Communication **Open Access**

Amanda M. Naaum*, Jason St.Jaques, Kimberly Warner, Linda Santschi, Ralph Imondi, Robert Hanner

Standards for Conducting a DNA Barcoding Market Survey: Minimum Information and Best Practices

DOI 10.1515/dna-2015-0010

Received February 26, 2015; accepted July 6, 2015

Abstract: DNA barcoding has been applied as a method to test seafood authenticity in numerous market surveys. This trend is continuing to gain momentum as DNA barcoding is employed as a regulatory tool, by the media, and by students to test seafood products, in addition to its use by scientific researchers to monitor seafood substitution. However, as market surveys documenting mislabeling continue to be published by both the press and scientific journals, there is a need for standardization in practices to aid in comparing and verifying results. This communication provides an overview of best practices for conducting and reporting DNA barcoding market studies for seafood. These standards can also be used as a guideline for other methods for conducting market surveys, or for market surveys employing DNA barcoding of other groups of organisms.

Keywords: DNA barcoding, seafood mislabeling, standardized methods

1 Introduction

Accurate identification of seafood is of global concern to consumers, businesses and regulators. Consumption of fish and other seafood products continues to rise [1], and globalization of the industry increases the complexity of the supply chain, allowing opportunities for species substitution to occur. Incidents of seafood mislabeling, specifically species substitutions, have been documented since the early 20th century [2]. In many cases, there is significant economic impact from seafood mislabeling as well as impacts to human health and to conservation efforts [3]. Improved technology has led to increased market testing to uncover substitution. Since 2006, there has been a major increase in published market surveys, largely due to the availability of simple and cost-effective DNA-based analyses [4].

DNA barcoding has been one of the most common and useful tools in conducting seafood market surveys and continues to be popular amongst the scientific community, in addition to its adoption as a regulatory tool by the FDA [5]. The efforts of FISH-BOL [6] have resulted in publically accessible DNA barcode sequences for over ten thousand fish species on the Barcode of Life Data System (BOLD) [7]. Although coverage of fish species is more complete than some other seafood types, more species continue to be added through other barcoding campaigns, like the Marine Barcode of Life, and BOLD continues to be populated with other seafood species [e.g. 8]. Notably, the availability of barcode reference sequences has allowed for the expansion of this method from the scientific community to the public with media outlets [e.g. 9,10] and high school students [e.g. 11,12] employing DNA barcoding for consumer studies to uncover seafood mislabeling. As consumer interest and knowledge about the topic of seafood authenticity increase, the trend of community engagement in seafood market surveys is likely to increase as well.

From a regulatory perspective, this increase of consumer awareness and participation is likely to impact policies in a positive way [13]. However, going forward, it will be imperative to provide guidelines for standardization of these studies so that the data generated can be aggregated, verified and compared more easily.

***Corresponding author: Amanda Madelaine Naaum,** University of Guelph, Biodiversity Institute of Ontario, 50 Stone Road E., Guelph, ON, Canada, Email: anaaum@uoguelph.ca
Robert Hanner, University of Guelph, Biodiversity Institute of Ontario, 50 Stone Road E., Guelph, ON, Canada
Jason St.Jaques , University of Guelph, School of Computer Science, 50 Stone Road E., Guelph, ON, Canada
Kimberly Warner, Oceana, 1350 Connecticut Ave., NW, 5th Floor, Washington, DC, 20036, USA
Linda Santschi, Ralph Imondi, Coastal Marine Biolabs, 1559 Spinnaker Drive, Ventura, CA, 93001, USA

The following perspective details standard methods for conducting a market survey, from developing a project to reporting of results that will help streamline the data from market surveys and help ensure their continued utility in monitoring the occurrence of seafood substitution. These apply not only to scientific articles, but also media articles and reports from consumer-led studies, and provide a framework for both conducting surveys and for reviewers tasked with considering articles for journal submission. While these guidelines are optimized for DNA barcoding, the protocol should prove helpful for standardizing methods using other genetic techniques for species authentication.

2 Market Survey Guidelines

2.1 General Information

The title of the survey should be simple, but also allow it to be differentiated from similar studies in the field of seafood substitution. It should reflect that the study represents a follow-up to an existing study if applicable. The name(s) and contact information for authors should be provided as well as the name of the organization that conducted the study. Any short forms or acronyms should be clearly defined. The location of the organization conducting the study should be provided, including address, city and country. The study start and end dates should be provided, and differentiate both the time period for collection and subsequent analysis of all samples.

It is important to understand the local and/or national regulatory framework for seafood labeling in the design of such studies. This includes all information legally required to be associated with sale of seafood in different retail types, such as small markets, large grocery stores and restaurants and for seafood with different levels of processing. These labeling regulations and their exceptions or exclusions (e.g. for heavily processed seafood) will help guide the selection of types of products targeted. These regulations should be clearly stated and referenced when discussing results. It is also important to understand the limitations of the analytical methods used when selecting the types of products to be targeted. For example, studies using DNA barcoding may not be able to identify species within a mixed sample if multiple species are present. There may also be limitations depending on the processing methods. Canning may degrade DNA to the point where standard DNA barcoding practices cannot be used. However, a shorter segment of the barcode region can be targeted using other approaches [e.g. 14]. Any modifications or additions to standard practices to be used in the case of failed samples should be clearly outlined in the study report, as should the specific methods used.

Cities and countries from where samples were collected should be included. This will impact the regulatory framework that should be used to identify mislabeling, and this should be determined before any analysis. Market names targeted by the study should be identified clearly as should the types of products purchased. For example, if the study only targets smoked salmon this should be indicated. If conducting a large-scale survey, particularly when many individuals will be collecting samples (e.g. a high school class), a sampling plan should be in place beforehand that includes which location(s) to visit and which market name(s) will be targeted by each individual participant. The type of location targeted should be part of this plan. Grocery stores, restaurants and fish markets may be common targets. This can be further subdivided, for example restaurants can be separated into groups by relative cost per meal. Any metadata (e.g. photographs) to be collected and their appropriate format should be decided upfront. Dated photographs of labels on packages or seafood displays and menus are very helpful in documenting and recalling how products purchased were labeled when sold. This information does not necessarily have to be in the report, but will improve survey design and recall. Including a standardized collection document (e.g. Supplementary Table 1), or the use of a mobile app (e.g. DNA Barcoding Assistant, see below) can help with standardization of information collected.

2.2 Sample Information

Depending on market names and types of products targeted, market samples collected may originate from a single individual or multiple individuals. Multi-individual specimens can include food items such as a mixed-maki sushi roll, a prepared fish ball, canned fish, surimi, or any other item that may contain more than one individual fish. It is important to designate between the two, and where possible identify the actual species to be examined. For example, artificial crab is almost always the by-product of several species of fish, commonly some variety of pollock. In these situations, the species listed in the ingredients list should be identified for comparison, not the presence of “crab” as it is accepted that the product is not crab. However, if the product is sold as crab (e.g. if the menu lists crab, but artificial crab is found on the premises instead), then the species for comparison to determine mislabeling would be crab. For cases of multi-species specimens where there is a distinction between

pieces, the items should be analyzed separately. For sushi that may contain roe on top of a portion, the roe should be treated separately from the other seafood portion and given a separate sample ID number. Specific care should be taken with regards to contamination between portions of the same sample if these will be separately analyzed.

A unique sample ID should be given to each sample tested. This ID should be unique for each sample and follow a consistent format throughout the study, and any subsequent follow-up studies. The collection date should be provided for each sample. The location where the sample was collected should be provided including: name and address, GPS coordinates if available (identifying information may be withheld for confidentiality reasons if necessary), and location type (e.g. restaurant). The market name from the label should be recorded and reported for each sample and should match exactly as found on the sample. Retention of sales receipts with sample records is recommended. The means by which the label was affixed or the name communicated should also be reported. (e.g. menu item, label on packaging, verbally told by employee, etc.). Details on the type of sample should include the processing level of sample (e.g. whole, whole – head off, fillet, canned, prepared meal, sashimi, smoked, dried, etc.), including whether the sample was frozen, previously frozen or never frozen, how and where it was caught and if labeled as farmed or wild. It should also include whether the sample is mixed or not (for those samples where the mixture cannot be separated and separately analyzed; e.g. fish stick, artificial crab, etc.).

Metadata can be included, but should appear in separate columns from other data. While not required, information such as sample price and photographs of specimens and/or labels can be useful in a study report. This information should be included in the online database where the sequences are deposited, if possible. Photographs should be labeled with the sample ID and a description of the photo. If price is reported, the unit of measurement should be included (e.g. per kg, portion, etc.). Including this information can aid in adding validity to market studies and may help in potential regulatory use of results: therefore it is strongly recommended that it be collected. However, if photographs reveal vendor information, these records may be kept confidential in addition to location information.

2.3 Sample Analysis

The full chain of custody between sample collection and final analysis should be detailed from collection to final results. Methods of sample collection, preservation and shipment should be provided. The lab where sequencing was conducted should be named, together with its location. If a separate lab processed the samples before sequencing, this should be indicated, and information provided accordingly. The sequence returned from the DNA sequencing lab should be deposited in an online repository. Ideally BOLD should be used. Specifically, projects should be submitted under the "Barcoding Applications Campaign" so they will not be included in the reference database. Campaign selection is made when setting up a BOLD project, and full details on project submission can be found in the BOLD Handbook under "Resources" on the BOLD website (www.boldsystems.org). Using BOLD as a repository for sequences streamlines the process of using the reference database to identify samples. It also allows for easy addition of metadata and raw trace files to assist with external review of any results.

Because the true identity of a market sample is an hypothesis to be tested using barcoding, only minimal information on taxonomy for the sample should be entered. For example, fish should be entered only as Chordata, since the taxonomic identification of market samples is unknown. Names from market labels can be entered in the "Extra Info" field of BOLD. If market samples are included in a BOLD project outside of the Applications Campaign, additional information may result in erroneous information infiltrating the BOLD ID Engine, leading to incorrect or ambiguous identification for other users conducting searches to match sequences. Only taxonomically verified identifications should be entered for samples when creating a BOLD project.

Detailed methodology of all analyses should be provided. This should include sequences of primers used, protocols for PCR cleanup if any, protocols for DNA extraction, protocols for re-analysis of failed samples, thermocycling conditions, and equipment used for sequencing and nucleic acid quantification. Ideally nucleic acid concentration/quality should be assessed and reported. Reporting of the target region is particularly important as "DNA barcoding" is sometimes used informally to refer to studies where the COI-5P region is not used, and some studies may target the COI-5P region without mentioning DNA barcoding specifically. The use of the term DNA barcoding for animal identification should only be used for studies employing the standard DNA barcoding region as outlined by Hebert *et al.* [15].

The inferred scientific name as reported by the sequencing lab, or as identified using a search tool, should be reported. The search tool and any associated parameters used should be listed (e.g. BOLD, species ID database). Particular attention should be paid to spelling

and formatting of scientific names to ensure ease of data recovery and analysis by others. The total number of samples analyzed should be reported, including any sequencing failures, and any failures to find an identity match in BOLD (or other database). The criteria for a match (e.g. divergence threshold, monophyly) should be explicitly mentioned. If other studies are used to determine this, they should be cited. The regulatory framework used for identification of mislabeled sample should be provided (e.g. FDA Seafood List) and a reference provided. If an official regulatory list is not available, an explanation should be provided as to how samples were identified as mislabeled.

2.4 Educational Projects

Growing interest in DNA barcoding as a pedagogical tool to improve life science teaching and learning has motivated the involvement of high school and undergraduate students in a spectrum of DNA barcoding projects ranging from market surveys to the creation of parameterized barcode libraries for commercially and ecologically important fish species [e.g. 12,16]. Given the rising popularity of DNA barcoding as an instructional tool, nowhere is the need for standardization more apparent than in traditional learning settings, where inexperienced participants operating with limited or no guidance from the scientific community are likely to overlook important practical considerations associated with project design and key issues related to appropriate data collection, data management, and data reporting practices. This is especially problematic when educational users deposit project-related information in publicly accessible online repositories without clearly articulated guidelines and supervision from the scientific community. In addition to submitting data that can inadvertently lead to erroneous search results for other users, student users may naively omit certain forms of crucial information and thereby diminish the current and future value of an otherwise important project.

It has been demonstrated that with the appropriate scientific oversight and resources, high school and undergraduate students are capable of collecting, generating, and sharing publication quality DNA barcode data in connection with market surveys and related efforts and that this can be a valuable learning tool [16,17]. To encourage a level of student engagement in DNA barcoding that simultaneously serves scientific, educational, and potentially regulatory interests, two key forms of technology have been developed, and are accessible through the Education and Barcode of Life (eBOL) Community Web Portal (http://www.educationandbarcoding.org/). The first consists of a free utility application, DNA Barcoding Assistant, for iPhone and Android smartphones that helps streamline and standardize the collection of sample information by students. The second consists of a customized, classroom-focused interface to the BOLD researcher workbench and data repository, the BOLD Student Data Portal (BOLD-SDP) [16]. In addition to simplifying the assembly of DNA barcode records that aggregate the various forms of sample information, sample analysis details, and other data elements described here, BOLD-SDP was designed with a hierarchical data validation system. This system engages educators, scientists, and professional data managers in a sequential, 3-tier vetting process that ensures student compliance with established data standards and streamlines the publication of student barcode data in BOLD and potentially other online repositories. BOLD-SDP was designed with additional features that enable educators to conveniently manage and monitor the activities of project participants. It also contains an integrated suite of online data analysis, data visualization, and sequence editing tools that aid the data validation process. In light of these important considerations and their implications for effective project management in challenging educational contexts and settings, we advocate the use of BOLD-SDP as an additional guideline for educators, students and citizen scientists seeking to organize, manage, and share their project data with the broader DNA barcoding community.

3 Conclusion

In general, adherence to the above guidelines will allow better standardization and facilitate ease of interpretation and future use of data generated from DNA barcoding market surveys. As this tool continues to be widely accessed for seafood market surveys by both the scientific community and the public, this type of standardization will streamline collection and analysis of the increasing volume of data. Clear guidelines may also help with community engagement, thereby increasing public knowledge about the important issues related to seafood mislabeling. Standardized methods and careful collection and reporting of information may also help increase the possibility for regulatory use of data collected.

Conflict of interest: Dr Naaum has nothing to disclose.

References

[1] United Nations Food and Agriculture Organization (FAO). The state of world fisheries and aquaculture, 2014, http://www.fao.org/3/a-i3720e.pdf

[2] Anonymous. Shark Meat in Market, New York Times, 10 September 1915, http://query.nytimes.com/mem/archive-free/pdf?res=9A03E7DF1239E333A25753C1A96F9C946496D6CF

[3] Warner K., Timme W., Lowell B., Hirshfield M., Oceana Study Reveals Seafood Fraud Nationwide, Oceana, 2013, http://usa.oceana.org/sites/default/files/reports/National_Seafood_Fraud_Testing_Results_FINAL.pdf

[4] Golden R. E., Warner K., The Global Reach of Seafood Fraud: a Current Review of the Literature, Oceana, 2014, https://s3.amazonaws.com/s3.oceana.org/images/Seafood_Fraud_Map_White_paper_new.pdf

[5] Handy S.M., Deeds J.R., Ivanova N. V., Hebert P.D.N., Hanner R. H. Ormos S., Weigt *et al.*, Single-Laboratory Validated Method for the Generation of DNA Barcodes for the Identification of Fish for Regulatory Compliance, J. AOAC Int., 2011, 94, 201-210

[6] Ward R. D., Hanner R., Hebert P.D.N., The campaign to DNA barcode all fishes, FISH-BOL. J. Fish Biol., 2009, 74, 329–356

[7] Ratnasingham S., Hebert P.D.N., BOLD: The Barcode of Life Data System (www.barcodinglife.org), Mol. Ecol. Notes., 2007, 7, 355-364

[8] Lobo J., Costa P.M., Teixeira M.A.L., Ferreira M.S.G., Costa M.H., Costa F.O, Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans, BMC Ecology, 2013, 13, doi:10.1186/1472-6785-13-34

[9] Abelson J., Daley B., On the menu, but not on your plate, The Boston Globe, 23 October, 2011, http://www.bostonglobe.com/business/2011/10/22/menu-but-not-your-plate/NDbXGXd-PR6O37mXRSVPGlL/story.html

[10] Hanner R.H., Becker S., Ivanova N.V., Steinke D., FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada, Mitochond. DNA, 2011, 22, 106-122

[11] Stoeckle K. & Strauss L., 2008, High school students track down fish fraud. From phe.rockefeller.edu/docs/**pacific-fishing**sept2008.pdf

[12] Naaum A.M., Hanner R., Community engagement in seafood identification using DNA barcoding reveals market substitution in Canadian seafood, DNA Barcodes, (in press)

[13] Mariani S., Ellis, J., O'Reilly A., Brechon A.L., Sacchi C., Miller D.D., Mass media influence and the regulation of illegal practices in the seafood market, Conserv. Lett., 2014, doi: 10.1111/conl.12085

[14] Rasmussen R.S., Hellberg M.T., Hanner R.H., A Multiplex PCR Method for the Identification of Commercially Important Salmon and Trout Species (*Oncorhynchus* and *Salmo*) in North America, J. Food Sci., 2010, 75, c595-c606

[15] Hebert P.D.N., Cywinska,A., Call S.L., DeWaard J.R., Biological identifications through DNA barcodes, Proc. R. Soc. of Lond. B., 2003, 270, 313-321

[16] Santschi L., Hanner R.H., Ratnasingham S., Riconscente M., Imondi R., Barcoding Life's Matrix: translating biodiversity genomics into high school settings to enhance life science education, PLoS Biology, 2013, 11, e1001471. doi:10.1371/journal.pbio.1001471

[17] Naaum A.M., Frewin A., Hanner, R., DNA Barcoding as an Educational Tool: Case Studies in Insect Biodiversity and Seafood Identification, Teach. Learn. Innov., 2014, 16, journal.lib.uoguelph.ca/index.php/tli/article/view/2790

Supplemental Material: The online version of this article (DOI: 10.1515/dna-2015-0010) offers supplementary material.