Data Management Plan

Both PIs of this proposal, Cordes and Kulathinal, understand that research has become both increasingly data-intensive and highly collaborative. They firmly believe in the open-access principle and will provide all generated data to the coral and genomics communities in a presentable and usable form, prior to publication. Both PIs come from exceptionally strong collaborative communities. As an active member of the growing deep coral community, Cordes has led many initiatives, including organizing collaborative grants and cruises, throughout his academic career. Kulathinal comes from the Drosophila community that has a century-old tradition of sharing resources with researchers. Kulathinal holds much experience in interacting with the fly, genetics and genomics communities during his postdoctoral fellowship at FlyBase at Harvard University, one of the leading model organism databases, as a developer and biocurator from 2005-2008.

Genomic Data

This proposal will generate an enormous amount of sequence data for high-throughput analyses. We have budgeted for a small web server that will provide access to data for collaborators and the community. This web server will contain ample hard-disk space for curated sequence files and documentation. To backup these and all other files generated from this project, we have already secured 10TB of hard-drive space on our department's file server specifically for this project, and have the capability to add more. Kulathinal brings to the proposal his experience as an evolutionary geneticist and bioinformatician. Kulathinal has extensive resources in his lab including access to Temple University's new CPU/GPU cluster, a local 2x12 node 128GB workstation, as well as ample computational personnel including a Postdoctoral Fellow trained in computer science, a bioinformatician with extensive next-generation sequence assembly and mysql database experience, and undergraduate (1) and Masters-level (1) computer scientists.

Raw transcriptome data, quality files and sample metadata will be expeditiously deposited to NCBI's GEO database. Once assembled, the genome will be submitted to NCBI's Genome db. As part of a parallel effort to provide these data to the community, we will also make these data available on our local server (as mentioned above). This server will provide users with raw, assembled, and aligned sequence data via an ftp site linked to each of our websites. Our functional annotations will be provided to users as flatfiles, mysql tables, and GFF3 files and will provide easy-to-use documentation in addition to suggestions on how to handle, manage, and analyze these massive datasets.

The Kulathinal lab has successfully developed a database resource for the lizard community (*lizardbase.org,* also hosted at Temple University) that includes visual tools (i.e., a genome browser) to explore the recently sequenced green anole lizard genome and its annotations, information on museum collections and lab specimens, a GIS-mapper, and educational curricula used for K-12 students. While developing a similar database for these coral data could be a priority, the PIs also understand the importance of developing a strong centralized resource, and realize that excellent coral databases already exist. The PIs plan to support these resources as much as possible and will actively solicit guidance from these databases on how to help build community-capacity, as soon as the first sequences are generated.

Oceanographic Data

In addition to the sequence data produced from this proposal, a large amount of oceanographic data will be generated. We plan to submit all amassed carbonate chemistry data to the Global Ocean Data Analysis Project (GLODAP; http://cdiac.ornl.gov/oceans/glodap/), maintained by the Carbon Dioxide Information Analysis Center (CDIAC) at the Oak Ridge National Laboratory, in Oak Ridge, Tennessee. Presently, GLODAP contains no oceanographic data from the Gulf of Mexico. In addition, we will submit our oceanographic data to the National Oceanographic Data Center (NODC, http://www.nodc.noaa.gov/). Finally, all protocols for total alkalinity and pH measurements used for this project will be posted on the Cordes lab website (http://astro.temple.edu/~ecordes).

Summary of Data Management Plan:

The PI's, Cordes and Kulathinal, are deeply committed to disseminating all generated data—in both raw and polished forms—to the community in an expeditious manner. They also realize that these data provide a unique opportunity to encourage researchers from other fields as well as students interested in marine ecological genomics to utilize the growing resources of deep coral biology. Both Kulathinal and Cordes will make it a priority to not only have these data readily available, but to ensure that their utility is well advertised through their own websites and especially through other community websites and databases.