

Data Management Plan update:

Raw sequence data associated with the project will be made available on NCBI where possible. Some owners of the data wish to contribute to the project, but not to make all raw data publicly available. However, we will continue to pursue consent from all and make as much of the data publicly available as possible. At a minimum, a subset of the sequence and genotype data can be made publicly available for each species. We would like to use the PopSet Database on NCBI but, format is not designed for genotype frequency data combining sequence, microsatellite and genomic scale data as will be the case for our dataset (see below).

NCBI is not ideal for archiving this type of linked population genetic/phylogeographic data, and because of the lack of an accepted universal database for these type of data, we have already applied for pending funding by NESCent to develop a collaborative project among researchers across the entire Indo-Pacific to develop a shared database and data sharing network. We expect all data owners from this project to be members and make the raw data available on the public data repository developed by this complementary project focused specifically on the marine population genetics of the Indo-Pacific. This or other funding (such as an RCN proposal) will be pursued to develop a lasting data repository for future and complimentary efforts to investigate multi-species connectivity using the available data.

Storage of the diversity of data used and produced by this project will require a flexible data repository platform. One such option is The Knowledge Network for Biocomplexity (KNB) and its user-friendly Morpho software to create data storage packages for both internal data management and data sharing. We will create a data storage package to place on the KNB using Morpho for this project that houses metadata and datatables for all components of the project, including single species and multi-species summary statistics and analyses outputs of spatial genetic structure and diversity, and associated life history data, ecological data and environmental data used for all analyses (e.g., oceanographic connectivity matrices, maps and geo-referenced data on multi-species connectivity analyses). Several examples of similar collections of genetic and environmental datasets exist in the KNB database, such as this entry: "*Vellend M., Parallel effects of land-use history on species diversity and genetic diversity of forest herbs*" which contains community data and frequencies of both alleles and haplotypes.

Finally, the oceanographic connectivity matrices created for the project will also be stored in the BCO-DMO repository, similar to this example of mussel connectivity matrices for the Southern California Bight: <http://osprey.bcodmo.org/dataset.cfm?id=14208&flag=viewwt>. Directions on where to find downloadable genetic and environmental data associated with each matrix will be provided in the dataset description, and on our project web page.