Data Management Plan

Research Cruise Data

MBARI

MBARI's Video Annotation and Reference System (VARS) is a software interface and SQL database system that provides tools for describing, cataloging, retrieving, and viewing the visual, descriptive, and quantitative data associated with MBARI's deep-sea video archives. MBARI provides the data 'as is' to the public. The VARS Public Query database is updated monthly and contains data spanning back through 20 years of observations. (The most recent 2 years of observational data remain embargoed for internal use only.)

Haddock Lab maintains an internal FileMaker database for collected, preserved, and frozen specimens, and stores genetic data and a BLAST resource on their server (see Facilities), software under development is stored on the labs bitbucket repository where it can also be transitioned to a public repo for release. Sanger sequence data in the program Geneious as a MySQL database using the Shared Server feature and BioCode plug-ins.

Illumina data and assembled transcriptomes are stored on our servers in a RAID6 array (dual failure tolerance).

Voucher data

Digitized data of physical specimens from this project will be made available through the online National Resource for Digitized Collections (iDigBio.org), located at the University of Florida and funded by the ADBC program at NSF. The teams will collaborate on this aspect, and Evergreen undergraduates and the "Partners in Science" high school teachers will devote time to this project during our UNOLS research cruises.

Functional Diversity

Physiological Data

All physiological data will be formatted to be consistent with `ecoretriever.org`, and deposited into Dryad. This data entry will be done by Evergreen students supervised by Thuesen. In addition, Haddock and Thuesen are curators for Encyclopedia of Life which is launching a central data repository, and our data will be served there as well. Haddock has contribute to EOL's TraitBank. Ryan has extensive experience setting up and managing web databases (see references) and will insure that data processing and compatibility are maintained.

Taxonomic Diversity

Morphological traits will be coded and entered into a data file using the Lucid Builder software. Lucid Builder allows for the quick and easy construction of multimedia identification keys as well as output of data files. This will be served to the public through an interactive app on the JellyWatch website. Haddock is the director of jellywatch.org, which is currently maintained through students and volunteers and citizen scientists. A private prototype can be viewed at `http://jellywatch.org/ctenokey`

Trees will be posted to Dryad and Open Tree of Life. Our colleague Claudia Mills is the editor for ctenophores at WoRMS (World Registry of Marine Species), which is used as a taxonomic reference for other major databases. We will incorporate our tree revisions into that database as well.

Molecular Diversity

Annotated molecular data will be deposited in GenBank, and all sequencing reads will be added to the NCBI's Short Read Archive (`www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?`). Together, Ryan and Haddock have submitted more than 6000 nucleotide sequences to GenBank stemming from over 25 projects dating back to 2000. Ryan, Haddock, and their students and trainees will be responsible for submitting this data on a quarterly basis.

Database References

Haddock, S.H.D and C.W. Dunn (2010). *Practical Computing for Biologists*. Chapter 15.

Szabo, C., Masiello, A., Ryan, J. F., & Brody, L. C. (2000). The breast cancer information core: database design, structure, and scope. *Human mutation*, *16*(2), 123-131.

Moreland, R. T., Ryan, J. F., Pan, C., & Baxevanis, A. D. (2009). The homeodomain resource: a comprehensive collection of sequence, structure, interaction, genomic and functional information on the homeodomain protein family. *Database: the journal of biological databases and curation*, *2009*.

Ryan, J. F., & Finnerty, J. R. (2003). CnidBase: the cnidarian evolutionary genomics database. *Nucleic acids research*, *31*(1), 159-163.

Sullivan, J. C., Ryan, J. F., Watson, J. A., Webb, J., Mullikin, J. C., Rokhsar, D., & Finnerty, J. R. (2006). StellaBase: the *Nematostella vectensis* genomics database. *Nucleic acids research*, *34*(suppl 1), D495-D499.

Moreland, R., Nguyen, A.-D., Ryan, J.F., Schnitzler, C.E., Koch, B.J., Siewert, K., Wolfsberg, T.G., Baxevanis, A.D. 2014. A customized Web portal for the genome of the ctenophore *Mnemiopsis leidyi*. *BMC Genomics* 15: 316.

Moreland R. T., Ryan, J. F., Nguyen A., Schnitzler C.E., Baxevanis A. D. *Mnemiopsis* Genome Project Portal: `http://research.nhgri.nih.gov/mnemiopsis/`