DATA MANAGEMENT PLAN

RAPID: Collaborative Proposal: Dynamics of storm-mediated asexual reproduction in Florida Keys corals post-Hurricane Irma

PI: Carly D. Kenkel[1] (Univ. of So. California), Co-PI: Mikhail Matz (Univ. of Texas Austin)

(1) Products of Research

1) Site map survey data (timepoint, colony/fragment identification, size, relative position, growth, survival, bleaching, disease status)
2) Ethanol preserved tissue samples
3) Raw sequencing data (2bRAD FASTQ files)
4) Processed sequence data (tables of SNP counts)
5) Statistical analysis R scripts
6) Results from site surveys (summary files and figures)
7) Results from sequence analyses (summary files and figures)
8) Written reports and manuscript drafts

(2) Data Format

In-water and laboratory protocols and field-collected organismal and environment data will be stored in notebooks that remain at all times in the PI's labs, and also recorded and archived in electronic form (text files or Excel spreadsheets). Tissue samples will be stored in labeled boxes at -20C and storage information will be recorded on the lab freezer organization spreadsheet. Sequence read data will be stored as compressed FASTQ files. Processed 2bRAD data will be stored as tab-delimited text files and/or Excel spreadsheets. Statistical analyses data will be stored as tab-delimited input files and R source files. Results summaries will be stored as Word documents and PDF figure files.

(3) Access to Data and Data Sharing

Results will be published in open access research journals which have a wide distribution (e.g. Scientific Reports, PLoS ONE) to facilitate uptake of results by all end-users. Manuscripts will also be uploaded to open-access pre-print servers, such as BioRxiv prior to formal review and publication. All primary data that have been included in the results section of the published manuscripts and statistical analysis scripts sufficient to reproduce reported results will be submitted to the journal as Supplemental Information or archived in open-access databases, such as DRYAD or GitHub.

Sequencing data will be released through archiving on NCBI's Sequence Read Archive. The release of the data will be announced through the Coral-List email server maintained by NOAA (coral-list@coral.aoml.noaa.gov) and through social media platforms.

---

[1] Responsible for compliance with Data Management Plan

The 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.

(4) Re-Use, Re-Distribution and Production of Derivatives

Primary data and other supporting materials will be available for unrestricted use to all academic researchers upon request. USC's policy is to encourage, wherever appropriate, research data to be shared with the general public through internet access. Terms of use will include requirements of attribution along with disclaimers of liability in connection with any use or distribution of the research data. The data will be available for at least three years beyond the award period.

(5) Archiving Data

Final peer-reviewed journal manuscripts and supplemental information, such as data tables for graphical information in manuscript figures, will be posted in the USC Digital Repository and will be available on this publically available website no later than 12 months after publication. These records will be durable, accessible through web protocols and made safe from tampering or falsification. The storage media will be updated as necessary to keep it current.

Kenkel: Prior to publication, data stored in notebooks will be kept strictly in the lab at USC. Digital data will stored in four locations: 1. Personal computers; 2. Kenkel lab file storage / ftp server; 3. USC Digital Repository file server maintained on site; (4) weekly and monthly back-ups of the USCDR, stored off-campus.

Matz: 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.

(6) Prior Data Sharing Record

Through posting manuscripts to preprint servers, such as BioRxiv, announcements on the NOAA-maintained "Coral-list" and other social media platforms, Lab websites, presentations at scientific conferences and invited lectures, the PI's are committed to making data available prior to publication for unconditional use by the broader community.

PI Kenkel also has a strong track record of immediate and unrestricted data sharing. At present, she has uploaded 86 entries to NCBI's Sequence Read Archive for various projects, contributed six data packages to the DRYAD Digital Repository, which include both data and statistical analysis scripts, uploaded manuscripts on BioRxiv ahead of peer-review and shared transcriptome resources on her lab website prior to formal publication.

Co-PI Matz initiated his lab's policy of rapid genomics data sharing in 2007. Currently his lab's web server (http://matzlab.weebly.com/data--code.html) hosts six annotated transcriptomes and one genome, four of which have not yet been published. Matz also maintains a series of GitHub repositories, including scripts and data for early bioRxiv preprints (https://github.com/z0on?tab=repositories). Matz lab has uploaded 10 preprints to bioRxiv (https://www.biorxiv.org/search/mikhail%252Bmatz) corresponding to manuscripts in submission.