Contribution No. AED-09-046

Bridging the Gap Between Large-Scale Data Sets and Analyses: Semi-Automated Methods to Facilitate Amplified Fragment Length Polymorphism Scoring and Data Analyses.

McGreevy Jr TJ[1], Markert JA[1,2], Grear J[1], and Nacci D[1]. 1US EPA, Atlantic Ecology Division, Population Ecology Branch, Narragansett, RI 02882 USA; 2US EPA, Ecological Effects Research Division, Population Ecology Branch, Cincinnati, OH 45219 USA

Amplified fragment length polymorphism (AFLP) markers can be developed more quickly and at a lower cost than microsatellite and single nucleotide polymorphism markers, which makes them ideal markers for large-scale studies of understudied taxa — such as species at risk. However, manual scoring of AFLP markers is prone to data entry errors, time intensive, and subjective. Recently, the objectivity of scoring AFLP DNA fingerprinting data produced from automated sequencers has been greatly improved with the development of AFLPScore v1.3 (Whitlock et al., 2008). We developed an R script to expedite the transformation of the raw peak intensity data output from GeneMarker® v1.6 (SoftGenetics LLC®, State College, PA) to a format compatible for AFLPScore. AFLPScore was used to minimize the genotype scoring error and maximize the number of AFLP markers retained. We developed a second R script to transform the binary genotype output generated by AFLPScore to a format compatible for AFLP-Surv v1.0 (Vekemans, 2002). AFLP-Surv was used to calculate the proportion of polymorphic loci and expected heterozygosity (H_j) using a large AFLP data set for mysid shrimp (*Americamysis bahia*). We investigated the correlation between *A. bahia* AFLP genetic diversity values and extinction risk using replicated experimental populations with manipulated levels of genetic diversity subjected to environmental stress. We also tested the reliability of estimating the initial H_j for both the control and experimental population lines using the harmonic mean effective population size for each population and only the corresponding control population's ending H_j. The two R scripts we developed will greatly reduce data entry error and facilitate the analyses of large-scale data sets. The line code for the two R scripts also could be manually adjusted to accommodate the transformation of AFLP data between other commonly used computer programs.

KEYWORDS: Genetic diversity; Heterozygosity; Mysids; Population stress