5/22/12

“Read me” file for DRYAD data deposit, file name “Microsatellite genotypes for Mol. Ecology environmental stress study in C gigas.xlsx”.

# *Molecular Ecology* manuscript entitled “Environmental stress increases selection against and dominance of deleterious mutations in inbred families of the Pacific oyster *Crassostrea gigas”* by Louis V. Plough

This excel file contains 7 tabs each with a different file. Four tabs contain the genotype data used in the study (named “07x5 1-algal genotypes”, “07x5 3-algal genotypes”, “08x3 1-algal genotypes”, and “08x3 3-algal genotypes”). In these genotype files, the first column contains the identification of the individual, either the sire or dam (the parents of the cross) or a number, 1-96, which are the ID’s of the progeny. The next columns (41 for 07x5, 39 for 08x3) represent the different microsatellite markers and their respective genotypes. The marker name is in row 2. The order of the alleles (*e.g.* either “ab” or “ba”) in the parents have been entered according to the phase of the alleles on a given linkage group, which was determined from the linkage mapping procedure in Joinmap 3.0. These genotype files are mostly ready as input files in the mapping program ProcQTL and the exact format is described in the third paragraph.

The two map files (“08x3 Map file” and “07x5 Map file”) are also setup for input into ProcQTL. The first column is the marker number, which is required by ProcQTL. The second column is the centimorgan position of each marker, and the third column is the linkage group. The fourth column must be deleted before running the model, but corresponds with the Marker number in column one. Notice the marker order in column 4 of the mapping file matches the marker order in the genotype file for a given family. The two files must correspond in this fashion when running the mapping procedure.

The last file, “example input genotype file” is an example of the input format required for the genotype file in ProcQTL. The “example” dataset is for the 07x5 1-algal genotypes. Column 1 contains the “id”, a number that must start at “1” in row 1. “ID” # 1 and 2 (i.e. rows 1 and 2) must be the parents, then numbers 3 and above for however many progeny are analyzed (in this case we have 96 progeny which have IDs 3-98). The next two columns (columns 2 and 3) are to designate whether individuals are parents or offspring. “-1” should be in both columns for the sire and dam (which are ID 1 and 2 respectively). For all progeny (ID’s 3-98) column 2 should have a ‘1’ and column 3 should have a ‘2’. Column 4 and above represent each microsatellite marker, and the genotypes of each row correspond to each individual.Note there are no column headings on this file, and thus no marker names. The marker order must still match the mapping file as described above. For more help regarding the format of the data files, please refer to the PROC QTL manual online: <http://statgen.ucr.edu/download/software/PROC%20QTL%201.5/PROC%20QTL.pdf>