Plan Overview

A Data Management Plan created using DMPTool

Title: Barriers to cross-shelf coral connectivity in the Florida Keys

Creator: Mikhail Matz - **ORCID:** 0000-0001-5453-9819

Affiliation: The University of Texas at Austin

Principal Investigator: Mikhail Matz

Data Manager: Mikhail Matz

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: PD 98-1650

Grant: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1737312

Template: BCO-DMO NSF OCE: Biological and Chemical Oceanography

Last modified: 06-22-2017

Barriers to cross-shelf coral connectivity in the Florida Keys

Data Policy Compliance

Identify any published data policies with which the project will comply, including the NSF OCE Data and Sample Policy as well as other policies that may be relevant if the project is part of a large coordinated research program (e.g. GEOTRACES).

Principal Investigator agrees to comply with the Division of Ocean Sciences Sample and Data Policy.

Pre-Cruise Planning

If the proposed project involves a research cruise, describe the cruise plans. (Skip this section if it is not relevant to your proposal.) Consider the following questions: (1) How will pre-cruise planning be coordinated? (e.g. email, teleconference, workshop) (2) What types of sampling instruments will be deployed on the cruise? (3) How will the cruise event log be recorded? (e.g. the Rolling Deck to Repository (R2R) event logger application, an Excel spreadsheet, or paper logs) (4) Will you prepare a cruise report?

Question not answered.

Description of Data Types

Provide a description of the types of data to be produced during the project. Identify the types of data, samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Include a description of the location of collection, collection methods and instruments, expected dates or duration of collection. If you will be using existing datasets, state this and include how you will obtain them.

Field data:

- GPS coordinates of sampled reefs;
- biological samples preserved in 100% ethanol for two coral species (*Montastrea cavernosa*, *Porites astreoides*);
- data on growth and survival of sampled coral colonies.

Sequence data:

- genome sequences for *Porites astreoides* and *Montastrea cavernosa* (fasta format);
- annotations for genome sequences (gff format);
- genome-wide genetic variation data for *Porites astreoides*, *Montastrea cavernosa* and *Orbicella faveolata* (vcf format).

Scripts and simulations:

- bioinformatics analysis walkthroughs (plain text format)
- inferred migration rates (tab-delimited text tables)
- simulations of population adaptation in SLiM software (Eidos/SLiM code format)

Curriculum materials:

- selected datasets to serve as training examples during Ecological Genomics workshops led by the PI.

Data and Metadata Formats and Standards

Identify the formats and standards to be used for data and metadata formatting and content. Where existing standards are absent or deemed inadequate, these formats and contents should be documented along with any proposed solutions or remedies. Consider the following questions: (1) Which file formats will be used to store your data? (2) What type of contextual details (metadata) will you document and how? (3) Are there specific data or metadata standards that you will be adhering to? (4) Will you be using or creating a data dictionary, code list, or glossary? (5) What types of quality control will be used? How will data quality be assessed and flagged?

Fieldwork metadata will be stored in Excel format.

Genome sequence will be in FASTA format; annotations will be in GFF format.

Variation data will be in VCF format.

Raw sequence reads will be stored in compressed fastq format.

Statistical analysis pipelines will be recorded as R and perl scripts and accompanies by detailed instructions and comments within the scripts.

Daily protocols and organismal data will be stored in a notebook that remains at all times in the PI's lab.

Data Storage and Access During the Project

Describe how project data will be stored, accessed, and shared among project participants during the course of the project. Consider the following: (1) How will data be shared among project participants during the data collection and analysis phases? (e.g. web page, shared network drive) (2) How/where will data be stored and backed-up? (3) If data volumes will be significant, what is the estimated total file size?

The metadata accompanying specific publications stemming from this project will be deposited on Dryad server. Scripts and bioinformatics instructions will be made available on PI's GitHub page, https://github.com/z0on . Large datasets (such as genomes and genetic variation data) will be available through PI's lab data page, http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html

Location of all datasets will be registered with the Biological and Chemical Oceanography Data Management Office (BCO-DMO), providing links to the locations of specific datasets.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

---

Describe mechanisms for data access and sharing, and describe any related policies and provisions for re-use, re-distribution, and the production of derivatives. Include provisions for appropriate protections of privacy, confidentiality, security, intellectual property, or other rights or requirements. Consider the following: (1) When will data be made publicly available and how? Identify the data repositories you plan to use to make data available. (2) Are the data sensitive in nature (e.g. endangered species concerns, potential patentability)? If so, is public access inappropriate and how will access be provided? (e.g. formal consent agreements, restricted access) (3) Will any permission restrictions (such as an embargo period) need to be placed on the data? If so, what are the reasons and what is the duration of the embargo? (4) Who holds intellectual property rights to the data and how might this affect data access? (5) Who is likely to be interested in re-using the data? What are the foreseeable re-uses of the data?

> Early availability of datasets will be announced through email-list servers (coral-list, ECOLOG), through PI's professional twitter feed, and eventually through forthcoming papers.
>
> Our data will be freely available to any interested party, primarily other researchers interested in our genetic work. The new R packages will be available under GPL-3 license. All data files will be freely available for at least three years beyond the award period, as per NSF guidelines.

Plans for Archiving

Describe the plans for long-term archiving of data, samples, and other research products, and for preservation of access to them. Consider the following: (1) What is your long-term strategy for maintaining, curating, and archiving the data? (2) What archive(s) have you identified as a place to deposit data and other research products?

Data stored in notebooks will be kept strictly in the lab at the University of Texas at Austin. Monthly, these notebooks will be photocopied and the copies will be kept at Dr. Matz's personal residence. Digital data on personal laptops will be backed up continuously using MacBook's TimeMachine and weekly to the RANCH storage server at the Texas Advanced Computer Center (TACC). Additionally, all the data and manuscript files related to this project will be synchronized with the Box (UT-approved online storage service analogous to Dropbox). The data acquired and preserved as part of the proposed research will be governed by the University of Texas' policies regarding intellectual property, record retention, and data management.

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data. Consider the following: (1) If there are multiple investigators involved, what are the data management responsibilities of each person? (2) Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

The PI Mikhail V. Matz will be responsible for compliance wiht the Data Management Plan.
