2015

Food Fish Identification from DNA Extraction Through Sequence Analysis

Heather E. Hallen-Adams
University of Nebraska–Lincoln, hhallen-adams2@unl.edu

Published in *Journal of Food Science Education* 14 (2015), pp. 116–120; doi: 10.1111/1541-4329.12063

Submitted October 22, 2014; accepted April 17,2015.

Innovative Laboratory Exercises

Food Fish Identification from DNA Extraction Through Sequence Analysis

Heather E. Hallen-Adams

Department of Food Science and Technology, University of Nebraska–Lincoln, Lincoln, NE 68583-0919, USA; email hhallen-adams2@unl.edu

Abstract

This experiment exposed 3rd and 4th y undergraduates and graduate students taking a course in advanced food analysis to DNA extraction, polymerase chain reaction (PCR), and DNA sequence analysis. Students provided their own fish sample, purchased from local grocery stores, and the class as a whole extracted DNA, which was then subjected to PCR, gel electrophoresis, and Sanger sequencing. Students retrieved their sequences and identified their fish using the NCBI BLAST nucleotide database. Slightly more than half of the samples yielded sequences identical or close to expected (based on the identification of the fish on the packaging); some other samples matched unanticipated fish or other organisms, due to an incomplete database, minor sequencing errors, or laboratory contamination (human and fungal sequences); 1 canned tuna sample identified as hake could represent food fraud.

Introduction

Food fraud occurs when food is somehow misrepresented to customers and consumers, and encompasses a variety of activities, including adulteration. In economic adulteration, superior or more expensive ingredients are fraudulently replaced by less expensive ones without the consumers' awareness (Everstine and others 2013). Foods commonly subjected to economic adulteration include honey (which may be replaced, in part or whole, by corn syrup or other sweeteners; Everstine and others 2013), oils (in which premium oils, such as extra virgin olive oil, may be blended with or replaced by lower grade oils; Everstine and others 2013), and seafood. Processed fish products, such as fish sticks, nuggets, and patties, are particularly susceptible to economic adulteration, as the end product has been minced, reshaped, and frequently breaded, obscuring any morphological features of the original fish (Pepe and others 2005; Everstine and others 2013).

Polymerase chain reaction (PCR) provides the substrate, in the form of billions of copies of a gene or DNA region of interest, for numerous biochemistry and molecular biology techniques. Downstream techniques include gel electrophoresis, various fingerprinting techniques, and gene sequencing. Many PCR primers are available to amplify conserved regions of DNA. The 28S region of the nuclear ribosomal RNA gene possesses both sufficiently conserved regions to allow a single primer pair to amplify DNA from a wide range of species and sufficient diversity to enable species-level identification (Bruns and others 1991; Sonnenberg and others 2007). Similarly, the mitochondrial gene encoding cytochrome *b* is both sufficiently conserved and sufficiently diverse to allow amplification and identification from across the vertebrates (Zardoya and Meyer 1996; Baharum and Nurdalia 2012). Both have been extensively used in food fish (Chen and others 2003; Pepe and others 2005).

DNA- and PCR-based techniques have found wide application in forensic contexts, and experiments using such methodologies are becoming increasingly popular in the classroom as they provide an alternative to more cut-and-dried experiments with a single outcome or "right" answer and engage students in the chance to solve a mystery. PCR-based forensic investigations designed for use in the classroom include fingerprinting methods based on tandem repeats (Carson and others 2009) or restriction fragment length polymorphisms (Millard and others 2013); the use of geneor species-specific primers (Taylor and Sajan 2005, Childs-Disney and others 2010); and DNA sequencing followed by database comparisons (Chao and others 2012). Despite the ready availability of samples and the scope of possible experiments, the use of DNA for product identification has been underutilized in food science teaching—but see Taylor and Sajan (2005) for a food-based example. The experiment described herein uses DNA extraction, PCR, and sequence analysis to identify food fish samples to species. Students are engaged by the opportunity to solve a mystery and the possibility of detecting fraud; asking students to provide their own samples provides a sense of ownership in the experiment.

Background to the Experiment

Upper-level undergraduates and graduate students in Food Science at the University of Nebraska-Lincoln (UNL) take Advanced Food Analysis, which introduces them to a variety of chromatographic, spectrographic, and biochemical techniques. The course consists of 2 lectures and 1 lab per week. In 2014, a lab was introduced in which a variety of DNA-centered tech-

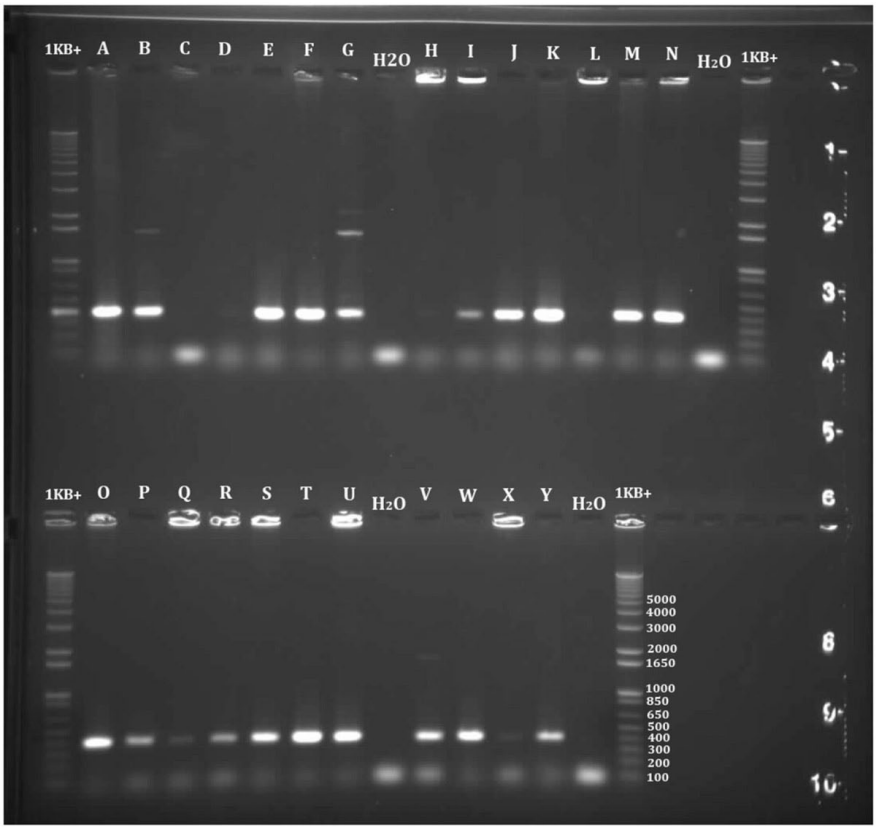

Figure 1. PCR products from fish samples amplified with C1′ and C2 primers (amplify a portion of the 28S rRNA gene) on a 0.7% agarose gel. DNA was used at 1/10 dilution in the PCR reaction. Negative control lanes are labeled "H_2O"; the size standard is the 1KB+ ladder from Life Technologies, with band sizes given in base pairs. A product of approximately 400 bp amplified from most samples. Sample C (pollock fillet) did not amplify under any conditions, while other samples with little or no product (faint band or no band) amplified when undilute DNA was used in the reaction (not shown).

niques were demonstrated and utilized to identify fish products.

The generation of DNA sequence data is arguably less important than the ability to interpret such data; consequently, this exercise focused largely on understanding the sequencing results. This laboratory was carried out by a comparably large class facing significant time restraints, so some portions were run as demonstrations, or conducted behind the scenes by instructors. It would lend itself well to a format in which multiple lab sessions could be devoted to carrying out the experiments, to afford the students more hands-on experience. The protocol given the students is available in the Appendix, and provides detailed information on all steps, including those conducted "behind the scenes."

Materials and Methods

Each student was requested to provide a fish sample of _0.25 g. The classwas divided roughly in thirds, with different students providing frozen, processed fish products (fish sticks, nuggets, or patties), fresh or frozen fillets, or canned products. DNA was extracted using a standard hexadecyltrimethyl ammonium bromide (CTAB)–phenol–chloroform method (Hallen and others 2003). Briefly, instructors incubated the samples in 700 μL of a 2% CTAB solution at 65 °C for 2 h prior to class. During class, the students extracted the DNA by a succession of phenol and chloroform: isoamyl alcohol (IAA; 24:1) additions, followed by vortexing, centrifugation, and collection of the aqueous upper layer. Following the final chloroform:IAA step, ice cold ethanol was added to the aqueous layer to precipitate the DNA, which was briefly dried, then resuspended in 50 μL molecular biology grade water.

This class consisted of 26 students, with slightly under 3 h allocated for lab time. To prevent the bottlenecks that would ensue if each student prepared their own sample start to finish, ensure more consistent extractions, and allow everyone to participate, each student performed a single task (adding reagents, transferring samples to new tubes) in the DNA extraction.

The instructors demonstrated making a 0.7% agarose gel, and discussed the gel electrophoresis and PCR steps; these steps were performed after class. DNA was diluted to 1/10, and 1 μL of this dilution was used in a 25 μL PCR reaction including 0.25 U Phusion High Fidelity Master Mix (Thermo Scientific, Waltham, Mass., U.S.A.), 10 μM each forward and reverse primers, and molecular biology grade water. Each DNA sample was subjected to 2 PCR reactions. Primers C1′ (5′-ACCCGCTGAATTTAAGCAT-3′) and C2 (5′-TGAACTCTCTCTTCAAAGTTCTTTTC-3′; Lê and others 1993; Chen and others 2003) amplify a ~320 bp region of the 28S nuclear ribosomal RNA gene, whereas CYTB1 (5′-CCATCCAACTCTCAGCATGATGAAA-3′) and CYTB2 (5′-GCCCCTCAGAATGATATTTGTCCTCA-3′; Carr and Marshall 1991; Pepe and others 2005) amplify a ~300 bp region of the mitochondrial cytochrome *b* gene. PCR began with a 94 °C melting step, followed by 30 cycles of 94 °C 30 s, 55 °C 30 s, 72 °C 1 min, and a final 10 min elongation at 72 °C (Chen and others 2003; Pepe and others 2005) on an Eppendorf Mastercycler (Hauppauge, N.Y., U.S.A.). Ten microliters of each PCR product was run on a 0.7% agarose gel containing 0.5 μg/mL ethidium bromide for 90 min at 70 V and visualized with UV light. Images of the gels were sent to the students (Figure 1).

PCR products were sent to the Michigan State University Research Technology Support Facility for Sanger sequencing using the appropriate forward primer (C1′ or CYTB1). Results were made available to the students online. During lecture, students were shown the gels and interpretation was discussed;

Table 1. Student samples and top BLAST hits to sequencing results.

Sample, as identified by student	28S sequence top BLAST result[a]	Cytochrome *b* sequence top BLAST result[b]
Frozen, processed fish		
Fish sticks (species unspecified)	*Gadus morhua* (Atlantic cod), 98%	
Fish sticks (species unspecified)	Bad sequencec *Homo sapiens,* 99%[d]	
Fish sticks (species unspecified)	*Gadus morhua* (Atlantic cod), 100%	
Fish sticks (species unspecified)	Bad sequence	*Merluccius productus* (North Pacific hake), 99%
Pollock sticks (*Pollachius* species)	Bad sequence	
Pollock sticks (*Pollachius* species)	Bad sequence	
Pollock patty (*Pollachius* species)	*Gadus morhua (Atlantic cod), 90%*	
Pollock patty (*Pollachius* species)	Bad sequence	
White fish patty	*Gadus morhua* (Atlantic cod), 94%	*Merluccius productus* (North Pacific hake), 99%
Fillets		
Cod (*Gadus* species)	*Gadus morhua* (Atlantic cod), *99%*	
Cod (*Gadus* species)	*Melanogrammus aeglefinus* (Haddock), *97%*[f]	
Fresh cod (*Gadus* species)	*Gadus morhua* (Atlantic cod), 99%	*Sus scrofa* (Pig), 91%[g]
Mahi mahi (*Coryphaena hippurus*)	*Trigia lucerna* (Tub gunnard), 98%[e]	*Coryphaena hippurus* (Mahi mahi), 100%
Mahi mahi (*Coryphaena hippurus*)	*Antennarius striatus* (Striated frogfish), 99%[e]	*Coryphaena hippurus* (Mahi mahi), 99%
Pollock (*Pollachius* species)	No PCR product	
Salmon (*Oncorhynchus* and *Salmo* species)	*Oncorhynchus mykiss* (Rainbow trout), 98%	*Oncorhynchus gorbuscha* (Pink salmon), 100%
Salmon (*Oncorhynchus* and *Salmo* species)	*Oncorhynchus mykiss* (Rainbow trout), 98%	
Tilapia (3 genera, including *Oreochromis*)	*Oreochromis aureus* (Blue tilapia), 99%	
Tilapia (3 genera, including *Oreochromis*)	*Oreochromis aureus* (Blue tilapia), 99%	
Canned fish		
Canned tuna (*Thunnus* species)	Bad sequence	*Homo sapiens,* 99%[d]
Canned tuna (*Thunnus* species)	Bad sequence	
Canned tuna (*Thunnus* species)	Bad sequence	
Canned tuna (*Thunnus* species)	*Aspergillus flavus* (mold), 98%[h]	
Canned tuna (*Thunnus* species)	Bad sequence	*Merluccius productus* (North Pacific hake), 94%[i]

a. The top BLAST hit for each sample is given. In some cases, the top hit may not provide the most accurate identification, as discussed below.
b. Not all samples amplified using the cytochrome *b* primers CYTB1–CYTB2.
c. Bad sequences were those for which a PCR product was obtained and sequenced, but DNA sequence fell below the quality cutoff scores implemented in the Finch Server.
d. Sequences with cytochrome *b* top matches to human are most likely explained by contamination. If BLAST is constrained to search only fish sequences, the top fish hits contain only one more mismatch than the hits to *Homo sapiens*; however, food fish are not among the top piscine hits.
e. There was only one Mahi mahi 28S sequence in the NCBI databases as of May 8, 2014, and this sequence did not provide significant overlap with the sequences produced by the C1'-C2 primers used in our study.
f. Haddock was the top BLAST hit for this sequence; however, a cod sequence actually had higher sequence identity (99%), but less overlap, which resulted in BLAST assigning a higher score to the haddock.
g. When BLAST was constrained to search only fish sequences, the top hit was *Parablennius sanguinolentus* (rusty blenny; not a food fish) at 85% sequence identity; if constrained to cod, the sequence identity was 72%.
h. DNA was extracted and PCR performed in laboratories that conduct research on food molds; *Aspergillus flavus* is likely a contaminant from the laboratory environment.
i. Sequence identity was 83% when BLAST was constrained to search only tuna (*Thunnus*) sequences.

then students were shown how to access their sequences, evaluate the quality, and obtain an identification for their fish using NCBI BLAST (http://blast.ncbi.nlm.nih.gov ; Altschul and others 1990). Students used a nucleotide query to search the nucleotide databases (nucleotide BLAST), using megablast (optimized for highly similar sequences). In total, 114 samples were submitted for sequencing; this included resequencing some products with different quantities of DNA and rerunning some PCR reaction under different annealing conditions.

Students were asked to include the following in their lab reports: the brand, type, and stated species of their fish product; PCR gel image(s) showing their samples; the sizes of their PCR products; their DNA sequence; the top 5 BLAST hits for their DNA sequences. They were asked to discuss whether 28S and cytochrome *b* sequences both matched the same species (when applicable; not everyone had both sequences), and whether BLAST results matched the species their fish was sold as. Finally, students were asked to evaluate whether the DNA results unequivocally identified their fish and, if not, what was uncertain about the identification and what were some possible reasons for the uncertainty?

Hazards

Phenol and chloroform are toxic, and pose inhalation and contact risks. Students wore gloves for all steps involving student participation, and all steps involving phenol and chloroform were performed in a fume hood.

Results and Discussion

In 2014, 26 students provided 24 samples: 9 frozen, processed fish samples (3 patties and 6 fish sticks); 10 fillets; and 5 canned samples. All samples yielded DNA; 1 sample (a pollock fillet) did not amplify with either PCR primer pair under the conditions tested, despite attempts to amplify 3 different dilutions from 2 DNA extractions. All other samples yielded PCR products with the C1'–C2 primer pair, and 10 samples produced PCR products with the CYTB1–CYTB2 primer pair.

Fillets yielded the most satisfactory results, with all but one producing high quality DNA sequences matching food fish in the BLAST results (Table 1). 5 out of 9 processed fish samples generated high-quality DNA sequences matching food fish. 4 out of 9 of the processed samples (fish sticks and patties) produced a bright band on the gel of the PCR products, but yielded poor-quality sequence (Figure 2). The values given for signal strengths ("A = 468, C = 433 …") were consistent with the values obtained in quality DNA sequences, indicating that neither too little nor too much DNA was provided for sequencing. The problem became clearer when the graphical trace was examined; in many places, peaks of multiple colors (representing different DNA bases) occupied the same space (for example, around positions 120 to 140 [peaks shown below the corresponding bases and numbers]). This suggested that multiple distinct DNA sequences were amplified in the same PCR reaction—entirely possible in a fish stick, where different species of fish may be commingled. Intriguingly, when the small portions

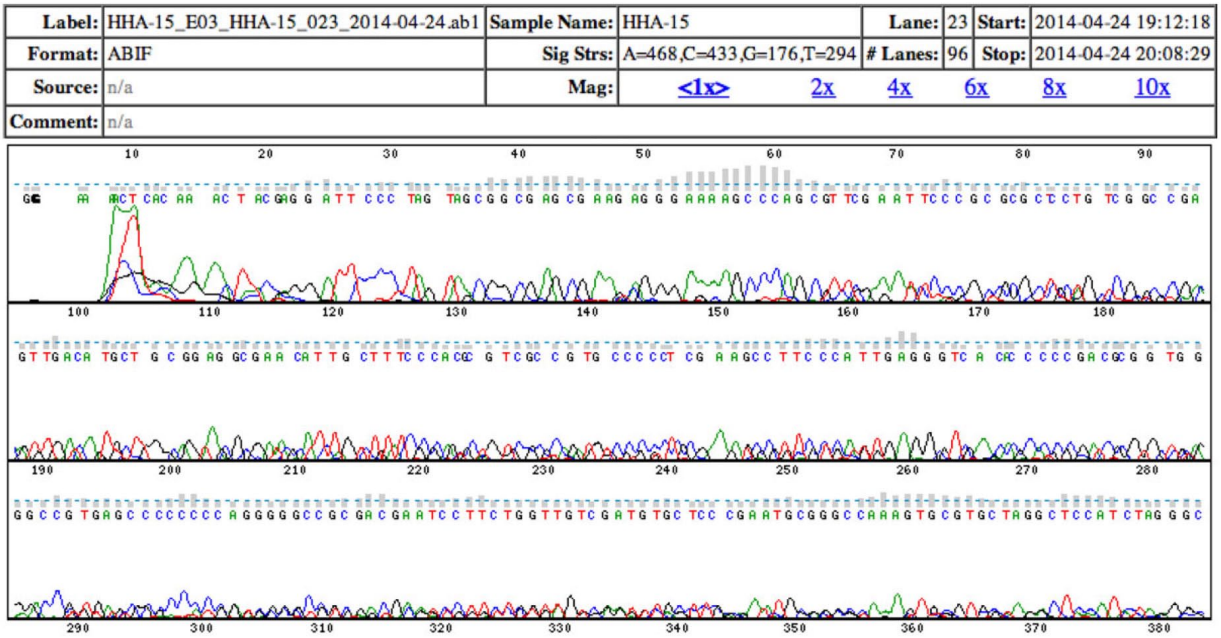

Figure 2. Chromatographic trace of DNA of a low quality DNA sequence (shown in Figure 1, lane V). Signal strengths ("Sig Strs," above) are consistent with high quality DNA sequence on the ABI 3730xl DNA sequencer; however, in many cases multiple nucleotides were detected at the same position (for example, around 116, where the printed sequence reads "GAACATT", but there are red, blue, and black peaks (representing T, C, and G) intermingled amongst the "AAC"). This pattern suggests the presence of multiple distinct DNA templates in the same sequencing reaction.

of quality sequence were examined from such samples, some of them yielded BLAST hits to plants, indicating that the breading may be amplifying! If one wants unequivocal, easy to interpret results, it would be advisable to seek an explicitly fish-specific primer pair (although this would not get around the problem of potentially commingled species). I find the multiple species possibility leads to valuable classroom discussion (and a more realistic lab), and will continue use of these primers. Canned fish yielded poor results, with only 1 sample generating high-quality DNA sequence matching a fish. One sample produced high quality fungal DNA sequence whereas another produced high-quality human sequence, both most likely due to contamination during handling and DNA extraction.

As a result of the size of the class (26 students) and time constraints (2 h 50 min), the DNA extraction, PCR, and agarose gel portions of this lab were more of the nature of demonstrations than of hands-on experiments. This could easily be adjusted in a smaller class, or one that could devote multiple class sessions to the lab. For this class, the more valuable experience was in data collection and interpretation. The BLAST results raised a lot of questions and valuable discussion about the limitations of the technology as well as its possibilities. Guided discussion was useful in helping students understand unexpected results. The likelihood of contamination was not immediately apparent to students; after considering all of the human intervention during the DNA extraction and PCR process, it became clearer that the odds of skin cells or other human contamination occurring in the lab probably outweighed the odds of a terrible accident in the canning factory as an explanation for matches to human DNA. Unexpected results also allowed a discussion of the limitations of the technology employed. A BLAST search is only as good as its database; putative Mahi mahi may come back as a striated frogfish if the database lacks Mahi mahi sequence for the gene in question. In a highly conserved gene, like cytochrome *b*, 1 or 2 nucleotides can be the difference between hake, cod, and haddock; could sequencing errors have played a role? Graduate students and even primary investigators whose main area of expertise is not nucleotide analysis may accept BLAST results uncritically, and this laboratory exercise hopefully serves to counteract that tendency, and encourage critical evaluation of results.

PCR primers were chosen for this lab due to their demonstrated use in fish phylogenetics and concomitant ability to both amplify DNA from amajority of fish species and yield species-level differentiation. Sequencing 2 genes allowed discussion on the congruence (or lack thereof) of sequencing results, and increased the chances of obtaining usable sequence. Unfortunately, fewer than half of our samples amplified with the CYTB1–CYTB2 primer pair in time to be used by the class; however, use of a 60 to 45 °C touchdown protocol amplified products from the majority of samples, andwill be incorporated in future labs. As discussed above, the low-quality sequences obtained with the C1′–C2 primer pair are largely explicable by comingling of DNA from different sources, and selection of a different primer pair may not yield a significant improvement; under optimal conditions, Sonnenberg and others (2007) report favorably on the ability of 28S sequence to identify fish to species. Canned fish samples yielded little usable sequence, and will not be used in future labs.

Conclusions

This lab, with numerous uncertainties involved and "problems" such as commingled DNA sequences and unexpected BLAST results, exposed students to a more realistic picture of science as it is practiced than they frequently obtain in prepared laboratories. Students enjoyed this lab, and took ownership of their fish samples, with many including pictures and brief discussions of the fish identified by BLAST in their lab reports.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 5:403–10.

Baharum SN, Nurdalia AA, 2012. Application of 16s rDNA and cytochrome b ribosomal markers in studies of lineage and fish populations structure of aquatic species. Mol Biol Rep 39:5225–32.

Bruns TD, White TJ, Taylor JW. 1991. Fungal molecular systematics. Ann Rev Ecol Syst 22:525–64.

Carr SM, Marshall HD. 1991. A direct approach to the measurement of genetic variation in fish population: application of the polymerase chain reaction to studies of Atlantic cod *Gadus morhua* L. J Fish Biol 39(Suppl A):101–7.

Carson TM, Bradley SQ, Fekete BL, Millard JT, LaRiviere FJ. 2009. Forensic analysis of canine DNA samples in the undergraduate biochemistry laboratory. J Chem Educ 86:376–8.

Chao T-C, Bhattacharya S, Ros A. 2012. Microfluidic gel electrophoresis in the undergraduate laboratory applied to food analysis. J Chem Educ 89:125–9.

Chen W-J, Bonillo C, Lecointre G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol Phylogenet Evol 26:262–88.

Childs-Disney JL, Kauffmann AD, Poplawski SG, Lysiak DR, Stewart RJ, Arcadi JK, Dinan FJ. 2010. A metabolic murder mystery: a case-based experiment for the undergraduate biochemistry laboratory. J Chem Educ 87:1110–2.

Everstine K, Spink J, Kennedy S. 2013. Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. J Food Prot 76:723–35.

Hallen HE, Watling R, Adams GC. 2003. Taxonomy and toxicity of *Conocybe lactea* and related species. Mycol Res 107:969–79.

Lê HLV, Lecointre G, Perasso R. 1993. A 28S rRNA based phylogeny of the Gnathostomes: first steps in the analysis of conflict and congruence with morphologically based cladograms. Mol Phylogenet Evol 2:31–51.

Millard JT, Chuang E, Lucas JS, Nagy EE, Davis GT. 2013. Case-study investigation of equine maternity via PCR-RFLP: a biochemistry laboratory experiment. J Chem Educ 90:1518–21.

Pepe T, Trotta M, di Marco I, Cennamo P, Anastasio A, Cortesi ML. 2005. Mitochondrial cytochrome *b* DNA sequence variations: an approach to fish species identification in processed fish products. J Food Prot 68:421–5.

Sonnenberg R, Nolte AW, Tautz D. 2007. An evaluation of LSU rDNA -D1-D2 sequences for their use in species identification. Front Zool 4:6.

Taylor A, Sajan S. 2005. Testing for genetically modified foods using PCR. J Chem Educ 82:597–8.

Zardoya R, Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol 13:933–42.

Supporting Information

Supplemental information (Lab protocol given to students: Food fish identification from DNA extraction through sequence analysis [Hallen-Adams]) is available to academic partners upon request. Address requests to the author.