Microsatellite genotypes, and cytb sequences for Etelis coruscans, Etelis carbunculus, and Etelis sp. and sample collection information in the Indo-Pacific Ocean from 1997 and 2012

Website: https://www.bco-dmo.org/dataset/873568
Data Type: Other Field Results
Version: 1
Version Date: 2022-04-26

Project
» Origins of Hawaiian Reef Fishes (Hawaiian Fish Origins)

Program
» Indo-Pac Research Coordination Network (Indo-Pac RCN)

Contributors	Affiliation	Role
Bowen, Brian	University of Hawaii	Principal Investigator
Andrews, Kimberly R.	University of Hawaii	Scientist
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Sample collection locations and microsatellite genotypes, and cytb sequences for Etelis coruscans, Etelis carbunculus, and Etelis sp. samples were collected in the Indo-Pacific Ocean between 1997 and 2012. Microsatellite analysis performed in 2015 - 2017. These data were published in Andrews et al. (2020).

Table of Contents

Coverage

Spatial Extent: **N**:32 **E**:-120.1551564 **S**:-35 **W**:29.8892
Temporal Extent: 1997 - 2012

Acquisition Description

Specimen collection and DNA extraction

A total of 1,153 specimens were collected from 15 geographic regions for *E. coruscans*, 1,064 specimens from 11 regions for *E. carbunculus*, and 590 specimens from 16 regions for *E. sp*. Some of these specimens were used in two previous population genetic studies, as described below in the sections on mitochondrial DNA sequencing and microsatellite genotyping (Andrews *et al.* 2014; Andrews *et al.* 2016); in one of these studies, *E. carbunculus* was called *Etelis marshi* (Andrews *et al.* 2014) due to nomenclatural confusion and the

existence of the cryptic species *E. sp.* (Andrews et al. 2016). Tissue specimens consisted of fin clips or muscle tissue collected by commercial fishers, purchased in fish markets, or collected on research cruises, and stored in salt-saturated DMSO buffer (Seutin *et al.* 1991). Within the Hawaiian Archipelago, specimens were collected between 1997 and 2012. For all other locations, specimens were collected between 2004 and 2012. Genomic DNA extractions were conducted using a phenol chloroform method (Cummings & Thorgaard 1994), DNeasy extraction kits (Qiagen, Valencia, CA, USA), or the Hotshot method (Meeker *et al.* 2007).

Mitochondrial DNA sequencing

We used cytochrome *b* (cyt*b*) sequence data from two previous studies for *E. coruscans* and *E. carbunculus* collected throughout the Hawaiian Archipelago (Andrews *et al.* 2014), and for *E. carbunculus* and *E. sp.* collected from 13 additional regions in the Indo-Pacific (Andrews *et al.* 2016). In this study, we also generated cyt*b* sequence data for *E. coruscans* specimens collected from 11 additional regions in the Indo-Pacific. PCR conditions and DNA sequencing methods were consistent across all studies. For *E. coruscans*, a 560bp fragment was amplified using the primers Cyb-05 L15020 (GCCAACGGCGCATCCTTCTTCTT; Meyer 1993) and Cyb-07 H15573 (AATAGGAAGTATCATTCGGGTTTGATG; Taberlet *et al.* 1992). For *E. carbunculus* and *E. sp.*, a 524bp fragment was amplified using the primers EhucybF (TCAGTCGCACACATCTGCCG) and EhucybR (AGTGCAACAAGGACGGCTGC), both from Andrews *et al.* (2014); this region overlaps the region amplified for *E. coruscans*. PCR products were sequenced in one direction using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), and sequences were edited and aligned using GENEIOUS PRO 5.6.2 (Biomatters, LTD, Auckland, NZ).

Microsatellite genotyping

We used microsatellite data obtained from the study of Andrews *et al.* (2014) for *E. coruscans* and *E. carbunculus* specimens collected throughout the Hawaiian Archipelago. We also generated microsatellite data from 10 additional Indo-Pacific locations for *E. coruscans* and three additional Indo-Pacific locations for *E. carbunculus*. Microsatellite PCR and genotyping protocols followed those described in Andrews *et al.* (2014), and were conducted by the same person as in the previous study for each species. A total of 10 microsatellite loci were analyzed for *E. coruscans* and 11 microsatellite loci for *E. carbunculus*. PCR products were analyzed on ABI 3730XL or ABI 3130XL genetic analyzers, with all fragments from each primer set run on one machine to avoid bias in fragment length estimates across sequencers. Fragments were scored using GENEMAPPER 4.0 (Applied Biosystems).

GENALEX 6.5 (Peakall & Smouse 2006) was used to identify specimens with identical microsatellite genotypes to determine whether any individual fish were represented more than once in the dataset. Each microsatellite locus was tested for deviations from Hardy Weinberg Equilibrium (HWE) and linkage equilibrium using ARLEQUIN 3.5.2.2 (Excoffier *et al.* 2005) for all locations outside of the Hawaiian Archipelago with $n \geq 10$. Tests for deviation from HWE and linkage equilibrium for the samples from the Hawaiian Archipelago were reported in Andrews *et al.* (2014). Bayesian clustering analyses (described below) were also conducted removing one locus at a time to investigate the influence of each locus on the results.

For a description of the population structure and demographic history analyses using these data see Andrews et al. (2020).

Species List:

ScientificName,AphiaID,LSID
Etelis carbunculus,212545,urn:lsid:marinespecies.org:taxname:212545
Etelis coruscans,212544,urn:lsid:marinespecies.org:taxname:212544

Processing Description

BCO-DMO Data Manager Processing Notes:
* The region is known ("Indo-Pacific"). Bounding box coordinates were calculated in python using the multipolygon from marineregions.org (MRID: 14289).
* Corrected column alignment caused by the locations with multiple words parsed into separate columns which pushed allele columns to the right. Used version Brian Bowen submitted on 2022-03-22.
* Sheet 1 in two excel files imported into the BCO-DMO Data system (Microsats_Ecoruscans_columns_realigned.xlsx, Microsats_Ecarbunculus_columns_realigned.xlsx) and tables combined
* Table transformed (unpivoted) from two columns per locus containing the first and last allele IDs, to one

column for Locus_Name, Allele1_ID, and Allele2_ID.

[table of contents | back to top]

Data Files

File	Version
Etelis carbunculus mtDNA cytb sequence data filename: Ecarbunculus_mtDNA_cytb.fasta (FASTA, 475.49 KB) MD5:feb10ea4cacf2093f800d59f9ceb607b *Etelis carbunculus mitochondrial DNA cytochrome b (cytb) sequence data in .fasta format. The description for each sequence within the fast file contains the sample name and location (e.g. >ECACH1213_ChristmasIsland). For a description of fasta format see https://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml*	original
Etelis coruscans mtDNA cytb sequence data filename: Ecoruscans_mtDNA_cytb.fasta (FASTA, 591.14 KB) MD5:4e1937b80d0cd30cd5f098c81e3a8e0b *Etelis coruscans mitochondrial DNA cytochrome b (cytb) sequence data in .fasta format. The description for each sequence within the fast file contains the sample name and location (e.g. >ECACH1213_ChristmasIsland). For a description of fasta format see https://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml*	original
Etelis spp. mtDNA cytb sequence data filename: Esp_mtDNA_cytb.fasta (FASTA, 266.68 KB) MD5:73e04441879a3a5e454de6203fa4369b *Etelis spp. mitochondrial DNA cytochrome b (cytb) sequence data in .fasta format. The description for each sequence within the fast file contains the sample name and location (e.g. >ECACH1213_ChristmasIsland). For a description of fasta format see https://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml*	original

[table of contents | back to top]

Related Publications

Andrews, K. R., Copus, J. M., Wilcox, C., Williams, A. J., Newman, S. J., Wakefield, C. B., & Bowen, B. W. (2020). Range-Wide Population Structure of 3 Deepwater Eteline Snappers Across the Indo-Pacific Basin. Journal of Heredity, 111(5), 471–485. https://doi.org/10.1093/jhered/esaa029
Results

Andrews, K. R., Moriwake, V. N., Wilcox, C., Grau, E. G., Kelley, C., Pyle, R. L., & Bowen, B. W. (2014). Phylogeographic Analyses of Submesophotic Snappers Etelis coruscans and Etelis “marshi” (Family Lutjanidae) Reveal Concordant Genetic Structure across the Hawaiian Archipelago. PLoS ONE, 9(4), e91665. https://doi.org/10.1371/journal.pone.0091665
Methods

Andrews, K. R., Williams, A. J., Fernandez-Silva, I., Newman, S. J., Copus, J. M., Wakefield, C. B., Randall, J. E., & Bowen, B. W. (2016). Phylogeny of deepwater snappers (Genus Etelis) reveals a cryptic species pair in the Indo-Pacific and Pleistocene invasion of the Atlantic. Molecular Phylogenetics and Evolution, 100, 361–371. https://doi.org/10.1016/j.ympev.2016.04.004
Methods

Cummings SA, Thorgaard GH (1994) Extraction of DNA from fish blood and sperm. Biotechniques 17, 426-430. PMID: 7818889. https://pubmed.ncbi.nlm.nih.gov/7818889/
Methods

Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47-50. PMID: 19325852. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2658868/
Methods

Meeker, N. D., Hutchinson, S. A., Ho, L., & Trede, N. S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. BioTechniques, 43(5), 610–614. https://doi.org/10.2144/000112619
Methods

Meyer A (1993) Molecular phylogenetic studies of fishes. In: Evolution and genetics of aquatic organisms (ed. Beaumont AR). Chapman and Hall, London.
Methods

PEAKALL, R., & SMOUSE, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6(1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x
Methods

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. In Canadian Journal of Zoology (Vol. 69, Issue 1, pp. 82–90). Canadian Science Publishing. https://doi.org/10.1139/z91-013
Methods

TABERLET, P., MEYER, A., & BOUVETV, J. (1992). Unusual mitochondrial DNA polymorphism in two local populations of blue tit Parus caeruleus. In Molecular Ecology (Vol. 1, Issue 1, pp. 27–36). Wiley. https://doi.org/10.1111/j.1365-294x.1992.tb00152.x
Methods

U.S. National Library of Medicine. (n.d.). Fasta format description. National Center for Biotechnology Information (NCBI). Retrieved April 26, 2022, from https://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml
Methods

[table of contents | back to top]

Related Datasets

IsRelatedTo

Bowen, B., Andrews, K. R. (2022) **Genetic sequence identifiers for Etelis samples collected in the Indo-Pacific Ocean between 1997 and 2012.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2022-04-21 http://lod.bco-dmo.org/id/dataset/873427 [view at BCO-DMO]
Relationship Description: The Etelis range-wide data set was designed to resolve population structure and management units for Etelis carbunculus. In the course of this population genetic study, it became apparent that the putative Etelis carbunculus actually contained two species, the foundation for the "Etelis genetics" and "Etelis morphology" data sets.

Bowen, B., Andrews, K. R. (2022) **Morphological measurements and meristics for Etelis samples collected in the Indo-Pacific Ocean between 1997 and 2012.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2022-04-14 http://lod.bco-dmo.org/id/dataset/873174 [view at BCO-DMO]
Relationship Description: The Etelis range-wide data set was designed to resolve population structure and management units for Etelis carbunculus. In the course of this population genetic study, it became apparent that the putative Etelis carbunculus actually contained two species, the foundation for the "Etelis genetics" and "Etelis morphology" data sets.

[table of contents | back to top]

Parameters

Parameter	Description	Units
SampleName	Sample name	unitless
Species	Species ID (Genus species)	unitless
SampleLocation	Geographic sampling location	unitless
Locus_Name	Locus name	unitless
Allele1_ID	ID of the first allele for the locus.	unitless
Allele2_ID	ID of the first allele for the locus.	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	ABI 3730XL or ABI 3130XL genetic analyzers
Generic Instrument Description	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

[table of contents | back to top]

Project Information

Origins of Hawaiian Reef Fishes (Hawaiian Fish Origins)

Coverage: Central and West Pacific Ocean

Project summary:
This research is designed to resolve the origins of Hawaiian reef fishes. All living inhabitants of the Hawaiian archipelago necessarily originate elsewhere, due to the volcanic history of the island arc. Hawaii also has the highest endemism (native species) in the Pacific, with 25% of the 625 near-shore fish species found nowhere else. Where did these fishes come from? Two prominent hypotheses regarding the origins of Hawaiian marine species maintain that colonists arrive either from the south (via the Line Islands and Johnston Atoll) or from the west (via Japan). Previous research has shown that Hawaiian endemic limpets (genus *Cellana*) colonized from Japan (Bird et al. 2011 Mol. Ecol. 20:2128 – 2141). Andrews et al. (2014; PLoS One 9: e91665) report evidence for a colonization pathway from the south (Johnston Atoll) to the middle of the archipelago in the Papahanaumokuakea Marine National Monument (PMNM). In this project, we will sample locations to the south of Hawaii (Johnston and Line Islands) and to the west of Hawaii (Ogasawara and Ryukyu Islands) for a suite of 20 reef fishes in order to resolve the origins of Hawaiian biodiversity. Advanced rebreather technology allows dives with longer bottom time and more efficient sample collection, and our program is pioneering the applications of this advance diving technology. To test alternate hypotheses in the lab, we will employ both population genetics (shifts in genotype frequencies) and phylogenetics (DNA sequence divergence) for more ancient separations. Restriction-digest associated DNA sequencing (RAD-seq) is the best method for studies of phylogeography, phylogenetics, and population biology because it provides high coverage of homologous portions of the genome from multiple individuals for comparatively low cost and effort. We use the ezRAD approach developed in the shared Bowen-Toonen Lab.

Description from NSF award abstract:
The Hawaiian Islands are the product of a volcanic hot spot in the middle of the North Pacific. Hence every living thing on this isolated archipelago has origins elsewhere. This project will investigate the origins of Hawaiian reef fishes, which are important both as a food source and a cultural touchstone in native Hawaiian communities. Two prominent hypotheses maintain that marine fish originally arrived from the south (Line Islands and Johnston Atoll) or from the west (Japan). To test these hypotheses, this research will augment existing specimens from Hawaii with expeditions to Johnston Atoll (closest shallow habitat to the south), the northern Line Islands (Palmyra), southern Line Islands (Christmas Island), and Ryukyu Islands and Ogasawara Islands in Japan. Advanced genetic techniques will be used to resolve the closest relatives to the Hawaiian fish

species and the pathways by which reef species colonize Hawaii and help establish patterns of biodiversity. In cases where Hawaiian species are closely related to widespread sister species, this project will detect hotspots of genetic divergence. Because this research will reveal the sources of Hawaiian marine biodiversity, results can be used to help define priorities for reef protection. The project will support two graduate students and train at least two more in all aspects of the project from rebreather diving, specimen collection and curation, information management, and advanced genetic techniques. There will be outreach efforts to schools through existing programs, and expedition teams will include a videographer to provide footage for the award-winning Voice of the Sea program, broadcast locally. Expeditions will also include an outreach specialist to handle media reports and promote awareness and concern for reefs in the communities surrounding study sites.

The investigators will sample a suite of 20 reef fishes at locations to the south (Johnston and Line Islands) and west (Ogasawara and Ryukyu Islands) of Hawaii to resolve the origins of Hawaiian biodiversity. The investigators will employ both population genetics (shifts in genotype frequencies) and phylogenetics (DNA sequence divergence) for more ancient separations to test their hypotheses. Restriction-digest associated DNA sequencing (RAD-seq) will be employed for the phylogeography, phylogenetics, and population biology studies because it provides high coverage of homologous portions of the genome from multiple individuals for comparatively low cost and effort.

[table of contents | back to top]

Program Information

Indo-Pac Research Coordination Network (Indo-Pac RCN)

Website: https://indopacificnetwork.wikispaces.com/

Description from NSF award abstract:
The objective of this Research Coordination Network project is to develop an international network of researchers who use genetic methodologies to study the ecology and evolution of marine organisms in the Indo-Pacific to share data, ideas and methods. The tropical Indian and Pacific Oceans encompass the largest biogeographic region on the planet, the Indo-Pacific. It spans over half of the Earth's circumference and includes the exclusive economic zones of over 50 nations and territories. The Indo-Pacific is also home to our world's most diverse marine environments. The enormity and diversity of the Indo-Pacific poses tremendous logistical, political and financial obstacles to individual researchers and laboratories attempting to study the marine biology of the region. Genetic methods can provide invaluable information for our understanding of processes ranging from individual dispersal to the composition and assembly of entire marine communities.

The project will:
(1) assemble a unique, open access database of population genetic data and associated metadata that is compatible with the developing genomic and biological diversity standards for data archiving,
(2) facilitate open communication and collaboration among researchers from across the region through international workshops, virtual communication and a collaborative website,
(3) promote training in the use of genetic methodologies in ecology and evolution for researchers from developing countries through these same venues, and
(4) use the assembled database to address fundamental questions about the evolution of species and the reservoirs of genetic diversity in the Indo-Pacific.

The network will provide a model for international collaborative networks and genetic databasing in biodiversity research that extends beyond the results of this Research Coordination Network effort.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1558852

[table of contents | back to top]