scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Biodiversity of Philippine marine fishes: A DNA barcode reference library based on voucher specimens

Katherine E. Bemis[1,2,8]✉, Matthew G. Girard[2,3,8]✉, Mudjekeewis D. Santos[4], Kent E. Carpenter[5], Jonathan R. Deeds[6], Diane E. Pitassy[2], Nicko Amor L. Flores[4], Elizabeth S. Hunter[6], Amy C. Driskell[7], Kenneth S. Macdonald III[7], Lee A. Weigt[7] & Jeffrey T. Williams[2]✉

Accurate identification of fishes is essential for understanding their biology and to ensure food safety for consumers. DNA barcoding is an important tool because it can verify identifications of both whole and processed fishes that have had key morphological characters removed (e.g., filets, fish meal); however, DNA reference libraries are incomplete, and public repositories for sequence data contain incorrectly identified sequences. During a nine-year sampling program in the Philippines, a global biodiversity hotspot for marine fishes, we developed a verified reference library of cytochrome c oxidase subunit I (COI) sequences for 2,525 specimens representing 984 species. Specimens were primarily purchased from markets, with additional diversity collected using rotenone or fishing gear. Species identifications were verified based on taxonomic, phenotypic, and genotypic data, and sequences are associated with voucher specimens, live-color photographs, and genetic samples catalogued at Smithsonian Institution, National Museum of Natural History. The Biodiversity of Philippine Marine Fishes dataset is released herein to increase knowledge of species diversity and distributions and to facilitate accurate identification of market fishes.

Background & Summary

In 2007, the United States Senate requested the Government Accountability Office (GAO[1]) investigate several federal agencies responsible for seafood and determine actions necessary to detect and prevent seafood fraud. GAO (2009) found that Americans consumed almost 5 billion pounds of seafood annually, more than 80% of which was imported, and determined that current practices were not sufficient to assure consumers that the seafood they purchased was correctly labeled. The Presidential Task Force on Combating Illegal, Unreported, and Unregulated (IUU) Fishing and Seafood Fraud identified products like grouper and snapper as particularly at risk because the same species are marketed under different names in different countries, and extensive processing, common for imported products, removes diagnostic characters required for morphological species identification[2]. As part of its summary report, GAO[1] recommended "a federal agency wide library of seafood species standards" be developed. In response, the United States Food and Drug Administration (FDA), which is responsible for the identification of seafood distributed through interstate commerce, established a collaboration with the Smithsonian Institution's National Museum of Natural History (NMNH), Division of Fishes and Laboratories of Analytical Biology (LAB) to develop a DNA-barcode reference library for commercial seafood as a regulatory tool for species identification[3–5].

[1]National Systematics Laboratory, Office of Science and Technology, NOAA Fisheries, Washington, D.C., 20560, USA. [2]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, D.C., 20560, USA. [3]Biodiversity Institute, University of Kansas, Lawrence, Kansas, 66045, USA. [4]Genetic Fingerprinting Laboratory, National Fisheries Research and Development Institute, Quezon City, 1103, Philippines. [5]Department of Biological Sciences, Old Dominion University, Norfolk, Virginia, 23529, USA. [6]Center for Food Safety and Applied Nutrition, United States Food and Drug Administration, College Park, Maryland, 20740, USA. [7]Laboratories of Analytical Biology, National Museum of Natural History, Smithsonian Institution, Washington, D.C., 20560, USA. [8]These authors contributed equally: Katherine E. Bemis, Matthew G. Girard. ✉e-mail: bemisk@si.edu; girardmg@si.edu; jtwfish@gmail.com

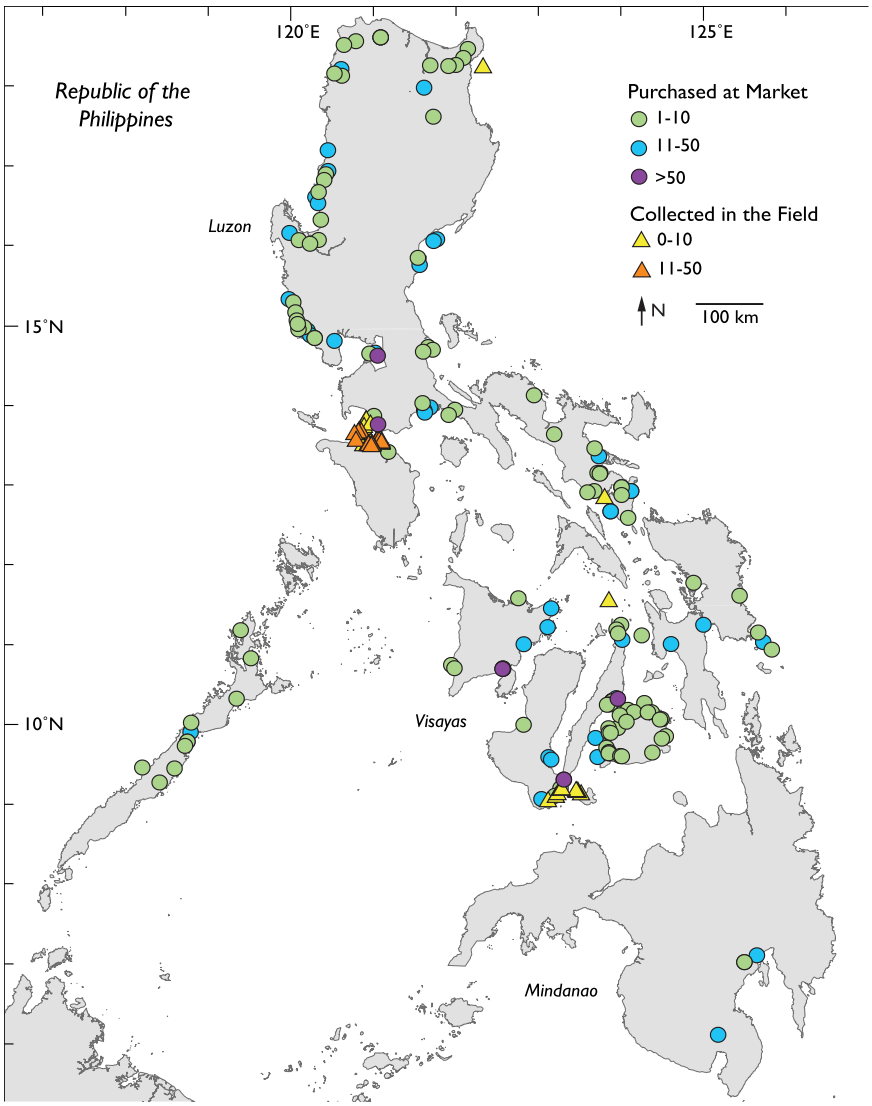

Fig. 1 Map of sampling localities. For fishes purchased from markets, the market location is shown; vendors stated that fishes were captured from local coastal waters in the vicinity of the markets. For fishes collected in the field using rotenone while snorkeling, SCUBA diving, or using fishing gear, points on the map represent precise collection localities.

During the past 30 years, genetic sequencing has improved the identification of global biodiversity. Commonly known as DNA barcoding, sequencing one or more loci and comparing these data to those in reference libraries is an efficient way to identify or verify species. DNA barcoding has been used to identify species diversity of a region (e.g.[6–9]), describe new species (e.g.[10]), link different ontogenetic stages (e.g.[11,12]), detect illegally traded wildlife (e.g.[13]), and confirm the identity of animals sold in markets (e.g.[14]). The FDA now uses DNA barcoding as its primary tool for species identification of seafood products and has used this technique successfully both in seafood-related-illness investigations[15,16] and in cases of species substitution for economic gain[17].

Success of species-level identification using barcodes depends on completeness and accuracy of DNA reference libraries[8,18,19]. Public sequence repositories such as GenBank (https://www.ncbi.nlm.nih.gov) and the Barcode of Life Database (BOLD; https://www.boldsystems.org/) are essential resources for DNA barcoding, yet they contain misidentified sequences (e.g.[20–23]). Misidentifications can happen because some species are difficult to identify morphologically or because of errors in data management. In other cases, sequence data does not align with currently accepted morphological identifications for legitimate reasons including cryptic diversity, mixed genealogies between sister species, incomplete lineage sorting, and introgressive hybridization[8]. Incongruent sequences that lack associated voucher specimens, have incomplete collection data, or lack protocols for revising identifications can make problematic sequences in public repositories difficult to resolve. Thus, reference libraries greatly increase in scientific and practical value when sequences are linked to cataloged voucher specimens in natural history museums that can be reexamined to verify or revise identifications.

With more than 7,600 islands and 36,000 kilometers of coastline in the Coral Triangle, the Philippines is the epicenter of marine shorefish biodiversity and home to more than 2,600 species of marine fishes[24–26]. Shorefish

	Purchased from market	Collected in field by rotenone or fishing gear	Combined datasets
Number of specimens	1,956 (77% of samples)	569 (23% of samples)	2,525
Number of species	760	340	984
Top three families for species diversity	Labridae, Epinephelidae, Lutjanidae	Gobiidae, Labridae, Pomacentridae	Labridae, Gobiidae, Epinephelidae
Number of newly sequenced species	22 (40%)	33 (60%)	55
Number of newly sequenced species for COI	9 (31%)	20 (69%)	29

Table 1. Overview of 2,525 verified specimens in the dataset highlighting that different collecting methods are important for completing reference libraries. Most, 760 species, were purchased from markets, 340 species were collected in the field, and 116 species were collected both from the market and field. The majority of the newly sequenced (60%) and newly sequenced for COI species (69%) were collected in the field, likely because market collections emphasized commercial species with large maximum sizes, whereas less well-known cryptobenthic fishes were more often collected in the field.

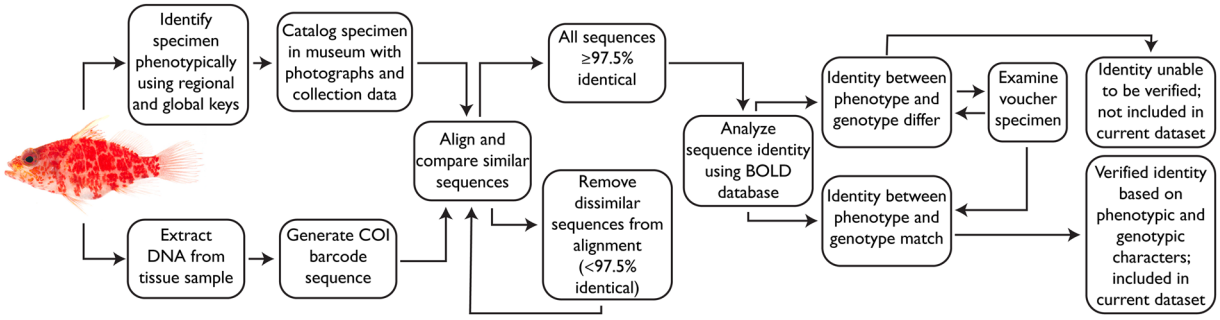

Fig. 2 Methods for sequence identification and verification. Photograph of *Plectranthias inermis* USNM 431978, 33 mm SL.

biodiversity is a critical resource to Filipinos, with the Philippines ranking eleventh in global marine capture production[27] and Philippine fish markets are some of the most species diverse markets in the world[10]. Roughly 70% of Filipino nutritional protein comes from fishes and more than 1.6 million people depend on the fishing industry for their livelihood[28]. Not all fishes caught are sold within the country; the Philippines exported $46 million in fish filets and other fish meat in 2021 and 21% of that was imported by the United States[29]. Given the remarkable biodiversity of the region and the high species diversity sold in markets, accurate identification of seafood is essential to food safety and management for both the Philippines and the countries that import fish products from the Philippines. To inventory species sold in Philippine fish markets and additional diversity of the region, the Bureau of Fisheries and Aquatic Resources-National Fisheries Research and Development Institute (BFAR-NFRDI), Department of Agriculture, Philippines, and the NMNH developed a collaboration in 2011 to generate a genetic reference library based on voucher specimens to quickly and efficiently identify fishes, regardless of whether they are whole or processed (e.g., filets, fishmeal).

Our dataset, Biodiversity of Philippine Marine Fishes, is the result of nine years (2011–2019) of sampling (Fig. 1) and includes 2,525 specimens representing 984 species. Seventy-seven percent of specimens in our dataset were purchased from fish markets; the remaining 23% were collected using rotenone or fishing gear to capture additional diversity of the region (Table 1). Specimens were sequenced for a ~655 base pair portion of the mitochondrial cytochrome c oxidase I locus (COI) to develop a barcode reference library linked to voucher specimens (Fig. 2). The dataset represents the most comprehensive DNA reference library of Philippine marine fishes, and includes vouchered museum specimens, collection data, live color photographs (e.g., Figs. 3, 4), and genetic samples for future analyses. Species identifications were verified based on taxonomic, phenotypic, and genotypic data (see Methods and Technical Validation sections; Fig. 2). Among the 135 families of fishes included in the verified dataset, the families Labridae (78 species), Gobiidae (69 species), Epinephelidae (53 species), and Pomacentridae (53 species) have the greatest species diversity (Fig. 5a). The dataset includes sequences for 55 species in 25 families that were not previously publicly available on GenBank or BOLD for any loci, and an additional 29 species in 19 families that represent the first publicly available COI barcode sequence for the taxon (as of September 28, 2022, Table 1, Fig. 5b; see Verified specimen records deposited at FigShare[30]). The Biodiversity of Philippine Marine Fishes dataset represents ~50% of the estimated Philippine market fish diversity[10] and will serve as a foundational checklist and reference library to improve knowledge of species diversity and distributions, to identify and describe new species, and to confirm the identity of commercially caught fishes in this global biodiversity hotspot.

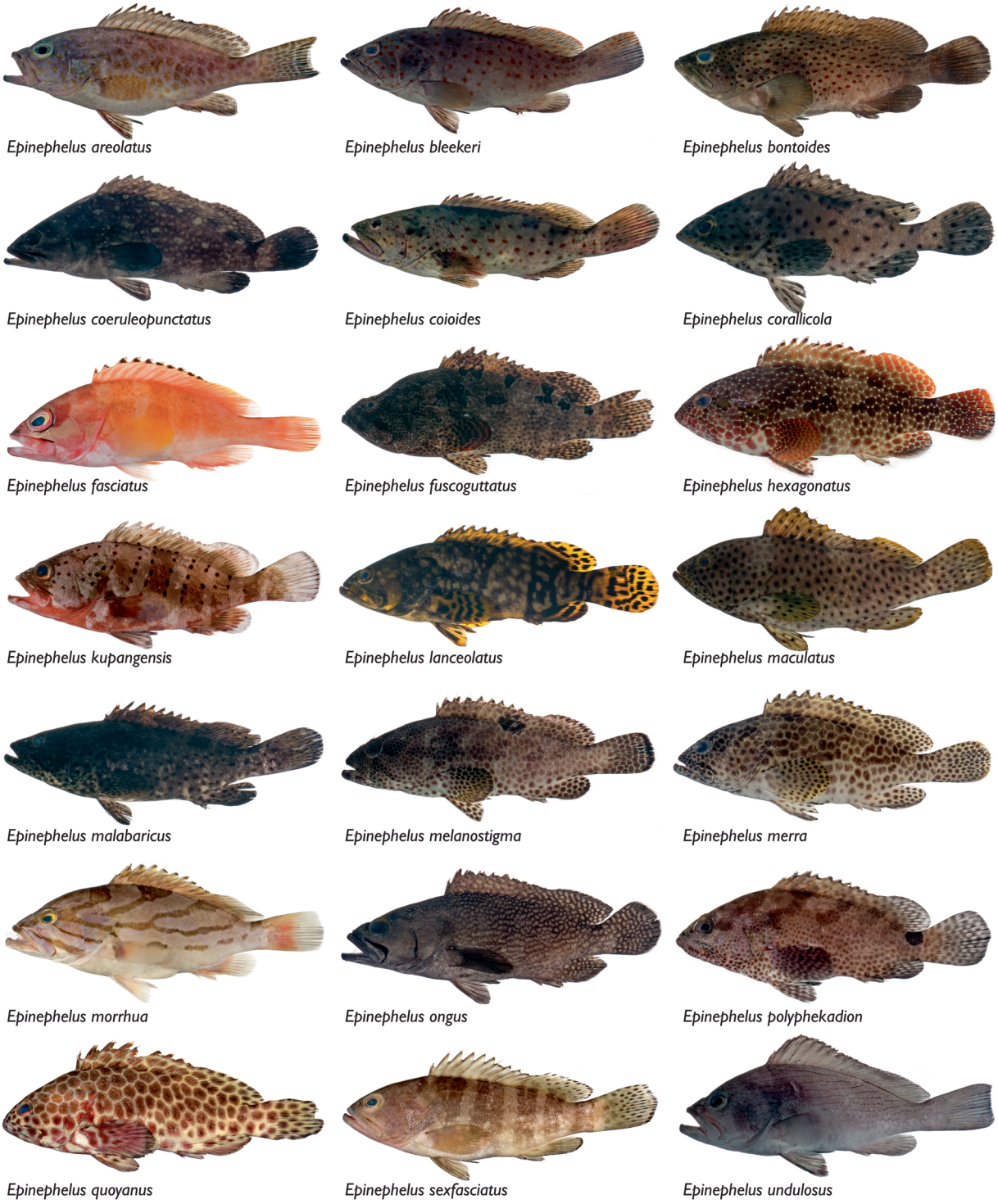

Fig. 3 Photographs of groupers (*Epinephelus*), a challenging genus to identify and one that is frequently mislabeled. Live-color photographs, voucher specimens, and molecular sequence data enabled verified identifications of these specimens collected from Philippine fish markets. *Epinephelus areolatus* USNM 435700, 163 mm SL; *E. bleekeri* USNM 443235, 209 mm SL; *E. bontoides* USNM 403254, 160 mm SL; *E. coeruleopunctatus* USNM 431516, 223 mm SL; *E. coioides* USNM 435656, 237 mm SL; *E. corallicola* USNM 431624, 143 mm SL; *E. fasciatus* USNM 403060, 108 mm SL; *E. fuscoguttatus* USNM 431625, 261 mm SL; *E. hexagonatus* USNM 443283, 152 mm SL; *E. kupangensis* USNM 443558, 172 mm SL; *E. lanceolatus* USNM 431627, 204 mm SL; *E. maculatus* USNM 431623, 250 mm SL; *E. malabaricus* USNM 431611, 261 mm SL; *E. melanostigma* USNM 431622, 204 mm SL; *E. merra* USNM 443288, 144 mm SL; *E. morrhua* USNM 403199, 252 mm SL; *E. ongus* USNM 435567, 170 mm SL; *E. polyphekadion* USNM 423651, 154 mm SL; *E. quoyanus* USNM 445240, 215 mm SL; *E. sexfasciatus* USNM 445239, 151 mm SL; and *E. undulosus* USNM 431568, 181 mm SL.

Fig. 4 Photographs of snappers (*Lutjanus*) a commercially important genus; live-color photographs, voucher specimens, and molecular sequence data enabled verified identifications of these specimens collected from Philippine fish markets. *Lutjanus argentimaculatus* USNM 403082, 182 mm SL; *L. biguttatus* USNM 443245, 145 mm SL; *L. bohar* USNM 445453, 185 mm SL; *L. carponotatus* USNM 424827, 110 mm SL; *L. decussatus* USNM 445218, 170 mm SL; *L. dodecacanthoides* USNM 443545, 142 mm SL; *L. fulviflamma* USNM 403054, 134 mm SL; *L. fulvus* USNM 435687, 166 mm SL; *L. gibbus* USNM 403099, 223 mm SL; *L. johnii* USNM 423648, 139 mm SL; *L. lutjanus* USNM 403104, 127 mm SL; *L. monostigma* USNM 403313, 168 mm SL; *L. quinquelineatus* USNM 423610, 149 mm SL; *L. rivulatus* USNM 438092, 205 mm SL; *L. rufolineatus* USNM 445298, 148 mm SL; *L. russellii* USNM 423649, 144 mm SL; *L. sebae* USNM 403135, 256 mm SL; and *L. timoriensis* USNM 403119, 160 mm SL.

Methods

Specimen collection. Between 2011 and 2019, we collected and verified 2,525 specimens, representing 135 families, 445 genera, and 984 species of fishes (see Verified specimen records deposited at FigShare[30]). Of these, 1,956 (77%) were purchased from fish landings, roadside stalls, and municipal and city markets (155 localities; Fig. 1; see Verified specimen records deposited at FigShare[30]). The remaining 569 fishes included in the dataset were collected from near-shore habitats using rotenone while snorkeling or SCUBA diving, or using fishing gear (71 localities; Fig. 1; see Verified specimen records deposited at FigShare[30]. Specimens were photographed in the

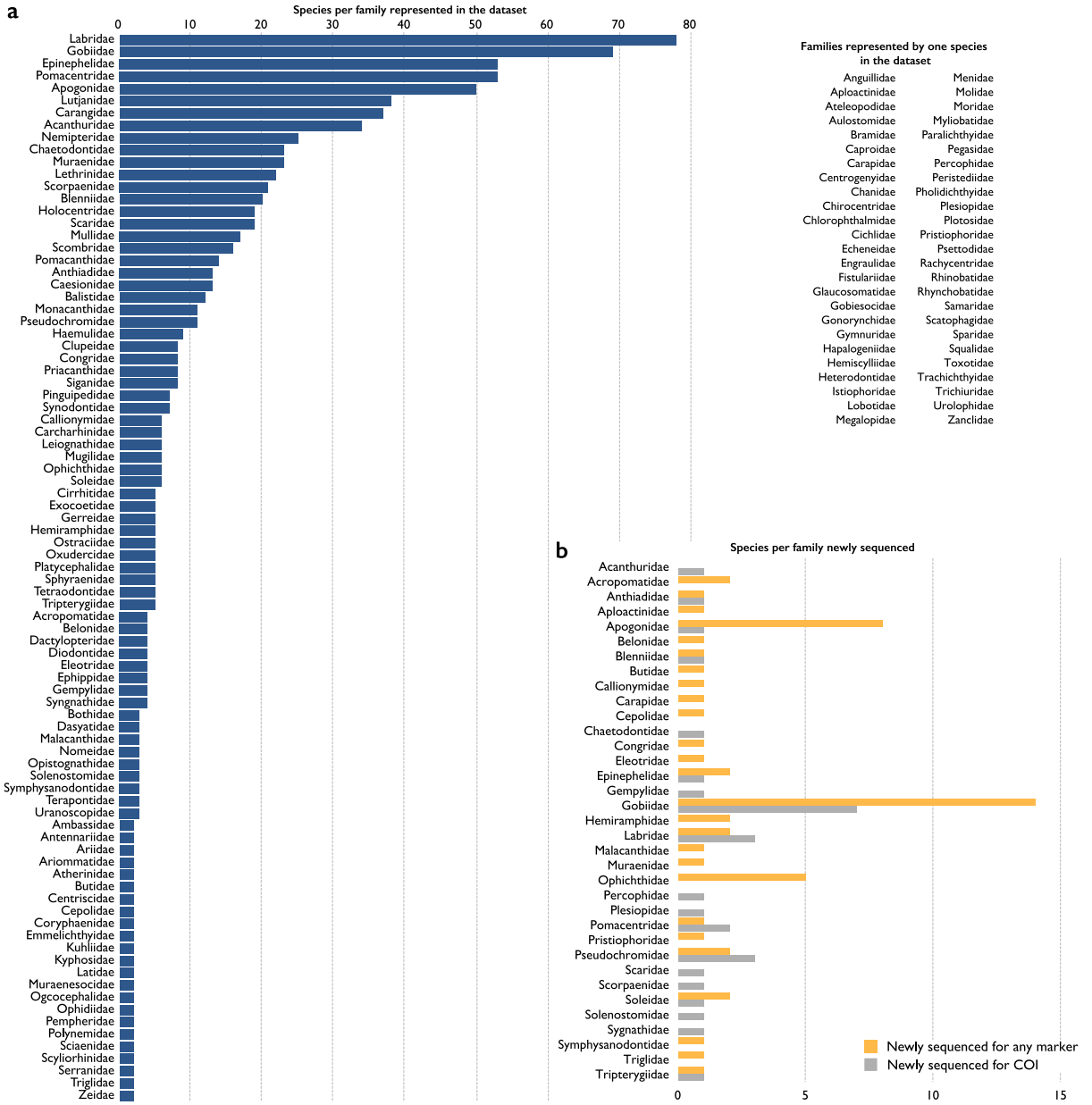

Fig. 5 Species diversity included in the Biodiversity of Philippine Marine Fishes dataset. (**a**) Number of species by family; families represented by a single species shown to the right of the graph. (**b**) Number of newly sequenced species within each family in the dataset. Gold bars represent species publicly available for the first time; gray bars represent species for which some genetic data are available publicly but are first publicly available herein for COI.

field to capture live colors by J. T. Williams using Fujifilm and Nikon camera bodies with 105 mm macro lenses under flash or LED daylight lighting. Tissue samples from each specimen were preserved in ethanol and M2 lysis buffer (AutoGen Inc.), and whole voucher specimens were fixed using 10% formalin. Voucher specimens were transferred to 75% EtOH for long-term storage and cataloged in the NMNH Fish Collection and associated tissues and extractions were archived in the NMNH Biorepository (see Verified specimen records deposited at FigShare[30]. Specimen lengths used herein are reported as standard length (SL).

Specimen identification and validation. Morphological identifications were made in the field on fresh specimens, and preserved specimens further examined at NMNH by J. T. Williams, K. E. Carpenter, K. E. Bemis, M. G. Girard, and D. E. Pitassy using global, regional, or taxon-specific keys (e.g.[25,31,32]). Voucher identities were verified using molecular characters by comparing newly generated sequences with those available through public repositories (e.g., GenBank, BOLD) as described in the Technical Validation section below and in Fig. 2. To obtain DNA sequence data, tissues were extracted using the AutoGenPrep 965 (AutoGen, Holliston, Massachusetts,

USA) following manufacturer protocols. We targeted the 655 base-pair (bp) barcode of COI following Weigt *et al.*[7] using the primers from Baldwin *et al.*[33]. Sequencing of PCR products was done on an Automated ABI 3730xl at the NMNH, with sequence trace files trimmed of low-quality ends and forward and reverse reads assembled into contigs using Sequencher 5.4 (Gene Codes). Resulting sequences were 515–686 bps in length (mean: 651.7, median: 655, mode: 655, standard deviation: 12.7). Sequence alignments were performed using MAFFT 7.475[34] within Geneious 11.1.5[35] (see Technical Validation section). Once sequences were confirmed (see Technical Validation section), taxonomic names were validated using either Eschmeyer's Catalog of Fishes[36] or FishBase[37] using the 'rfishbase' 4 module[38] within R 4.2.1.

Data Records

The verified COI sequence library for Biodiversity of Philippine Marine Fishes includes (1) voucher specimens (2), tissues samples and DNA extracts (3), voucher collection information (4), live-color photographs, and (5) COI sequences of at least 500 bp. All photographs, voucher catalog numbers, DNA sequences, and collection data are publicly available through FigShare[30]. Data is also available on BOLD[39], GenBank (BioProject PRJNA947503[40]), through the Fish Collection at the National Museum of Natural History Smithsonian Institution (https://collections.nmnh.si.edu/search/fishes/), and the FDA Reference Standard Sequence Library for Seafood Identification (RSSL; https://www.fda.gov/food/dna-based-seafood-identification/reference-standard-sequence-library-seafood-identification-rssl). The library follows the BARCODE data standard requirements[41,42] for (1) species name (2), voucher data (3), collection data (4), sequence length (5), PCR primers used to generate the amplicon, and (6) trace files.

Technical Validation

In addition to morphological identification, all vouchers were validated based on molecular characters using COI data from BOLD. The BOLD database was used for verification because it has more barcode sequences (when including those that are "private") than available on GenBank and because barcodes published on GenBank are actively extracted and included within the BOLD database. Sequences generated from voucher specimens were imported into Geneious and multiple-sequence alignments were generated using MAFFT for each morphologically identified taxon. All taxon-specific alignments were checked individually to ensure each represented a single taxon and to identify divergent sequences and associated vouchers. Sequences were considered to be the same taxon if sequence identity was $\geq$97.5%. Sequences with similarity $\leq$97.4% were aligned with additional sequences and vouchers were examined to determine a revised identification.

Once each multiple-sequence alignment represented a single taxon based on our criteria, a representative sequence was submitted to the BOLD SYSTEMS Identification Engine via the web portal and searched against "All Barcode Records on BOLD." The results, including the 101-terminal Neighbor-Joining (NJ) phylogeny, were examined to confirm identity of submitted sequences based on monophyletic groups. In total, 2,371 of the 2,525 (93.9%) sequences in our dataset were confirmed using publicly available sequences (see Verified specimen records deposited at FigShare[30]). For the remaining 154 sequences, representing 84 species, a matching COI barcode was not publicly available on GenBank or BOLD. These 154 sequences represent the first publicly available barcode for their identified taxon. Sequences in this category are either "Newly sequenced" or "Newly sequenced for COI" based on if they have been sequenced for one or more locus previously (Fig. 5b, see Verified specimen records deposited at FigShare[30]); determinations as of 28 September 2022).

Usage Notes

The Biodiversity of Philippine Marine Fishes dataset is freely available to use for DNA barcoding or metabarcoding surveys, specimen identification, or other purposes (see Data Records). Additional specimens were collected that could not be verified based on currently available taxonomic, phenotypic, and genotypic information. Although not included in the verified sequence dataset released herein, these additional voucher specimens all have associated collection metadata, live-color photographs, and archived genetic samples available for future analyses (see Unverified specimen records deposited at FigShare[30]). As our understanding of the taxonomy of fishes in the Philippines increases, additional sequences from these collections will be verified and incorporated into the GenBank and BOLD projects.

Code availability

No custom code was used.

Received: 8 February 2023; Accepted: 12 June 2023;
Published online: 24 June 2023

References

1. GAO. Seafood fraud: FDA program changes and better collaboration among key federal agencies could improve detection and prevention. GAO-09-258. United States Government Accountability Office, Washington, D.C., USA. (2009).
2. NOAA. Presidential Task Force on Combating Illegal Unreported and Unregulated (IUU) Fishing and Seafood Fraud Action Plan. *Federal Register* **80**, 66867–66879 (2015).
3. Handy, S. M. *et al.* A single laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *J. AOAC Int.* **94**, 201–210 (2011).
4. Deeds, J. R. *et al.* Protocol for building a reference standard sequence library for DNA-based seafood identification. *J. AOAC Int.* **97**, 1–8 (2014).
5. Eischeid, A. C., Stadig, S. R., Handy, S. M., Fry, F. S. & Deeds, J. R. Optimization and evaluation of a method for the generation of DNA barcodes for the identification of crustaceans. *LWT Food Sci. Tech.* **73**, 357–367 (2016).

6. Ribeiro, A. *et al*. DNA barcodes identify marine fishes of São Paulo State, Brazil. *Mol. Ecol. Res.* **12**, 1012–1020 (2012).
7. Weigt, L. A., Driskell, A. C., Baldwin, C. C. & Ormos, A. DNA barcoding fishes. pp. 109–126 in Kress, W. J. & Erickson, D. L. (Eds.), *DNA barcodes: Methods and protocols*. (Humana Press, 2012).
8. Delrieu-Trottin, E. *et al*. A DNA barcode reference library of French Polynesian shore fishes. *Sci. Data* **6**, 114 (2019).
9. Amir, S. A. *et al*. Deeper in the blues: DNA barcoding of fishes from Pakistani coast of the Arabian Sea reveals overlooked genetic diversity. *Mar. Biodiv.* **52**, 1–22 (2022).
10. Carpenter, K. E., Williams, J. T. & Santos, M. D. *Acanthurus albimento*, a new species of surgeonfish (Acanthuriformes: Acanthuridae) from northeastern Luzon, Philippines, with comments on zoogeography. *J. Ocean Sci. Found.* **25**, 33–46 (2017).
11. Nonaka, A., Milisen, J. M., Mundy, B. C. & Johnson, G. D. Blackwater diving: an exciting window into the planktonic arena and its potential to enhance the quality of larval fish collections. *Ichthyol. Herpetol.* **109**, 138–156 (2021).
12. Girard, M. G. *et al*. Cusk-eel confusion: revisions of larval *Luciobrotula* and *Pycnocraspedum* and re-descriptions of two bythitid larvae (Ophidiiformes). *Ichthyol. Res*. (2023).
13. Poniente, J. A., Pereda, J. M. R., Dela Peña, J. T., Ventolero, M. F. H. & Santos, M. D. Mitochondrial DNA-based species testing of confiscated aquatic wildlife in the Philippines. *Forensic Sci. Internat. Anim. Environ.* **2**, 100051 (2022).
14. Maralit, B. A. *et al*. Detection of mislabeled commercial fishery by-products in the Philippines using DNA barcodes and its implications to food traceability and safety. *Food Control* **13**, 119–125 (2013).
15. Cohen, N. J. *et al*. Public health response to puffer fish (tetrodotoxin) poisoning. *J. Food Protect.* **72**, 810–817 (2009).
16. Cole, J. B., Heegaard, W. G., Deeds, J. R., McGrath, S. C. & Handy, S. H. Tetrodotoxin poisoning from imported dried puffer fish – Minneapolis, Minnesota, 2014. *Morb. Mortal. Wkly. Rep.* **63**, 1222–1225 (2015).
17. Windsor, A. M., Moore, K., Warner, K. A., Stadig, S. R. & Deeds, J. R. Evaluation of variation within the barcode region of Cytochrome c Oxidase I (COI) for the detection of commercial *Callinectes sapidus* Rathbun, 1896 (Blue Crab) products of non-US origin. *PeerJ* **7**, e7827 (2019).
18. Silva, A. J. & Hellberg, R. S. DNA-based techniques for seafood species authentication. *Advan. Food Nutri. Res.* **95**, 207–255 (2021).
19. Hoban, M. L. *et al*. Skimming for barcodes: rapid production of mitochondrial genome and nuclear ribosomal repeat reference markers through shallow shotgun sequencing. *PeerJ* **10**, e13790 (2022).
20. Phillips, M. J., Westerman, M. & Cascini, M. The value of updating GenBank accessions for supermatrix phylogeny: The case of the New Guinean marsupial carnivore genus. *Myoictis. Mol. Phylogen. Evol.* **166**, 107328 (2021).
21. Tang, K. L., Stiassny, M. L. J., Mayden, R. L. & DeSalle, R. Systematics of damselfishes. *Ichthyol. Herpetol.* **109**, 258–318 (2021).
22. Girard, M. G. *et al*. Molecular phylogeny of the threadfin fishes (Polynemidae) using ultraconserved elements. *J. Fish Biol.* **100**, 793–810 (2022).
23. Mulcahy, D. G. *et al*. DNA barcoding of the National Museum of Natural History reptile tissue holdings raises concerns about the use of natural history collections and the responsibilities of scientists in the molecular age. *PLoS ONE* **17**, e0264930 (2022).
24. Carpenter, K. E. & Springer, V. G. The center of the center of marine shore fish biodiversity: the Philippine Islands. *Environ. Biol. Fishes* **72**, 467–480 (2005).
25. Allen, G. R. & Erdmann, M.V. *Reef fishes of the East Indies, Volumes I-III* (University of Hawai'i Press, Tropical Reef Research, 2012).
26. Sanciangco, J. C., Carpenter, K. E., Etnoyer, P. J. & Moretzsohn, F. Habitat availability and heterogeneity and the Indo-Pacific warm pool as predictors of marine species richness in the tropical Indo-Pacific. *PLoS One* **8**, e56245 (2013).
27. FAO. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. Rome, FAO (2022).
28. Santos, M. D., Dickson, J. O., & Velasco, P. L. Mitigating the impacts of climate change: Philippine fisheries in focus in Pongsri, C., (ed) *Fish for the People*, ed. (Bangkok: Southeast Asian Fisheries Development Center), 101–110 (2011).
29. United Nations International Trade Statistics. 2021. https://unstats.un.org/UNSDWebsite/Publications/ (accessed December 2022).
30. Bemis, K. E. *et al*. Biodiversity of Philippine marine fishes: A DNA barcode reference library based on voucher specimens. *FigShare* https://doi.org/10.25573/data.c.6484798.v2 (2023).
31. Carpenter, K. E. & Niem, V. H. *FAO Species Identification Guide for Fishery Purposes The Living Marine Resources of the Western Central Pacific*. (Food and Agriculture Organization of the United Nations, 1999).
32. White W.T., *et al*. Market fishes of Indonesia (= Jenis-jenis ikan di Indonesia). ACIAR Monograph No. 155. (2013).
33. Baldwin, C. C., Mounts, J. H., Smith, D. G. & Weigt, L. A. Genetic identification and color descriptions of early life-history stages of Belizean *Phaeoptyx* and *Astrapogon* (Teleostei: Apogonidae) with comments on identification of adult *Phaeoptyx*. *Zootaxa* **2008**, 1–22 (2009).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. Kearse, M. *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformat.* **28**, 1647–1649 (2012).
36. Fricke, R., Eschmeyer, W. N. & Van Der Laan, R. (eds.). Eschmeyer's Catalog of Fishes: Genera, species, references. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (2022).
37. Froese, R. & Pauly, D. (eds.). FishBase, version 08/2022. www.fishbase.org (2022).
38. Boettiger, C., Lang, D. T. & Wainwright, P. C. rfishbase: exploring, manipulating and visualizing FishBase data from R. *J. Fish Biol.* **81**, 2030–2039 (2012).
39. Bemis, K. E. *et al*. Barcode of Life Data Systems, http://www.boldsystems.org/index.php/Public_SearchTerms?query=DS-PHILFISH (2023).
40. Bemis, KE. *et al*. *GenBank*, https://identifiers.org/ncbi/bioproject:PRJNA947503 (2023).
41. Hubert, N. *et al*. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One* **3**, e2490 (2008).
42. Hubert, N. & Hanner, R. DNA Barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes* **3**, 44–58 (2015).

Acknowledgements

The study of Philippine market fishes was conducted and funded by Memorandum of Agreements (dated 1 March 2011, 26 March 2013, and 21 October 2016) between the Department of Agriculture, Bureau of Fisheries and Aquatic Resources-National Fisheries Research and Development Institute (BFAR-NFRDI), Philippines, and the National Museum of Natural History Smithsonian Institution Department of Vertebrate Zoology and Laboratories of Analytical Biology, USA (project title: *Collaboration on the Inventory and DNA Barcoding of Commercial Fishes of the Philippines for Food Safety and Biodiversity*) and by an Interagency Agreement between the National Museum of Natural History Smithsonian Institution and the United States Food and Drug Administration. National Science Foundation Award DEB 1257632 supported rotenone and SCUBA collections of Philippine fishes in collaboration with the California Academy of Sciences. Fishes were collected under Gratuitous Permit Numbers: GP-0077-14, GP-0085-15, and GP-0112-16 and their collection approved by the NMNH Institutional Animal Care and Use Committee and Old Dominion University IACUC Protocol #13-013.

David Catania and Terrence Goseliner assisted in permits and transfer of materials to the NMNH. Abby Reft, NOAA National Systematics Lab, provided data retrievals from the NMNH database. We are grateful to Daniel DiMichele, Chris Huddleston, and Kathy Adia of the NMNH Biorepository for curating tissues and extracts for the project. Dan Cole, NMNH Cartography prepared the initial map used in Fig. 1 and Beth Gamble assisted with archiving photographs. Sara Handy, Sarah Stadig, and Fred Fry, FDA and Hannah Evans, American University assisted in formatting and uploading sequences and metadata to the FDA RSSL. We acknowledge Apollo Marco Lizano, Aaron Macaspac, Toby Potenciana, Mikel Mendiola, Philip Jessele, Jay Dicdiquin, Jin Kevin Dionisio, Jasmin Meren, and Clien Ivan Ongkiatco for their diligent administrative and field assistance. Don Carpenter, Nicolas Minsalan V, and Nicolas Minsalan VI provided collection support on Panay and Negros. Brian Stockwell and Abner Bucol assisted with collections in Negros, and the latter contributed many valuable specimens from his own market collections. Hilconida Calumpong assisted us in Dumaguete. Cleto 'Ting' Nañola assisted during the sampling at Mindanao, and Mia Comeros helped with logistical support in Cebu. We thank Marivene M. Santos, Museum Curator, of the Zoology Division of the Philippine National Museum for her support of this project. We thank Don Dumale of the Philippine National Museum who assisted in curation and field collections and provided access to the facilities at the Philippine National Museum for packing and storing specimens. We especially thank BFAR Director Asis G. Perez and NFRDI Officer-in-Charge Melchor M. Tayamen for their support of this project. We also thank BFAR Director Eduardo Gongona and NFRDI Director Drusila Esther Bayate. Through the National Stock Assessment Program, numerous BFAR-NFRDI staff at national and regional offices facilitated collections of specimens and provided space to conduct lab work in Metro Manila and Provinces of the Philippines, including the following: Director Nestor D. Domenden, Rosarion Segundina Gaerlan, staff and enumerators of BFAR 1; Director Milagros Morales, Angel Encarnacion, staff and enumerators of BFAR 2 (Cagayan); Director Wilfredo Cruz, Romina Yutuc, staff and enumerators of BFAR 3; Director Lilian Garcia, Maribeth Ramos, staff and enumerators of BFAR 4a; Director Ruben Jardin, Myrna Candelario, Elmar Villaflor, staff and enumerators of BFAR 4b; Director Dennis del Socorro, Virgina Olano, Aida Andayog, Noime Lanzuela, staff and enumerators of BFAR 5 (Sorsogon), Director Remia Aparri, Drusila Bayate, Sheryll Mesa, staff and enumerators of BFAR 6 (Iloilo); Director Andres Bojos Prudencio Belga, staff and enumerators of BFAR 7 (Cebu); Director John Albaladejo, Miriam Francisco, staff and enumerators of BFAR 8 (Tacloban); Director Isidro Velayo, Pedling Munap, staff and enumerators BFAR 9; Director Visa Tan Dimerin, Gigi Albor, staff and enumerators of BFAR 10; Director Fatima Idris, Jose Villanueva, staff and enumerators of BFAR 11; Director Sammy Malvas, Laila Emperua, staff and enumerators of BFAR 12; Director Nerio Casil, Joyce Baclayo, staff and enumerators of BFAR 13; Nonita Cabacaba, and staff of NFRDI lab at Guiuan, BFAR Fisheries Regulatory and Quarantine Division (Palawan); Glenda Cadigal of PCSDS (Palawan); Ludivina Labe and Edwyn Alesna of BFAR-FRQD (Quezon City); and the lab assistants of the NFRDI Genetic Fingerprinting Laboratory (Quezon City). We thank Philippine Game Fishing Foundation (PGFF) President Robert Nery, Jr., Roland Zamora, Merwin Asuncion, PGFF staff, and participants of the game fishing tournament (Cagayan) for their support and for allowing us to obtain tissue samples from their catch. We thank Jerry Bico for pointing out the fish market in Cagayan, which yielded the last specimen of *A. albimento*. We thank Diane Tyler for editing the manuscript.

Author contributions

K.E.B. Co-first author with M.G.G.; Conceptualization, Data curation, Project administration, Visualization, Writing original draft, Writing review and editing. M.G.G. Co-first author with K.E.B; Conceptualization, Data curation, Formal Analysis, Methodology, Validation, Visualization, Writing original draft, Writing review and editing. M.D.S. Methodology, Validation, Resources, Supervision, Project Administration, Funding Acquisition, Writing-review and editing. K.E.C. Conceptualization, original investigation design, methodology development, funding acquisition, resources, organization of collecting trips, participation in most market fish collecting trips,, data curation, formal analysis, methodology, validation, writing review and editing. J.R.D. Conceptualization, funding acquisition, project administration, writing original draft, writing review and editing. D.E.P. Conceptualization, data curation, project administration, supervision, writing review and editing. N.A.L.F. Field research assistant, field administration, data curation, writing-review and editing. E.S.H. Data curation, writing review and editing. A.D. Data curation, investigation, methodology, writing review and editing. K.S.M. Data curation, investigation, methodology, writing review and editing. L.A.W. Conceptualization, funding, project administration, resources, methodology, supervision, writing review and editing. J.T.W. Conceptualization, original investigation design, methodology development, funding acquisition, resources, organization of collecting trips, project administration and supervision of all market fish collecting trips, data curation, formal analysis, methodology, validation, visualization, writing review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.E.B., M.G.G. or J.T.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023