

Title:

geneGIS: computational tools for spatial analyses of DNA profiles with associated photo-identification records

Authors:

Dori Dick¹, Shaun Walbridge², Dawn Wright^{1,2}, John Calambokidis³, Erin Falcone³, Debbie Steel⁴, Tomas Follett⁴, Jason Holmberg⁵, C. Scott Baker⁴

¹College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR

²Esri, Redlands, CA

³Cascadia Research Collective, Olympia, WA

⁴Marine Mammal Institute, Oregon State University, Newport, OR

⁵Shepherd Project, Portland, OR

Contact: dickd@geo.oregonstate.edu

Abstract Body (max word count 300):

Spatially explicit, individual-based records from photo-identification and genetic samples are increasingly common for many long-term or large-scale studies of whales and dolphins. Such spatio-temporal data are used to track migration and life history parameters of individuals, to estimate abundance and trends of populations and, in the case of genetic markers, to infer close kinship and define management units, or Distinct Population Segments. These data however, are rarely integrated into one database and few tools exist to explore the spatial patterns of the integrated data. To enhance understanding of cetacean population structure, ecosystem relationships and predictive models of human impact, we developed and implemented 'geneGIS', computational tools to facilitate visual exploration and spatial analyses of individual-based records from DNA profiles and photo-identification records. Two different environments were created: 1.) a Java-based relational database with a web interface supported by the Shepherd Project (<http://www.ecoceanusa.org/shepherd/doku.php>) and 2.) ArcGIS 10.1 with a menu driven toolbar and geoprocessing Python toolbox (<http://genegis.org>). Tools implemented in geneGIS provide the ability to spatially view, filter and summarize combined genetic and photo-identification records and map their spatial/temporal relationships. GeneGIS permits export of spatially selected data in standard tabular formats (e.g., XLS, CSV) and spatial formats (e.g.,

shapefile, KML), as well as to specialized programs for statistical analyses of genetic diversity and differentiation (e.g., Genepop and GenAlEx) and for capture-recapture estimates of abundance (e.g., Capture and Mark). We demonstrate the utility of geneGIS using an integrated database of more than 18,000 records of humpback whales (*Megaptera novaeangliea*) in the North Pacific collected during the Structure of Populations, Levels of Abundance and Status of Humpback Whales in the North Pacific (SPLASH) program. These records represent more than 8,000 naturally marked individuals and 2,700 associated DNA profiles, including 10 microsatellites, mtDNA and sex.

Total word count = 289