

EAGER: Can ancient DNA illuminate the fate of Caribbean reefs?

Data Policy Compliance

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

Description of Data Types

Sample data:

- Reef core sampling location and storage location;
- Age of samples (according to radiocarbon dating)

Sequence data:

- Sequencing reads for raw libraries (for microbiome analysis)
- Sequencing reads for hybridization-captured libraries (for coral genetics analysis)

Scripts and simulations:

- bioinformatics analysis and walkthroughs

Curriculum materials

Data and Metadata Formats and Standards

All metadata will be stored in tab-delimited (tab) or comma-delimited (csv) format.

Raw sequence reads will be stored in compressed fastq format.

Statistical analysis pipelines will be recorded as R, python, and bash scripts and accompanied by detailed instructions and comments within the scripts.

Laboratory protocols 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. Sequencing reads will be stored at the NCBI SRA archive. A copy of metadata, scripts and bioinformatics instructions will be made available on PI's GitHub page, which will also be duplicated to a Zenodo page to obtain a stable DOI. Select datasets, useful for training purposes, will be available through PI's lab data page.

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

We believe in early sharing 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 will be made freely publicly available within two years of collection.**

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 to the Box (UT-approved cloud storage service analogous to Dropbox). Copies of large raw datasets (sequencing reads) will be also kept at the Ranch server of the Texas Advanced Computing Center (TACC). 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 with the Data Management Plan.