Voucher summary of invertebrates and barcoded OTU's with field identifications, collected from Palau marine lakes

Website: https://www.bco-dmo.org/dataset/768196
Data Type: Other Field Results
Version: 1
Version Date: 2019-05-13

Project

» Do Parallel Patterns Arise from Parallel Processes? (PaPaPro)

Program

» Dimensions of Biodiversity (Dimensions of Biodiversity)

Contributors	Affiliation	Role
Dawson, Michael N.	University of California-Merced (UC Merced)	Principal Investigator
Copley, Nancy	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Voucher summary of invertebrates and barcoded OTU's with field identifications, collected from Palau marine lakes.

Table of Contents

Coverage

Spatial Extent: **N**:7.3237 **E**:134.5089 **S**:7.1506 **W**:134.3447
Temporal Extent: 2014-07-11 - 2015-07-27

Dataset Description

Voucher summary of invertebrates and barcoded OTU's with field identifications, collected from Palau marine lakes. Summaries by major taxon (e.g. phylum) are provided.

* NOTE: The P.I.'s are using this dataset to write papers. Please contact them before using these data to make sure you are not duplicating efforts.

Methods & Sampling

After completion of fieldwork, a subset of specimens from the transect surveys were chosen for DNA barcoding to confirm or amend field identifications. These specimens included (i) at least one specimen from

each field-ID (except obvious species such as *Mastigias papua*) and (ii) several specimens representing the range of phenotypic variation of field-IDs that showed considerable variation or were challenging to distinguish (e.g. small sponge specimens of similar color and texture). Additionally, specimens from a previously collected voucher collection (indicated with "V_" in prefix of sequence ID) were barcoded and identified by taxonomic experts. Specimens from population genetic collections (indicated with "PG_" in prefix of sequence ID) were also barcoded. DNA was purified using a modified phenol-chloroform CTAB extraction protocol (1) or AcroPrep PALL 5053 glass fiber plates procedure (2, 3). We amplified the Cytochrome c Oxidase subunit I (COI) barcode locus using 0.5 µL of purified DNA in a 25-µL polymerase chain reaction (PCR) with 0.05 µL AMPLITAQ (Applied Biosystems, Foster City, California, USA), 2.5 µL 10x buffer (Applied Biosystems), 0.63 µL of 20 µM primers (Operon Biotechnologies Inc., Huntsville, Alabama, USA), 2.5 µL of 25 mM MgCl2 (Applied Biosystems), 0.5 µL of 10 mg/mL bovine serum albumin (BSA) and 0.5 µL of 10 mM dNTPs. Several primer sets were used (Table 1). Amplicons were sequenced at the University of California Berkeley DNA Sequencing Facility (Berkeley, California, USA). Base calls in electropherograms were visually checked and manually corrected for errors and forward and reverse reads were assembled in Sequencher 4.8 (GeneCodes, Ann Arbor, Michigan, USA). We used Basic Local Alignment Search Tool (BLASTn) to determine the higher level taxonomic assignment for each sequence (which we used to process batches of similar sequences) — ascidians, bivalves, bryozoans, cnidarians, crustaceans, echinoderms, gastropods, polychaetes, and poriferans. Sequences organized by these broad groups were then aligned using Muscle v3.8.425 (4). For each group, alignments were manually adjusted and trimmed to the same length in Mesquite v3.5 (5) to balance total individuals retained and sequence length. The resulting alignment lengths were: ascidians 395bp, bivalves 567bp, bryozoans 622bp, cnidarians 612bp, crustaceans 299bp, echinoderms 357bp, gastropods 562bp, polychaetes 509bp, and poriferans 688bp. Sequences were translated to amino acid sequence to confirm an open reading frame. Short sequences were excluded from further analysis, but percent pairwise identity with the closest match was recorded for each based on the shortest sequence. Pairwise sequence distance was calculated using dist.dna with Kimura's 2-parameter distance model of evolution (6) in the ape package v4.1 (7) in R (8). OTUs, or clusters of sequences, similar at 97% were identified using tclust in the spider package v1.5.0 (9) in R (8) for each taxonomic group, except for poriferans, which were clustered at 99% sequence similarity given their slow sequence evolution (10).

1.   Dawson MN, Raskoff KA, Jacobs DK (1998) Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotechnol 7(2):145–52.

2.   Ivanova N V., Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes 6(4):998–1002.

3.   Schiebelhut LM, Abboud SS, Gómez Daglio LE, Swift HF, Dawson MN (2017) A comparison of DNA extraction methods for high-throughput DNA analyses. Mol Ecol Resour 17(4):721–729.

4.   Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797.

5.   Maddison WP, Maddison DR (2018) Mesquite: a modular system for evolutionary analysis.

6.   Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120.

7.   Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.

8.   R Core Team (2018) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).

9.   BROWN SDJ, et al. (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12(3):562–565.

10.   Huang D, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J Mol Evol 66(2):167–174.

See Table 1. Primers and thermocycle conditions used for PCR of macroinvertebrates by taxonomic group in Supplemental Documents, below.

For the sequence alignment files (.fas) mentioned in the methods above, see the Supplemental Files section below.

Data Processing Description

BCO-DMO Processing:
- added conventional header with dataset name, PI name, version date
- replaced blanks cells with nd

[table of contents | back to top]

Data Files

File
5_invert_voucher_summary.csv(Comma Separated Values (.csv), 60.49 KB) MD5:36c95be3794ddd1bca32a63d270fe258 Primary data file for dataset ID 768196

[table of contents | back to top]

Supplemental Files

File
Table 1 filename: Dataset768138_Table1.pdf   (Portable Document Format (.pdf), 410.57 KB) MD5:d46391bfb4c5dd39afd47a8e3b813538 Primers and thermocycle conditions used for PCR of macroinvertebrates by taxonomic group

[table of contents | back to top]

Related Publications

BROWN, S. D. J., COLLINS, R. A., BOYER, S., LEFORT, M.-C., MALUMBRES-OLARTE, J., VINK, C. J., & CRUICKSHANK, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources, 12(3), 562–565. doi:10.1111/j.1755-0998.2011.03108.x
Methods

Dawson, M. N., Raskoff, K. A., & Jacobs, D. K. (1998). Field preservation of marine invertebrate tissue for DNA analyses. Molecular marine biology and biotechnology, 7(2), 145-152.
Methods

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. doi:10.1093/nar/gkh340
Methods

Huang, D., Meier, R., Todd, P. A., & Chou, L. M. (2008). Slow Mitochondrial COI Sequence Evolution at the Base of the Metazoan Tree and Its Implications for DNA Barcoding. Journal of Molecular Evolution, 66(2), 167–174. doi:10.1007/s00239-008-9069-5
Methods

IVANOVA, N. V., DEWAARD, J. R., & HEBERT, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes, 6(4), 998–1002. doi:10.1111/j.1471-8286.2006.01428.x
Methods

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution, 16(2), 111-120.
Methods

Maddison, W. P., & Maddison, D. R. (2018). Mesquite: a modular system for evolutionary analysis. version 0.992, 2002.
Software

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, 20(2), 289–290. doi:10.1093/bioinformatics/btg412
Methods

R Core Team (2018) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria). https://www.R-project.org
Software

Schiebelhut, L. M., Abboud, S. S., Gómez Daglio, L. E., Swift, H. F., & Dawson, M. N. (2016). A comparison of DNA extraction methods for high-throughput DNA analyses. Molecular Ecology Resources, 17(4), 721–729. doi:10.1111/1755-0998.12620
Methods

[table of contents | back to top]

Parameters

Parameter	Description	Units
CRRF_ID	internal ID number for voucher sample	unitless
Phylum	Phylum assigned by taxonomic expert	unitless
Class	Class assigned by taxonomic expert	unitless
Order	Order assigned by taxonomic expert	unitless
Family	Family assigned by taxonomic expert	unitless
Genus	Genus assigned by taxonomic expert	unitless
Species	Species assigned by taxonomic expert	unitless
ID_name	taxonomic identification	unitless
OTU_id	number of Operational Taxonomic Units individuals in sample	specimens
lake_code	3-letter code for sampled lake name	unitless
comment	Comments pertaining to percentage similarity with other organisms.	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	
Generic Instrument Name	Automated DNA Sequencer
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.

Dataset-specific Instrument Name	
Generic Instrument Name	Thermal Cycler
Generic Instrument Description	A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

Deployments

Palau_lakes

Website	https://www.bco-dmo.org/deployment/542180
Platform	Small boats - CRRF
Start Date	2010-08-21
End Date	2016-06-14
Description	Palau marine lakes

Project Information

Do Parallel Patterns Arise from Parallel Processes? (PaPaPro)

Website: http://marinelakes.ucmerced.edu/

Coverage: Western Pacific; Palau; Indonesia (West Papua)

This project will survey the taxonomic, genetic, and functional diversity of the organisms found in marine lakes, and investigate the processes that cause gains and losses in this biodiversity. Marine lakes formed as melting ice sheets raised sea level after the last glacial maximum and flooded hundreds of inland valleys around the world. Inoculated with marine life from the surrounding sea and then isolated to varying degrees for the next 6,000 to 15,000 years, these marine lakes provide multiple, independent examples of how environments and interactions between species can drive extinction and speciation. Researchers will survey the microbes, algae, invertebrates, and fishes present in 40 marine lakes in Palau and Papua, and study how diversity has changed over time by retrieving the remains of organisms preserved in sediments on the lake bottoms. The project will test whether the number of species, the diversity of functional roles played by organisms, and the genetic diversity within species increase and decrease in parallel; whether certain species can greatly curtail diversity by changing the environment; whether the size of a lake determines its biodiversity; and whether the processes that control diversity in marine organisms are similar to those that operate on land.

Because biodiversity underlies the ecosystem services on which society depends, society has a great interest in understanding the processes that generate and retain biodiversity in nature. This project will also help conserve areas of economic importance. Marine lakes in the study region are important for tourism, and

researchers will work closely with governmental and non-governmental conservation and education groups and with diving and tourism businesses to raise awareness of the value and threats to marine lakes in Indonesia and Palau.

[table of contents | back to top]

Program Information

Dimensions of Biodiversity (Dimensions of Biodiversity)

Website: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503446

Coverage: global

(adapted from the NSF Synopsis of Program)
Dimensions of Biodiversity is a program solicitation from the NSF Directorate for Biological Sciences. FY 2010 was year one of the program. [MORE from NSF]

The NSF Dimensions of Biodiversity program seeks to characterize biodiversity on Earth by using integrative, innovative approaches to fill rapidly the most substantial gaps in our understanding. The program will take a broad view of biodiversity, and in its initial phase will focus on the integration of genetic, taxonomic, and functional dimensions of biodiversity. Project investigators are encouraged to integrate these three dimensions to understand the interactions and feedbacks among them. While this focus complements several core NSF programs, it differs by requiring that multiple dimensions of biodiversity be addressed simultaneously, to understand the roles of biodiversity in critical ecological and evolutionary processes.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1241255

[table of contents | back to top]