy of food labels for these tuna samples.

Keywords: food mislabeling, tuna, *COI*, cytochrome oxidase

Citation: Martin MP, Sobonya M, and Sheil CA. 2024. Using DNA barcoding to examine mislabeling in tuna. Article 33 In: Boone E and Thuecks S, eds. *Advances in biology laboratory education*. Volume 44. Publication of the 44th Conference of the Association for Biology Laboratory Education (ABLE). DOI: https://doi.org/10.37590/able.v44.art33

Correspondence to: **Christopher Sheil, csheil@jcu.edu**

INTRODUCTION

The issue of honest and accurate labeling of foods can have broad impacts and implications for society, and consumers should have reasonable expectations that the food items they purchase are, in fact, correctly labeled. Food mislabeling is the false or misleading representation with respect to another food (US Food & Drug Administration 2023). Consumer concerns about accurate food labeling revolve around desires to: avoid violating religious/cultural taboos; eat sustainably or with an awareness towards consuming ethically-harvested foods; avoid complications of eating foods that would trigger allergies or dietary sensitivities; or avoid paying premium prices for inferior alternatives (Filonzi et al. 2023). However, food mislabeling occurs for a variety of reasons, including those that are intentional (to increase profits or offset low supplies) or unintentional (relaxed labeling standards, poor naming practices, globalization, or carelessness). A recent issue of purported food mislabeling involved Subway™ tuna sandwiches, involving "The Big Tuna Sandwich Mystery" (Carmel 2021), leading to a series of newspaper editorials, and lawsuits against the largest US sandwich chain, all centering around the question, "Is Subway selling tuna?".

DNA barcoding is a technique whereby the probable identity of an unknown biological sample may be inferred by comparing DNA sequences to a database of known samples. The high degree of sequence conservation within a species is reflected in high similarity among individuals, whereas mutation accumulation results in a higher degree of variation between species. In short, DNA sequence matches within a specific threshold of similarity may be considered evidence of identity for an unknown sample. DNA barcoding of seafood has been successfully performed by utilizing a short (~650 bp) fragment of the mitochondrial *COI* gene, which encodes the cytochrome oxidase c subunit 1. DNA barcoding of unknown samples requires extracting genomic DNA followed by PCR amplification and DNA sequencing. Publicly-available databases (e.g., National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/; and BoldSystems Identification Engine,

https://www.boldsystems.org/index.php/databases) and search algorithms facilitate easy comparisons among unknown and known samples to infer the identity of unknown samples with reasonable certainty. In August 2023, the BoldSystems database held more than 240,000 species-level *COI* DNA barcodes for animals. The principles of DNA barcoding provide an opportunity for students to practice genomic DNA extraction, PCR amplification, DNA sequencing, and database searches in the context of food mislabeling.

This laboratory exercise explores food mislabeling via DNA barcoding and revolves around the research question, "Is sushi-grade tuna correctly labeled?" The scope of the project can be chosen by the instructor, focusing on a smaller scale to simply introduce students to the techniques involved with DNA barcoding, or focusing on a broader scale to examine the issue of food mislabeling and marketing at multiple restaurants or fish markets. In either case, this project is designed to be completed across 4 two-hour laboratory meetings, with a break/interval of time during which amplified DNA products are being sequenced by a third-party lab. Introductory biology majors are best suited to complete this project, assuming that they have studied and reviewed the principles of DNA, genes, evolution, and DNA extraction; this project is suitable for demonstrating and practicing hands-on skills for introductory-level biology students as they learn the principles behind the techniques of PCR and DNA sequencing.

STUDENT OUTLINE

Objectives

Describe food mislabeling and DNA barcoding
Perform DNA extraction and amplification by PCR
Demonstrate DNA barcoding by use of tuna samples
Demonstrate the NCBI database and BLAST analyses
Analyze FASTA files to infer identity of unknowns
Interpret results of BLAST search
Evaluate identity as inferred by two sources

Introduction

The issue of honest and accurate labeling of foods can have broad impacts and implications for society, and consumers should have reasonable expectations that the food items they purchase are, in fact, correctly labeled. Food mislabeling is the false or misleading representation with respect to another food (US Food & Drug Administration 2023). Consumer concerns about accurate food labeling revolve around desires to: avoid violating religious/cultural taboos; eat sustainably or with an awareness towards consuming ethically-harvested foods; avoid complications of eating foods that would trigger allergies or dietary sensitivities; or avoid paying premium prices for inferior alternatives (Filonzi et al. 2023). However, food mislabeling occurs for a variety of reasons, including those that are intentional (to increase profits or offset low supplies) or unintentional (relaxed labeling standards, poor naming practices, globalization, or carelessness). A recent issue of purported food mislabeling involved Subway™ tuna sandwiches, involving "The Big Tuna Sandwich Mystery" (Carmel 2021), leading to a series of newspaper editorials, and lawsuits against the largest US sandwich chain, all centering around the question, "Is Subway selling tuna?".

In order to determine the identity of a seafood sample, comparison of the sample's DNA to a database of known species is required. The mitochondrial *COI* gene encodes the cytochrome oxidase c subunit 1 protein that acts in the electron transport chain, and it is the standard marker for comparison. Intraspecies variation in the *COI* gene is low, whereas greater sequence diversity is observed when comparing two species.

In this multi-week exercise, you will be isolating genomic DNA from tuna samples that are acquired from restaurants and grocery stores. DNA amplification of the *COI* gene will be performed by PCR prior to samples being sequenced. Comparison of samples to two publicly-available databases will allow you to assess the species-level identity of your sample, and determine whether any mislabeling has occurred.

Methods and Data Collection

Pre-lab part A: Tuna sample collection

THIS PART OF THE PROJECT WILL TAKE 1 OR 2 WEEKS.

"Field Work" for this project involves visiting a local sushi restaurant, fishmonger, or grocery store, and purchasing samples of raw "tuna". Document the putative identity of your samples by noting how the fish is being labeled on the packaging or restaurant menu—what is it being sold as? Is it simply labeled as "tuna" or "yellowfin" or "bluefin". Does any of the labeling or marketing suggest where the fishes were caught, processed, frozen, distributed, or sold? Give each sample a unique code or number, and keep good records in your lab notebook for future reference, (e.g., "Sample 010: sold as "bluefin tuna" from Restaurant X"). Your tuna samples can be small as only 25 mg (0.0009 oz) are needed for DNA isolation. Tuna samples should be stored at -20°C.

Pre-lab part B: Prepare tissue for DNA isolation according to QIAgen DNeasy Blood & Tissue Kit

Your instructor has performed the following steps in order to break down tissue and lyse cells prior to isolating DNA:

1. Obtain 1.5 ml microcentrifuge tubes for each sample. Label the tubes with a sample number. Take the mass of the tube and label the mass on the tube.
2. Cut fish tissue into small pieces (≤25 mg) and place in the microcentrifuge tube. Take the mass of the tube again and determine the mass of tissue.
3. Add 180 µl Buffer ATL and 20 µl proteinase K.
4. Mix by vortexing for 15 seconds.
5. Incubate at 56℃ overnight with occasional vortexing.

Day 1: DNA isolation

Genomic DNA will be isolated using the QIAgen DNeasy Blood & Tissue Kit according to the manufacturer's instructions with an overnight incubation. Buffers AL and AW1 contain guanidine salts. No bleach should be used in the lab as combining guanidine salts and bleach will produce highly reactive compounds. Disposal of solutions containing AL and AW1 should be done according to local guidance.

1. Mix by vortexing for 15 seconds after incubation.
2. Incubate for 1 hour at 56℃ and vortex again (optional).
3. Add 200 µl Buffer AL and mix by vortexing for 15 seconds.
4. Incubate at 56℃ for 10 minutes.
5. Add 200 µl ethanol (96%-100%). Mix by vortexing for 15 seconds.
6. Place a DNeasy Mini spin column into a 2 ml collection tube. Label the top of the spin column.
7. Pipet the mixture (600 µl) into the spin column.
8. Centrifuge at 8000 rpm for 1 minute. Discard the flow-through and collection tube (no bleach).
9. Place the spin column in a new 2 ml collection tube.
10. Add 500 µl Buffer AW1.
11. Centrifuge at 8000 rpm for 1 minute. Discard the flow-through and collection tube (no bleach).
12. Place the spin column in a new 2 ml collection tube.
13. Add 500 µl Buffer AW2.
14. Centrifuge at 14,000 rpm for 3 minutes. Discard the flow-through and collection tube.
15. Label a new 1.5 ml microcentrifuge tube with sample number, sample type, the date, and your initials.
16. Transfer the spin column to this new 1.5 ml microcentrifuge tube.
17. Elute the DNA by adding 100 µl Buffer AE to the center of the spin column membrane while being careful not to touch the pipet tip to the membrane.
18. Incubate for 1 minute at room temperature. Centrifuge for 1 minute at 8,000 rpm.
19. Repeat steps 17-18 once.
20. Determine DNA concentration according to the instructor's directions.

Day 2: PCR of cytochrome c oxidase subunit 1 (COI) gene

Groups should share genomic DNA samples with 2 other groups in order to have replication of the PCR. In this case, each group should amplify their own sample, two others, and have a no DNA control.

A. Prepare for PCR

1. While wearing gloves, number four PCR tubes with the sample number (or no DNA) on the side of the tube immediately below the lip.
2. Place tubes on ice.
3. Add 2 µl genomic DNA or deionized water to the appropriate tubes. Micropipette tips should be discarded after a single use throughout the preparation of PCR reactions.

B. PCR amplification of a 655 bp portion of the *COI* gene

1. Each group will assemble a mix containing all reagents except the genomic DNA/water from part A (Table 1).

Table 1. COI PCR reaction components

	stock conc.	final conc.	amount per reaction (µl)	tuna PCR mix (µl), 4.5 reactions
deionized water			11.7	52.7
5X GoTaq buffer	5X	1X	5	22.5
$MgCl_2$	25 mM	3 mM	3	13.5
FISHCOIBC_ts new	10 µM	0.5 µM	1.25	5.6
FishCOIHBC-deg	10 µM	0.5 µM	1.25	5.6

dNTPs	10 mM	0.2 mM	0.5	2.3
template DNA	50 ng/µl	4 ng/µl	2	NONE
GoTaq DNA polymerase	5 units/µl	0.025 units/µl	0.125	0.6

2. Briefly vortex the tuna PCR mix and centrifuge to collect the liquid at the bottom of the tube.
3. Add 23 µl of tuna PCR mix to each tube. Centrifuge if any liquid is not at the bottom of the PCR tubes.
4. Tightly cap the PCR tubes and place them in the thermocycler.
5. Products will be amplified with the following conditions (35 cycles of steps b-d):

    a. 94°C for 2 minutes
    b. 94°C for 1 minute
    c. 50°C for 1 minute
    d. 72°C for 1 minute
    e. 72°C for 5 minutes
    f. 4°C indefinitely

Day 3: Gel electrophoresis of COI amplicon

1. Prepare 500 mL of 1X TBE buffer (enough for pouring agarose gel and electrophoresis running buffer).
    - Add 100 mL 5X TBE to a graduated cylinder
    - Add deionized water to a total of 500 mL
    - Cover with parafilm and mix gently
2. Prepare a 1% agarose/1X TBE solution
    - Measure 40 mL of 1X TBE and place into an Erlenmeyer flask
    - Add 0.4 g low-EEO agarose to the flask
    - Cover with plastic wrap and poke holes for ventilation
    - Dissolve agarose in a microwave according to the instructor's directions
    - Cool flask for 1 minute in a plastic beaker of cold water
    - Remove the flask from the water and add 2 µL of 0.5 µg/mL ethidium bromide and dispose of the tip in the ethidium bromide solid waste container
    - Swirl solution gently
3. Pour 1% agarose/1X TBE gel
    - Place gel tray in the casting tray
    - Pour the 1% agarose/1X TBE solution into the gel tray and wait for gel to solidify (~20 minutes)
4. Load the gel
    - Place the gel tray into the gel rig, pour 1X TBE running buffer into gel rig until the gel is covered by running buffer
    - Remove the combs
    - Load 3 µL of 100 bp DNA standard into the first well (on the left) in the top row
    - Load 10 µL of COI PCR reactions into separate wells
    - Record sample loading order in your laboratory notebook
5. Run the agarose gel
    - Connect the electrodes to a power source
    - Run at a constant voltage of 10 volts/cm of gel length for 30 minutes
    - While wearing gloves, turn off the power supply and remove the gel for evaluation (compare to Fig. 1)
    - Photograph the gel by exposing it to UV light and save the resulting image in your group's folder
    - Print a copy of the gel image and put in your laboratory notebook

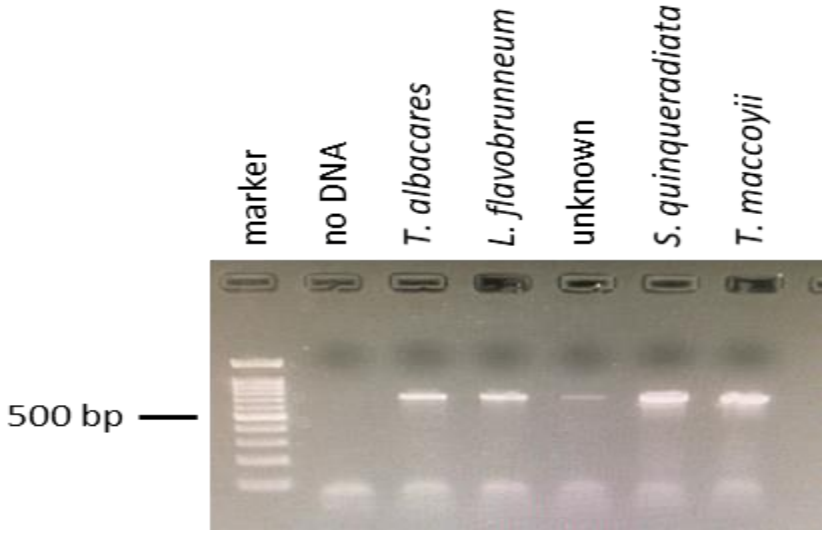


Figure 1. Sample data for amplification of COI gene. The negative control is indicated with "no DNA". Identity of the sample is listed above the gel, and "unknown" represents a sample that did not yield sequencing data.

Post-Day 3: Samples are sent out for DNA sequencing

Day 4: Evaluate/Infer the Identity of Unknown Samples

A: Receiving and Cleaning FASTA Files:

Students should open their FASTA files containing DNA sequences of their samples. Sequence data that is unreliable should be removed. These unreliable sequences are far from the primer, and hence the beginning of the sequencing reaction, and do not contain enough signal to be registered accurately. Depending on the sequence provider, these sequences may appear as lower-case letters. Chromatographs may be examined to confirm the need to remove these unreliable sequences, and evidence will be small peaks for each position. The cleaned sequence should be copied and moved into a new file.

The NCBI database (National Center for Biological Information; https://www.ncbi.nlm.nih.gov/) is part of the US National Library of Science, supported by the National Institutes of Health (https:// https://www.nih.gov/). NCBI houses and supports publicly-available, online databases (e.g., Genbank, PubMed, the NCBI Epigenomics database, and the BLAST sequence alignment program). The BLAST (Basic Local Alignment Search Tool) algorithm allows the user to search a database of nucleotide and/or amino acid sequences to infer near matches of target sequences with those of known identity, thereby allowing the user to infer the taxonomic identity of unknown samples. BLAST searches are used in this exercise to determine if correct labels have been given to tuna samples sourced from restaurants and grocery stores.

B. Performing a BLAST Search:

1. Access the text version of your FASTA file and copy the nucleotide sequence for your sample (Fig. 2). Copy the entire block of nucleotides, being sure to also Include the Sample Title (e.g., "**>Tuna001_FishCOIBCNEW**"), which will become your Job Title when you view your search results.

```
>Tuna001_FishCOIBCNEW
GTTGGCACGGCCTTAAGCTTGCTCATCCGAGCTGAACTAAGCCAACCAGG
TGCCCTTCTTGGGGACGACCAGATCTACAATGTAATCGTTACGGCCCATG
CCTTCGTAATGATTTTCTTTATAGTAATACCAATTATGATTGGAGGATTT
GGAAACTGACTTATTCCTCTAATGATCGGAGCCCCCGACATGGCATTCCC
ACGAATGAACAACATGAGCTTCTGACTCCTTCCCCCCTCTTTCCTTCTGC
TCCTAGCTTCTTCAGGAGTTGAGGCTGGAGCCGGAACCGGTTGAACAGTC
TACCCACCCCTTGCCGGCAACCTGGCCCACGCAGGGGCATCAGTTGACCT
AACTATTTTCTCACTTCACTTAGCAGGGGTTTCCTCAATTCTTGGGGCAA
TTAACTTCATCACAACAATTATCAATATGAAACCTGCAGCTATTTCTCag
tATCAAACACCACTGTTTGTATGaGCTGTACTAATTACAGCTGTTCTTCT
CCTACTTTCCCTT
```

Figure 2. Sample FASTA file for nucleotide sequence of tuna sample "**>Tuna001–FishCOIBCNEW**". The Sample Title will automatically become the Job Title when conducting your BLAST search.

2. Access the NCBI BLAST search portal by visiting: https://blast.ncbi.nlm.nih.gov/Blast.cgi. When you arrive at the portal, you will select "Nucleotide BLAST" (Appendix:Fig. A).

3. Parameters for a Standard Nucleotide BLAST search (blastn) will need to be set (Fig. 3).

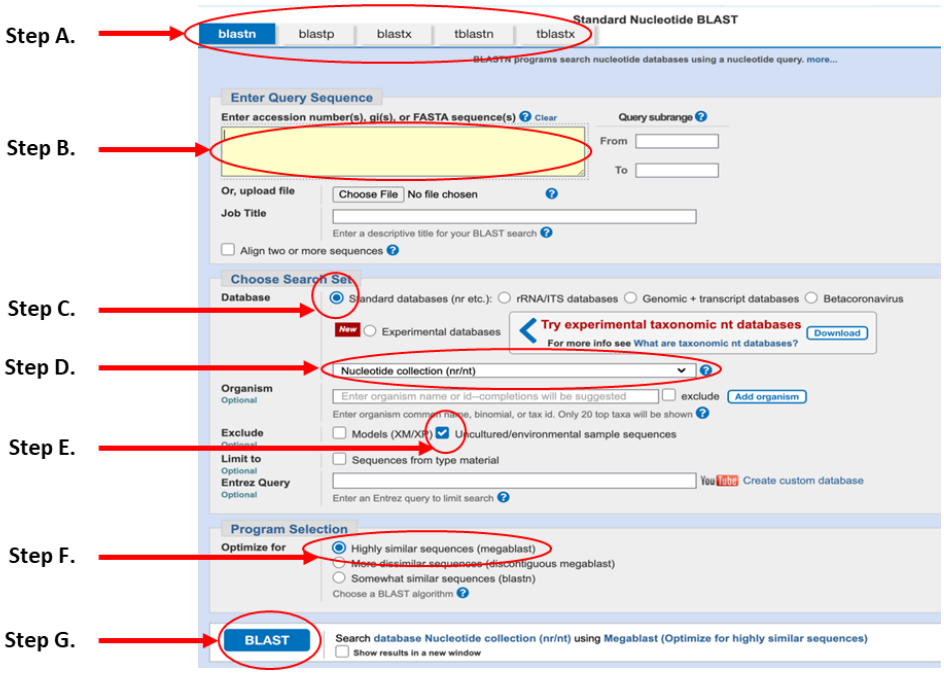
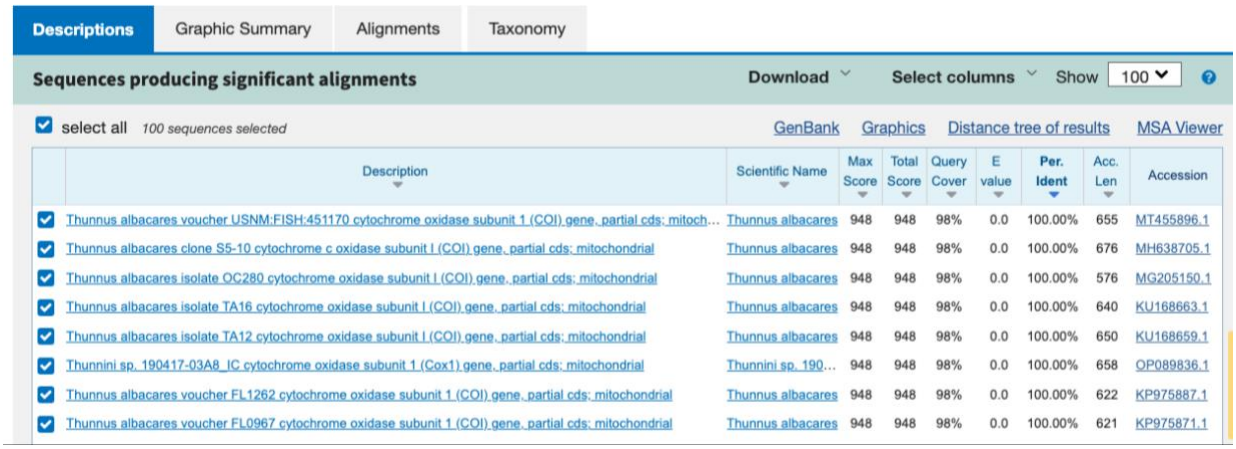


Figure 3. Interface that will be used for a Standard Nucleotide BLAST search.

Step A. Select the **blastn** tab (typically as the default for nucleotide sequences). This standard database search is used to compare the nucleotide sequence of your unknown sample against those sequences that exist in the database.
Step B. Paste your FASTA file into this field, including the Sample Title (e.g., "**>Tuna001–FishCOIBCNEW**"). If you had included a very large nucleotide sequence (e.g., that for an entire gene) but only wanted to conduct a query based on a subset of the entire sequence, enter values for the point to start ("From") and stop ("To") the search.
Step C. Databases. In the Choose Search Set field, we will select the "Standard databases (nr etc.) option.
Step D. Organism. Select the "Nucleotide collection (nr/nt)" database search from the pulldown menu. This means that we will be searching only those aspects of the database relating to nucleotide sequence (nt). This field would allow you to search many other subsets of the entire database.
Step E. Exclude Option. Select the option for excluding "Uncultured/Environmental Sample Sequences". This will eliminate from consideration all nucleotide sequences that were collected from environmental sampling, and for which there is no definitive identity to the source sequence (i.e., eliminating nucleotide sequences that came from unknown sources).
Step F. Optimize for. Select the option for "Highly similar sequences (megablast)". This constrains the algorithm to find the nucleotide sequences to which the sequence from your unknown sample are most similar. The algorithm will search only against sequences that are continuous (i.e., not discontinuous with missing regions within the entire sequence).
Step G. Search Database: Select the "BLAST" button. At this point, the algorithm will begin comparing your FASTA file against the nucleotide sequences that exist in the database, under the parameters you selected (Appendix: Fig. B). **Be Patient.** Searches typically take less than two minutes to complete.

4. Exploring your search results. The output of your BLAST search (Fig. 4) provides a visual summary that can be viewed in several formats (see tabs across top of Figure 4). View the "Descriptions" tab. The order of presentation of your BLAST results (i.e., rows in the Descriptions tab) will match the order in the other tabs.



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

select all 100 sequences selected — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Thunnus albacares voucher USNM:FISH:451170 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitoch...	Thunnus albacares	948	948	98%	0.0	100.00%	655	MT455896.1
Thunnus albacares clone S5-10 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	676	MH638705.1
Thunnus albacares isolate OC280 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	576	MG205150.1
Thunnus albacares isolate TA16 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	640	KU168663.1
Thunnus albacares isolate TA12 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	650	KU168659.1
Thunnini sp. 190417-03A8_IC cytochrome oxidase subunit 1 (Cox1) gene, partial cds; mitochondrial	Thunnini sp. 190...	948	948	98%	0.0	100.00%	658	OP089836.1
Thunnus albacares voucher FL1262 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	622	KP975887.1
Thunnus albacares voucher FL0967 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	621	KP975871.1

Figure 4. BLAST search results that show the highest degree of similarity with the nucleotide sequence of your unknown sample. Tabs for viewing BLAST results in various formats shown here. Columns indicating "Description", "Scientific Name", and "Per. Ident" (Percent Identical) will be used to infer the possible identity of your unknown samples. This is the BLAST result for **Scenario 1** (below), in which the Per. Ident values all fall above the 98.00% value, and for which most all suggested scientific names indicate that the sample hits to *Thunnus albacares.*

"Description" This provides a brief summary of the entity that was submitted by the original author of that field of data. This field will present information such as the taxonomy (as proposed in the original GenBank entry), accession number or strain identification number, name of the gene or region of genome sampled, and notes about the source of the entry.
"Scientific Name" This is the scientific name (ideally the Latin binomial, such as Genus and species) for the strain, as proposed by the investigator(s) who submitted the genome to GenBank.
"Max Score" A calculation of the alignment score, based on summing the number of matched nucleotides and mismatches and gaps between the target sample (i.e., your FASTA file) and the sequence to which it is being compared.
"Total Score" A calculation of the alignment scores for all segments of the subject source.
"Query Coverage" A calculation of the percentage of the contig length (your FASTA file) that perfectly aligns with the NCBI sequence to which your unknown hits or matches.
"E Value" A description of the number of hits you can expect to see by random chance when searching the database.
"Per. Ident" A measure of the extent to which your FASTA file perfectly matches the nucleotides of a sequence to which your FASTA file hits (is being compared); this is the primary measure used to evaluate the hits from your BLAST search.
"Acc. Len" A measure of the number of nucleotide positions in a hit sequence that overlap or match your FASTA file, based on position, independent of which nucleotide resides in that position.
"Accession" The unique identifier given to GenBank records, once they are uploaded into the nucleotide database.

5. Using "Per. Ident" to infer Identity of Your Sample. The values shown in the "Per. Ident" column are used to assess the quality of a match (or hit) and provides a tool to determine how well a sample sequence matches an entry within the database. This value is compared to a taxon-specific threshold value to make this determination. For vertebrate zoologists, Per. Ident values greater than 98% are indicative of a likely match to a particular species, whereas Per. Ident values in the range of 97% are indicative that an unknown sample belongs to a particular genus. For this study of tuna samples, we will use a threshold of Per., Ident ≥ 98% to infer the possible species identity of our unknown samples. Note that different thresholds exist for recognizing species of different taxonomic groups [see Krishnamurthy and Francis (2012) and Ratnasingham and Hebert (2013)]. Below, you are presented three scenarios that are commonly encountered when comparing Per. Ident numbers from your BLAST results. In each scenario, compare your Per. Ident. values to the threshold of 98.0%. If your values of percent ident fall between 98–100.0%, we will infer that the blast hit is a "good" match to identity, and we will consider the taxonomy proposed by the GenBank reference.

Scenario 1. All values of Percent Ident are above a given threshold. Examine the BLAST hits presented in Figure 4, which is based on the BLAST search of the FASTA file for ">Tuna001_FishCOIBCNEW". All values of "Per. Ident" are 100.00%, suggesting that the nucleotide sequence in the FASTA file of the reference strain (your unknown) is 100.00% identical to the particular entry to which it hits in the BLAST search. Examine the entry for "Scientific Name" for the first row—because the

Per. Ident value for this entry is greater than the 98.00% threshold, evidence suggests that our unknown sample belongs to the species, *Thunnus albacares* (*Thunnus* = generic epithet; *albacares* = specific epithet; see column "Scientific Name"). Further examination of these 8 hits shows that all Per Ident values are above the 98% threshold (indicating good matches to a particular taxon), and all have the suggested taxonomy "*Thunnus albacares*". The only exception being for the sixth hit, which identifies the taxonomic identity as "*Thunnus* sp.", which suggests that the investigator who entered the identity for this GenBank entry felt confident that the sample came from a member of the genus *Thunnus,* but that they were not confident on the particular species identity of that sample. A general consensus of our results provides good evidence that our unknown sample certainly belongs to the genus *Thunnus,* and most likely to the species *Thunnus albacares* (common name = Yellowfin Tuna). In this case, our unknown sample is considered to have been taken from the Yellowfin Tuna (*T. albacares*).

Scenario 2. Some (but not all) values of Percent Ident are above a given threshold. There might be obvious breaks in the value of Per. Ident. for your hits, and some of those values might fall above the 98.00% threshold. The taxonomic identity of those hits for which Per. Ident. ≥98.00% should be considered as applicable to your unknown. Conversely, the taxonomic identity of those hits that fall below the threshold likely are not applicable to your unknown.

Scenario 3. All values of Percent Ident are above the reference threshold, but the taxonomy of our sample is uncertain. If all values of Per. Ident. are above your threshold, but each hit suggests a different scientific name, then the proposed identity of your unknown sample is uncertain with the available data and analytical tools. For example, one hit might suggest that your sample belonged to *Thunnus thynnus,* whereas another suggests that it belongs to *Thunnus orientalis,* and yet another might suggest that your unknown sample belongs to *Thunnus maccoyii.* In this scenario, there is a logical inconsistency in your results, because it is unlikely that your sample comes from more than a single species. Think about what statements can be made in this scenario.

6. Once you have inferred the taxonomic identity of your unknown (ideally to genus and species), perform a browser search to determine the common name of the species to which your unknown sample belongs.

Primary Research Question: Based on this, can you state whether the unknown sample was correctly labeled at its source (i.e., in the restaurant or grocery store)?

Discussion

1. Did the taxonomic identities (common and/or scientific names) inferred from your BLAST searches match the identity of your unknown samples at the source (e.g., the labels given to the samples by the restaurant or grocery store)? Were the unknown samples that you sequenced correctly labeled at the source from which they were collected?
2. Following these protocols, were your BLAST results for each sample more like Scenario 1, 2, or 3?
3. What are some possible explanations for why Scenario 3 might occur? What might explain why your unknown samples BLAST to more than one species?
4. Can you be certain that the GenBank samples are correctly identified? For example, the original investigator was required to input a taxonomic identity (e.g., scientific name, or Family, or Order) for their uploaded sequences—can we be certain that they provided a correct taxonomic identity for their samples?
5. It is possible for your FASTA files to be identical to those that exist for multiple species in the database. Under what conditions/situations might this scenario exist?
6. If your results match Scenario 3, what could be done to "break the ties"?
7. Why might cooked or processed tuna (or seafood) not be suitable for DNA barcoding?
8. What are some advantages and disadvantages of using a mitochondrial gene, such as *COI*, in DNA barcoding?
9. If you identified examples of food mislabeling, where along the supply chain might the mislabeling have occurred?
10. Can you think of other examples where DNA barcoding could be applied to questions of food labeling?

Cited References

Carmel J. 2021 Jun 19. The big tuna sandwich mystery. New York Times. Available from: https://www.nytimes.com/2021/06/19/style/subway-tuna-sandwich-lawsuit.html.

Filonzi L, Ardenghi A, Rontani PM, Voccia A, Ferrari C, Papa R, Bellin N, and Marzano FN. 2023. Molecular barcoding: a tool to guarantee correct seafood labelling and quality and preserve the conservation of endangered species. Foods. 12:2420.

Krishnamurthy PK, Francis RA. 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. Biodivers Conserv 21:1901–19. doi.: 10.1007/s10531-012-0306-2.

Ratnasingham S, Hebert PDN. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLOS One. 8(8):366213. doi: 10.1371/journal.pone.0066213.

United States Food & Drug Administration. 2023. Code of Federal Regulations Title 21 [Internet]. Silver Springs (MD). [cited 2023 Aug 14]. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=101.18

MATERIALS

Genomic DNA isolation, amplification, and sequencing
QIAgen DNeasy Blood & Tissue kits ($215 for 50 preps) were used for genomic DNA isolation. GoTaq polymerase ($51 for 100 reactions) and dNTP mix ($30 for 200 µl of 10 mM solution) was acquired from Promega (Madison, WI), and primers were purchased from Eurofins Genomics (Louisville, KY). PCR products were sequenced by Functional Biosciences (Madison, WI).

Primers:
FISHCOIBC_ts-new: 5'-CTCAACYAATCAYAAAGATATYGGCAC-3'
FishCOIHBC-deg: 5'-ACTTCYGGGTGRCCRAARAATCA-3'
Y = C or T, R = A or G

NOTES FOR THE INSTRUCTOR

1. Tuna samples can be acquired from local sushi restaurants, restaurants, or markets. Restaurants are ideal because tuna samples are relatively small and easy to transport, and printed menus provide a record of how tuna is being labeled and sold. In earlier studies, we solicited samples from local restaurants and fishmongers by written letter; however, knowledge of the study's intent could bias the sample collection as reluctance to provide samples that could be mislabeled is more likely.
2. Do not use bleach during the genomic DNA isolation procedure. Properly dispose of the flowthrough according to local guidelines.
3. PCR samples may be submitted for sequencing with the primers that are used for PCR according to the guidelines provided by your DNA sequencing company.
4. There are many DNA visualization techniques that do not use ethidium bromide, a known carcinogenic agent, and the authors encourage you to use reagents that are not harmful. Additionally, agarose gels can be used that are much larger and more appropriate for sharing by students in order to minimize cost. Instructors may choose to pour the gels and have them available for students to load.
5. Instructors are encouraged to perform the entire set of protocols prior to lab in case any optimization of conditions is needed.
6. Scenarios 1–3 FASTA files are present in Appendix A; instructors can choose to do these as practice or examples.
7. Overnight incubation at 56˚C is a modification to the manufacturer's genomic DNA isolation protocol.
8. DNA concentrations can be estimated by running samples on an agarose gel with a standard, and sample integrity can be assessed by this method. Alternatively, spectrophotometric readings at A_{260} can be used to determine DNA concentrations.
9. Lower case letters in FASTA files from Functional Bioscience indicates low quality bases.

CITED REFERENCES

Carmel J. 2021 Jun 19. The big tuna sandwich mystery. New York Times. Available from: https://www.nytimes.com/2021/06/19/style/subway-tuna-sandwich-lawsuit.html.

Filonzi L, Ardenghi A, Rontani PM, Voccia A, Ferrari C, Papa R, Bellin N, and Marzano FN. 2023. Molecular barcoding: a tool to guarantee correct seafood labelling and quality and preserve the conservation of endangered species. Foods. 12:2420.

Krishnamurthy PK, Francis RA. 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. Biodivers Conserv 21:1901–19. doi.: 10.1007/s10531-012-0306-2.

Ratnasingham S, Hebert PDN. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLOS One. 8(8):366213. doi: 10.1371/journal.pone.0066213.

United States Food & Drug Administration. 2023. Code of Federal Regulations Title 21 [Internet]. Silver Springs (MD). [cited 2023 Aug 14]. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=101.18.

ACKNOWLEDGEMENTS

We would like to thank the National Science Foundation Scholarships in STEM program for summer research support to Madison Sobonya (DUE 1741814). In addition, the Colleran-Weaver Fund at John Carroll University provided support for this work.

About the Authors

Mike Martin is a Professor of Biology. He teaches introductory-to-graduate courses in Clinical Microbiology, Molecular Biology, and Genetics. His undergraduate degree is in Biology from the University of Cincinnati, and he earned his PhD in Cell and Molecular Biology from the University of Wisconsin-Madison. He has published research in laboratory education, yeast genetics, and molecular systematics of cyanobacteria.

Madison Sobonya earned a BS in Biology from John Carroll University where she was a National Science Foundation S-STEM Scholar. She is currently attending the Ohio University Heritage College of Osteopathic Medicine.

Christopher A. Sheil is a Professor of Biology. He teaches introductory-to-graduate courses in Principles of Biology II (Animal & Plant Anatomy & Physiology), Principles of Biology III (Biodiversity & Evolution), Vertebrate Anatomy, Biology of the Reptilia, Scientific Illustration, and Introduction to Systematics. He earned a BS in Systematics & Ecology and a PhD in Ecology & Evolutionary Biology from The University of Kansas. He has published research in areas of herpetology, developmental biology, evolutionary anatomy, taxonomy, and phylogenetic systematics.

APPENDIX A

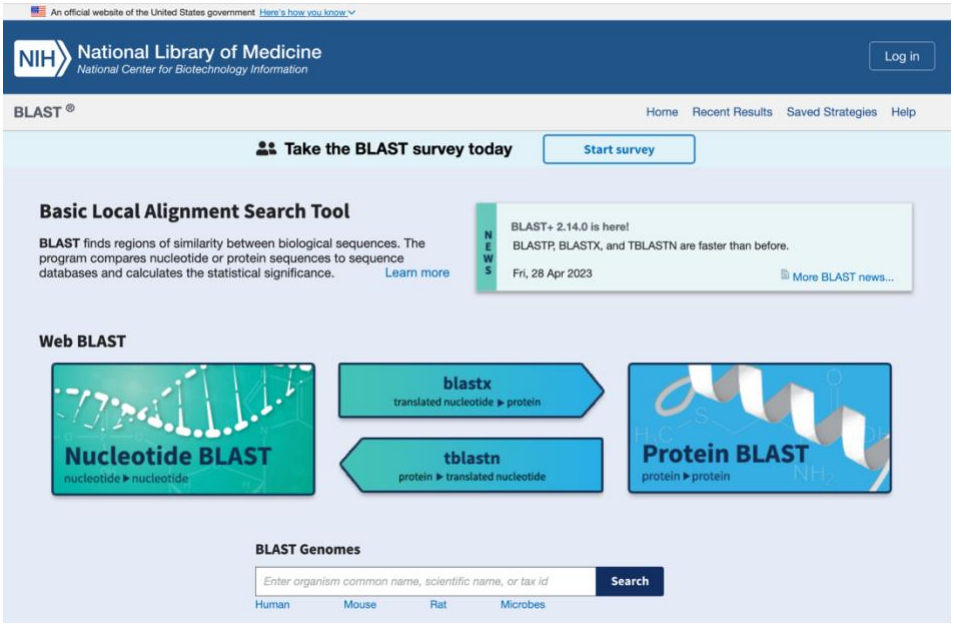


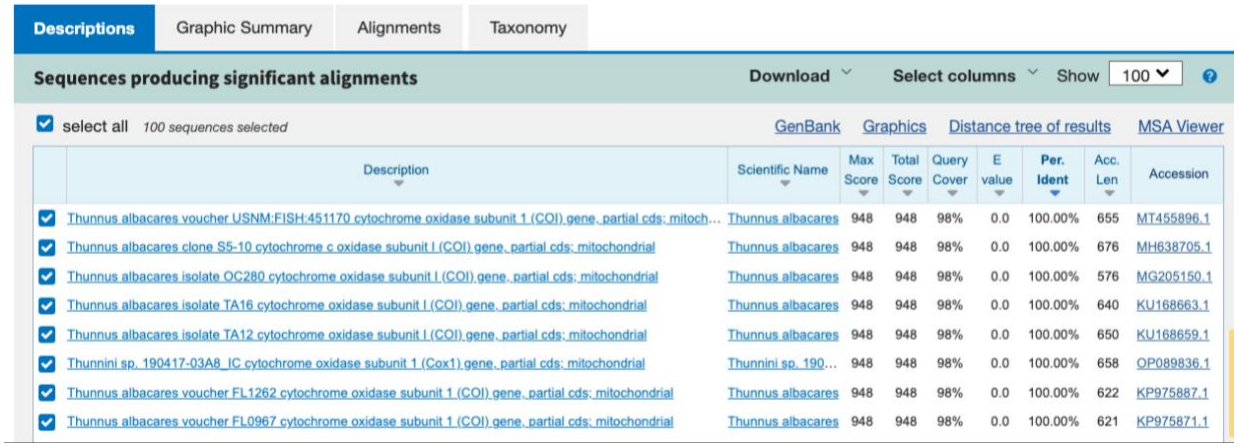
Figure A. NCBI Basic Local Alignment Search Tool (NCBI) portal. To compare the nucleotide sequence of your sample of DNA to those of known samples in the NCBI database, Select "Nucleotide BLAST".

Figure B. Job Status for BLAST search of your sample. The time since submission number will often be static, with only periodic updates. Be patient! Searches typically take less than two minutes to complete.

Browser link for BLAST search:
https://blast.ncbi.nlm.nih.gov/Blast.cgi

FASTA File: Scenario 1

```
>Tuna001_FishCOIBCNEW
GTTGGCACGGCCTTAAGCTTGCTCATCCGAGCTGAACTAAGCCAACCAGG
TGCCCTTCTTGGGGACGACCAGATCTACAATGTAATCGTTACGGCCCATG
CCTTCGTAATGATTTTCTTTATAGTAATACCAATTATGATTGGAGGATTT
GGAAACTGACTTATTCCTCTAATGATCGGAGCCCCCGACATGGCATTCCC
ACGAATGAACAACATGAGCTTCTGACTCCTTCCCCCCTCTTTCCTTCTGC
TCCTAGCTTCTTCAGGAGTTGAGGCTGGAGCCGGAACCGGTTGAACAGTC
TACCCACCCCTTGCCGGCAACCTGGCCCACGCAGGGGCATCAGTTGACCT
AACTATTTTCTCACTTCACTTAGCAGGGGTTTCCTCAATTCTTGGGGCAA
TTAACTTCATCACAACAATTATCAATATGAAACCTGCAGCTATTTCTCag
tATCAAACACCACTGTTTGTATGaGCTGTACTAATTACAGCTGTTCTTCT
CCTACTTTCCCTT
```



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | Select columns | Show 100

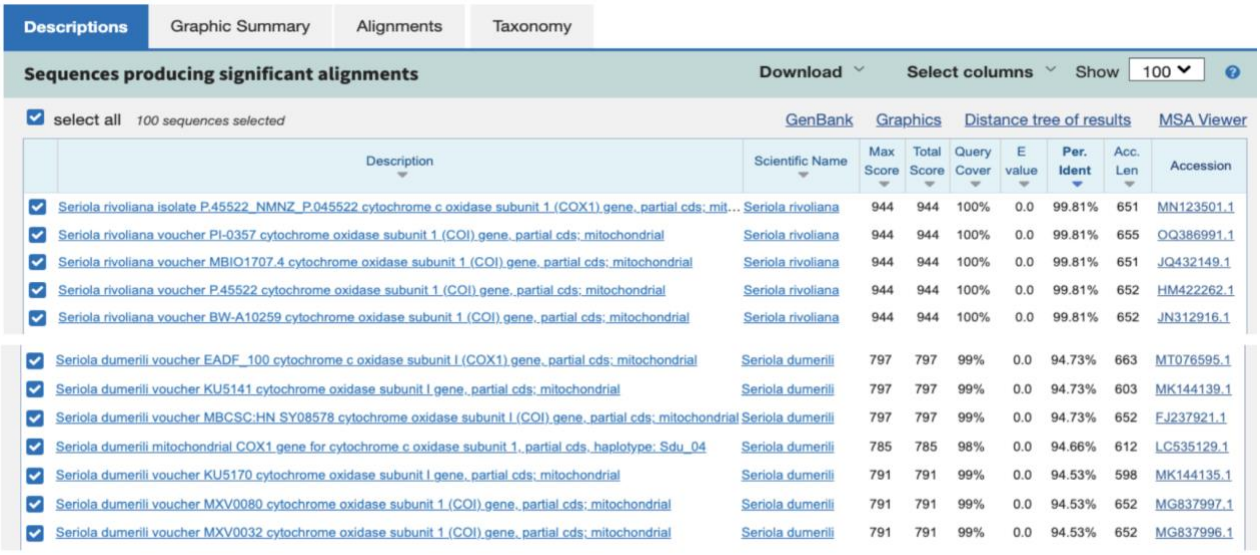
select all *100 sequences selected* — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Thunnus albacares voucher USNM:FISH:451170 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitoch...	Thunnus albacares	948	948	98%	0.0	100.00%	655	MT455896.1
Thunnus albacares clone S5-10 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	676	MH638705.1
Thunnus albacares isolate OC280 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	576	MG205150.1
Thunnus albacares isolate TA16 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	640	KU168663.1
Thunnus albacares isolate TA12 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	650	KU168659.1
Thunnini sp. 190417-03A8_IC cytochrome oxidase subunit 1 (Cox1) gene, partial cds; mitochondrial	Thunnini sp. 190...	948	948	98%	0.0	100.00%	658	OP089836.1
Thunnus albacares voucher FL1262 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	622	KP975887.1
Thunnus albacares voucher FL0967 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Thunnus albacares	948	948	98%	0.0	100.00%	621	KP975871.1

For the instructor. This is the BLAST result from Scenario 1 (using sample ">Tuna001_FishCOIBCNEW", described in body of manuscript). These BLAST search results (same as Figure 4 in text) show the highest degree of similarity with the nucleotide sequence of your unknown sample. In this BLAST result, all Per. Ident values fall above the 98.00% value, and for each hit there is strong evidence to support that the sample hits to *Thunnus albacares.* Tabs for viewing BLAST results in other formats (e.g., Graphic Summary, Alignments, and Taxonomy) would be available by selecting those tabs. Columns indicating "Description", "Scientific Name", and "Per. Ident" (Percent Identical) will be used to infer the possible identity of your unknown samples. A general consensus of our results suggests good evidence that our unknown sample belongs to the genus *Thunnus,* and most likely to the species *Thunnus albacares* (common name = Yellowfin Tuna). In this case, our unknown sample is considered to have been taken from the Yellowfin Tuna (*T. albacares*).

FASTA File: Scenario 2

>Tuna007_FishCOIBCNEW
GTCGGTACAGCCCTAAGTTTACTCATCCGAGCAGAACTAAGTCAACCTGG
AGCTCTCCTGGGAGACGATCAGATTTACAACGTAATCGTTACGGCACACG
CGTTTGTAATAATTTTCTTTATAGTAATGCCAATTATGATTGGAGGGTTT
GGAAACTGACTCATCCCTTTAATGATTGGAGCTCCCGATATAGCATTCCC
TCGAATGAACAATATGAGCTTCTGACTCCTCCCCCCTTCATTCCTTCTGC
TCCTGGCCTCTTCAGGTGTTGAAGCCGGAGCCGGAACAGGTTGAACAGTT
TACCCGCCCCTAGCCGGCAACCTTGCCCACGCAGGAGCATCCGTAGACTT
AACGATTTTCTCTCTTCATCTAGCTGGGATCTCCTCAATTCTAGGAGCTA
TTAACTTTATCACAACCATCGTCAATATGAAACCCCACGCCGTTTCCATG
TACCAAATCCCCCTGTTTGTCTGagcTGTCCTAATTACGGCTGTGCTTCT
ACTCCTATCACTCC



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

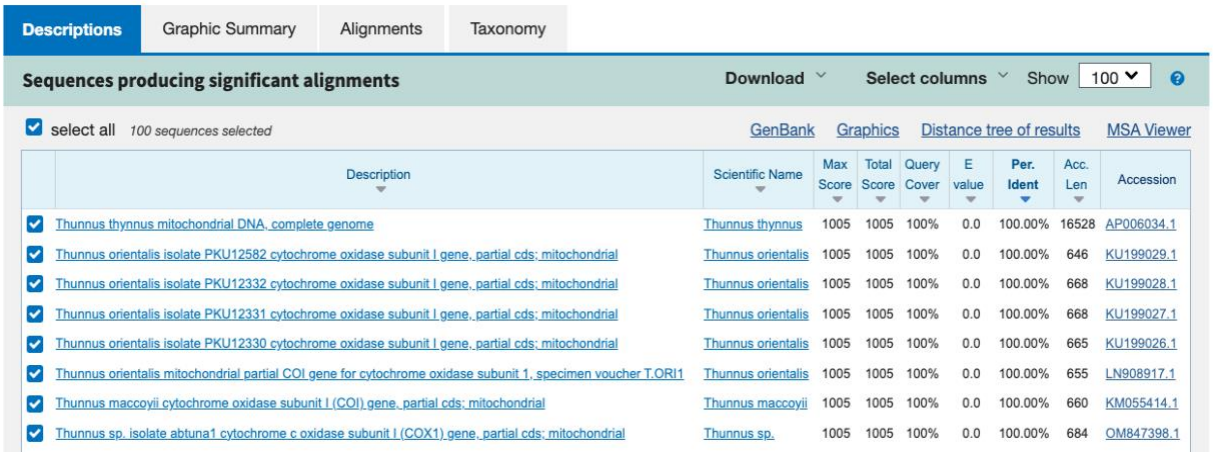
select all *100 sequences selected* — GenBank — Graphics — Distance tree of results — MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Seriola rivoliana isolate P.45522_NMNZ_P.045522 cytochrome c oxidase subunit 1 (COX1) gene, partial cds; mit...	Seriola rivoliana	944	944	100%	0.0	99.81%	651	MN123501.1
Seriola rivoliana voucher PI-0357 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola rivoliana	944	944	100%	0.0	99.81%	655	OQ386991.1
Seriola rivoliana voucher MBIO1707.4 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola rivoliana	944	944	100%	0.0	99.81%	651	JQ432149.1
Seriola rivoliana voucher P.45522 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola rivoliana	944	944	100%	0.0	99.81%	652	HM422262.1
Seriola rivoliana voucher BW-A10259 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola rivoliana	944	944	100%	0.0	99.81%	652	JN312916.1
Seriola dumerili voucher EADF_100 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Seriola dumerili	797	797	99%	0.0	94.73%	663	MT076595.1
Seriola dumerili voucher KU5141 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Seriola dumerili	797	797	99%	0.0	94.73%	603	MK144139.1
Seriola dumerili voucher MBCSC:HN SY08578 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Seriola dumerili	797	797	99%	0.0	94.73%	652	FJ237921.1
Seriola dumerili mitochondrial COX1 gene for cytochrome c oxidase subunit 1, partial cds, haplotype: Sdu_04	Seriola dumerili	785	785	98%	0.0	94.66%	612	LC535129.1
Seriola dumerili voucher KU5170 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Seriola dumerili	791	791	99%	0.0	94.53%	598	MK144135.1
Seriola dumerili voucher MXV0080 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola dumerili	791	791	99%	0.0	94.53%	652	MG837997.1
Seriola dumerili voucher MXV0032 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	Seriola dumerili	791	791	99%	0.0	94.53%	652	MG837996.1

For the instructor: This is the BLAST result for Scenario 2, using ">Tuna007–FishCOIBNEW". The values of "Per. Ident" for the first five hits are 99.81%, which falls above the 98.00% threshold and suggests that the unknown belongs to *Seriola rivoliana* (the Longfin Yellowtail; Carangidae; Carangiformes). However, the values for Per. Ident for hits 6–12 are ≤ 94.73%. Because these values fall below the 98.00% threshold, we would not consider our unknown to belong to *Seriola dumerilii* (the Greater amberjack).

FASTA File: Scenario 3

```
>Tuna011_FishCOIBCNEW
GTTGGCACGGCCTTAAGCTTGCTCATCCGAGCTGAACTAAGCCAACCAGG
TGCCCTTCTTGGGGACGACCAGATCTACAATGTAATCGTTACGGCCCATG
CCTTCGTAATGATTTTCTTTATAGTAATACCAATTATGATTGGAGGATTT
GGAAACTGACTTATTCCTCTAATGATCGGAGCCCCCGACATGGCATTCCC
ACGAATGAACAACATGAGCTTCTGACTCCTTCCTCCCTCTTTCCTTCTGC
TCCTAGCTTCTTCAGGAGTTGAGGCTGGGGCCGGAACCGGTTGAACAGTC
TACCCTCCCCTTGCCGGCAACCTAGCCCACGCAGGGGCATCAGTTGACCT
AACTATTTTCTCACTTCACTTAGCAGGGGTTTCCTCAATTCTTGGGGCAA
TTAACTTCATCACAACAATTATCAATATGAAACCTGCAGCCATCTCTCAA
TATCAAACACCACTGTTTGTATGAGCTGTACTAATTACAGCTGTTCTTCT
TCTACTTTCCCTTCCAGTCCTTGCCGCTGGTATTACAATGCTcC
```



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | Select columns | Show 100

select all *100 sequences selected* — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Thunnus thynnus mitochondrial DNA, complete genome	Thunnus thynnus	1005	1005	100%	0.0	100.00%	16528	AP006034.1
Thunnus orientalis isolate PKU12582 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Thunnus orientalis	1005	1005	100%	0.0	100.00%	646	KU199029.1
Thunnus orientalis isolate PKU12332 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Thunnus orientalis	1005	1005	100%	0.0	100.00%	668	KU199028.1
Thunnus orientalis isolate PKU12331 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Thunnus orientalis	1005	1005	100%	0.0	100.00%	668	KU199027.1
Thunnus orientalis isolate PKU12330 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Thunnus orientalis	1005	1005	100%	0.0	100.00%	665	KU199026.1
Thunnus orientalis mitochondrial partial COI gene for cytochrome oxidase subunit 1, specimen voucher T.ORI1	Thunnus orientalis	1005	1005	100%	0.0	100.00%	655	LN908917.1
Thunnus maccoyii cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Thunnus maccoyii	1005	1005	100%	0.0	100.00%	660	KM055414.1
Thunnus sp. isolate abtuna1 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Thunnus sp.	1005	1005	100%	0.0	100.00%	684	OM847398.1

For the instructor: BLAST results for Scenario 3 (using ">Tuna011–FishCOIBNEW", in which the Per. Ident values all fall above the 98.00% for all 8 blast hits shown, but for which the identity of the proposed hits are complicated. The first hit indicates that our unknown belongs to *Thunnus thynnus* (the Atlantic Bluefin Tuna), whereas hits 2–6 suggest that the unknown is *Thunnus orientalis* (the Pacific Bluefin Tuna), hit 7 suggests the identity is *Thunnus maccoyii* (the Southern Bluefin Tuna), and hit 8 suggests simply that the hit is to *Thunnus sp* (falling within the genus *Thunnus,* but with uncertainty about the specific epithet).

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 https://www.ableweb.org/.

Papers published in *Advances in Biology Laboratory Education: Peer-Reviewed Publication of the Conference of the Association for Biology Laboratory Education* 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.