Emlet, Shanks, and Sutherland

Data Management Plan

Types of Data and Data Processing

This research will produce a number of different types of data including;

1) Each year we will produce a roughly three month time series of temperature, salinity and Chl A collected during high tides at the Oregon Institute of Marine Biology (OIMB) pier near the mouth of Coos Bay. Salinity will be calculated using conductivity and temperature.

2) Each year we will produce a time series collected at the same site and time of zooplankton samples. The zooplankton will be filtered from the seawater system intake located at the OIMB pier. Data will take the form of counts of organisms per volume of water per day.

3) Each year we will produce a time series collected during the same period of settling invertebrates. Newly settled juveniles will be collected in Tuffy-type settlement substrates placed at three intertidal sites, the rocky shore near the pier and Sunset Bay and South Cove, Cape Arago. Data will take the form of counts of organisms per collector per day. There will be at least three collectors per site.

4) Physical oceanographic data collected on cruises in the coastal ocean. Weather permitting we will sample 5 to 7 stations along a 10 to 15 km transect, which will be located about 6 km north of the mouth of Coos Bay and will extend from just outside the surf zone (0.5 to 1 km offshore) to the end of the transect. Station position will be determined from GPS readings. At each station we will make a CTD (conductivity, temperature, depth) cast measuring temperature, salinity and Chl A vs. depth. We estimate that we will sample the transect 5 to 10 times per year.

5) At each station we will collect current speed and direction data with an ADCP (acoustic Doppler current profiler).

6) At each station we will collect depth stratified zooplankton samples at 3 to 5 depths (number of sample depths will increase with station bottom depth). Samples will be collected with 120 um mesh quarter meter tucker trawl. The volume filtered through the net will be determined from flow meters mounted in the mouth of the net. Data will include flow meter readings, sample depth as determined by wire angle and wire out, zooplankton counts, and zooplankton counts corrected for the volume of water filtered. Counts will be screened for outliers and these samples will be re-counted.

7) DNA from embryonic stages, larvae, and juveniles of marine invertebrates will be extracted and amplified. In order to identify species, mitochondrial genes COI and 16S will be targeted for sequencing. Target DNA samples from various embryonic and larval stages will be sent out for sequencing (e.g. Sequetech, Inc) and will be returned in the form of nucleotide sequences for each gene (COI and 16S). These raw data will be examined with software for quality of the sequences read, to trim ends, and to create consensus sequences from forward and reverse sequences of the same gene from one individual. For species identification, these sequence data will be compared with known material identified and curated in NCBI GenBank or from sequences of the same genes obtained from adult animals collected near study sites.

8) Digital photographs of live embryonic stages and larvae will be made with the purpose of providing visual descriptions of stages and will be cross-referenced with gene sequence data.

All physical and biological (e.g., CTD, ADCP, fluorometers) sensors will be calibrated both pre- and post-deployment. Corrections for sensor drift will be applied as necessary. Data will be screened for outliers. Processing will be performed using Matlab. Preserved samples of embryos and larvae will not be retained for future study because alcohol or formalin storage will change the color and shape of the larvae

or embryos so will not be useful for identification purposes. Further more formalin preservation would negate future genetic work and ethanol preservation has a limited period of preservation unless refrigeration is constant and curation timely.

During and upon completion of the project we will work with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at OCE Biological and Chemical Oceanography Programs to ensure that our data is of the appropriate quality and format for easy sharing with the scientific community and for long term storage.

Data Formats, Storage, and Metadata

Data will be saved as raw data and processed data. Both formats will be stored on multiple hard drives, e.g., each PIs workstation, as well as their long-term backup storage drives. In addition, all data will be backed up essentially continuously at Carbonite Inc., an offsite data backup company. *Raw data* will be stored as instrumentation output, raw plankton and settler counts, and nucleotide sequences sent from the offsite sequencing company. *Processed data* will be converted to ASCII and mat (Matlab) file formats. Matlab files are specific to Matlab; however, this program is commonly used by the oceanographic community. ASCII data are easily read via text editing programs. ASCII file headers will contain all information needed to interpret the file. The ASCII format is non-proprietary and interfaces with a variety of software. All changes to *Raw data* will be documented and described and included within the metadata. Metadata will be stored within headers for ASCII files and within structures for mat files. A readme text file containing pertinent information and metadata will be stored with the data files. Processed nucleotide sequence data trimmed and consensus contigs of the forward and reverse sequences will be stored as FASTA files. A readme file will be created to document and describe how each FASTA sequence was determined.

Distribution and Archiving of Data All project investigators will work together to ensure the preservation and timely distribution of data. Upon completion of the project and with the help of BCO-DMO, all data and metadata will be put into the appropriate format for submission to the National Oceanographic Data Center (NODC, www.nodc.noaa.gov) for archiving. Within two years of the completion of the project, data and metadata will be submitted to BCO-DMO for submission to NODC**.** Nucleotide sequence data of identified larval stages will be submitted to NCBI GenBank for permanent curation. Photographs of specimens will be deposited with the University of Oregon's Scholars' Bank, which uses Dublin Core metadata. Data submitted with Scholars' Bank is registered with DataCite, which creates unique identifiers (DOIs, just like there are for journal articles) for each data object. The DOI(s) can then be linked to in articles and other publications.