DNA Barcoding analysis of seafood accuracy in Washington, DC restaurants

David B Stern 1, Eduardo Castro Nallar 2, Jason Rathod 3, Keith A Crandall Corresp. 1, 4

1 Computational Biology Institute, Milken Institute School of Public Health, George Washington University, Washington, DC, USA
2 Center for Bioinformatics and Integrative Biology, Universidad Andrés Bello, Santiago, Chile
3 Migliaccio & Rathod LLP, Washington, DC, USA
4 Department of Invertebrate Zoology, Smithsonian Institution, Washington, DC, USA

Corresponding Author: Keith A Crandall
Email address: kcrandall@gwu.edu

In Washington DC, recent legislation authorizes citizens to test if products are properly represented and, if they are not, to bring a lawsuit for the benefit of the general public. Recent studies revealing the widespread phenomenon of seafood substitution across the United States make it a fertile area for consumer protection testing. DNA barcoding provides an accurate and cost-effective way to perform these tests, especially when tissue alone is available making species identification based on morphology impossible. In this study, we sequenced the 5’ barcoding region of the Cytochrome Oxidase I gene for 12 samples of vertebrate and invertebrate food items across six restaurants in Washington, DC and used multiple analytical methods to make identifications. These samples included several ambiguous menu listings, sequences with little genetic variation across closely related species and one sequence with no available reference sequence. Despite these challenges, we were able to make identifications for all samples and found that 33% were potentially mislabeled. Despite this high degree of mislabeling, the errors involved closely related species and we did not identify egregious substitutions as have been found in other cities. This study highlights the efficacy of DNA barcoding and robust analyses in identifying seafood items for consumer protection.
DNA Barcoding analysis of seafood accuracy in Washington, DC restaurants

David B. Stern¹, Eduardo Castro-Nallar², Jason Rathod³, and Keith A. Crandall¹,⁴

¹Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA
²Center for Bioinformatics and Integrative Biology, Universidad Andrés Bello, Santiago, Chile
³Migliaccio & Rathod LLP, Washington, DC, USA
⁴Department of Invertebrate Zoology, Smithsonian Institution, Washington, DC, USA

Corresponding Author:
Keith A. Crandall, PhD
Computational Biology Institute
George Washington University
800 22nd Street, NW
Washington, DC 20052
kcrandall@gwu.edu
Abstract

In Washington DC, recent legislation authorizes citizens to test if products are properly represented and, if they are not, to bring a lawsuit for the benefit of the general public. Recent studies revealing the widespread phenomenon of seafood substitution across the United States make it a fertile area for consumer protection testing. DNA barcoding provides an accurate and cost-effective way to perform these tests, especially when tissue alone is available making species identification based on morphology impossible. In this study, we sequenced the 5’ barcoding region of the Cytochrome Oxidase I gene for 12 samples of vertebrate and invertebrate food items across six restaurants in Washington, DC and used multiple analytical methods to make identifications. These samples included several ambiguous menu listings, sequences with little genetic variation across closely related species and one sequence with no available reference sequence. Despite these challenges, we were able to make identifications for all samples and found that 33% were potentially mislabeled. Despite this high degree of mislabeling, the errors involved closely related species and we did not identify egregious substitutions as have been found in other cities. This study highlights the efficacy of DNA barcoding and robust analyses in identifying seafood items for consumer protection.
Introduction

Seafood products are some of the most frequently mislabeled and misrepresented food items across the globe. The DNA-based investigation by Oceana, taking place from 2011 to 2013, revealed that nearly 33% of seafood is mislabeled in retail outlets (Warner et al. 2013). Snapper, tuna, and shrimp were the most frequently substituted and often were replaced with species carrying health advisories or of conservation concern (Warner et al. 2013). State and federal consumer protection laws prohibit food mislabeling of this kind. One of the strongest such laws is Washington DC’s Consumer Protection Procedures Act (“CPPA”). The CPPA is remarkably comprehensive in its protections of consumers and includes expansive remedies if the law is violated. Specifically, the statute allows for individuals to purchase goods for the express purpose of testing their contents to determine whether they comport with the DCCPA. It is illegal under the DCCPA to “represent that goods or services have a source, sponsorship, approval, certification, accessories, characteristics, ingredients, uses, benefits, or quantities that they do not have” (D.C. Code § 28-3904). Violations of the DCCPA entitle plaintiffs to $1500 in statutory damages per violation and the ability to recover not just for themselves, but on behalf of the general public, enabling citizens to act as “private attorneys general” (Rubenstein 2004).

DNA barcoding is a particularly useful tool for making species-level identifications when other data (morphological, geographical, ecological, etc.) are not available or may mislead identifications. The 5’ end of the Cytochrome Oxidase I (COI) mitochondrial gene has been established as the “barcode” sequence for metazoan taxa due to its high variability and conservation of PCR primer sites (Hebert et al. 2003). Due to the efforts of biologists around the world, there is a large database of barcode DNA sequences from 260,654 species of animals, plants, fungi and other life (Barcode of Life Database, http://www.boldsystems.org/, Accessed December 1, 2016), facilitating identification. The approach has been proven to be a valuable forensic tool in identifying substitution in the seafood market (Ardura et al. 2010; Hanner et al. 2011; Lowenstein et al. 2009; Torres et al. 2013; Vandamme et al. 2016; Wong & Hanner 2008),
as well as identifying seafood that carry conservation and human health concerns (Lowenstein et al. 2010). Although sophisticated uses of DNA barcoding exist to characterize large dataset of DNA barcodes, (e.g Brown et al. 2012; Zhang et al. 2012), a typical barcoding analysis consists of querying a sampled sequence against a database such as the National Center for Biotechnology Information (NCBI) GenBank or BOLD using BLAST (Altschul et al. 1990; Altschul et al. 1997) or a similar algorithm. The potential pitfalls of using this method to identify species have been explored and include sequencing errors, lack of variation in the barcode marker, incorrect identification of voucher sequences, arbitrary similarity cutoffs, unavailability of references and accidental sequencing of nuclear mitochondrial genes, in addition to methodological and theoretical shortcomings (Collins & Cruickshank 2013).

Even beyond seafood and species identification, phylogenetic methods have a history of use in US court cases (e.g. Metzker et al. 2002). For our purposes, phylogenetics presents a potential to increase and supplement the strength of similarity-based identifications by setting the tests in a statistical framework (Huelsenbeck & Crandall 1997). When 100% identical matches are not found or when multiple reference sequences are similarly distant from the query, phylogenetic topology tests can garner statistical support to reject alternative hypotheses. We can test a priori hypotheses based solely on the labeling of the seafood item, or test if an identification made by some other method is significantly better than a competing one.

In this study, as part of a preliminary legal investigation exploring the viability of a lawsuit under Washington DC’s CPPA, we sample the barcoding region of the COI gene to test the identity of commonly mislabeled seafood products from six restaurants in Washington, DC. We take multiple approaches to identify the sampled specimens. While we report multiple potentially mislabeled food items, the majority of our tested items were either correctly labeled or were identified as legally acceptable species for the food labeling.

Materials and Methods
Tissue and Data collection

Six restaurants in Washington, DC (Table 1) were visited in March of 2015. Two seafood dishes were sampled from each restaurant, targeting menu-listings with potential human health or conservation concerns. One tissue sample per menu item was stored in 95% ethanol. Care was taken to only collect tissue from the main dish from the center of the food item after rinsing away other substances. Total DNA from each sample was extracted with the Qiagen DNeasy DNA extraction kit. PCR amplification of COI was performed using the LCO1490/HCO2198 (Folmer et al. 1994) primers for invertebrate species and FishF2_t1/FishR2_t1 (Ivanova et al. 2007) for vertebrate species. 25 µl PCR reactions contained 2.5 µl 10X PCR Buffer, 0.75 µl 50 mM MgCl₂, 0.5 µl 10 mM dNTPs, 0.5 µl 10 µM forward primer, 0.5 µl 10 µM reverse primer, 0.1 µl Platinum Taq DNA Polymerase (Thermo Fisher Scientific, Waltham, MA) and 1 µl DNA template. After initial denaturation step of 95°C for five minutes, PCR took place in 35 cycles of 95°C for 30 seconds, 50°C for 45 seconds and 72°C for 60 seconds with a final extension of 72°C for five minutes. PCR products were cleaned with ExoSAP-IT (Affymetrix, Santa Clara, CA). Each product was sequenced in both directions using the Big Dye (Life Technologies) cycle sequencing protocol and an ABI 3730XL sequencer. We followed Song et al. (2008) to avoid sequencing nuclear copies of COI by checking for high-quality base calls and assuring the absence indels and stop codons in sequence alignments. Sequences have been deposited to NCBI GenBank for public access (Accession numbers: KY656473 - KY656484)

Database searches and reference sequence collection

Sequences were searched against the Barcode of Life Database (http://www.boldsystems.org) for initial species identification. BOLD contains barcode COI sequences from NCBI as well as from other researchers and institutions who deposit sequences along with species identifications and metadata. We also collected COI sequences from all species present in the top 100 hits of the NCBI BLAST searches and BOLD searches, if available.
Multiple reference COI sequences for the reported species or taxonomic group based on the menu listings were collected from NCBI GenBank. Two types of situations arose in which a proper reference sequence was not available. One was where the menu listing makes reference to a particular species, but a COI sequence was not available on NCBI or BOLD (e.g. Rock Shrimp, *Sicyonia brevirostris*). In this situation, we collected sequences from related species based on NCBI taxonomy. The other was where the menu listing does not reference one particular species (e.g. Snapper). In this situation, we used only the top BLAST hits from NCBI and BOLD in subsequent analyses.

**Phylogenetic analysis**

Two alignments were generated, one each for invertebrate and vertebrate sequences. Sequences were aligned using TranslatorX (Abascal et al. 2010) which translates protein-coding sequences to amino acids, aligns protein sequences with MAFFT v7.305b (Katoh & Standley 2013) and then back-translates to nucleotides. This aids in the identification and removal of sequences with premature stop codons and indels that are indicative of numts (Song et al. 2008).

Best-fit models of evolution and an optimal data-partitioning scheme were chosen using PartitionFinder v1.1.1 (Lanfear et al. 2012) with each codon position chosen as *a priori* data subsets and using the Bayesian Information Criterion (BIC) for model selection (Posada & Crandall 2001). A maximum-likelihood tree was estimated in RAxML 8.2 (Stamatakis 2014) using the partitioning scheme selected with PartitionFinder. 1000 bootstrap replicates were followed by 10 maximum likelihood tree searches under the GTRCAT model with final optimization under GTRGAMMA. We used MrBayes 3.2.2 (Ronquist et al. 2012) to asses the posterior probabilities of nodes recovered in the maximum likelihood estimate using the partitioning scheme and models suggested by PartitionFinder. Four independent runs of four MCMC chains were run for 10,000,000 generations, sampling the cold chain every 1000 steps for a total of 40,000 samples. The first 10% of samples were discarded as burn-in. Convergence was
assessed by assuring the standard deviation of split frequencies across runs was below 0.01 and that ESS values for all parameters were above 10,000.

Topology tests

A strength of phylogenetic analysis comes from the ability to compare statistical support for alternative topologies (Huelsenbeck & Crandall 1997). In this way, we can test alternative \textit{a priori} hypotheses, where appropriate, in which the sampled sequences are constrained to form monophyletic groups with reference sequences or constrained to not form monophyletic groups with those sequences. While monophyly does not equate to identification, strong support against monophyly would suggest that the query sequence is not the same species as the target group. We formed the alternative hypotheses based on the menu listings (Table 2). For example, we were able to test if our Chilean Seabass sequence formed a monophyletic group with other Chilean Seabass sequences, but could only test if “Snapper” formed a monophyletic group with other sequences that could be labeled “Snapper,” since the menu listing was not species-specific. Where appropriate, we also tested hypotheses for alternative groupings based on the maximum likelihood phylogeny. In a Bayesian framework, model selection is performed using the Bayes factor, which is the ratio of marginal likelihoods of two competing hypotheses. Marginal likelihoods were estimated for each hypothesis using stepping-stone analysis in MrBayes 3.2.2 (Ronquist et al. 2012) with the same models and partitioning scheme as above. Stepping-stone analyses were executed with 4 independent runs of 4 chains each for 50 steps of 200,000 generations each (sampled every 1000 generations) each for a total of 10,000,000 generations in each run for each hypothesis. The first step (200,000 generations) was discarded as burn-in. We considered a Bayes factor greater than 5 as strong support for one hypothesis over another (Kass & Raftery 1995).

Character-based tests
Fish species of the genus *Thunnus* have been shown to have considerably low levels of COI diversity (Ward et al. 2005) potentially as the result of rapid speciation and large effective population sizes that make distance-based identifications challenging (Elias et al. 2007). In this situation, a query sequence can have a nearly identical distance to multiple different reference sequences. While model-based phylogenetic approaches can be effective in this case, we also took advantage of the character-based identification key of Lowenstein et al. (2009) to identify tuna samples. This key consists of 14 nucleotide characters at specific positions across the COI barcoding region developed from an alignment of 87 reference sequences with diagnostic states that are specific to each of the eight tuna species.

**Results**

**Database searches**

BOLD searches were able to make species-level identifications for six of the 12 sampled specimens and genus-level identifications for the other six (Table 1). BOLD did not make species-level identifications for “Australian Barramundi,” “Snapper Salsa Verde” or any of the tuna samples. The “Australian Barramundi” sample was a 100% identical match with *Lates uwisara* and *L. calcarifer*. The BOLD records for *Lates uwisara* linked to GenBank records that were named as *L. calcarifer* and were therefore treated as *L. calcarifer* in phylogenetic analyses. Of the six species-level identifications, two did not match the species reported by the restaurants (Table 1). The COI sequence generated from Bobby Van’s “Rock Shrimp Tempura” sample matched the Whiteleg shrimp (*Litopenaeus vannamei*) sequence with 100% identity (Table 1). The Chilean Seabass sequence from The Oceanaire Seafood Room matched the Antarctic Toothfish (*Dissostichus mawsoni*) with 100% similarity.

While the labeling of “Gulf shrimp”, “Calamari,” “Everything Tuna” and “Snapper Salsa Verde” are not species-specific, the “Gulf shrimp” from Gordon Biersch matched *Farfantepenaeus aztecus* with up to 99.85% similarity and the “Calamari” from Bobby Van’s
Steakhouse matched *Uroteuthis edulis* with 100% identity. The “Everything Tuna” sample matched three different tuna species with *Thunnus obesus* (Bigeye Tuna) as the best hit and “Snapper Salsa Verde” matched three different species in the genus *Lutjanus* with *L. guttatus* as the best hit.

**Phylogeny-based tests**

The final alignments for vertebrate and invertebrate taxa were 699 and 708 bp long, respectively, and contained no gaps. Inspecting placement of query sequences in the maximum likelihood COI trees supported the BOLD identifications with high bootstrap support and posterior probability, except in the case of “tuna” samples, where the phylogeny failed to resolve relationships with high support and numerous species were not monophyletic (Figures 1 & 2). The failure to resolve a number of species and genera as monophyletic was likely the result of relying on a single mitochondrial marker to estimate a phylogeny of many distantly related groups. This is expected given the numerous biological and methodological reasons that a given gene tree may not match the true “species” tree (Edwards 2009).

Bayesian topology tests were able to garner support for identifications where BOLD searches, based on percent similarity, could not (Table 2). Our analysis of Joe’s Seafood’s “Ahi Tuna” suggested with strong support (BF>15) that the sequence was not monophyletic with other Ahi (*Thunnus albacares*) sequences. The maximum-likelihood phylogeny placed this sequence sister to a clade of six other tuna species and our Bayesian topology test supported its grouping with Albacore tuna (*Thunnus alalunga*) with strong support over its grouping with Yellowfin/Ahi (BF~25). BOLD analysis suggested that the Yellowfin tuna from Gordon Biersch could have been one of three species (Table 1) and the maximum-likelihood phylogeny narrowed this down to a likely grouping with Yellowfin (*T. albacares*). The topology test agreed (BF=8.36) that the sequence formed a monophyletic group with other Yellowfin sequences. We found strong support for the hypothesis that our sample from McCormick and Schmick’s “Albacore Tuna” did
not form a monophyletic group with the other Albacore (*Thunnus alalunga*) sequences (BF=20.95). The maximum-likelihood phylogeny placed this sequence sister to a *Thunnus albacares* sequence and our topology test supported this grouping over *Thunnus alalunga* (BF=24.55). The maximum-likelihood phylogeny grouped the “Snapper Salsa Verde” sequence with an assemblage of *Lutjanus guttatus* and *L. synagris*. While the topology tests suggested strongly that this sequence grouped with other *Lutjanus* sequences (BF > 55), we could not find significant support for its grouping with *Lutjanus guttatus* or *L. synagris* over the other (BF < 1).

Regardless of the fact that there was no reference COI sequence for Rock Shrimp (*Sicyonia brevirostris*) available for download, we rejected the hypothesis that this sequence formed a group with other *Sicyonia* sequences and found strong support for its grouping with the Whiteleg Shrimp (*Litopenaeus vannamei*, BF>40).

**Character-based identification**

Using the character-based identification scheme of Lowenstein et al. (2009), we were able to make species-level identifications of all of the tuna specimens (Table 3). Only one of the four tuna specimens matched the menu-listed species, although one did not make a species-level listing. The character-based identifications matched our phylogenetic hypotheses in all cases.

**Discussion**

DNA barcoding is a powerful and cost-effective method for making species-level identifications of seafood items. Here we take advantage of multiple analytical methods (database searches, phylogeny estimate, topology tests, character-based ID) using DNA barcoding to test for seafood substitutions in six restaurants in Washington, DC. Perhaps the clearest act of seafood substitution was the sale of Whiteleg shrimp (*Litopenaeus vannamei*) as “Rock Shrimp.” This was a commonly observed trend in Oceana’s 2014 study (Warner et al. 2014). One of the three “Chilean seabass” we tested was identified as Antarctic toothfish, the sister species to the
“Chilean seabass” also known as the “Patagonian toothfish.” In general there was agreement, where possible, across multiple methods, and where BOLD was unable to make a species-level identification, another method was able to narrow the identification further. Three of the major issues confronted in this analysis were 1) lack of available reference sequence, 2) ambiguous menu listings, and 3) lack of sequence variation among congeneric species.

The lack of available reference sequences was largely an issue that was discovered through database searches and phylogeny estimation. Particularly for the “Rock Shrimp” specimen, there are no available reference COI sequences for *Sicyonia brevirostris* on GenBank or BOLD. Therefore, we were not able to explicitly test if our sampled “Rock Shrimp” sequence was *Sicyonia brevirostris* even though BOLD does have sequences from several *Sicyonia* species including *S. brevirostris*. It is unfortunate that these are not currently made available to the public especially considering the widespread sale and consumption of “Rock Shrimp” across the country (Warner et al. 2014). We were able to test if our sample belonged to the genus *Sicyonia* because COI sequences from *S. carinata* and *S. lancifer* were available. Nevertheless, the sequence matched with 100% identity to the Whiteleg Shrimp, *Litopenaeus vannamei*.

Many of the menu listings were vague enough to allow an array of acceptable species. For example, “snapper” can refer to any of the 113 species in the family Lutjanidae or to species from a number of other families with a common name that includes “snapper.” “Red snapper” is approved in the US to only refer to the species *Lutjanus campechanus* and has been involved in a number of cases of seafood fraud (Marko et al. 2004). The phylogenetic analyses strongly supported our sample’s placement in the genus *Lutjanus*, but we were not able to distinguish between an identification of *L. guttatus* or *L. synagris*. Either these species are very closely related (resulting in little variation among COI sequences) or reference sequences were not properly identified. Our analyses strongly suggested that the “Calamari” specimen was *Uroteuthis edulis*, which is one of many squid species approved to be sold as “Calamari” by the FDA (http://www.accessdata.fda.gov, Accessed August 25, 2016). *Farfantepenaeus* shrimp species
are commercially important with millions of pounds being caught and sold each year (Warner et al. 2014). *Farfantepenaeus aztecus*, also known as “brown shrimp”, is the most widely fished species in the genus. These species are found in the Gulf of Mexico and are therefore often referred to as “Gulf Shrimp.”

There are eight recognized species in the genus *Thunnus*, but all can be sold under the name “tuna” or its synonyms in the US according to the FDA (Lowenstein et al. 2009). Database searches and phylogeny estimation failed to make species-level identifications for our “tuna” samples based on the COI sequence. The very low variability among tuna COI sequences has been observed before and results in low support for phylogenetic groupings (Ward et al. 2005). Phylogenetic topology tests did prove more successful in this manner, because we were able to directly compare support for alternative hypotheses, even if the probability of one particular phylogeny or grouping is not objectively high. These results agreed with the character-based identifications, which proved to be the most useful for tuna samples in combination with phylogenetic tests.

The “Everything Tuna” sample was identified as *Thunnus obesus*, which is given a status of “Vulnerable” by the International Union for Conservation of Nature Redlist (IUCN 2015). This was the only one of our samples that was identified as a species with a potential conservation concern. This was certainly lower than the rate of threatened or endangered species found in other cities that have been surveyed, although we did not test grouper, halibut or eel samples, which are frequently substituted with fish from threatened populations (Vandamme et al. 2016; Warner et al. 2013). While none of the substituted seafood items we found carry well-known health risks or conservation concerns *per se*, the labeling of farmed Whiteleg Shrimp as “Rock Shrimp” is worrisome as wild-caught shrimp are not required to be screened for veterinary drug residue levels like farmed shrimp (FDA 2011).

Conclusions
Of our 12 samples from six restaurants, we found that four menu items, one “Chilean Seabass,” two “Tuna” and one “Rock Shrimp,” were potentially mislabeled, albeit with species that are either closely related or typically considered acceptable for the menu listing. This is consistent with the 33% average rate found across United States cities by Oceana and lower than seven of the 12 cities surveyed in that study (although higher than the 26% found in Washington, DC by Oceana in 2013) (Warner et al. 2013).

As always, these results rely on the quality of the reference database used, especially the identification of voucher specimens for reference sequences. The potential mislabeling we identified here requires further investigation in order to pinpoint the source of the substitution.

Our study highlights the utility of using multiple analytical methods to identify specimens with standard DNA barcoding and especially that of statistical phylogenetics.

Table 1. Sampling information and identification with BOLD and the maximum-likelihood phylogeny

| Restaurant         | Menu Listing  | Putative Genus | Putative Species | BOLD ID (Top percent identity) | ML COI Phylogeny ID (Bootstrap support/ PP of node) |
|--------------------|---------------|----------------|-----------------|--------------------------------|---------------------------------------------------|
| Bobby Van’s Steakhouse | Rock Shrimp Tempura | Sicyonia brevirostris | Litopenaeus vannamei (100) | Litopenaeus vannamei (88/87) |
| Gordon Biersch      | Yellowfin Tuna    | Thunnus albacares   | Thunnus albacares (100), Thunnus atlanticus (100), Thunnus obesus (100) | Thunnus albacares (11/6) |
|                      | NA             | NA               | Farfantepenaeus aztecsus (99.85) | Farfantepenaeus aztecsus (100/100) |
| The Oceanaire Seafood | Chilean Seabass | Dissostichus eleginoides | Dissostichus mawsoni (100) | Dissostichus mawsoni (99/100) |
| Room | Australian Barramundi | Lates calcarifer | Lates uwisara (100), Lates calcarifer (100) |
|------|-----------------------|-----------------|---------------------------------------------|
| Joe’s Seafood, Prime Steak and Snow Crab | Ahi Tuna Tartare | Thunnus albacares | Thunnus alalunga (100), Thunnus obesus (99.83), Thunnus orientalis (99.67), Thunnus maccouyi (99.67) |
|      |                       |                 | Thunnus sp. (31/9)                          |
|      | Chilean Seabass       | Dissostichus eleginoides | Dissostichus eleginoides (100) |
|      |                       |                 | Dissostichus eleginoides (99/99)            |
| Legal Sea Foods | Snapper Salsa Verde | Lutjanus | Lutjanus guttatus (100), Lutjanus sp. (100), Lutjanus synagris (98.31) |
|      |                       | NA              | Lutjanus guttatus, Lutjanus synagris (85/41) |
|      | Everything Tuna       | Thunnus | NA | Thunnus obesus (100), Thunnus albacares (99.84) |
|      |                       |                 | Thunnus obesus (61/52)                      |
| McCormick and Schmick’s | Sesame Crusted Albacore Tuna | Thunnus alalunga | Thunnus albacares (100), Thunnus atlanticus (100), Thunnus obesus (100) |
|      |                       |                 | Thunnus albacares (14/6)                    |
|      | Pesto Chilean Seabass | Dissostichus eleginoides | Dissostichus eleginoides (100) |
|      |                       |                 | Dissostichus eleginoides (99/99)            |

Table 2. Tested hypotheses and marginal log likelihoods used for Bayes Factors

| Restaurant          | Menu Listing          | Tested Hypotheses (Marginal LOG Likelihood)                                                                 |
|---------------------|-----------------------|-------------------------------------------------------------------------------------------------------------|
| Bobby Van's Steakhouse | Rock Shrimp Tempura  | Monophyletic with Sicyonia (-6984.27), Not monophyletic with Sicyonia (-6951.31), Monophyletic with Whiteleg Shrimp (-6940.03) |
| Gordon Biersch       | Yellowfin Tuna        | Monophyletic with Yellowfin Tuna (-5463.35), Not-Monophyletic with Yellowfin Tuna (-5471.71)              |
| Restaurant             | Menu Listing            | Character positions (Lowenstein et al. 2009) | ID                        |
|------------------------|-------------------------|---------------------------------------------|---------------------------|
| Gordon Biersch         | Yellowfin tuna          | CCCC ACCT ATTG AC                            | Yellowfin tuna (T. albacares) |
| Joe’s Seafood, Prime Steak and Snow Crab | Ahi Tuna Tartare | CTCC GCAT ATCA AT                           | Albacore (T. alalunga)    |

**Table 3.** Character based identification for tuna species.
| Legal Sea Foods | Everything Tuna | CCCT ACGG ATTG AC | Bigeye tuna  
(T. obsesus) |
|----------------|-----------------|-------------------|------------------|
| McCormick and Schmick’s | Sesame Crusted Albacore Tuna | CCCA ACGT ATTG AC | Yellowfin tuna  
(T. albacares) |

**Figure 1.** Maximum-likelihood phylogeny of Vertebrate COI sequences. Names of samples from this study are colored according to the “target” sequences of the menu-listing. Reference sequence labels include GenBank accession numbers. Sequences that are only found on BOLD are labeled as such before the accession number.

**Figure 2.** Maximum-likelihood phylogeny of Invertebrate COI sequences.

**Acknowledgements**
We thank the Smithsonian’s Laboratories for Analytical Biology (LAB) for lab space and sequencing services and the George Washington University’s Colonial One High-Performance Computing Cluster for computational support.

**References**

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. BASIC LOCAL ALIGNMENT SEARCH TOOL. *Journal of Molecular Biology* 215:403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, and Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25:3389-3402.
Ardura A, Linde AR, Moreira JC, and Garcia-Vazquez E. 2010. DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation* 143:1438-1443.

Brown SD, Collins Ra, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, and Cruickshank RH. 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular ecology resources* 12:562-565.

Collins Ra, and Cruickshank RH. 2013. The seven deadly sins of DNA barcoding. *Molecular ecology resources* 13:969-975.

Edwards SV. 2009. IS A NEW AND GENERAL THEORY OF MOLECULAR SYSTEMATICS EMERGING? *Evolution* 63:1-19.

Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AV, Mallet J, and Jiggins CD. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London B: Biological Sciences* 274:2881-2889.

FDA. 2011. Fish and fishery products hazards and controls guidance. US Department of Health and Human Services Food and Drug Administration Center for Food Safety and Applied Nutrition.

Folmer O, Black M, Hoeh W, Lutz R, and Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 3:294-299.

Hanner R, Becker S, Ivanova NV, and Steinke D. 2011. FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* 22:106-122.

Hebert PDN, Cywinska A, Ball SL, and DeWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences* 270:313-321.

Huelsenbeck JP, and Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28:437-466.
IUCN. 2015. IUCN Red List of Threatened Species. Available at http://www.iucnredlist.org (accessed 3/29/2015 2015).

Ivanova NV, Zemlak TS, Hanner RH, and Hebert PD. 2007. Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes 7:544-548.

Kass RE, and Raftery AE. 1995. Bayes factors. Journal of the american statistical association 90:773-795.

Katoh K, and Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30:772-780.

Lanfear R, Calcott B, Ho SY, and Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution 29:1695-1701.

Lowenstein JH, Amato G, and Kolokotronis S-O. 2009. The real maccoyii: identifying tuna sushi with DNA barcodes—contrasting characteristic attributes and genetic distances. PloS one 4:e7866.

Lowenstein JH, Burger J, Jeitner CW, Amato G, Kolokotronis SO, and Gochfeld M. 2010. DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. Biology letters 6:692-695.

Marko PB, Lee SC, Rice AM, Gramling JM, Fitzhenry TM, McAlister JS, Harper GR, and Moran AL. 2004. Mislabelling of a depleted reef fish. Nature 430:309-310.

Metzker ML, Mindell DP, Liu X-M, Ptak RG, Gibbs Ra, and Hillis DM. 2002. Molecular evidence of HIV-1 transmission in a criminal case. Proceedings of the National Academy of Sciences of the United States of America 99:14292-14297.

Posada D, and Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. Systematic biology 50:580-601.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, 
Suchard MA, and Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic 
inference and model choice across a large model space. Systematic biology 61:539-542.
Rubenstein WB. 2004. On what a private attorney general is and why it matters. Vanderbilt Law 
Review 57:2129.
Song H, Buhay JE, Whiting MF, and Crandall KA. 2008. Many species in one: DNA barcoding 
overestimates the number of species when nuclear mitochondrial pseudogenes are 
coamplified. Proceedings of the National Academy of Sciences of the United States of 
America 105:13486-13491.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large 
phylogenies. Bioinformatics 30:1312-1313.
Torres Ra, Feitosa RB, Carvalho DC, Freitas MO, Hostim-Silva M, and Ferreira BP. 2013. DNA 
barcoding approaches for fishing authentication of exploited grouper species including 
the endangered and legally protected goliath grouper <i>Epinephelus itajara</i>. Scientia 
Marina 77:409-418.
Vandamme SG, Griffiths AM, Taylor SA, Di Muri C, Hankard EA, Towne JA, Watson M, and 
Mariani S. 2016. Sushi barcoding in the UK: another kettle of fish. Peerj 4.
Ward RD, Zemlak TS, Innes BH, Last PR, and Hebert PDN. 2005. DNA barcoding Australia’s 
fish species. Philosophical Transactions of the Royal Society B-Biological Sciences 
360:1847-1857.
Warner K, Golden R, Lowell B, Disla C, Savitz J, and Hirshfield M. 2014. Shrimp: Oceana 
Reveals Misrepresentation of America’s Favorite Seafood. Oceana, http://oceana 
org/sites/default/files/reports/oceana_reveals_misrepresentation_of_americas_favorite_s 
eafood pdf.
Warner K, Timme W, Lowell B, and Hirshfield M. 2013. Oceana study reveals seafood fraud 
nationwide. Oceana Retrieved on August 11:2014.
Wong EHK, and Hanner RH. 2008. DNA barcoding detects market substitution in North American seafood. *Food Research International* 41:828-837.

Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, Wan P, Feng J, and Ward RD. 2012. A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular Ecology* 21:1848-1863.
