# Barriers to cross-shelf coral connectivity in the Florida Keys

*A Data Management Plan created using DMPTool*

Creator: Mikhail Matz

Affiliation: University of Texas at Austin (UT)

Template: National Science Foundation (NSF)

Last modified: 06-22-2017

Copyright information:

# Barriers to cross-shelf coral connectivity in the Florida Keys

## Data Policy Compliance

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

## Pre-Cruise Planning

Question not answered.

## Description of Data Types

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

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

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

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

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

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