8-16-2017

Evaluation of DNA Barcoding Methodologies for the Identification of Fish Species in Cooked Products

Sophia J. Pollack
Chapman University

Michael D. Kawalek
U.S. Food and Drug Administration

Donna M. Williams-Hill
U.S. Food and Drug Administration

Rosalee S. Hellberg
Chapman University, hellberg@chapman.edu

Recommended Citation

Pollack S.J., Kawalek M.D., Williams-Hill D.M. & Hellberg R.S. (2017). Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products. *Food Control, 84*, 297-304. doi: 10.1016/j.foodcont.2017.08.013

Evaluation of DNA Barcoding Methodologies for the Identification of Fish Species in Cooked Products

Accepted Manuscript

Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products

Sophia J. Pollack, Michael D. Kawalek, Donna M. Williams-Hill, Rosalee S. Hellberg

PII: S0956-7135(17)30409-7

DOI: 10.1016/j.foodcont.2017.08.013

Reference: JFCO 5747

To appear in: *Food Control*

Received Date: 14 March 2017

Revised Date: 7 July 2017

Accepted Date: 13 August 2017

Please cite this article as: Pollack S.J., Kawalek M.D., Williams-Hill D.M. & Hellberg R.S., Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products, *Food Control* (2017), doi: 10.1016/j.foodcont.2017.08.013.

Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products

Sophia J. Pollack[a], Michael D. Kawalek[b], Donna M. Williams-Hill[b], Rosalee S. Hellberg[a*]

[a]Chapman University, Schmid College of Science and Technology, Food Science and Nutrition, One University Drive, Orange, CA 92866

[b]U.S. Food and Drug Administration, Office of Regulatory Affairs, Pacific Regional Laboratory Southwest, 19701 Fairchild, Irvine, CA 92612

***Corresponding Author**:

Rosalee S. Hellberg, Ph.D.

Chapman University

Ph: 714-628-2811

Fax: 714-289-2041

E-mail: hellberg@chapman.edu

Disclaimer: The views in this publication represent those of the authors. The inclusion of specific trade names or technologies does not imply endorsement by the U.S. Food and Drug Administration nor is criticism implied of similar commercial technologies not mentioned within.

Abstract

DNA barcoding is a powerful sequencing-based tool for the detection of fish species substitution. However, various cooking methods have the potential to reduce the quality and success of DNA sequencing. The objective of this study was to determine the effects of common cooking methods on DNA sequencing results with both full-length (655 bp) and mini-barcodes (208-226 bp), and to determine the optimal methodology to use for species identification of various fish products. Six types of fish (salmon, tuna, scad, pollock, swai and tilapia) were prepared in triplicate using the following methods: uncooked, baked, fried, broiled, acid-cooked, smoked and canned. DNA was extracted from each sample and tested using full and mini-barcoding of the cytochrome *c* oxidase subunit I (COI) gene. The resulting sequences were compared based on quality parameters, success rates, and genetic identifications. SH-E mini-barcoding showed the highest overall success rates (92-94%), followed by full barcoding (90%), and SH-D mini-barcoding (67-90%). Across the individual cooking methods, SH-E mini-barcodes performed as well or better than full barcodes for most samples. The sequencing results were fairly consistent across cooking methods with the exception of canning, which showed marked decreases in sequencing success, quality, and length. Despite the reduced sequence length of mini-barcodes compared to full barcodes, identification of fish species was largely consistent across the methods. Overall, the results of this study show that DNA barcoding is a robust tool for fish species identification, and that mini-barcoding has high potential for use as a complement to full barcoding.

Keywords: DNA barcoding; fish; species identification; mislabelling; mini-barcodes; species substitution

1. Introduction

Fish is an important staple of the world's food supply, accounting for ~17% of the global population's intake of animal protein in 2013 (FAO, 2016). Globally, aquaculture and fisheries production has been increasing at an average annual rate of 3.2% over the past five decades, with a combined production of 167.2 million tonnes in 2014. The United States is the top importer of fish and fishery products, totalling $20.3 billion in 2014 (FAO, 2016). Fish and seafood prices are volatile because they are susceptible to a variety of constantly changing factors, such as product quality and supply and demand. These price differentials, combined with factors such as increased consumption of processed fish, as well as increases in international trade, have increased the vulnerability of fish to fraudulent market practices (Hellberg & Morrissey, 2011). One type of economic fraud affecting the seafood industry is the occurrence of species substitutions (NOC, 2016). This practice is largely motivated by the economic benefits of substituting inexpensive species for advertised and labelled premium species. There have been numerous reports of mislabelling of fish species in the United States, including Atlantic salmon (*Salmo salar*) mislabelled as Pacific salmon (*Oncorhynchus* sp.) (Cline, 2012), striped bonito (*Sarda orientalis*) mislabelled as tongol tuna (*Thunnus tonggol*) (Mitchell & Hellberg, 2016), and Indian scad (*Decapterus russelli*) mislabelled as mackerel (unspecified) (Shokralla, Hellberg, Handy, King, & Hajibabaei, 2015).

Besides economic deception, fish species substitution is problematic from the standpoint of food allergies and other health risks. Allergies to specific varieties of seafood, including fish, crab and other shellfish can be life-threatening (Sicherer, Munoz-Furlong, & Sampson, 2004) and put consumers of adulterated fish and seafood products at increased risk. Proper labelling of fish species is also important so that at-risk consumers, such as pregnant women and young

children, can avoid fish that contain concerning levels of mercury, a potent neurotoxin (EPA/FDA, 2014). Another health concern associated with mislabelling is the exposure to tetrodotoxin, a neurotoxin found in certain species of puffer fish. In one instance an individual purchased what was labelled as "monk fish, gutted and head off, product of China," from an Asian market in Chicago, IL, and became ill soon after (Cohen et al., 2009). The FDA field office analyzed the purchased fish to discover that it was not monk fish, but puffer fish of the toxic variety. Furthermore, wax esters, which cause gastrointestinal discomfort, are found at high levels in escolar (*Lepidocybium flavobrunneum*), a common substitute for "white tuna" sushi products (Lowenstein, Amato, & Kolokotronis, 2009; Warner, Timme, Lowell, & Hirshfield, 2013).

Fish identification is often reliant on taxonomic features; however, these features are removed during processing, making it challenging to accurately identify fish to the species level. DNA barcoding is a common method used for species identification in these situations and has been adopted by the FDA for use in testing regulatory fish samples (Handy et al., 2011a). This method is a DNA sequencing-based technique in which a standardized genetic region is targeted across multiple species and queried against an existing library of reference sequences (Hebert, Cywinska, Ball, & DeWaard, 2003). The standard DNA barcode for identification of animal species is a ~650-bp region of the gene coding for cytochrome *c* oxidase subunit 1 (COI). DNA barcoding of this region has been successful in identifying myriad fish species around the world (Hubert et al., 2008; Kim et al., 2012; Landi et al., 2014; Steinke, Zemlak, Boutillier, & Hebert, 2009; Ward, Zemlak, Innes, Last, & Hebert, 2005; Yancy et al., 2008; Zhang & Hanner, 2012). Whilst DNA barcoding is known to be widely successful with uncooked fish, various cooking methods can potentially affect the quality and length of DNA sequences. Subjecting a sample to

high temperatures, pressure and other forms of processing is known to degrade DNA, making it more difficult to successfully identify a species (Hellberg & Morrissey, 2011). To aid in the identification of fish that have undergone processing, a mini-barcoding system has been developed (Shokralla et al., 2015). These mini-barcodes target 127–314 bp regions of the COI gene and have been shown to be more successful in species identification for certain fish products compared to the full-length barcode. Specifically, Shokralla et al. (2015) reported a sequencing success rate of 20.5% when using the full-length DNA barcode with heavily processed fish products, while individual mini-barcode primer sets achieved success rates of 27.3-88.6%.

Although fish mini-barcodes have been developed, they have not yet been extensively researched for use with regulatory samples. Furthermore, there is currently a lack of information regarding the most appropriate technique to use for fish samples that have been cooked in different ways. Therefore, the objective of this study was to determine the effects of common cooking methods on DNA sequencing results using both full-length and mini-barcodes, and to determine the optimal methodology to use for species identification of various fish products. The two mini-barcodes (SH-D and SH-E) that showed the greatest success rates in Shokralla et al. (2015) were selected for use in this study.

2. Materials and Methods

2.1. Sample collection

Six common types of fish were collected for testing in this study representing a cross-section of ocean and fresh water fishes with either oily or white flesh. These included: salmon, tuna, scad, pollock, swai, and tilapia. All samples were collected fresh/frozen either as whole fish or as fillets. Uncooked tissue was obtained from each species and tested in triplicate to serve as a

baseline sequencing control. Following collection, each fish sample was stored frozen at -20°C in a Whirl-pak bag (Nasco, Fort Atkinson, WI).

2.2. Cooking methods

Prior to cooking, fish samples were thawed overnight at 4°C and whole fish were filleted. Then, each fillet was cut into portions weighing approximately 100 g and the portions were prepared in triplicate using six common cooking methods: acid (ceviche), baking, broiling, canning, frying, and smoking. Whenever possible, portions cooked using the different methods were confined to a single fish. If portions had to be prepared from multiple individuals of a particular species, uncooked tissue samples of each individual first underwent full-length DNA barcoding as described below to ensure that all the individuals within a species had identical DNA sequences. Taking all replicates into account, a total of 18 fish samples were tested with each preparation method, for an overall total of 126 samples (including the uncooked controls).

For acid cooking, fish portions were submerged in 5% acetic acid and held for 4 h at 4°C in sealed plastic bags. Upon removal from the acid, the portions were rinsed one time with distilled water to stop the cooking process. For baking, the portions were placed on aluminium foil in a metal baking sheet and baked at 180°C for 30 min, or until the internal temperature reached 63°C (USDA, 2015b). For broiling, the fish portions were placed on aluminium foil in a metal baking dish and placed 10 cm directly below a gas broiler flame set on high for approximately 20 min, or until the internal temperature reached 63°C (USDA, 2015b). To pressure-can the fish, the portions were placed in 250-mL glass jars with screw-cap metal lids. Water was added to bring the total volume of material in the jars to approximately 10 mm from the lip of the jars. The jars were canned in a Presto brand pressure canner (Eau Claire, WI) operated at 118°C for 100 min (USDA, 2015a). Digital thermocouples placed inside the jar indicated that the fish were exposed

to approximately 17.5 F_o of heat (1.0 F_o = 1 min at 121°C). For deep frying, vegetable oil was heated in a saucepan to 180°C and then the portions were added until fully cooked, with an internal temperature of 63°C (USDA, 2015b). Finally, for smoking, each portion received an even coating of table salt (sodium chloride) and was held at 4°C for 4 h. Next, the portions were rinsed briefly with distilled water to remove the surface salt and smoked in a Masterbuilt Electric Smoker (Columbus, GA) at 93.3°C, with an internal temperature of at least 71.1°C for 30 min (Hilderbrand, 1999).

Once cooked, the prepared fish samples were stored inside wire-closed Whirl-pak collection bags at 4°C for two days prior to the start of analysis. This storage method simulated the collection of a consumer complaint sample that would be transferred to the laboratory and analyzed after arrival.

2.3. DNA extraction

DNA was extracted from fish samples using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), Spin-Column protocol following the modifications described in Handy et al. (2011b). Tissue samples (~10 mg) were mixed with 50 µL Buffer ATL and 5.56 µL Proteinase K and then incubated at 56˚C in a dry heat block. Each set of extractions included a reagent blank without sample tissue as a negative control. The samples were incubated for 3 h, with vortexing at 30 min intervals. Following incubation, equal parts (55.6 µL) Buffer AL and 95% ethanol were added to the sample tubes. Samples were vortexed immediately to yield a homogenous solution and then transferred by sterile pipette into DNeasy Mini spin columns placed inside 2 mL collection tubes. The samples were centrifuged (6,000 x g) for 1 min and the columns were placed inside new collection tubes. Next, 140 µL AW1 Buffer was added to each column and the centrifugation process was repeated. Columns were placed inside new collection tubes and

140 µL AW2 Buffer was added prior to centrifugation at 20,000 x g for 3 min. Following centrifugation, each column was placed inside a sterile 1.5 mL microcentrifuge tube and 50 µL of AE buffer preheated to 37˚C was pipetted gently over the column membrane. The samples were incubated for 1 min at room temperature (20-25˚C) and then centrifuged (6,000 x g) for 1 min. The eluted DNA was used in the polymerase chain reaction (PCR) and DNA sequencing, as described below.

2.4. PCR and DNA sequencing

All samples (n = 126) underwent PCR and DNA sequencing using both full barcoding (655 bp) and mini-barcoding (208-226 bp) of the COI gene. Full barcoding was carried out as described by Handy et al. (2011b) while mini-barcoding was carried out as described by Shokralla et al. (2015) using primer sets Mini_SH-D (208 bp) and Mini_SH-E (226 bp), with some modifications. For full barcoding, each reaction tube contained 6.25 µL 10% trehalose, 3.0 µL molecular grade H_2O, 1.25 µL 10X buffer, 0.625 µL $MgCl_2$ (50 nM), 0.062 µL dNTPs (10 mM), 0.06 µL Platinum Taq (5U/µL; Invitrogen, Carlsbad, CA), 0.125 µL of each 10 uM primer, and 1 µL of DNA template. Cycling conditions for the full barcode were carried out as in Handy et al. (2011b): 94°C for 2 min; 35 cycles of 94°C for 30 s, 55°C for 40 s, and 72°C for 1 min; and a final extension of 72°C for 10 min. For mini-barcoding, each reaction tube contained 16.0 µL molecular grade H_2O, 2.5 µL 10X buffer, 2.5 µL $MgCl_2$ (50 nM), 0.5 µL dNTPs (10 mM), 0.5 µL Platinum Taq (5U/µL; Invitrogen), 0.5 µL of each10 uM primer, and 2 µL of DNA template. The cycling conditions for amplification of the mini-barcodes were carried out as follows: 95°C for 5 min; 35 cycles of 94°C for 40 s, 46-50°C for 1 min, 72°C for 30 s; and a final extension of 72°C for 5 min. An annealing temperature of 46°C was used for primer set Mini_SH-E and an

annealing temperature of 50°C was used for primer set Mini_SH-D. Thermocycling was carried out with a Mastercycler pro S (Eppendorf, Hamburg, Germany).

PCR product confirmation for both full and mini-barcodes was carried out according to Handy et al. (2011b). PCR products (4 µL) were loaded onto pre-cast E-Gels (Life Technologies, Carlsbad, CA) and the total volume was brought up to 20 µL with dd H_2O. The gel was run for 20 min using an E-Gel iBase (Life Technologies). Images were captured using the Bio-Rad Imaging System with Quantity One v4.6.2 software. All PCR products were cleaned up using ExoSAP-IT (Affymetrix, Santa Clara, CA) following the manufacturer's instructions. Next, bidirectional cycle sequencing was carried out using M13 primers as described in Handy et al. (2011b). Sequencing clean-up was performed using an Edge Pro Bio PERFORMA DTR V3 96-well short plate (Edge Bio, Gaithersburg, MD) and the samples were run on a 3500xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA) using POP7 polymer (Thermo Fisher Scientific).

2.5. Analysis of DNA sequences

Following sequencing, the raw data was assembled and edited using Geneious v.5.4.7 (Biomatters Ldt., Auckland, New Zealand). Full-length barcode sequences were analyzed in accordance with quality control (QC) parameters established by Handy et al. (2011b), which call for bidirectional sequences with $\geq$ 500 bp and < 2% ambiguities or single reads with $\geq$ 500 bp and $\geq$ 98% high quality bases (HQ). Because QC parameters have not yet been established for mini-barcodes, the data were examined in two ways: (1) all sequences that were successfully assembled were examined and (2) similar QC parameters established for full-length barcodes were applied to the mini-barcodes (i.e., bidirectional sequences that are $\geq$ 76% of the target length and have < 2% ambiguities or single reads that are $\geq$ 76% of the target length and have $\geq$

98% HQ). Consensus sequences were generated for all successful files and aligned in Geneious using the "Muscle Alignment" default settings. The consensus sequences were then queried against the Barcode of Life Database (BOLD; http://www.boldsystems.org/index.php/IDS_OpenIdEngine), using the Public Record Barcode Database. If an identification could not be made with BOLD, the sequence was then searched using GenBank, using the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.6. Statistical analysis

The sequence lengths, quality scores (% HQ), and percent ambiguities within each set of full and mini-barcodes were compared across cooking methods using the Kruskal-Wallis H test (one-way ANOVA on ranks), using a significance level of $p < 0.05$. Statistically significant results were then analyzed with the post-hoc Dunn's test, using the Bonferroni correction for multiple tests ($p < 0.007$). To compare the sequencing success and failure rates of full barcodes, SH-D mini-barcodes, and SH-E mini-barcodes, Cochran's Q test was used with a predetermined significance level of $p < 0.05$, two-tailed. Analyses were carried out using IBM SPSS Statistics 23 (Armonk, New York, USA).

3. Results and Discussion

3.1. Full barcodes

Full barcoding showed an overall sequencing success rate of 90% (113 of 126 samples) when the results of all species and cooking methods were combined. Sequencing success rates varied by species, with the highest rate (100%) observed for swai and salmon, followed by 86% for scad, pollock and tilapia, and 81% for tuna. When compared based on cooking methods, the success rate was highest (100%) for samples that were baked, broiled, fried and smoked,

followed by 94.4% for uncooked and acid-treated samples and 39% for canned samples (Fig. 1a). Similarly, Shokralla et al. (2015) reported a low sequencing success rate (20.5%) for full barcoding of heavily processed, shelf-stable commercial fish products and Armani et al. (2015) reported a 0% success rate for full COI barcoding of canned seafood samples. The low success rates found with canned samples can be attributed to the degradation of DNA that occurs during the canning process, which includes high pressure and temperature. Previous studies have found that the DNA is degraded into fragments with maximum lengths of approximately 250-350 bp during canning (Chapela et al., 2007; Hsieh, Chai, & Hwang, 2007; Pardo & Pérez-Villarreal, 2004).

As shown in Table 1, the average full barcode length obtained for successfully sequenced samples was equal to or near the target length of 655 bp for most of the cooking methods. The two cooking methods that showed notable reductions in sequence length were canning (644 ± 25 bp) and broiling (646 ± 35 bp). These results were likely due to DNA degradation from the high heat treatments used with these cooking methods. As shown in Table 2, the average sequence quality was relatively high for all cooking methods tested with full barcoding, ranging from 95.6 ± 5.6% HQ for broiling to the highest score of 97.9 ± 2.7% HQ for frying. The average percent ambiguities among the full barcodes was very low, ranging from 0.02 ± 0.05% for the fried and smoked samples to 0.32 ± 0.50% for the uncooked samples (Table 3). According to the Kruskal-Wallis H test, there were no significant differences ($p > 0.05$) in the cooking methods when the full barcoding results were compared based on sequence length, quality scores, or percent ambiguities. Overall, these results suggest that full barcodes are a robust tool for successfully sequencing fish products for most cooking methods, with reduced success observed for canned samples.

Besides the ability to obtain a high quality sequence, it is also important that the resulting DNA barcode enables genetic identification of the fish sample. As shown in Table 4, full barcoding resulted in species-level identifications for four of the six types of fish tested in this study, with no other species matching at levels greater than 98% similarity. Specifically, the salmon was identified as Atlantic salmon (*Salmo salar*), the scad was identified as mackerel scad (*Decapterus macarellus*), the pollock was identified as walleye pollock (*Gadus chalcogrammus*), and the swai was identified as *Pangasius hypophthalmus*. On the other hand, the tilapia showed top matches to numerous species of commonly farmed tilapia species (*Oreochromis* spp. and *Coptodon zillii*). The inability of DNA barcoding to identify these samples at the species level is likely a result of the use of tilapia hybrids in aquaculture (Fitzsimmons, 2000). Due to its reliance on mitochondrial DNA, COI DNA barcoding cannot be used to differentiate hybrid species (Hellberg, Pollack, & Hanner, 2016). In the case of tuna, all samples tested matched multiple species within the *Thunnus* genus with genetic similarity of 100%. These results were consistent with previous research, which has reported challenges in discriminating closely related *Thunnus* species using COI-based DNA barcoding combined with BOLD (Lowenstein et al., 2009).

3.2. Mini-barcodes with QC parameters

As mentioned previously, the mini-barcodes were analyzed in two ways: with and without QC parameters. When QC parameters were applied to the mini-barcodes, SH-E mini-barcoding and full barcoding outperformed SH-D mini-barcoding across all cooking methods (Fig. 1a). SH-E mini-barcoding showed the highest overall success rate (92%), followed by full barcoding (90%), and then SH-D mini-barcoding (67%). According to Cochran's Q test, the success rate for SH-D mini-barcoding was significantly lower than the success rates for both full barcoding

and SH-E mini-barcoding ($p < 0.05$). There was no significant difference between the success rates of SH-E and full barcoding ($p > 0.05$). The success rate for SH-D mini-barcoding varied greatly by species, with swai having the lowest success (14%), followed by pollock (52%), scad (62%), tilapia/tuna (86%), and salmon (100%). On the other hand, SH-E performed relatively well across species, with 81% success for tilapia samples, 86% for tuna and scad samples, and 100% success for salmon, pollock and swai samples. Mini-barcodes also varied in terms of success rate by cooking method. As expected, SH-E mini-barcoding showed increased success in recovering sequences from canned products (50%) as compared to full barcoding (39%). Interestingly, SH-E mini-barcoding also outperformed full barcoding based on sequencing success for fish samples that were uncooked, acid-cooked, baked and broiled, with 100% success for each group. Unexpectedly, SH-D mini-barcoding did not perform well with canned samples and had the lowest success rate (28%) of all three barcoding methods. In comparison, Shokralla et al. (2015) reported success rates of 63.6% for SH-D mini-barcoding and 88.6% for SH-E mini-barcoding when tested with a variety of heavily processed commercial fish products. The rates reported in the current study were likely lower due to the use of QC parameters as well as differences in the types of fish tested. For example, Shokralla et al. (2015) did not test products labelled as containing swai, which showed low success rates in the current study for SH-D mini-barcoding.

As shown in Table 1, the average sequence length for SH-E was equal to the target length of 226 bp for all cooking methods, and close to the target of 208 bp for SH-D mini-barcoding. The canned samples showed the shortest average length across the SH-D sequencing results (200 ± 31). As shown in Table 2, the SH-E mini-barcodes had higher average sequence quality scores, ranging from 88.9 ± 6.2% for canned samples to 98.7 ± 0.8% for fried samples. In comparison,

301 the SH-D mini-barcodes ranged in quality from 79.8 ± 11.0% for uncooked samples to 90.0 ±
302 13.8% for fried samples. Similarly, SH-E mini-barcoding outperformed SH-D mini-barcoding
303 on the basis of percent ambiguities, with overall average values of 0.02 ± 0.11% and 0.35 ±
304 0.68%, respectively (Table 3).

305 Based on the results of the Kruskal-Wallis H test, there were no significant differences ($p >$
306 0.05) when SH-D mini-barcodes were compared across cooking methods for sequence lengths,
307 quality scores or percent ambiguities. Also, SH-E mini-barcodes also did not show significant
308 differences across cooking methods for sequence lengths; however, quality scores were found to
309 be significantly lower for canned samples as compared to those from all other sample groups
310 except acid cooking (Table 2). Percent ambiguities were significantly higher in canned samples
311 as compared to the other cooking methods (Table 3), according to the Kruskal-Wallis H test,
312 which was followed by Dunn's post-hoc test with the Bonferroni correction for multiple
313 comparisons ($p < 0.007$).

314 As shown in Table 4, the top species matches obtained with both SH-D and SH-E mini-
315 barcoding were very similar to those obtained for the full barcodes, meaning that a similar level
316 of discrimination was achieved despite the reduced barcode coverage. As with full barcoding,
317 both SH-D and SH-E mini-barcoding identified the species for four of the six fish types
318 analyzed. Although some of the SH-D mini-barcoding results showed a top species match to a
319 single tuna species (*T. albacares*), the COI mini-barcode has been determined previously not to
320 be a reliable indicator of tuna species and additional genetic markers have been recommended
321 for this purpose (Lowenstein et al., 2009; Mitchell & Hellberg, 2016; Shokralla et al., 2015).

322 Overall, when QC parameters were applied, SH-E mini-barcoding showed the greatest
323 sequencing success of the three methods and the same level of genetic discrimination as full

barcodes. These results indicated a strong potential for the use of SH-E mini-barcodes as a complementary method to full-length DNA barcoding, especially when analyzing fish that have been canned, acid-cooked, baked or broiled.

3.3. Mini-barcodes with no QC parameters

Because QC parameters have not yet been established for mini-barcodes, the data were also analyzed without standards for sequence quality, length, or percent ambiguities. As shown in Fig. 1b, when no QC parameters were applied to the mini-barcodes, SH-E mini-barcoding showed the highest overall success rate (94%) followed by SH-D (90%) and full barcoding (90%). There were no significant differences in these success rates ($p > 0.05$), according to Cochran's Q test. The removal of QC parameters had the greatest effect on the overall success rate of the SH-D mini-barcodes, which was 67% with QC parameters. In comparison, the removal of QC parameters did not have a major effect on the SH-E mini-barcoding success rate, which was 92% with QC parameters.

Interestingly, both SH-D and SH-E mini-barcodes outperformed full barcodes for uncooked, acid cooked, and canned samples, while SH-E and full barcoding showed the greatest success with the other cooking methods (Fig. 1b). The cooking method with the greatest disparity in success between full and mini-barcoding was canning, which showed 39% for full barcoding, 67% for SH-D mini-barcoding (no QC), and 56% for SH-E mini-barcoding (no QC). These results are improved as compared to SH-D and SH-E mini-barcoding with QC parameters, which showed success rates of 28% and 50%, respectively. Shokralla et al. (2015) reported similar sequencing success rates for heavily processed commercial fish products as compared to the current study for mini-barcode primer set SH-D (63.6%), but higher rates for primer set SH-E (88.6%). Similar to current study results, Armani et al. (2015) reported greater sequencing

success for canned seafood samples when a COI mini-barcode (139 bp) was used, as compared to the full COI barcode.

As shown in Tables 1 and 3, in the absence of QC parameters applied, SH-D mini-barcodes showed significant differences across cooking methods in terms of sequence length and percent ambiguities, based on the Kruskal-Wallis H test and Dunn's post hoc test with the Bonferroni correction ($p < 0.007$). Specifically, samples that had been canned (158 ± 68 bp) showed a significant reduction in sequence length, as compared to samples that were fried (201 ± 17 bp). These results are consistent with those found in previous studies, in that canned products had reduced sequencing success rates than other cooking methods (Armani et al., 2015; Chin, Adibah, Hariz, & Azizah, 2016). In terms of percent ambiguities, there were statistically significant differences between fried (0.39 ± 0.87%) and uncooked samples (0.78 ± 0.88%), but not in any of the other samples. As shown in Table 2, there were no significant differences among the sequence quality scores, which ranged from an average of 65.9 ± 38.5% for canned samples, to 86.0 ± 10.9% for broiled samples and 86.0 ± 14.7% for fried samples. The lower quality scores for canned samples are likely due to the degradation of DNA during processing.

SH-E mini-barcodes showed no significant differences in length across cooking methods according to the Kruskal-Wallis H test (Table 1). The average sequence length was consistently at the target length of 226 bp for all cooking methods except canning, which showed an average length of 213 ± 39 bp. Average quality scores were consistently higher than those found with SH-D mini-barcoding across all cooking methods, ranging from 84.6 ± 14.7% for canned samples to 98.6 ± 1.3% for uncooked samples (Table 2). According to the Kruskal-Wallis H test and Dunn's post hoc test with the Bonferroni correction ($p < 0.007$), the SH-E quality scores for canned samples were significantly lower than those of all other sample groups, except acid

cooking (Table 2) and the percentage of ambiguities was significantly higher for canned samples as compared to the other cooking methods (Table 3). However, the average percent ambiguity values obtained with SH-E mini-barcoding were consistently lower than those obtained with SH-D mini-barcoding across all cooking methods.

As shown in Table 4, there was one instance in which the lack of QC parameters led to inclusion of a sequence in the dataset with a lower level of species discrimination as compared to data with QC parameters applied. In this case, a successfully assembled canned tilapia sequence obtained with SH-D mini-barcoding could not be identified in BOLD and showed 100% genetic similarity to multiple species in GenBank, in addition to *Oreochromis* spp. and *Coptodon zillii*. This sequence was only 31 bp and showed a quality score of 0%, meaning that it was only analyzed in the data set that did not apply QC parameters.

Overall, the application of QC parameters reduced the rate of sequence recovery for both SH-D (26% decrease) and SH-E mini-barcodes (2% decrease). However, it also resulted in the exclusion of a low-quality SH-D sequence that could not be identified genetically. While the use of QC parameters allows for a standardized method of analyzing sequences, in some instances it may be desirable to analyze the data without QC parameters in order to increase sequencing success (e.g., for research purposes).

4. Conclusions

Overall, this study shows the robustness of full barcodes and mini-barcodes across many different cooking methods. Mini-barcoding was found to be advantageous over full barcoding for the analysis of canned samples and showed similar or improved sequencing success for many of the other cooking methods, with SH-E mini-barcoding showing the greatest overall success rates. Success rates were fairly consistent across cooking methods with the exception of canned

samples, which showed a marked reduction in success for both full and mini-barcoding. Canned samples also showed some statistically significant differences in sequencing quality scores, percent ambiguities, and lengths as compared to the other cooking methods. The application of QC parameters to mini-barcodes was found to have varying effects on success rates and further research should include developing a defined range of QC parameters for mini-barcodes. While full barcoding continues to be the standard method for genetic identification of fish species, this study has shown potential advantages to including mini-barcoding as a complementary analytical tool.

5. Acknowledgements

The authors would like to thank Dr. Lillian Were for assistance with editing; Dr. Kenneth Murphy for help with statistics; and Amy Kha and Michelle Ecarma for helping to complete the laboratory work for this project. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6. References

Armani, A., Guardone, L., La Castellana, R., Gianfaldoni, D., Guidi, A., & Castigliego, L. (2015). DNA barcoding reveals commercial and health issues sold on the Italian market in ethnic seafood. *Food Control, 55*, 206-214.

Chapela, M. J., Sotelo, C. G., Perez-Martin, R. I., Pardo, M. A., Perez-Villareal, B., Gilardi, P., & Riese, J. (2007). Comparison of DNA extraction methods from muscle of canned tuna for species identification. *Food Control, 18*(10), 1211-1215.

Chin, T. C., Adibah, A. B., Hariz, Z. A. D., & Azizah, M. N. S. (2016). Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving transparency in food market. *Food Control, 64*, 247-256.

Cline, E. (2012). Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. *Food Research International, 45*(1), 388-393.

Cohen, N. J., Deeds, J. R., Wong, E. S., Hanner, R. H., Yancy, H. F., White, K. D., . . . Gerber, S. I. (2009). Public health response to puffer fish (Tetrodotoxin) poisoning from mislabeled product. *Journal of Food Protection, 72*(4), 810-817.

EPA/FDA. EPA-FDA Advisory on Mercury in Fish and Shellfish. (2014). https://www.epa.gov/fish-tech/epa-fda-advisory-mercury-fish-and-shellfish Accessed 16.03.28.

FAO. (2016). The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Rome, Italy: Food and Agriculture Organization of the United Nations. 200 pp.

Fitzsimmons, K. (2000). Future trends of tilapia aquaculture in the Americas. In B. A. Costa-Pierce & J. E. Rakocy (Eds.), *Tilapia Aquaculture in the Americas* (Vol. 2, pp. 252-264). Baton Rouge, Louisiana: The World Aquaculture Society.

Handy, S. M., Deeds, J. R., Ivanova, N. V., Hebert, P. D. N., Hanner, R., Ormos, A., . . . Yancy, H. F. (2011a). A single-laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *Journal of AOAC International, 94*(1), 201-210.

Handy, S. M., Deeds, J. R., Ivanova, N. V., Hebert, P. D. N., Hanner, R. H., Ormos, A., . . . Yancy, H. F. Single Laboratory Validated Method for DNA-Barcoding for the Species Identification of Fish for FDA Regulatory Compliance. U.S. Food and Drug Administration Standard Operating Procedure. (2011b).

http://www.fda.gov/food/foodscienceresearch/dnaseafoodidentification/ucm237391.htm Accessed 16.11.01.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences, 270*(1512), 313-321.

Hellberg, R. S., Pollack, S. J., & Hanner, R. H. (2016). Seafood species identification using DNA sequencing. In R. H. Hanner & A. M. Nauum (Eds.), *Seafood Authenticity and Traceability: A DNA-based Perspective* (pp. 113-132). San Diego, CA, USA: Academic Press/Elsevier.

Hilderbrand, K. S., Jr. Smoking Fish at Home - Safely. (1999). http://seafood.oregonstate.edu/.pdf%20Links/Smoking-Fish-at-Home-Safely.pdf Accessed 16.10.20.

Hsieh, H. S., Chai, T. J., & Hwang, D. F. (2007). Using the PCR-RFLP method to identify the species of different processed products of billfish meats. *Food Control, 18*, 369-374.

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., . . . Bernatchez, L. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *Plos One, 3*(6), 8.

Kim, D.-W., Yoo, W. G., Park, H. C., Yoo, H. S., Dang, D. W., Jin, S. D., . . . Lim, J. (2012). DNA barcoding of fish, insects, and shellfish in Korea. *Genomics & Informatics, 10*(3), 206-211.

Landi, M., Dimech, M., Arculeo, M., Biondo, G., Martins, R., Carneiro, M., . . . Costa, O. F. (2014). DNA barcoding for species assignment: The case of Mediterranean marine fishes. *PLOS One, 9*(9), e106135.

460 Lowenstein, J. H., Amato, G., & Kolokotronis, S. O. (2009). The real maccoyii: identifying tuna
461 sushi with DNA barcodes – contrasting characteristic attributes and genetic distances.
462 *PLoS ONE, 4*(11), e7866.
463 Mitchell, J. K., & Hellberg, R. S. (2016). Use of the mitochondrial control region as a potential
464 DNA mini-barcoding target for the identification of canned tuna species. *Food Analytical*
465 *Methods, 9(10)*, 2711-2720.
466 NOC. National Ocean Council's Committee on Illegal, Unreported and Unregulated (IUU)
467 Fishing and Seafood Fraud (NOC Committee). Seafood Fraud Frequently Asked
468 Questions. (2016). http://www.iuufishing.noaa.gov/FAQs/SeafoodFraudFAQs.aspx
469 Accessed 16.11.1.
470 Pardo, M. A., & Pérez-Villarreal, B. (2004). Identification of commercial canned tuna species by
471 restriction site analysis of mitochondrial DNA products obtained by nested primer PCR.
472 *Food Chemistry, 86*, 143-150.
473 Hellberg, R. S., & Morrissey, M. T. (2011). Advances in DNA-based techniques for the
474 detection of seafood species substitution on the commercial market. *Journal of*
475 *Laboratory Automation, 16*, 308-321.
476 Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., & Hajibabaei, M. (2015). A DNA mini-
477 barcoding system for authentication of processed fish products. *Scientific Reports, 5*(Article number: 15894), 1-11.
479 Sicherer, S. H., Munoz-Furlong, A., & Sampson, H. A. (2004). Prevalence of seafood allergy in
480 the United States determined by a random telephone survey. *Journal of Allergy and*
481 *Clinical Immunology, 114*(1), 159-165.

Steinke, D., Zemlak, T. S., Boutillier, J. A., & Hebert, P. D. N. (2009). DNA barcoding of Pacific Canada's fishes. *Marine Biology, 156*, 2641-2647.

USDA. Complete Guide to Home Canning. (2015a). http://nchfp.uga.edu/publications/usda/GUIDE01_HomeCan_rev0715.pdf Accessed 17.01.26.

USDA. Safe Minimum Internal Temperature Chart. (2015b). http://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/safe-food-handling/safe-minimum-internal-temperature-chart/ct_index Accessed 16.12.03.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1462), 1847-1857.

Warner, K., Timme, W., Lowell, B., & Hirshfield, M. Oceana Study Reveals Seafood Fraud Nationwide. (2013). http://oceana.org/sites/default/files/reports/National_Seafood_Fraud_Testing_Results_FINAL.pdf Accessed 16.11.1.

Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J., Nguyen, N. L. T., . . . Hebert, P. D. N. (2008). Potential use of DNA barcodes in regulatory science: Applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection, 71*(1), 210-217.

Zhang, J., & Hanner, R. (2012). Molecular approach to the identification of fish in the South China Sea. *Plos One, 7*(2), e30621.

505 **Figure Captions**

506 **Figure 1.** DNA barcoding success rates for fish samples tested in this project (n = 126) using
507 quality control parameters applied to (a) both full and mini-barcodes or (b) full barcodes only

508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524

Table 1. DNA barcode lengths for fish samples successfully sequenced using full DNA barcoding and mini-barcoding (SH-D and SH-E). Mini-barcodes were analyzed with and without quality control (QC) parameters. Results are reported as the average ± standard deviation for samples tested with each cooking method

Cooking method	Full barcode length (bp)	Mini-barcode length (bp) with QC		Mini-barcode length (bp) without QC	
		SH-D	SH-E	SH-D	SH-E
Uncooked	655 ± 1	204 ± 7	226 ± 0	199 ± 18[ab]	226 ± 0
Acid	655 ± 1	201 ± 14	226 ± 0	196 ± 19[ab]	226 ± 0
Baked	654 ± 2	203 ± 9	226 ± 0	192 ± 30[ab]	226 ± 0
Broiled	646 ± 35	206 ± 5	226 ± 0	202 ± 6[ab]	226 ± 0
Canned	644 ± 25	200 ± 31	226 ± 0	158 ± 68[a]	213 ± 39
Fried	655 ± 0	207 ± 3	226 ± 0	201 ± 17[b]	226 ± 2
Smoked	652 ± 11	205 ± 5	226 ± 0	193 ± 33[ab]	226 ± 0
Overall	652 ± 16	203 ± 11	226 ± 0	192 ± 34	225 ± 11

[ab]A different superscript letter in the same column indicates a significant difference, based on the Kruskal-Wallis H test and Dunn's post hoc test using the Bonferroni correction for multiple tests ($p < 0.007$). Columns with no superscript letters did not have significant differences across cooking methods.

Table 2. DNA barcode quality scores (HQ%) obtained in this project for fish samples successfully sequenced using full DNA barcoding and mini-barcoding (SH-D and SH-E). Mini-barcodes were analyzed with and without quality control (QC) parameters. Results are reported as the average ± standard deviation for samples tested with each cooking method.

Cooking method	**Full barcodes HQ%**	**Mini-barcodes HQ% with QC**		**Mini-barcodes HQ% without QC**	
		SH-D	SH-E	SH-D	SH-E
Uncooked	96.3 ± 3.3	79.8 ± 11.0	98.6 ± 1.3[a]	78.2 ± 18.7	98.6 ± 1.3[a]
Acid	97.0 ± 3.6	80.8 ± 20.2	96.0 ± 4.2[ab]	77.3 ± 23.2	96.0 ± 4.2[ab]
Baked	97.7 ± 3.1	87.7 ± 12.0	98.0 ± 1.9[a]	80.7 ± 23	98.0 ± 1.9[a]
Broiled	95.6 ± 5.6	88.9 ± 11.3	98.2 ± 1.8[a]	86.0 ± 10.9	98.2 ± 1.8[a]
Canned	96.4 ± 2.6	86.5 ± 8.5	88.9 ± 6.2[b]	65.9 ± 38.5	84.6 ± 14.7[b]
Fried	97.9 ± 2.7	90.0 ± 13.8	98.7 ± 0.8[a]	86.0 ± 14.7	97.1 ± 6.7[a]
Smoked	97.1 ± 3.5	86.8 ± 11.5	95.9 ± 6.4[a]	81.4 ± 20.7	95.9 ± 6.4[a]
Overall	96.9 ± 3.7	85.9 ± 13.5	96.9 ± 4.4	79.1 ± 23	96.2 ± 6.8

[ab]A different superscript letter in the same column indicates a significant difference, based on the Kruskal-Wallis H test and Dunn's post hoc test using the Bonferroni correction for multiple tests ($p < 0.007$). Columns with no superscript letters did not have significant differences across cooking methods.

Table 3. DNA barcode ambiguities obtained in this project for fish samples successfully sequenced using full DNA barcoding and mini-barcoding (SH-D and SH-E). Mini-barcodes were analyzed with and without quality control (QC) parameters. Results are reported as the average ± standard deviation for samples tested with each cooking method**.**

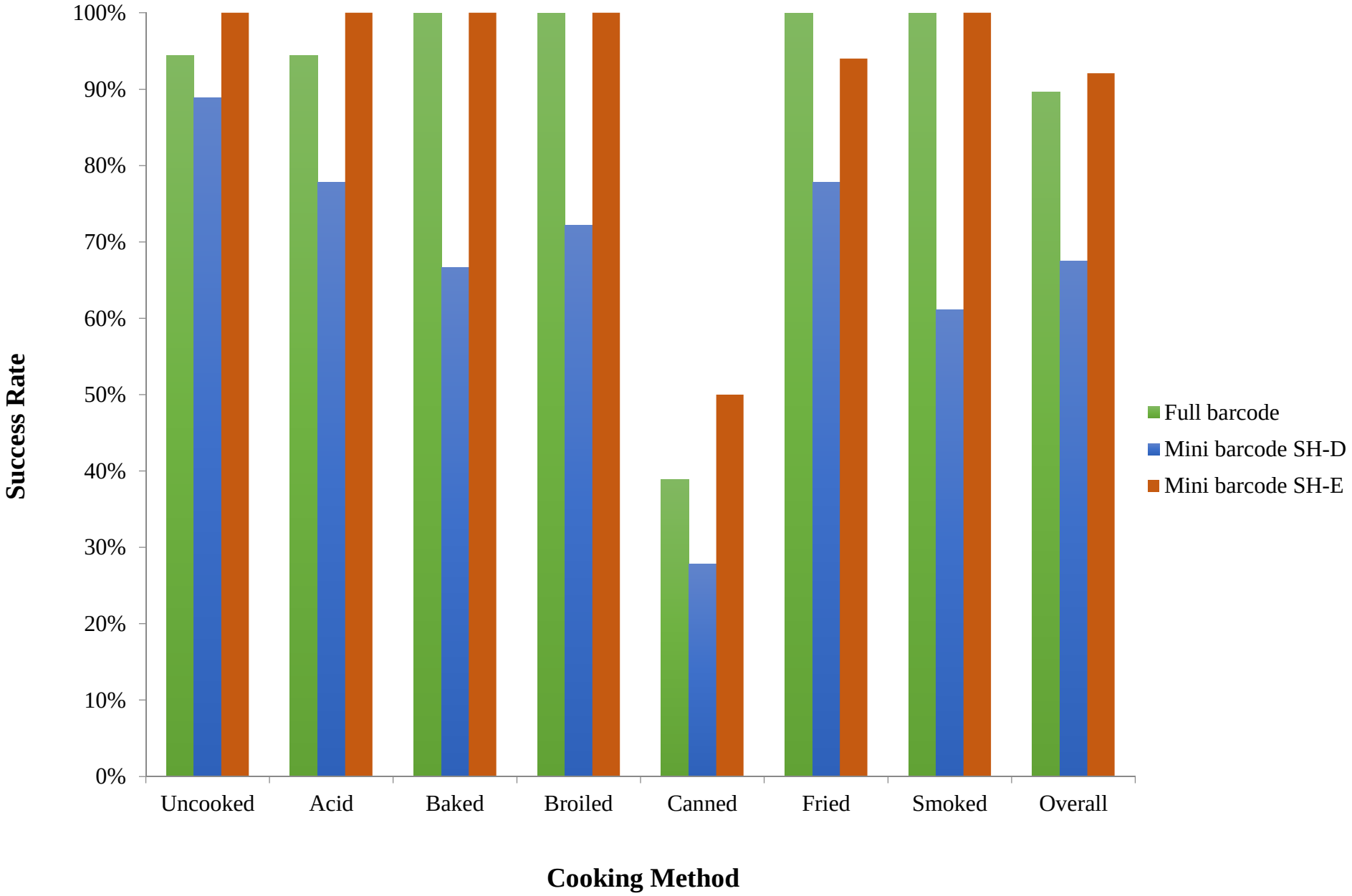
Cooking method	**Full barcode ambiguities (%)**	**Mini-barcode ambiguities (%) with QC**		**Mini-barcode ambiguities (%) without QC**	
		SH-D	SH-E	SH-D	SH-E
Uncooked	0.32 ± 0.50	0.95 ± 1.27	0.05 ± 0.21[a]	0.78 ± 0.88[a]	0.05 ± 0.21[a]
Acid	0.04 ± 0.07	0.40 ± 0.42	0.00 ± 0.00[a]	0.46 ± 0.58[ab]	0.00 ± 0.00[a]
Baked	0.07 ± 0.25	0.13 ± 0.34	0.00 ± 0.00[a]	0.24 ± 0.49[ab]	0.00 ± 0.00[a]
Broiled	0.03 ± 0.06	0.41 ± 0.46	0.00 ± 0.00[a]	0.39 ± 0.47[ab]	0.00 ± 0.00[a]
Canned	0.07 ± 0.13	0.08 ± 0.22	0.15 ± 0.22[b]	0.25 ± 0.49[ab]	0.23 ± 0.34[b]
Fried	0.02 ± 0.05	0.08 ± 0.18	0.00 ± 0.00[a]	0.05 ± 0.16[b]	0.00 ± 0.00[a]
Smoked	0.02 ± 0.05	0.26 ± 0.37	0.00 ± 0.00[a]	0.41 ± 0.70[ab]	0.00 ± 0.00[a]
Overall	0.08 ± 0.24	0.35 ± 0.68	0.02 ± 0.11	0.44 ± 0.89	0.03 ± 0.14

[ab]A different superscript letter in the same column indicates a significant difference, based on the Kruskal-Wallis H test and Dunn's post hoc test using the Bonferroni correction for multiple tests ($p < 0.007$). Columns with no superscript letters did not have significant differences across cooking methods.

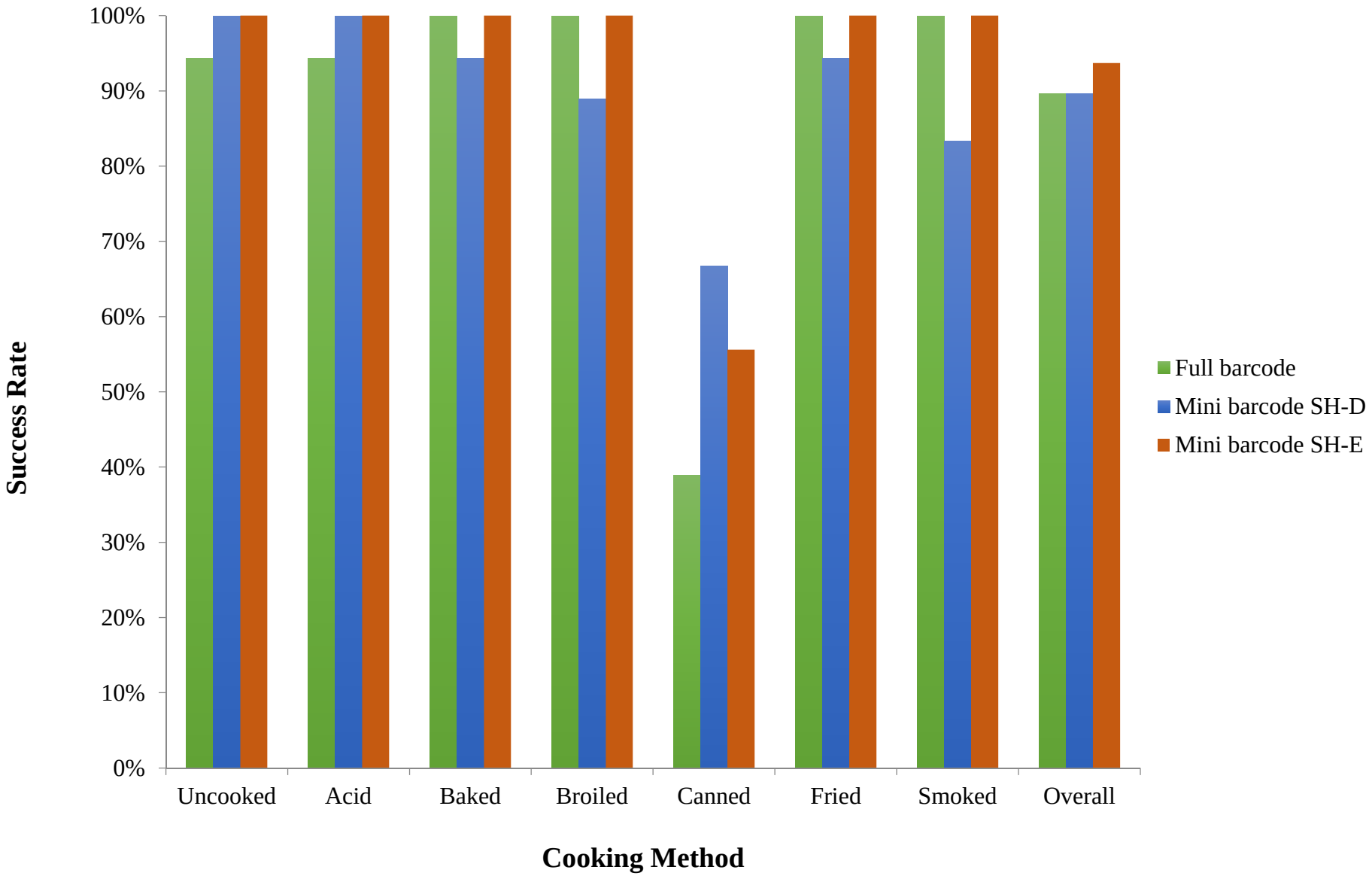
Table 4. Top species matches with genetic similarity >98% for samples that were successfully sequenced. All sequences were queried against the Barcode of Life Database (BOLD); in instances where BOLD was unable to identify a sequence, it was then queried against GenBank. Top species matches and genetic similarities were not affected by the application of quality control (QC) parameters to the mini-barcodes, unless otherwise noted.

Fish type	**Full barcoding/SH-E mini-barcoding results**		**SH-D mini-barcoding results**	
	Top species match	Genetic similarity	Top species match	Genetic similarity
Salmon	Atlantic salmon (*Salmo salar*)	100%	Atlantic salmon (*S. salar*)	100%
Tilapia	*Oreochromis* spp./Redbelly tilapia (*Coptodon zillii*)	99.8-100%	*Oreochromis* spp./Redbelly tilapia (*C. zillii*)[a]	100%
Tuna	*Thunnus* spp.	100%	*Thunnus albacares or Thunnus* spp.	99.5-100%
Scad	Mackerel scad (*Decapterus macarellus*)	99.5-100%	Mackerel scad (*D. macarellus*)	99.5-100%
Pollock	Walleye pollock (*Gadus chalcogrammus*)	100%	Walleye pollock (*G. chalcogrammus*)	99-100%
Swai	Swai (*Pangasianodon hypophthalmus*)	100%	Swai (*P. hypophthalmus*)	99.5-100%

[a]One canned tilapia sample analyzed with mini-barcode SH-D without QC parameters matched numerous additional species in other genera.



	Uncooked	Acid	Baked	Broiled	Canned	Fried	Smoked	Overall
Full barcode	94%	94%	100%	100%	39%	100%	100%	90%
Mini barcode	89%	78%	67%	72%	28%	78%	61%	67%
Mini barcode	100%	100%	100%	100%	50%	94%	100%	92%



	Uncooked	Acid	Baked	Broiled	Canned	Fried	Smoked	Overall
Full barcode	94%	94%	100%	100%	39%	100%	100%	90%
Mini barcode	100%	100%	94%	89%	67%	94%	83%	90%
Mini barcode	100%	100%	100%	100%	56%	100%	100%	94%

- DNA barcoding is a robust method for identification of species in processed fish
- Canned products showed marked decreases in sequencing success, quality, and length
- Mini-barcoding showed a slightly higher success rate than full barcoding
- Mini-barcoding and full barcoding showed similar results for species discrimination
- Mini-barcoding has high potential to be used as a complement to full barcoding