Abstract
The tools of molecular biology provide a rich platform for teaching the scientific process, as interesting questions pertaining to fields such as evolution and ecology can be pursued on short time scales. In this inquiry-based laboratory project, students investigate the authenticity of fish products purchased in local markets and restaurants by DNA sequence analysis of a segment of the mitochondrial cytochrome c oxidase subunit I (COI) gene. In the course of their investigation, students are exposed to fundamental molecular biology techniques such as DNA isolation, agarose gel electrophoresis, polymerase chain reaction, DNA sequencing mechanisms, and DNA database analysis. In addition, students will observe how the evolutionary relatedness of species is reflected in the genetic code, and consider how the ecology of fish species influences their product distribution and environmental impact. This project is suitable for advanced high school or undergraduate students.

Key Words: polymerase chain reaction; agarose gel electrophoresis; DNA sequence alignment; fish species identification.

Introduction
The Next Generation Science Standards promote the understanding of core ideas as well as the development of scientific practices, and particularly emphasize the use of evidence to generate explanations for scientific problems (NGSS Lead States, 2013). Inquiry-based activities are an excellent means of helping students understand and appreciate the scientific concepts and techniques used in modern biology, especially if they can be related to students’ lives outside the classroom. Furthermore, such activities allow students to engage in the scientific method directly by generating questions, evidence, and conclusions in a hands-on laboratory setting. As a means of introducing fundamental molecular biology tools such as DNA isolation, agarose gel electrophoresis, polymerase chain reaction (PCR), DNA sequencing, and DNA database analysis, we have designed a laboratory exercise to detect mislabeled fish (i.e., “fish fraud”) from local restaurants and supermarkets. Previous studies have used data collected in high school or undergraduate courses to investigate seafood mislabeling (Naaum & Hanner, 2015; Willette et al., 2017), and have described laboratory exercises in which students analyze fish DNA sequences (Hallen-Adams, 2015; Clime & Gogarten, 2012). Our article builds on this work by outlining a multiday project in which students perform inquiry-based research from start to finish, and by incorporating sequencing and phylogenetic analysis of fish species from multiple families and orders. We have also included a number of exercises designed to expand students’ understanding of primer design and DNA sequence structure and analysis, and to stimulate cross-disciplinary discussion of ecological sustainability and the economics involved in seafood mislabeling.

U.S. consumers spent an estimated $91.7 billion for fishery products in 2013, and consumed an average of 14.6 pounds of fish and shellfish per capita per year between 2000 and 2009 (NMFS, 2014). Unfortunately, recent investigations have discovered species substitution and mislabeling of seafood products. A nationwide study by Oceana from 2010 to 2012 found that 33 percent of 1,215 seafood samples were mislabeled (Warner et al., 2013). And recent DNA testing of 174 fish product lots by the Food and Drug Administration found that 15 percent were incorrectly labeled (FDA, 2013). Table 1 includes a list of commonly substituted fish.

Many fish species may look and taste similar, so DNA testing is often necessary to determine the true identity of the sample. The technique of DNA barcode identification is based on the comparison of a DNA sequence from a defined region of the genome to the database of genome sequences from known species of fish. For this project, we...
analyzed the approximately 700 base pair (bp) region of the mitochondrial cytochrome c oxidase subunit I gene (COI). This region of the genome has been used extensively in the DNA barcoding of animals (Hebert et al., 2003). Universal PCR primers have been developed to amplify a section of the COI gene in the fish genome (Ward et al., 2005), and we found these published primers to work very well in PCR amplification of all of our samples.

This investigative laboratory exercise was used successfully in a two-week summer course for high school students, and in a second-year undergraduate Genetics course at Carleton College. This project was designed to span four 3-hour laboratory periods (Table 2), but can be adjusted to fit other time frames. The exercises can be adapted to a variety of student audiences: from an introduction to molecular biology for high school students, to an advanced research project for an undergraduate laboratory course. In the Summary and Extensions section we describe activities for students needing additional practice, as well as more independent learning opportunities for advanced students. Our students worked in groups of 2–4, thus helping to foster scientific collaboration and communication skills. The goal of this lab is that students will achieve the following:

1. Appreciate the utility of DNA based testing.
2. Understand and perform a variety of molecular biology techniques: DNA extraction, agarose gel electrophoresis, and polymerase chain reaction.
3. Maintain a comprehensive and independent research notebook.
4. Navigate bioinformatics tools (e.g., BLAST, sequence alignment) to analyze DNA sequence data.
5. Synthesize results and present conclusions.
6. Appreciate the connections between investigative science and their daily lives.

○ Project Summary & Methods

Introduction

To get students motivated and prepared for the project, it is helpful to present the findings of previous studies on fish fraud. A good place to start is to have the students read Oceana’s online summary (Warner et al., 2013). Students can also search for articles online that deal with fish fraud as an assignment prior to the first laboratory period. It is important for the instructor to impress upon the students that fish mislabeling can happen anywhere along the supply chain, thus the final distributor may not be aware of mislabeling of their products. All of the supplemental protocols used in this project can be found at our Carleton College laboratory website (https://apps.carleton.edu/people/szweifel/fishfraud/).

Step 1: Obtaining Fish Samples

If time permits, the instructor can have the students collect fish samples from local vendors (a great way to engage them in the investigative project!). We collected samples from grocery stores, fish markets, campus dining hall, restaurants, and fast food establishments. Both fresh and cooked fish worked well in our DNA extraction. A sample of fish the size of a sugar cube can be wrapped in aluminum foil and then stored in the freezer. Fish such as red snapper, wild salmon, white tuna, or cod are often mislabeled and are thus good candidates for students to analyze. We recommend marking samples with a numbering system before distributing them to students, and emphasizing the importance of accurate recordkeeping. Maintaining a lab notebook is good practice for future research experiences, and enables the students to organize their experiments and data.

Step 2: DNA Extraction

Students used the DNeasy Blood and Tissue Kit (QIAGEN Inc., product # 69504) to extract DNA from their fish samples. This kit uses spin column-based DNA purification, but presumably any DNA extraction kit could be substituted. Only a small amount of fish tissue is necessary, about 25 mg (approximately the size of a corn kernel). The protocol requires a 2 hr incubation at 56°C for cell lysis, followed by a series of washes and microcentrifuge spins (Supplemental Protocol #1). We found that an explanation of column purification before the extraction and periodic reminders (“Where is the DNA?”) helped students to understand the rationale behind the extraction protocol. This DNA extraction method resulted in genomic DNA that was a reliable substrate for the subsequent PCR reaction (i.e., greater than 90% PCR success rate without needing to determine gDNA concentration).

Step 3: PCR Amplification of the COI Gene

After a discussion of the PCR technique in the classroom, we used PCR primers designed for fish DNA barcoding (Ward et al., 2005) and the HotStarTaq Master Mix Kit (QIAGEN Inc., product #

---

### Table 1. Common fish species substitutions. Adapted from the summary reports of Oceana and the FDA revealing seafood fraud (Warner et al., 2013; FDA, 2013).

<table>
<thead>
<tr>
<th>Labeled species</th>
<th>Actual species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaskan/Pacific cod</td>
<td>Pangasius (Asian “catfish”), Atlantic cod, threadfin, slickhead, Alaska pollock</td>
</tr>
<tr>
<td>Salmon (wild, king, sockeye)</td>
<td>Farmed Atlantic salmon</td>
</tr>
<tr>
<td>Sea bass</td>
<td>Antarctic toothfish, Patagonian toothfish</td>
</tr>
<tr>
<td>Red snapper</td>
<td>Various snappers, Pacific ocean perch, rockfish, tilapia, white bass, giltheaded seabream</td>
</tr>
<tr>
<td>Lemon sole</td>
<td>Flounder, flathead sole</td>
</tr>
<tr>
<td>White tuna</td>
<td>Escolar</td>
</tr>
<tr>
<td>Walleye</td>
<td>Sauger</td>
</tr>
</tbody>
</table>

---
203443) to amplify a 700 bp region of the COI gene. PCR primer concentrations were 10 μM and sequences were as follows:

- FishF1 forward primer 5′ TCAACCAACCACAAAGACATTGGC AC 3′, and
- FishR1 reverse primer 5′ TAGACTTCTGGGTGGCCAAAGA TCA 3′.

The PCR program contained an initial activation step at 95°C for 15 min., 35 cycles of denaturation at 94°C for 30 sec., annealing at 54°C for 30 sec., and elongation at 72°C for 1 min., and a final extension step at 72°C for 10 min. (See Supplemental Protocol #2 for further details.) PCR set-up and reaction time takes about 2 hours, but the PCR machine can also be run overnight and the tubes transferred to the freezer the next day for subsequent steps. For more advanced students, using a gradient PCR machine and running multiple reactions is a good way to emphasize the effect of varying annealing temperatures on the resulting PCR product (i.e., low temperatures may lead to nonspecific amplification, and high temperatures may lead to lack of amplification).

**Step 4: PCR Product Visualization**

To determine if the COI region was successfully amplified, students used agarose gel electrophoresis to detect the PCR fragment (Supplemental Protocol #3). Students were responsible for making the 1 percent agarose gels, including adding the fluorescent stain. We used GelStar because it is a highly sensitive dye for visualizing DNA using UV light, and it is less toxic than traditional ethidium bromide staining. (Other nucleic acid stains such as SybrSafe or GelRed could also be used.) While the agarose was cooling, students prepared their samples for loading. We used a 100 bp DNA marker ladder (New England Biolabs, product # N3231S) so that students could identify the amplified 700 bp segment. This laboratory session allowed us to discuss DNA migration in agarose gels and the use of a size standard to determine the length of our PCR bands. (Having the students prepare a graph of DNA length-vs-distance run for the marker lane is another useful exercise.) Figure 1 shows the results of the PCR products from six different fish samples separated on an agarose gel.

---

**Table 2. Synopsis of laboratory activities.** The timeline can be extended with additional activities (see Summary and Extensions section) and/or creation of posters summarizing the project.

<table>
<thead>
<tr>
<th>Before lab</th>
<th>Lab 1</th>
<th>Lab 2</th>
<th>Lab 3</th>
<th>Lab 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory activities</td>
<td>Obtain fish samples</td>
<td>DNA isolation from fish (~2hrs)</td>
<td>PCR amplification of COI region (~3hrs)</td>
<td>(1) Gel of PCR product (2) PCR product purification (3) DNA sequencing (3–4hrs)</td>
</tr>
<tr>
<td>Supplemental Protocols</td>
<td>—</td>
<td>#1</td>
<td>#2</td>
<td>#3, 4, 5</td>
</tr>
<tr>
<td>Reagents and expenses</td>
<td>—</td>
<td>DNeasy Blood and Tissue Kit ($3/sample)</td>
<td>PCR primers ($15 total) HotStarTaq Master Mix ($2/sample)</td>
<td>MinElute PCR Purification ($2.50/reaction) DNA seq. $3-(5/reaction)</td>
</tr>
<tr>
<td>Equipment</td>
<td>—</td>
<td>Dissecting tools Balance Vortexer Microcentrifuge</td>
<td>PCR machine Microwave Gel box NanoDrop</td>
<td>Computer</td>
</tr>
</tbody>
</table>

---

**Figure 1.** Image of an agarose gel showing the COI PCR products of six fish samples. (Each lane was loaded with 5μl of PCR product). The 100 bp DNA ladder in the first lane was used as a size standard reference.

**Step 5: PCR Sample Preparation and DNA Sequencing**

Before the PCR product can be sequenced, it must be purified to eliminate residual primers, unincorporated nucleotides, enzymes, and other reagents in the PCR reaction that might inhibit DNA sequencing. Many sequencing companies will purify the DNA.
sample as part of their preparation, but this must be noted in the request. (The added benefit is that the sequencing company will also optimize the DNA concentration for the reaction). Alternatively, the PCR sample can be purified with a kit such the MinElute PCR Purification Kit (QIAGEN Inc., product # 28004). This will also require access to a spectrophotometer to optimize the DNA concentration requested by the sequencing facility (Supplementary Protocols #4 and #5).

Our DNA samples were sent to Elim Biopharmaceuticals (https://www.elimbio.com/) for sequencing, but we suggest that the instructor contact DNA sequencing centers in their area to compare prices. The DNA sequencing service will outline how they wish the primers to be sent with the samples. Our success rate with obtaining high-quality DNA sequencing reads with either the forward (FishF1) or reverse (FishR1) primer was over 90 percent, and thus the instructor could opt for just one sequencing reaction per fish sample if time and/or budget is a factor. However, we recommend two DNA sequencing reactions for each fish PCR product, one reaction using FishF1 and the second reaction using FishR1. Having DNA sequence data for both the forward and reverse reactions has a number of benefits: (1) a back-up in case one reaction does not work; (2) emphasizing the reproducibility of sequencing both strands of DNA to determine the nucleotide sequence; and (3) the experience of comparing the forward and reverse sequences to identify possible errors in sequencing (especially when examining peaks on the chromatogram that might be ambiguous). There is often a 2–3 day delay in receiving the sequence information, and we used this time to explain the di-deoxy sequencing reaction and how to interpret DNA sequence information.

**Step 6: DNA Sequence Analysis**

Once the DNA sequence information is received (usually by downloading from the sequencing company’s website), there are several steps to the analysis process (see Supplemental Protocol #6). In brief, students should compare the forward and reverse sequences to ensure accuracy of sequencing. Prior to this stage, the concept of ‘5′ → 3′ directionality of the sequencing reaction, and the use of ‘reverse complement’ should be discussed (there are many free online programs that will convert a DNA sequence into its reverse complement). Using the DNA analysis program MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle) from EBI, we aligned the forward reaction DNA sequence and the reverse complement DNA sequence (it will take some patience in helping students keep this straight!). To resolve any discrepancies between the two sequences, we had students examine the histogram files in the 4Peaks application (a free online application). Although sequences are often not in 100 percent agreement due to sequencing errors, students should nevertheless have a decent sequence for the next step. (In some cases only the forward or reverse reaction produced a useful sequence, but we were able to proceed to the next stages.)

The bioinformatics tool Basic Local Alignment Search Tool (BLAST) from NCBI was used to determine those sequences in the public database that most closely matched the unknown fish COI DNA sequence (detailed instructions are in Supplemental Protocol #6). BLAST gives an output of many similar sequences and displays metrics that describe the quality of the alignment of each sequence with the query sequence (Figure 2A). The “Query cover” column indicates the percentage of the original sequence covered in

<table>
<thead>
<tr>
<th>Description</th>
<th>Max Score</th>
<th>Total Score</th>
<th>Query Cover</th>
<th>E Value</th>
<th>Identi</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmo salar cytochrome oxidase subunit I (COI) gene, partial cds: mitochondrial</td>
<td>1188</td>
<td>1188</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>KF597049.1</td>
</tr>
<tr>
<td>Salmo salar mitochondrial_complete genome</td>
<td>1182</td>
<td>1182</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
<td>KF792790.1</td>
</tr>
<tr>
<td>Salmo salar isolate #76 cytochrome oxidase subunit I (COI) gene, partial cds: mitochondrial</td>
<td>1182</td>
<td>1182</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
<td>JX960941.1</td>
</tr>
<tr>
<td>Salmo salar voucher US&lt;ZAF&gt;:09022 cytochrome oxidase subunit I (COI) gene, partial cds: mitochondrial</td>
<td>1182</td>
<td>1182</td>
<td>99%</td>
<td>0.0</td>
<td>99%</td>
<td>HM007799.1</td>
</tr>
</tbody>
</table>

**Figure 2.** (A) Screen shot of NCBI nucleotide BLAST hits. A 646 bp sequence obtained from an unknown fish sample was submitted to the online BLAST search tool. The descriptions and scores for the top four hits are shown, all of which indicate that Salmo salar is likely the species identity of the unknown fish sample. (B) and (C) Alignments comparing student-generated DNA sequence from a fish of an unknown species with sequences of known species. A 60 bp section of the cytochrome c oxidase gene is shown. Known sequences were obtained from the online NCBI database. (B) Sequence alignment of the unknown species (top) and Atlantic salmon (bottom) shows complete agreement over many base pairs, indicating a species match. (C) Sequence alignment of the unknown species (top) with Pacific salmon (bottom) shows mismatched nucleotides, indicating that the unknown sample is unlikely to be Pacific salmon.
the alignment, and the “Ident” column indicates the percent nucleotide identity (matching sequence) within that covered sequence. High values in these columns indicate a close alignment. The “E value” column represents the expected number of hits by chance; therefore low E values suggest the alignment is not a result of background noise. Students can use these metrics to determine the known fish species with the most similar sequence, which is assumed to be the species identity of the unknown sample. Students can align their sequence with the sequences from the BLAST search to see the specific base pair similarities and differences. (At this point we revealed to the students the species of fish that we thought we had purchased.) Figure 2 displays the sequencing results from an unknown fish sample, labeled as Pacific salmon from the vendor, aligned with Atlantic (B) or Pacific (C) salmon sequences obtained from the database. The alignment reveals that the genetic identity of the unknown fish sample is more similar to Atlantic than Pacific salmon, suggesting that the fish was mislabeled.

We finished the project by having the students work in groups to create a poster that summarized the project and presented the results. (An example student poster is included at our website.) This exercise proved to be an excellent addition to the project as it gave the instructor and teaching assistants an opportunity to work with the students on scientific writing methods, as well as reinforcing the main concepts and techniques of the project.

○ **Summary and Extensions**

We found that this inquiry-based project was an engaging and thought-provoking means of exposing our students to common molecular biology techniques. Aside from illustrating many key concepts, the project also serves to demonstrate the connection between fields of biology that students might initially assume to have little overlap. The project is very flexible, and allows the instructor the opportunity to add additional material as they see fit. Some features that an instructor may wish to consider including in the project follow.

**Sustainable Fish Harvesting**

One of the interesting tangents of this project was that it sparked a conversation about sustainable harvesting of our ocean resources. As students were investigating fish fraud on the internet, they stumbled upon plenty of sites dedicated to informing consumers about best choices for well-managed fish stocks (e.g., http://www.seafoodwatch.org, http://www.nrdc.org/oceans/seafoodguide). Some suppliers and restaurants have online resources that describe their fisheries’ practices, and our students enjoyed investigating the sources of food they eat. Discussing fisheries management practices served to help the students see the connections inherent in the study of biology.

**Advanced Exercise in Primer Design**

Though this project uses established primers to barcode fish, researchers often need to design PCR primers for their particular application. To give students experience in designing primers, and to become more comfortable working with DNA sequence data, we developed an additional activity based on evolutionary conservation of DNA sequence as the basis of primer design (Supplemental Protocol #1.3). In this activity, students align the COI DNA sequence from a number of known fish species, and then choose optimal forward and reverse primer sequences from those regions of consensus. This activity can help students gain a deeper understanding of PCR, bioinformatics, the importance of evolutionarily conserved regions, and even gene expression (for example, the degeneracy of the genetic code becomes very evident as students see the nonalignment of nucleotides among fish species almost always at the third position in the codon).

**Phylogenetic Analysis of Fish COI Sequences**

Besides using their sequencing results to determine whether the original fish sample was mislabeled, students can use the set of sequences generated by the class to develop a phylogenetic tree of commercial fish species. This activity helps students to make connections between molecular and evolutionary biology, and allows them to visualize the evolutionary information contained in their sequence data. We used the Mega6 software program for both alignment and phylogenetic analysis (Supplemental Protocol 6.1). Figure 3 shows an example phylogenetic tree inferred using eight student sequences. To obtain a discernable phylogenetic tree, it may be necessary for students to include additional COI sequences from the NCBI database in their sequence alignment and phylogenetic analysis. These additional DNA sequence files will allow students to better distinguish between closely related species, and can serve as an outgroup sequence to differentiate clades of the tree. In our example tree in Figure 3, the Litopenaeus vannamei sequence from whiteleg shrimp, from the phylum Artropoda, serves as an effective outgroup for the other chordate fish in the tree.

○ **Acknowledgments**

We would like to thank Jeremy Updike and the Carleton Summer Science Institute classes of 2013 and 2015 for their help in shaping this project (we had a “BLAST!”). Additional thanks to Kate Crofton and Gail Waltz for their help in the 2015 adaptation of this project.
References


MADELINE ARNOLD (madeline.l.arnold@gmail.com) and DREW HOLMEN (dholmen@gmail.com) are students, and STEPHAN ZWEIFEL (szweifel@carleton.edu) is a biology professor at Carleton College, One North College St., Northfield, MN 55057.

TK

NABT Awards