Title
Establishing a DNA Sequence Database for the Marine Fish Fauna of California

Authors
Hastings, Philip A.
Burton, Ron S.

Publication Date
2008-11-01

California Sea Grant Sea Grant
Final Project Progress Report

07/08/2008
03/1/2005–02/28/2008
R/F-194
Establishing a DNA sequence database for the marine fish fauna of California

Philip A. Hastings
SIO/UCSD
Marine Biology Research Division
phastings@ucsd.edu

Ron S. Burton
SIO/UCSD
Marine Biology Research Division
rburton@ucsd.edu

Project Hypotheses
Fishes can be identified to species using DNA sequence data IF an appropriate database of sequences from known species is available.

Project Goals and Objectives
The overall goal of this project was to develop the necessary infrastructure to permit species identification of California marine fishes using molecular sequence data. This involved the following efforts. 1) Compilation of an up-to-date list of marine fish species occurring in and around California. 2) Collecting and archiving in the Marine Vertebrate Collection, Scripps Institution of Oceanography, tissues and voucher specimens of California marine fish species. 3) Development of a standardized protocol for molecular identification of fishes, including sample preparation/DNA extraction, DNA amplification via polymerase chain reaction (PCR) and DNA sequencing, using three mitochondrial gene sequences (cytochrome b, 16S ribosomal DNA, cytochrome oxidase 1). 4) Making sequence data available on-line permitting remote comparison of sequence data with reference sequences from known species.

Briefly describe project methodology
Fish specimens were collected, identified, and catalogued into the SIO Marine Vertebrate Collection. For small specimens, the entire specimen was preserved in 95% ethanol. For larger specimens, blocks of muscle tissue were removed from the right side of the body and preserved in 95% ethanol, while the voucher specimen was fixed in 10% formalin and transferred to 50% isopropanol for long-term storage in the MVC.

DNA was extracted from small (approx. 10-20 mg) pieces of muscle tissue using a commercial kit (Qiagen's QIAamp DNA Mini Kit). The protocol included a Proteinase K digestion step followed by selective binding of DNA to a membrane in a spin column for washing before elution in 200 ul of deionized water. Amplification of 16S rDNA, cyt b and CO1 follow standard PCR protocols with "universal" primers (Palumbi et al. 1991) and fish specific primers (Ward et al. 2005). Following standard 35 cycle PCR reactions with 50 degree C annealing, 5 ul of the reaction products were run out on 2% agarose gels, stained in ethidium bromide solution and photographed under UV transillumination. Reactions

consisting of single bright bands of appropriate size (~640 bp for 16S rDNA, 875 bp for cyt b) progressed to sequencing. Failed PCR reactions were repeated, and typically were successful after dilution of the DNA extract; lowering annealing temperature was sometimes required. From successful reactions, the remaining 20 ul of the PCR reaction was purified using Qiagen╒s QIAquick PCR Purification kit, with final elution in 30 ul of deionized water.

DNA sequencing reactions used Amersham╒s ET Dye Terminator chemistry following the provided protocol with either the forward or reverse PCR primer; reaction volumes (and required reagents) were typically 10 ul (50% of the standard protocol). Sequencing reactions were purified using Sephadex G-50 spin columns and the resulting 10 ul volume was placed in a well of a 96-well plate. Sequences were then run (48 simultaneously) on the Amersham MegaBACE 500 DNA sequencer located in the Marine Biology Research Division. To insure sequence quality and to maximize sequence length, both strands of the PCR products were sequenced. Data were edited and aligned using Sequencher 4.0 software (Gene Codes, Inc.). The 16S sequences were submitted to the GenBank database. DNA isolates were sent to the Barcode of Life site (Guelph) were CO1 sequences were collected and posted on the Fish Barcode of Life project as well as on GenBank.

Describe progress and accomplishments toward meeting goals and objectives
This project made a concerted effort to collect and archive tissues of as many fish species as possible that have been recorded from California waters. These were deposited in the SIO Marine Vertebrate Collection along with voucher specimens from which the tissues had been taken. This included collecting efforts by SIO staff as well as contributions of specimens from colleagues including the National Marine Fisheries Service West Coast Groundfish survey. Tissues archived in the MVC serve as the source materials for generating sequence data for over 450 species. This involved establishment and implementation of standard techniques for DNA extraction, amplification and sequencing. Sequence data have been submitted to Genbank (http://www.ncbi.nlm.nih.gov/Genbank/index.html), a resource that permits remote comparison of sequence data with posted sequence data. In addition, Genbank sequence data are freely available for use by researchers for other applications such as phylogenetic analyses. Our data for CO1 gene are also available via the Fish Barcode of Life project (http://www.fishbol.org), a resource with similar functions for comparing sequences from unknown samples with reference sequences.

Project modifications
Initially we planned to sequence two genetic markers, 16S and cytochrome b. We successfully generated sequences of 16S for approximately 460 species and these sequences have been submitted to GenBank. Obtaining Cytochrome b sequences proved to be more inconsistent, although sequences for approximately 250 species were obtained. After the start of our project, the community of "barcode" researchers focused instead on the CO1 DNA marker. Consequently, in collaboration the FishBOL, we generated CO1 sequences for 463 species and these have been posted at GenBank and on the FishBOL web site/database where they contribute to the effort to barcode all species of fishes.

Project outcomes
All 16S sequence data have been submitted to GenBank. In addition, CO1 data have also contributed to the Barcode of Life database, specifically to the Fish Barcode of Life effort. Sequence data for additional species will continue to be added to the archive well beyond the end date as specimens become available.

Impacts of project
These sequence data are now available to all via the internet for use in molecular identification of fishes. They have already been used by students of fishes to identify unknown fish eggs and larvae, and stomach contents of other fishes.

Benefits, commercialization and application of project results
NA

Economic benefits generated by discovery
NA

Issue-based forecast capabilities
NA

Tools, technologies and information services developed
Genetic data are now freely available permitting molecular identification of fishes, their eggs and larvae and for fish products (e.g., fillets) for most species known from California waters via GenBank (all three sequences) and the Barcode of Life Project (CO1 only). The project will continue to add sequence data for rare species as specimens become available.

Publications
Conference papers, proceedings, symposia
Title: Barcoding the Fishes of North America (invited seminar; Available on-line at: http://www.fishbol.org/meeting_june05.php#workshop_program)
Authors: Hastings, P.A.
Date:June 5, 2005
Conference Title: Barcode of Fishes Conference
Location: Guelph, Ontario, Canada

Title: Establishing a DNA sequence database for the marine fish fauna of California (poster)
Authors: Ellison, C., K. Gruenthal, P. Hastings, R. Burton
Date:
Conference Title: Western Society of Naturalists
Location:

Title: Barcoding the Marine Fishes of California (invited seminar)
Authors: Hastings, P.A.
Date: May 5, 2008
Conference Title: North and Central American Linkages for the DNA barcoding of fish
Location: Chetumal, Yucatan, Mexico

Electronic publications
Title: Barcoding the Fishes of North America (Powerpoint presentation available on-line at: http://www.fishbol.org/meeting_june05.php#workshop_program
Authors: P. A. Hastings
Date: June 5, 2005

Please list any workshops/presentations given, type of audience (i.e., K-12 educators), location, date, number of attendees, and briefly describe content presented.

North and Central American Linkages for the DNA barcoding of fish, Jun 5-6, 2008, ECOSUR, Chetumal, Quintana Roo, Mexico: P. Hastings gave a talk entitled "Barcoding California Marine Fishes" to a group of 50+ researchers, fishery managers and conservation agents from Latin America reviewing this project with recommendations for future efforts to barcode fishes.

Dissemination of results
Applications of the techniques have been presented to UCSD undergraduate students in Introduction to Marine Biology (BIEB 132) by Burton and Hastings.

Cooperating organizations
Nongovernmental
Barcode of Life - provided technical assistance

Academic Institutions
University of Washington Fish collection - provided tissue samples
University of Kansas Fish Collection - provided tissue samples

International implications
Results from this study have been included in the Fish Barcode of Life Project (FishBOL), an international effort to compile sequence data for vouchered specimens of all fish species.

Awards none

Keywords barcode, California, cytochrome b, 16s, cytochrome oxidase I, database, DNA, eggs, finfish, fishes, larvae, molecular identification, GenBank, FishBOL, genetics, ichthyofauna, species identification

Patents or licenses none