

## Data Management Plan

### *Type of data and samples collection*

The raw data collected in this study include a large amount of 16S rDNA sequences from epibiotic and planktonic bacterial community, algal growth measurements as indicated by fluorescent assay under different biotic and abiotic conditions, domoic acid (DA) production by different species of *Pseudo-nitzschia* and the DA concentration under different conditions, measurements of bacteria and viruses growth and decline during co-incubation with algal culture, and bacterial extracellular enzymes and activities. The physical, chemical, and biological parameters will be collected at the field sampling-sites, including water temperature, salinity, rainfall, chlorophyll and phytoplankton species and abundance. Raw data will be recorded in the lab notebook and/or stored in lab computer files. The electronic data files will have at least one backup copy at a secondary computer or the storage area. The sequence data after validation and trimming will be submitted to the GenBank for public access.

In addition to data collection described above, this study will also collect epibiotic bacteria, multiple species of *Pseudo-nitzschia* and possibly bacteriophages that infecting epibiotic bacteria. The isolates will be made into frozen stock and stored in the laboratory deep freezers during the study. The subculture of some of the isolates will be deposited to the American Type Culture Collection once the organism has been characterized.

### *QA/QC plan and data standards*

For quality assurance and quality control (QA/QC) of the data collection, all field sampling and laboratory experiments will be conducted following protocols approved by the principal investigator. The procedures for each experiment will be documented in a laboratory notebook. The results from each experiment will be recorded in the notebook and in computer files. The principal investigator will review laboratory notes, procedures, and outcomes on weekly basis. All samples are collected and stored in laboratory-certified, contaminant-free, sterilized bottles. All samples are transported and stored under the best condition to minimizing the changes in microbial community. These records are completed by field personnel and become a permanent part of the data file.

In the laboratory, controls include chemical purity and media sterility checks, dilution water blanks, and equipment blanks are performed. Known negative samples are tested on each lot or batch of chemical and media as a negative standard control. Known positive samples are tested on each lot or batch of chemical and media as a positive standard. In addition, positive and negative process controls are performed with each set of experiments. Further procedures and formulas for calculating data quality indicators or applicable QC statistics, including precision, bias, outliers and missing data, are according to those found in the federal experimental conduct protocols.

All research personnel involved in this research will be trained for sampling and experimental techniques. The personnel are required to demonstrate their ability to independently perform experimental procedures accurately before their work and results are accepted. All lab personnel will receive training in biosafety through the UC Irvine Environmental Health and Safety program. All personnel are responsible for complying with QA/QC requirements that pertain to their technical function. Each technical staff member must have a combination of experience and education to adequately demonstrate a specific knowledge of their particular function and a general knowledge of laboratory operations, test methods, QA/QC procedures, and records management.

All equipment, reagents, and medium used in this project will be purchased from credible vendors. Calibration of equipment such as thermal cyclers and pipettes will be performed on a routine basis and documented.

Positive and negative controls will be built into the experimental design for all assays. For example, seeding studies will be performed to quantify the efficiency of nucleic acid extraction and PCR. Replications will be performed on samples whenever possible.

#### *Corrective action for laboratory measurements*

Laboratory staff members have the primary responsibility of reporting failures in the laboratory or measurement systems to the PI. Deviations from defined protocols are documented in the comment section of the analyst's notes and reported to the PI immediately. If any equipment fails, laboratory personnel report the problem in the comment section of their notes and do not record data values for the variables in question. Actions are taken to replace or repair broken equipment prior to the next use. No data is entered into the project database that is known to be collected with faulty equipment. It is the combined responsibility of all members of the laboratory staff to determine if the performance requirements of the specific laboratory method have been met and to repeat sample analysis if required. The PI determines if all analytical and data QC procedures for reasonableness, accuracy, and clerical errors have been met. The PI works with the analyst to solve any problems and prevent the reporting of suspect data.

The PI will serve as the quality control officer (QCO) for the lab, and is responsible for team training, research oversight, data review procedures, and research outcome assessment.

#### *Data Validation*

Data validation is the process whereby data are filtered and accepted or rejected based on a set of criteria. It is a systematic procedure of reviewing a body of data against a set of criteria to provide assurance of its validity prior to its intended use. All data is checked for accuracy and completeness. The data validation process consists of data generation, reduction, and review.

#### *Data sharing*

Synthesized data will be presented as reports, presentations and research journal articles and are archived on the project website, national and international conferences and scientific journals. Results from this research will also be distributed to a broad audience by lectures and seminars through the public educational program such as those conducted by Back Bay Science Center. The cultured organisms will be made available to researchers upon request. The sequence information obtained from this study will be deposited into the GenBank sequence database.