

geneGIS: Geoanalytical Tools and Arc Marine Customization for Individual-Based Genetic Records

Dorothy M. Dick,^{*} Shaun Walbridge,[†] Dawn J. Wright,^{*,†}
John Calambokidis,[‡] Erin A. Falcone,[‡] Debbie Steel,[§] Tomas Follett,[§]
Jason Holmberg[¶] and C. Scott Baker^{§,**}

^{*}College of Earth, Ocean, and Atmospheric Sciences, Oregon State University

[†]ESRI (Environmental Systems Research Institute)

[‡]Cascadia Research Collective

[§]Marine Mammal Institute, Oregon State University

[¶]Wild Me

^{**}Department of Fisheries and Wildlife, Oregon State University

Abstract

To improve understanding of population structure, ecosystem relationships and predictive models of human impact in cetaceans and other marine megafauna, we developed *geneGIS*, a suite of GIS tools and a customized Arc Marine data model to facilitate visual exploration and spatial analyses of individual-based records from DNA profiles and photo-identification records. We used the open source programming language Python 2.7 and ArcGIS 10.1 software to create a user-friendly, menu-driven toolbar linked to a Python Toolbox containing customized geoprocessing scripts. For ease of sharing and installation, we compiled the *geneGIS* program into an ArcGIS Python Add-In, freely available for download from the website <http://genegis.org>. We used the Lord-Castillo et al. (2009) Arc Marine data model customization as the starting point for our work and retained nine key base Arc Marine classes. We demonstrate the utility of *geneGIS* using an integrated database of more than 18,000 records of humpback whales (*Megaptera novaeangliae*) 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 biparentally inherited microsatellite loci, maternally inherited mitochondrial DNA, and genetic sex.

1 Introduction

Landscape genetics (or seascape genetics in the ocean) aims to study spatial ecological processes by combining knowledge from population genetics, landscape ecology and spatial analysis to quantify the influence of landscape features on population genetic structure (Manel et al. 2003; Storfer et al. 2007). Understanding the relationship between landscape and genetic connectivity can reveal new insights into biological processes and lead to detecting, predicting and mitigating the effects of anthropogenic landscape modification

Address for correspondence: Dorothy Dick, College of Earth, Ocean, and Atmospheric Sciences, 104 Wilkinson Hall, Oregon State University, Corvallis, OR 97331 USA. Email: dickd@geo.oregonstate.edu

Acknowledgements: We would like to thank all the researchers involved with the SPLASH program as well as Cascadia Research Collective and the Oregon State University Cetacean Conservation Genetics Laboratory (CCGL) for database management. Thanks also to members of the CCGL for beta testing various components of *geneGIS*. The comments from two anonymous reviewers significantly improved the manuscript. This research and computational development was funded by the Office of Naval Research contract N0270A awarded to CSB and DJW, Oregon State University.

and global climate change (Wagner et al. 2012). This knowledge can aid managers in conservation measures by identifying barriers to gene flow or genetic diversity and provide alternative management scenarios to predict consequences to genetic variation and population connectivity (Storfer et al. 2007). With the advent of global positioning system (GPS) technology, growing databases of spatially-explicit genetic data have opened novel analysis opportunities including the development of geographic information system (GIS) software packages such as the Landscape Genetics GIS Toolbox (Vandergast et al. 2011) and Landscape Genetics Toolbox (Etherington 2011) (see Table 1). Additional standalone software such as GenGIS2 (Parks et al. 2013) and *Wildbook* (<http://www.wildme.org/wildbook/>) allow for the integration of genetic data with digital maps to enhance geographic and/or ecological data visualization (Table 1). Although these packages help to visualize various genetic metrics across geographic space, none directly calculate genetic distance measures (e.g. F-statistics) or provide estimates of kinship and relatedness while also allowing for the visualization *and* spatial analysis of multiple records from known individuals, as is typical in the field of marine mammal science.

Table 1 Summary of spatially-based landscape genetic software packages developed to help analyze various genetic metrics in geographic space

Landscape Genetics Software Package	Analyses Performed	GIS based?
Genetic Landscapes GIS Toolbox (Vandergast et al. 2011)	Creates raster surfaces of genetic divergence and diversity for single species (or genetic marker) and summarizes multiple genetic divergence or diversity rasters as average and variance surfaces	Yes (ArcGIS 9.3+)
Landscape Genetics Toolbox (Etherington 2011)	Creates a polyline shapefile to visualize genetic relatedness, conducts least-cost modeling to measure landscape connectivity, creates a matrix of pairwise points separated by a known barrier (either lines or landscape polygons)	Yes (ArcGIS 9.3+)
GenGIS 2 (Parks et al. 2013)	Integrates molecular biodiversity data with digital maps and habitat parameters to visualize geographic and ecological factors that influence community composition and function	No (basic mapping capabilities, supported by GDAL and R Project)
Wildbook (http://www.wildme.org/wildbook/) Developed in parallel with <i>geneGIS</i>	A web-accessible Java-based relational database management framework supporting capture-mark-recapture and molecular ecology of marine megafauna. Integrates photo-identification and genetic records, some direct calculations of F-statistics	No (basic mapping using Google Maps and export functions)

Many whale and dolphin species (cetaceans) are the focus of large-scale, long-term field studies that include numerous spatially-explicit observations of recognizable individuals. Repeated sightings of known individuals over time can reveal information on site fidelity (e.g. Baker et al. 2013), habitat use (e.g. Rasmussen et al. 2007), life history parameters (e.g. Ford et al. 2000), social organization (e.g. Baird and Whitehead 2000), distribution (e.g. Dalla Rosa et al. 2012), abundance (e.g. Barlow et al. 2011), and population structure (e.g. Baker et al. 1986, 1998, 2013). Such information is critical for protecting cetaceans and their natural ecosystems from the cumulative and synergistic effects of habitat degradation, fisheries, pollution, vessel traffic and global climate change (Reeves et al. 2003; Würsig et al. 2009).

Individual identity in cetaceans is typically determined by either photo-identification or genetic analysis. Photo-identification uses 35-mm cameras with telephoto lenses to capture distinct natural markings, color patterns, and scarring on an animal's body and/or notches and nicks along fins and fluke edges to identify individuals (Hammond et al. 1990). The photographs are reconciled to unique individuals and compiled into catalogs with associated databases for analyses and future reference. The replacement of film cameras with high-resolution digital cameras increased the accuracy, speed and efficiency of photo-identification techniques (Markowitz et al. 2003). Alternatively, genetic analysis using non-lethal collection of tissue samples (e.g. biopsy dart deployed via a crossbow or rifle, Noren and Mocklin 2012) from animals in the wild and DNA markers are used to reveal a unique genetic identity (genotype or DNA profile) for each individual. In addition to obtaining a genotype, samples can also be used to determine population structure including kinship, prey preferences through stable isotope analysis, contaminant loads, and hormonal indicators of physiological processes (Noren and Mocklin 2012).

The number of records typically generated by the two approaches differs significantly. Photo-identification, especially when using digital cameras, generates large numbers of records (1,000s) because each time an individual is encountered there is an opportunity for many photographs (and associated spatio-temporal information) to be added to a database. Conversely, the number of genetic samples is typically far fewer because the genome of an individual does not change. Sampling, therefore, only needs to occur once to capture an individual's genetic identity in a database. It is critical, however, that a genetic sample and an associated identification photograph are collected simultaneously and recorded accurately to ensure that the two forms of individual identity are correctly associated in the database. Although linking the photographic and genetic databases via a common identity field is possible, it is often challenging. The lack of integration between the two data sources may be due to different research questions and subsequent data needs, permitting stipulations, or a lack of computational tools available to handle such data. Yet, from an analytical perspective, the extension of an individual's DNA profile to photo records where genetic data are lacking and their subsequent integration into one large database would enrich the information available that can be used for conservation and management decisions. Even when reconciled into a single database, few tools exist that enable a researcher to visualize the spatial pattern of such integrated data.

The Convention on Biological Diversity's recent call to improve biodiversity by safeguarding genetic diversity (CBD 2012) emphasizes the importance of its inclusion when planning conservation measures. A population or species with greater genetic variation should have higher resilience and be able to adapt to environmental changes and perturbations more readily (Primack 2010). The enrichment of a database by the addition of genetic information enables managers to factor in population structure and genetic diversity, and thus

maximize species resilience, when developing conservation actions. But, how can we best facilitate the exploration and visualization of *spatial* patterns of genetic variability in individual-based, long-term cetacean studies? To address this question, we develop *geneGIS*, a suite of GIS tools and a customized version of the Arc Marine data model (Wright et al. 2007) for spatially-explicit genetic and photo-identification records to enable: (1) data visualization; (2) spatial exploration, display and selection of data; (3) basic spatial analyses; (4) data extraction from relevant environmental layers; and (5) data export to specialized software packages for molecular ecology. We use data from a three-year humpback whale study in the North Pacific as our exemplar in the development and implementation of *geneGIS*. Although we focus here on the use of *geneGIS* for cetaceans, we envision *geneGIS* will be a powerful platform to enhance our understanding of population structure, ecosystem relationships and predictive models of human impact across species and ecosystems, while also contributing to the development of landscape and seascape genetics (Miller 2005; Etherington 2011; Vandergast et al. 2011; Parks et al. 2013).

2 Background Information

2.1 Humpback Whales of the North Pacific and the SPLASH Program

Humpback whales (*Megaptera novaeangliae*) occur in all major ocean basins and migrate seasonally between high latitude feeding grounds and low latitude breeding grounds (Johnson and Wolman 1984). Their coastal distribution enabled heavy exploitation by the whaling industry for several centuries (Clapham 2009) and severe depletion led to an endangered listing under the US Endangered Species Act of 1973 and endangered/vulnerable status (1986–1990/1990–2008, respectively) by the World Conservation Union (Stevick et al. 2003; Reilly et al. 2008). In 1966, the International Whaling Commission banned commercial humpback whale hunting in the North Pacific (Best 1993). Today, most studied populations are recovering (Barlow et al. 2011); however, their presence in coastal regions remains a concern because these areas tend to be the most heavily populated and modified by humans.

To better understand the abundance, distribution and population structure of humpback whale populations in the entire North Pacific, a three-year international collaborative effort including over 50 research groups and more than 400 researchers in 10 countries was conducted from 2004–2006 (Calambokidis et al. 2008). The Structure of Populations, Levels of Abundance and Status of Humpbacks (SPLASH) program targeted all known humpback whale winter breeding and summer feeding grounds. The SPLASH program yielded 18,640 quality photo-identification images representing 7,940 unique individuals. A total of 5,669 tissue samples were also collected; 2,703 of these were genotyped to resolve 2,161 individuals. Prior to beginning the *geneGIS* project, photographic and basic sample collection data (photoSPLASH) were stored in a Microsoft Access relational database, which serves as the primary data repository for the SPLASH program. In parallel development to *geneGIS*, photoSPLASH is also adapted to an online catalog and database repository (<http://www.splashcatalog.org>) hosted by *Wildbook* (<http://www.wildme.org/wildbook>), a Java-based software framework supporting capture-mark-recapture studies of marine megafauna. The SPLASH catalog allows users varying degrees of access to the photoSPLASH database (depending on authorization level) to search, filter, query and export records of individual-based humpback whale encounters made during the SPLASH project. Genetic analytical data from samples collected during SPLASH (*geneSPLASH*) including sex, maternally inherited

mitochondrial DNA (mtDNA) and 10 microsatellite loci were originally stored as tabular spreadsheets in Microsoft Excel. Although separate, the databases share several common fields including Occurrence ID (a point in time and space when one or more whales were observed), Encounter ID (a point in time and space at which a photograph and/or tissue sample of an individual was collected) and Individual ID (a unique number for each distinct individual based on photograph or genotype).

During 2011–2012, photoSPLASH and geneSPLASH were merged into a single database (hereafter referred to as SPLASH). The reconciliation extended the number of encounters to include 781 new identifications from whales with no photo record and extended 1,002 different genotypes to 3,189 encounters that previously only had a photo record. This resulted in 7,335 total encounters (roughly 40% of the database) for 2,151 whales with a unique genotype. The large increase in the number of spatially-explicit encounters now extended with genetic records provides an unprecedented opportunity to explore the spatial pattern of genetic diversity of North Pacific humpback whales using GIS.

2.2 Conceptual Framework

The conceptual framework for *geneGIS* relies upon three key components – the location of known individuals, the measured value of environmental variables at that location, and the DNA profiles of these individuals (Figure 1). The configuration of these components and the data available will determine the type of research questions that can be asked. For example, data on individual location and seascape covariates can lead to questions concerning habitat preference and habitat use. Individual location data combined with DNA profiles can be used to study population structure, relatedness and kinship. Finally, DNA profiles in combination with environmental variables can be used to focus on seascape genetics to determine how

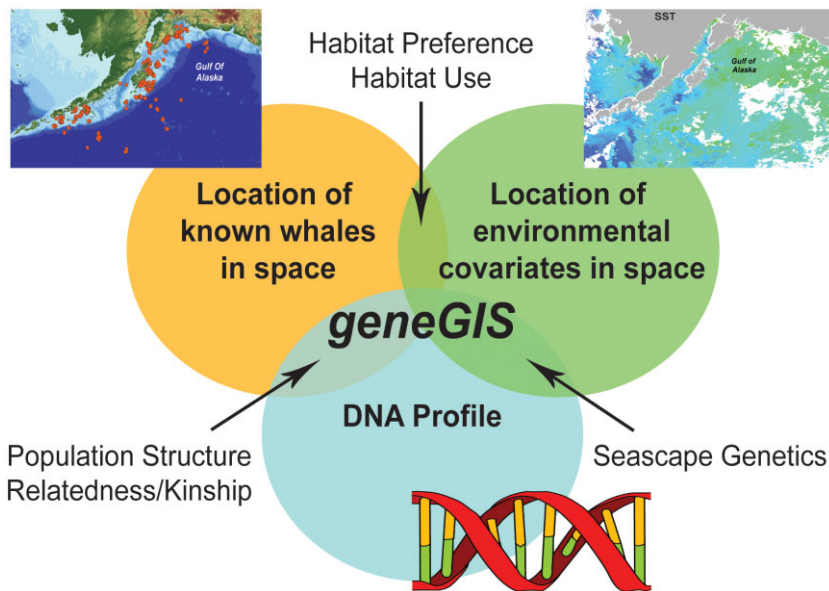


Figure 1 The conceptual framework used to illustrate how data for known individual locations, associated environmental variables, and DNA profiles can be integrated by the *geneGIS* initiative

the seascape may impact population structure. The point at which these pieces merge is *geneGIS*, an initiative that seeks to integrate spatially-explicit, individual-based data and seascape variables to better understand the patterns and processes of genetic variability in the marine environment.

2.3 Key Requirements

The success of the *geneGIS* initiative depends on several key requirements. To maximize the number of potential users, we target molecular ecologists and marine mammal scientists with little or no GIS background. Thus, tools must be easy to install and operate within ArcGIS. *geneGIS* must also be able to work with the various data types and data storage formats of our users. Most databases of genetic records are small (~ 100s to 1,000 records) and are stored in flat tabular formats such as Microsoft Excel spreadsheets. Data may consist of letters (e.g. genetic sex – M: male, F: female, U: unknown or nucleotide sequence – ATTGCAATGGCCTTA), numbers (e.g. microsatellite allele sizes – 122, 124), or alphanumeric sequences (e.g. mtDNA haplotype codes – F2, A+, E2). Photo-identification databases typically contain 1,000s of records and may or may not be stored in a relational database structure. Therefore, *geneGIS* must be able to function with the unique data types of genetic data stored in simple data tables and relational databases. For these reasons we chose a two pronged approach, a suite of GIS tools designed to function with flat tables and relational databases, plus an option to import data into a customized Arc Marine relational data model. The latter provides an additional opportunity to store and manage data in a relational database framework created specifically for marine data and can increase interoperability with other relational databases such as *Wildbook*.

3 *geneGIS* Tools

3.1 Software Platform and Tool Architecture

We developed *geneGIS* tools using the open source programming language Python (version 2.7) and ArcGIS software (version 10.1). Although a commercial product, ArcGIS is well known, widely used and considered the dominant platform used by GIS professionals (Roberts et al. 2010). As part of its built-in capabilities, Python scripts can be written to create customized geoprocessing tools that are run using simple dialog windows. This makes tools accessible to non-GIS experts and allows them to be combined with other standard ArcGIS tools for more complex spatial analyses. Moreover, because Python is an open source language, it allows GIS specialists to share and further customize scripts, a tradition that *geneGIS* builds upon by also being open source.

We used two new ArcGIS features released with ArcGIS version 10.1 – the Python toolbox and the Python add-in. A Python toolbox is a geoprocessing toolbox created entirely in Python and can be edited in any editor. Unlike script tools in custom ArcToolboxes, which are composed of three separate parts, a Python toolbox holds the parameter definitions, code validation and the source code in a single location using Python classes. From a developer's perspective, the Python toolbox provides a more streamlined environment for tool creation. Yet, from a user's perspective, a Python toolbox and tools look and function like any other.

A Python add-in is a customization that interfaces with ArcGIS for Desktop (e.g. ArcMap) to enable additional functionality for custom tasks. The add-in is created using a freely

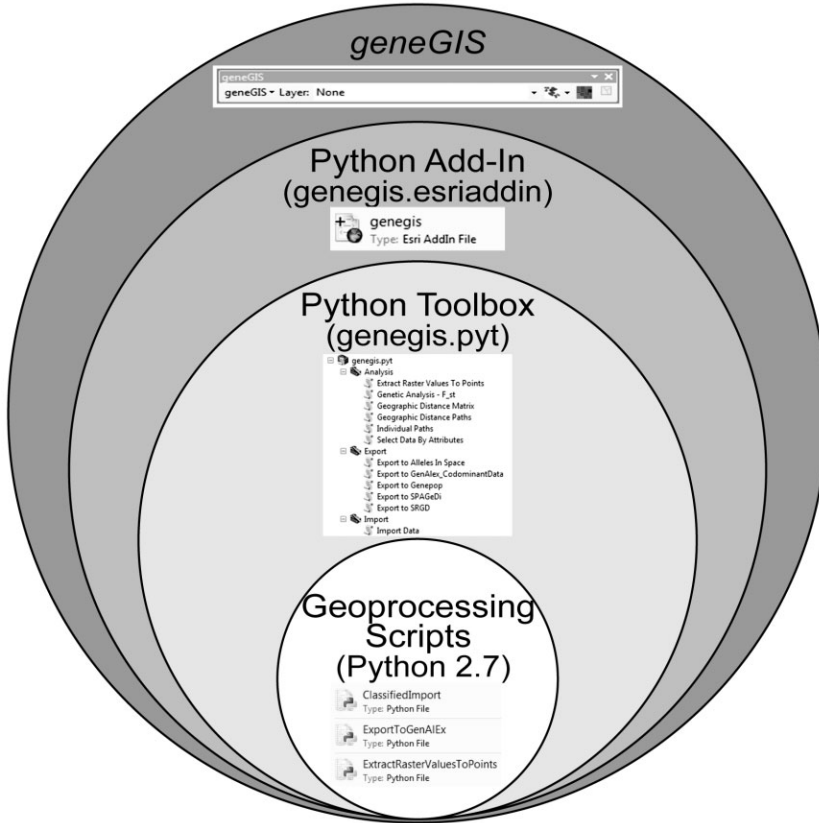


Figure 2 A diagram detailing the tool architecture of *geneGIS*. At its core are the customized geoprocessing scripts written in Python that are stored within a Python toolbox. The Python toolbox is then plugged into ArcGIS via a Python add-in to create a user-friendly, menu driven toolbar

available Python Add-In Wizard and is comprised of a single zipped package (with .esriaddin extension) containing a configuration XML file, the geoprocessing Python scripts, and any additional resource files necessary for the add-in. Add-ins are easily installed by downloading the add-in file to a user designated folder and double-clicking on the .esriaddin icon.

For our purposes, these two new features enabled an interactive environment for the user by developing a user-friendly menu driven toolbar that links to the Python toolbox containing the geoprocessing scripts (Figure 2). By containing the entire *geneGIS* program within a Python add-in, a non-GIS expert can easily download, install and use *geneGIS* with a few mouse clicks.

3.2 Standard Input File

To aid non-GIS users in importing data into *geneGIS*, we developed a standardized input file format, the Spatially Reference Genetic Data file or SRGD (Figure 3). The SRGD is a comma separated value (CSV) file and specifies the minimum data requirements necessary to use *geneGIS*. Based on expert opinion, we selected the most common data fields and formats

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Sample_ID	Individual_ID	Source_ID	Latitude	Longitude	Date_Time	Region	Sex	Haplotype	L_GATA417	L_GATA417	L_Ev37	L_Ev37	L_Ev96	L_Ev96
2	1	1001	Photo	13.9038	-90.7658	2004/03/09T11:39:00	Central America								
3	2	1001	Photo	20.6	-105.6	2006/02/16T07:40:00	Mainland Mexico								
4	3	1002	Photo	13.7991	-90.9955	2004/02/07T14:56:00	Central America	M		206		208	220	157	163
5	4	1002	Photo	10.9808	-85.8615	2004/02/16T16:25:00	Central America	M		206	222	208	220	157	163
6	5	1002	Photo/Genetic	11.0253	-85.9176	2005/01/25T07:41:00	Central America	M		206	222	208	220	157	163
7	6	1003	Photo	8.7206	-83.8183	2004/02/11T10:33:00	Central America	M	F2	206	226	196	208	147	163
8	7	1003	Photo	8.7396	-83.7873	2004/02/27T15:46:00	Central America	M	F2	206	226	196	208	147	163
9	8	1003	Photo	8.7122	-83.82	2004/03/13T00:00:00	Central America	M	F2	206	226	196	208	147	163
10	9	1003	Photo	10.96	-85.91	2004/03/18T11:35:00	Central America	M	F2	206	226	196	208	147	163
11	10	1003	Photo/Genetic	13.7851	-90.2733	2005/02/22T15:02:00	Central America	M	F2	206	226	196	208	147	163
12	11	1004	Photo/Genetic	8.708	-83.7341	2004/04/02T08:09:00	Central America	F		207	222	192	198	161	163
13	12	1004	Photo	13.769	-90.2985	2005/02/20T08:49:00	Central America	F		207	222	192	198	161	163
14	13	1004	Photo	13.8271	-90.4306	2005/02/20T09:15:00	Central America	F		207	222	192	198	161	163
15	14	1005	Photo	12.8006	-87.653	2005/01/16T15:23:00	Central America								
16	15	1006	Genetic	8.8413	-83.6498	2004/03/10T09:15:00	Central America	F	A+	214	214	192	210	161	163

Figure 3 An example of the SRGD.csv input file format for *geneGIS*. The individual-based photo and genetic databases were merged and the genetic information extended to the photo encounters. Columns A, B, and C represent fields considered to be identifiers and have the suffix *_ID*. Columns J–O represent three biallelic microsatellite loci, two columns per locus, each with a *L_* prefix. Individuals 1001 and 1005 have photo-only encounters; individual 1006 had a genetic-only encounter; and individuals 1002–1004 have had their genetic information extended to their photo-only encounters

used by molecular ecologists for inclusion in the SRGD file. Any additional data deemed necessary by the researcher (e.g. group size, behavior, etc.) may also be included. A complete description of the SRGD input file format, a sample dataset (courtesy of Cascadia Research Collective, <http://www.cascadiaresearch.org>) and tutorial are freely available for download from the *geneGIS* website, <http://genegis.org>.

3.3 *geneGIS* Tools

At the time of this writing, *geneGIS* consists of 12 tools grouped into four categories (Import, Export, Genetic Analysis and Geographic Analysis), plus a Help category that links to *geneGIS* website resources (Table 2). A key goal of *geneGIS* is to allow novel ways of data exploration through visualization, spatial selection, data extraction and basic analyses of genetic data in relation to the marine environment. This information is critical during hypothesis development for spatially-explicit analyses. We do not intend to duplicate the efforts of other specialized software packages for molecular ecology such as GenAIEx (Peakall and Smouse 2006, 2012), Genepop (Raymond and Rousset 1995; Rousset 2008), Alleles in Space (Miller 2005), and SPAGeDi (Hardy and Vekemans 2002), but instead enable exploratory analyses and data export in an appropriate format to those programs for further analyses. We also offer data export as a Keyhole Markup Language (KML) file for use with software such as Google Earth and a SRGD file format compatible for data upload into the *Wildbook* relational database management framework. In addition, we provide two tools (Summarize Encounters, Compare Encounters) invoked with buttons from the toolbar that allow the user to interactively spatially select up to two different groups of points and provide some basic statistics about that selection including the number of samples, the number of unique individuals and the number of unique individuals common to both selections.

3.4 *geneGIS* Application Examples

We use the reconciled and extended SPLASH data to illustrate a series of applications using the tools in *geneGIS* and ArcGIS to explore, develop and begin to answer spatially-explicit

Table 2 Suite of tools available in *geneGIS* release 0.2 for ArcGIS 10.1

Tool	Function	Reference Information
<i>Import</i> – Imports spatially referenced data into ArcGIS		
1. Import	Creates a new file geodatabase (if one does not exist) and imports individual-based genetic and photographic data from the SRGD.csv input file into a file geodatabase point feature class. A copy of the table is also placed in the geodatabase.	
<i>Export</i> – Exports data from ArcGIS feature class		
1. Export to Alleles in Space	Creates a file for use with Alleles in Space (AIS), software for the joint analysis of inter-individual spatial and genetic information.	Miller 2005 http://www.markgeneticsoftware.net
2. Export to GenALEX	Creates a text file formatted for use with GenALEX, an MS Excel Add-in for population genetic analyses.	Peakall and Smouse 2006, 2012 http://biology.anu.edu.au/GenALEX
3. Export to Genepop	Creates a text file formatted for use with Genepop, a web based program for population genetic analyses.	Raymond and Rousset 1995 Rousset 2008 http://genepop.curtin.edu.au
4. Export to KML	Creates a KML file, viewable in ArcGIS Explorer, ArcGlobe, and Google Earth.	
5. Export to SPAGeDi	Creates a file formatted for SPAGeDi, software for spatial pattern analysis of genetic diversity.	Hardy and Vekemans 2002 http://ebe.ulb.ac.be/ebe/SPAGeDi.html
6. Export to SRGD	Creates a SRGD formatted table (CSV) for data uploading to <i>Wildbook</i> , a software framework for mark-recapture studies.	http://www.wildme.org/wildbook
<i>Genetic Analysis</i>		
1. Calculate F-statistics	Uses SPAGeDi software to calculate a variety of F-statistics and outputs them to a text file. Text file opens upon completion.	Hardy and Vekemans 2002 http://ebe.ulb.ac.be/ebe/SPAGeDi.html

Geographic Analysis

1. **Compute Geographic Distance Matrix**
 Computes a full pairwise geodesic distance matrix between all input locations, such as encounters with individual whales. Calculations performed using Vincenty's formulae, accurate to within 0.5 mm. Output is a comma separated value (CSV) file.
2. **Compute Geographic Distance Paths**
 Computes pairwise geodesic arcs connecting all input points. The arcs represent the shortest distance (great circle distance) between locations. An attribute of the distance is also included as an output column, "Distance_in_km".
3. **Individual Paths**
 Creates individual paths, linking a selected set of individuals across all locations they have been encountered. Assumes linear movement between locations. Output is a new feature class.
4. **Extract Raster Values**
 Extracts values from one or more raster layers based on encounter (point) locations. Extracted values are added to the attribute table of the designated feature class. A new column is added for each input raster and based upon the raster name, prefixed with 'R_'.

Help

1. *geneGIS* Homepage Takes the user to the project website homepage
2. *geneGIS* Documentation Takes the user to the project documentation webpage

<http://genegis.org/>
<http://genegis.org/documentation.html>

research questions. Examples are broken down into the five current key functions of *geneGIS*: (1) data visualization; (2) spatially explore, display and select data; (3) export data; (4) data extraction from environmental layers; and (5) conduct basic spatial analyses.

3.4.1 Data Import and Data Visualization

The first step in any *geneGIS* application requires the import of georeferenced genetic data into ArcGIS. SPLASH data are formatted to meet the SRGD file specifications (Figure 3) and imported into a file geodatabase point feature class using the Import tool (Figure 4A). To provide geographical context, a base map layer can be added (Figure 4B). Initial data visualization provides the added benefit of quickly identifying and enabling the correction of questionable coordinates such as whales sighted on land.

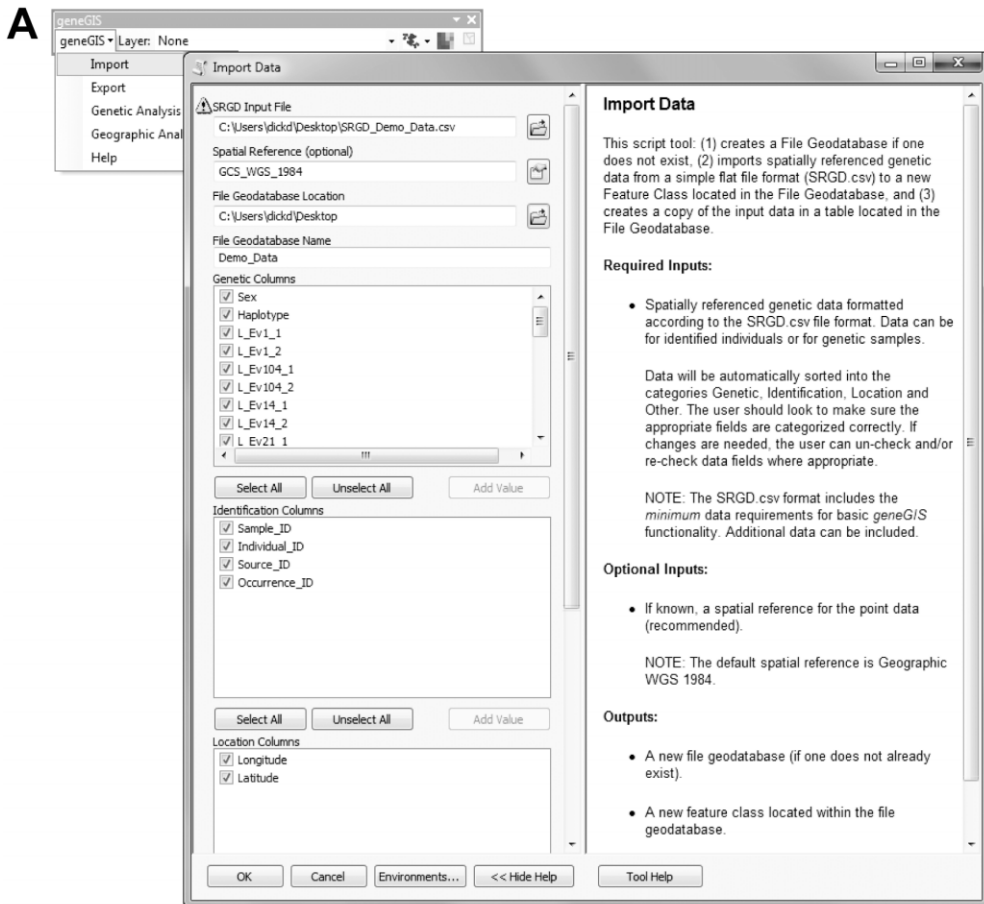


Figure 4A An example workflow using *geneGIS* – invoking the Import tool from the *geneGIS* toolbar and the dialog box filled according to the descriptive help text to the right. The warning icon next to the SRGD Input File provides a reminder that the microsatellite loci will be suffixed with an ‘_1’ or ‘_2’ when this tool is run

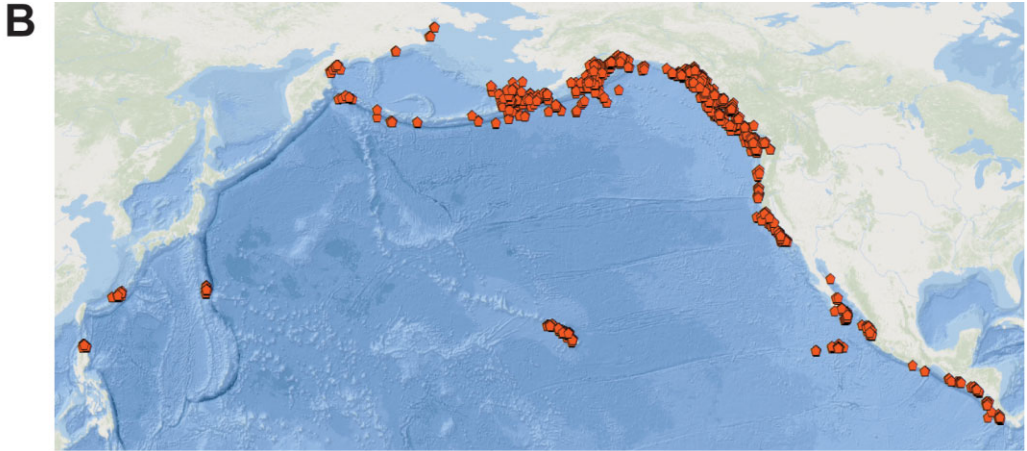


Figure 4B An example workflow for *geneGIS* – visualizing an output point feature class created from all individuals encountered during the SPLASH program from 2004–2006. For geographic context, the Esri Ocean Basemap (<http://esriurl.com/obm>, courtesy of Esri and its partners), is also added

3.4.2 Spatial Exploration, Data Selection and Data Export

These functions can assist with answering research questions such as: “Are humpback whale populations in the Western Gulf of Alaska and Southeast Alaska genetically differentiated?” Genetic differentiation between populations is a common question in molecular ecology; however, it is often limited to researcher-defined populations based on *a priori* knowledge and less often uses the specific spatial location of collected samples. To enhance the potential for using spatial exploration rather than *a priori* divisions, *geneGIS* enables the user to interactively spatially select points. Using the Summarize and Compare Encounter tools from the *geneGIS* toolbar, one group of points (“populations”) is spatially selected and briefly summarized for each area of interest (Figure 5A). Note, the text box reports on the total number of encounters and total individuals, as well as individuals found in both spatial strata (i.e. photo-ID resightings or genotypes recaptures). The Export to GenAIEx tool from the Export menu is used to export the selected data as a single file composed of the two selected populations to the text file input format required by GenAIEx v6.5 (Peakall and Smouse 2006, 2012) (Figure 5B). Additional analysis in GenAIEx, using mtDNA known to reflect maternal migration traditions, indicates the two populations are significantly differentiated ($F_{ST} = 0.197$, $p < 0.01$). To better illustrate this genetic differentiation, haplotype frequency pie charts were created within Excel (Figure 5C). Data can also be exported using the Export to Genepop tool and the output file meets the format requirements for Genepop (Raymond and Rousset 1995; Rousset 2008). In both cases, exports to GenAIEx and Genepop allow for microsatellite or mtDNA analyses of genetic differentiation.

An alternate analysis of genetic differentiation can be performed directly within ArcGIS using the program SPAGeDi (Hardy and Vekemans 2002), although currently limited to microsatellite data. In this instance, once the two spatial selections are completed, the standard ArcGIS Merge tool is used to merge the selected data into one feature class. The Calculate F-statistics tool from the Genetic Analysis menu invokes SPAGeDi to calculate F-statistics and create an output tab delimited text file that is opened directly within ArcGIS (Figure 5D).

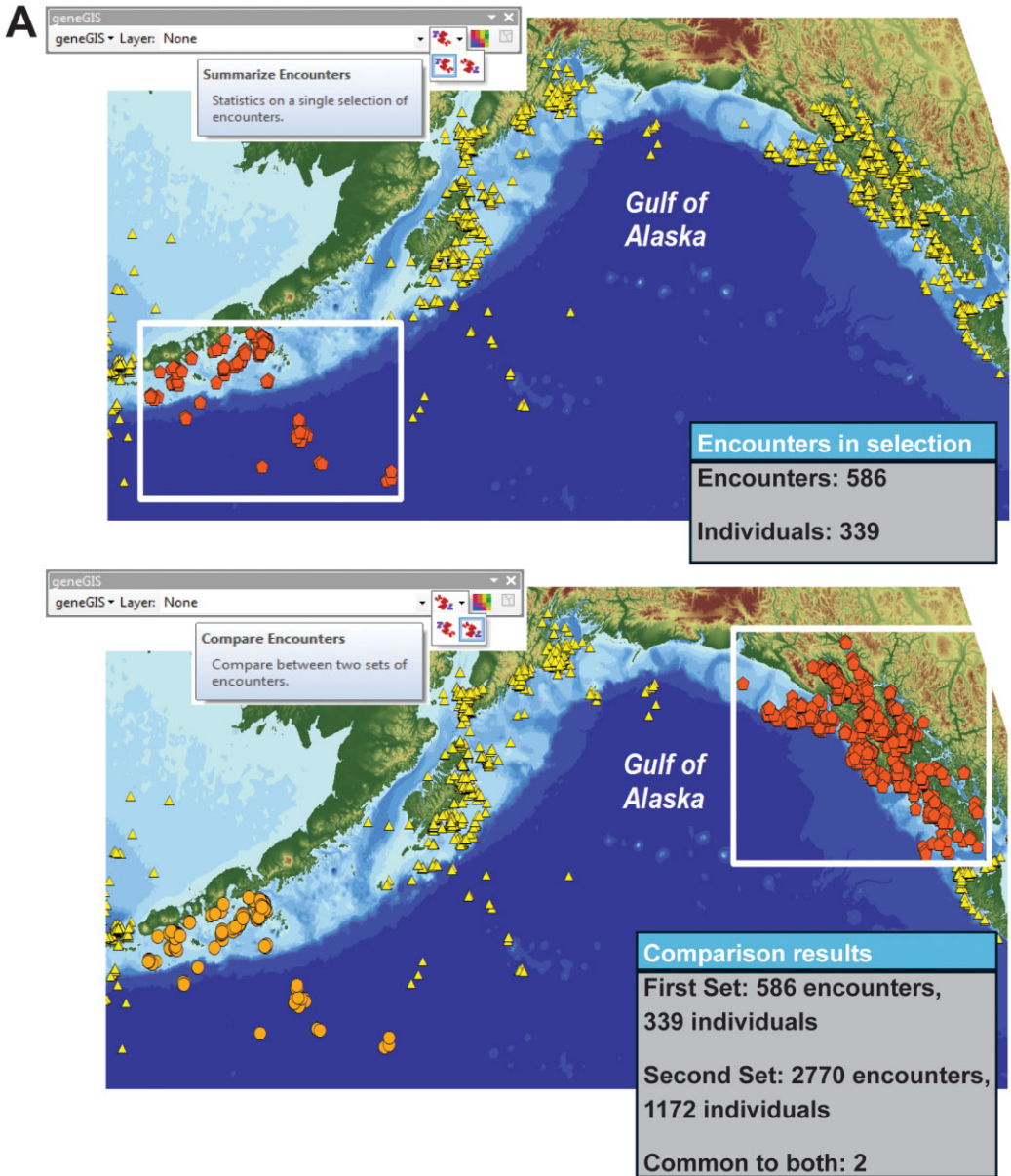


Figure 5A An example workflow using *geneGIS* – spatial selection using SPLASH data. Data are spatially selected using the Summarize Encounter (top) and Compare Encounter (bottom) tools on the *geneGIS* toolbar

3.4.3 Data Extraction from Environmental Layers

Data extraction can assist with exploring the relationship between environmental variables and individual presence/absence to assist with answering research questions such as: “Within a set of whales of known mtDNA haplotype is there any evidence of preference for particular

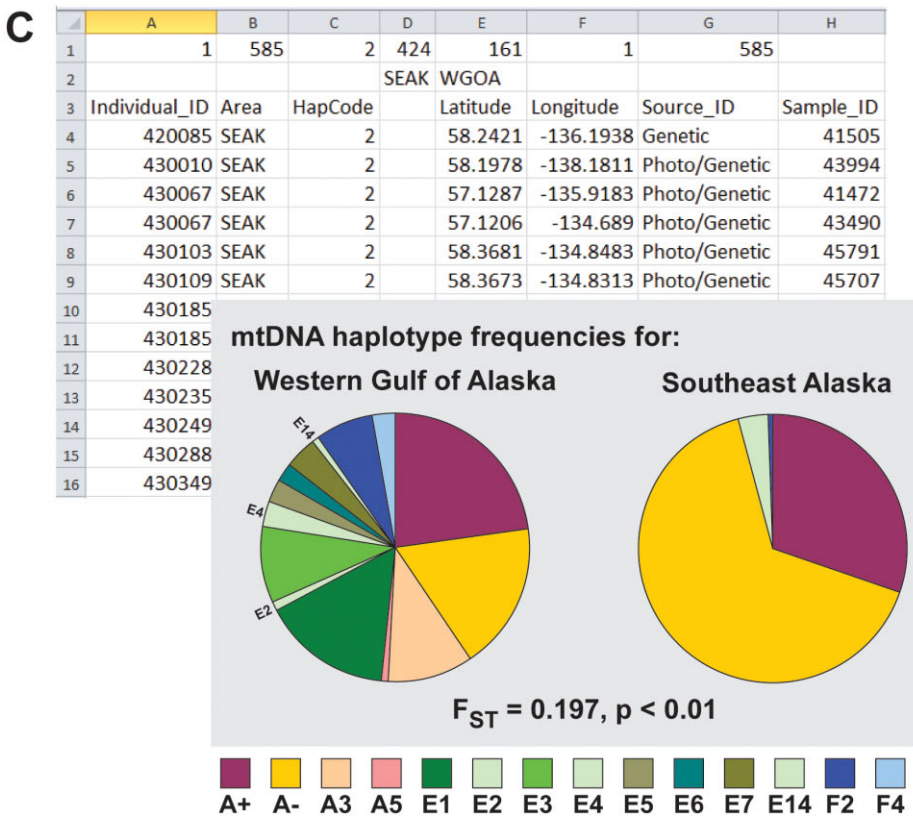
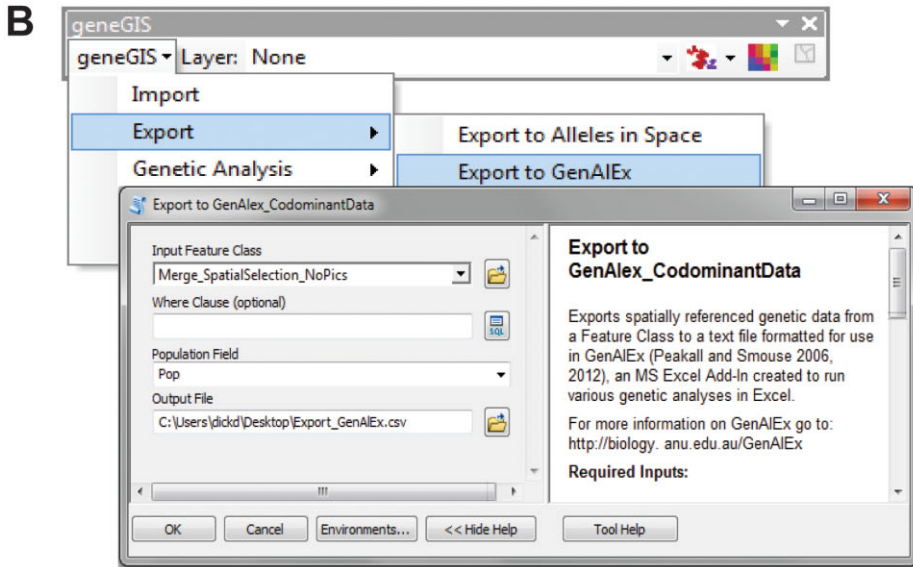


Figure 5B–C An example workflow in *geneGIS* – data export using SPLASH data. The Export to GenAIEx tool can be used for additional genetic analyses (B); a test for genetic differentiation in GenAIEx confirms the two “populations” are significantly differentiated based on mtDNA (C)

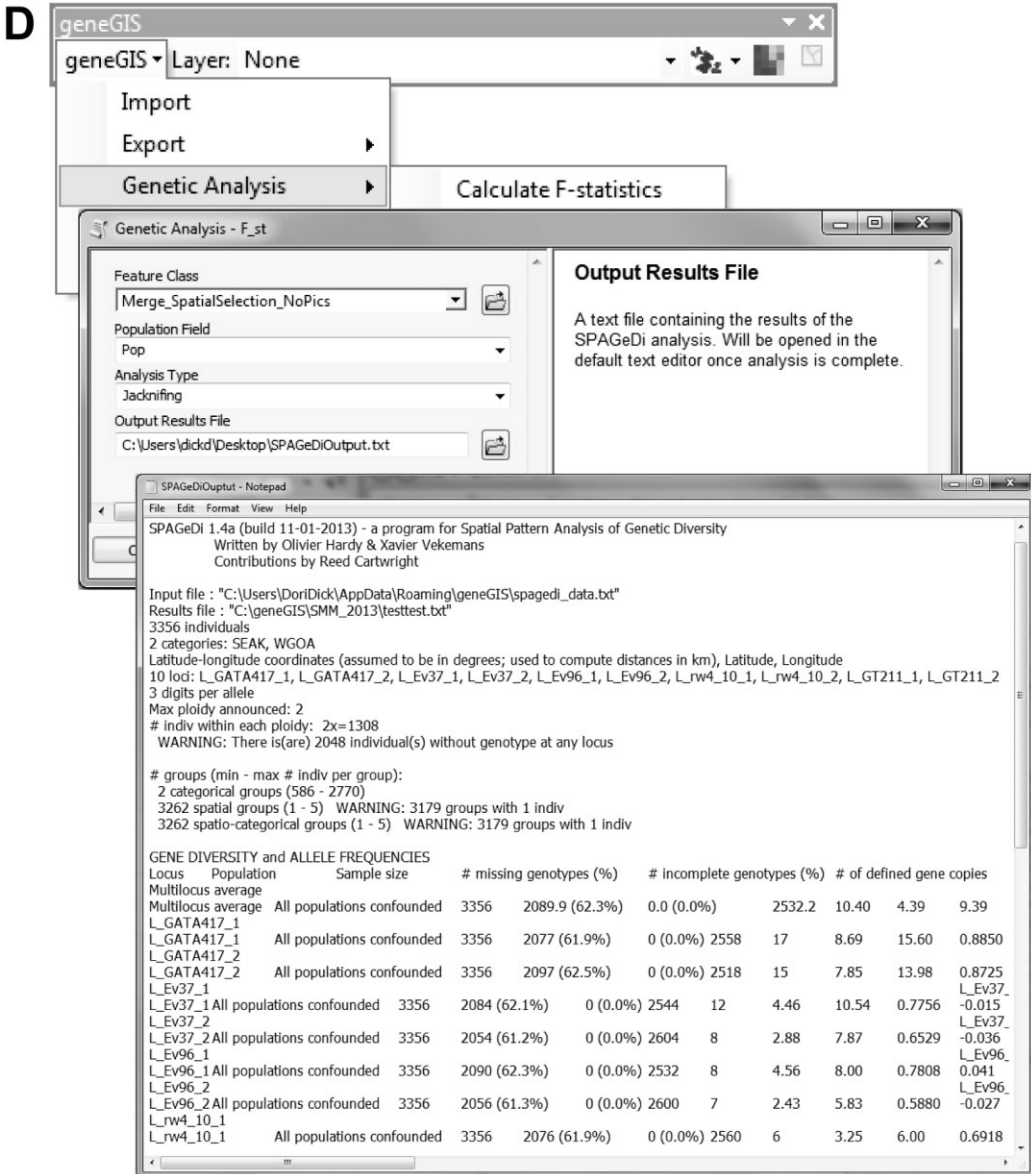


Figure 5D An example workflow in *geneGIS* – calculating F-statistics within ArcGIS using SPLASH data. The alternate method involves merging the two selected “populations” using the standard ArcGIS Merge tool followed by the Calculate F-statistics tool from the *geneGIS* toolbar to invoke SPAGeDi to test for genetic differentiation based on microsatellite DNA

depths?” In this instance, data from known individuals or mtDNA lineages can be mapped with one or more environmental raster layers such as bathymetry (e.g. the GEBCO_08 Grid, version 20100927, <http://www.gebco.net>) (Figure 6A). The Extract Raster Values tool from the Geographic Analysis menu (or toolbar button) is used to extract cell values of the bathymetry layer for each sample point location of the input feature class (Figure 6B). Extracted

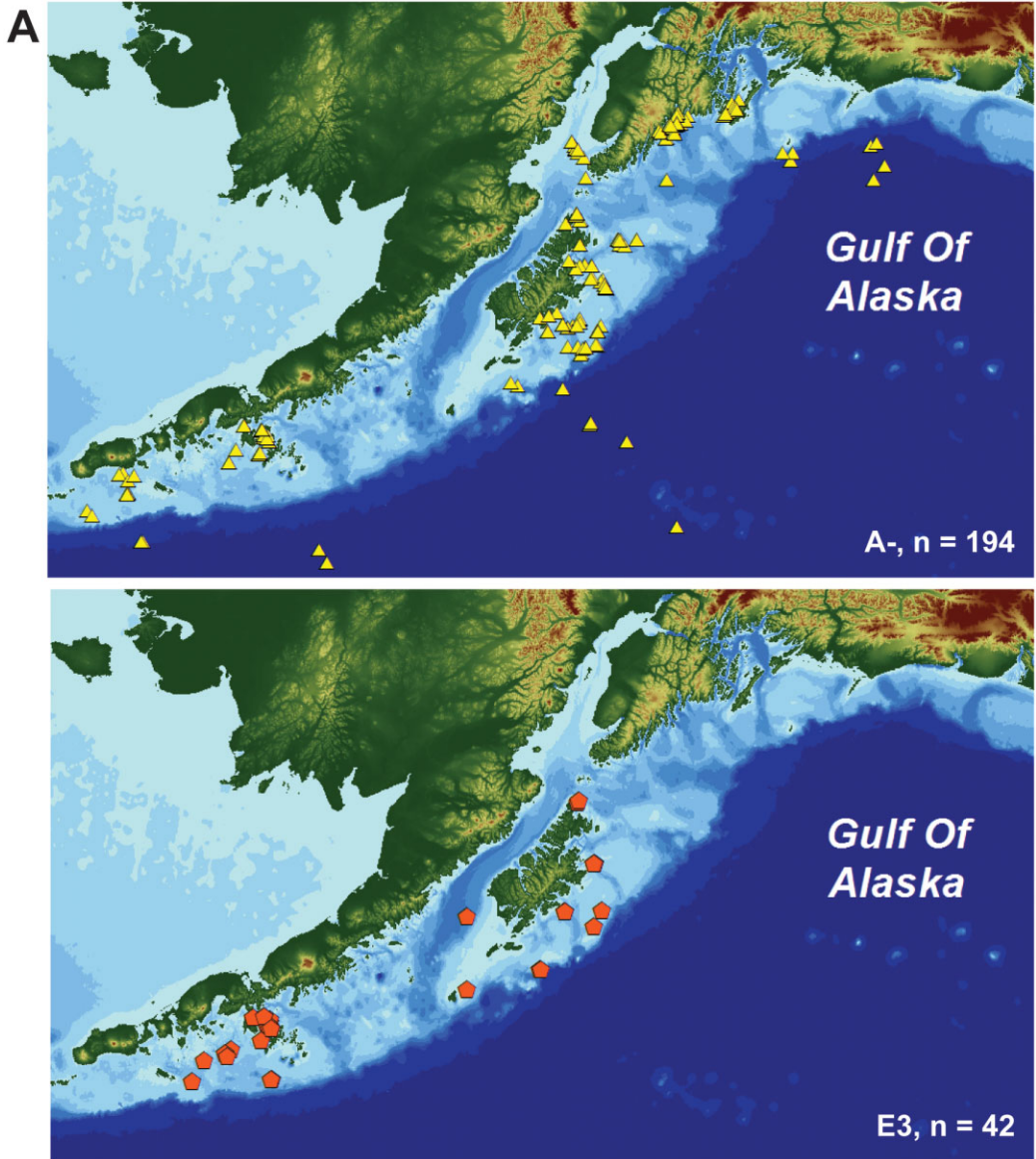


Figure 6A An example workflow in *geneGIS* – Extract Raster Values from bathymetry using SPLASH data. Individuals with known mtDNA haplotypes (A- top, E3 bottom) from the SPLASH data are mapped over a bathymetric raster layer

values are recorded to a new field in the attribute table of the feature class. New fields are named according to the raster layer used and prefixed with 'R_' (e.g. R_GEBCO). Using the standard ArcGIS table export option, the extracted values can be further analyzed in Excel or some other graphing software package to create a frequency histogram (Figure 6C). In this example, the results from the data extraction suggest that the A- and E3 mtDNA haplotypes occur more frequently at different modal depths, 110 m and 150 m, respectively.

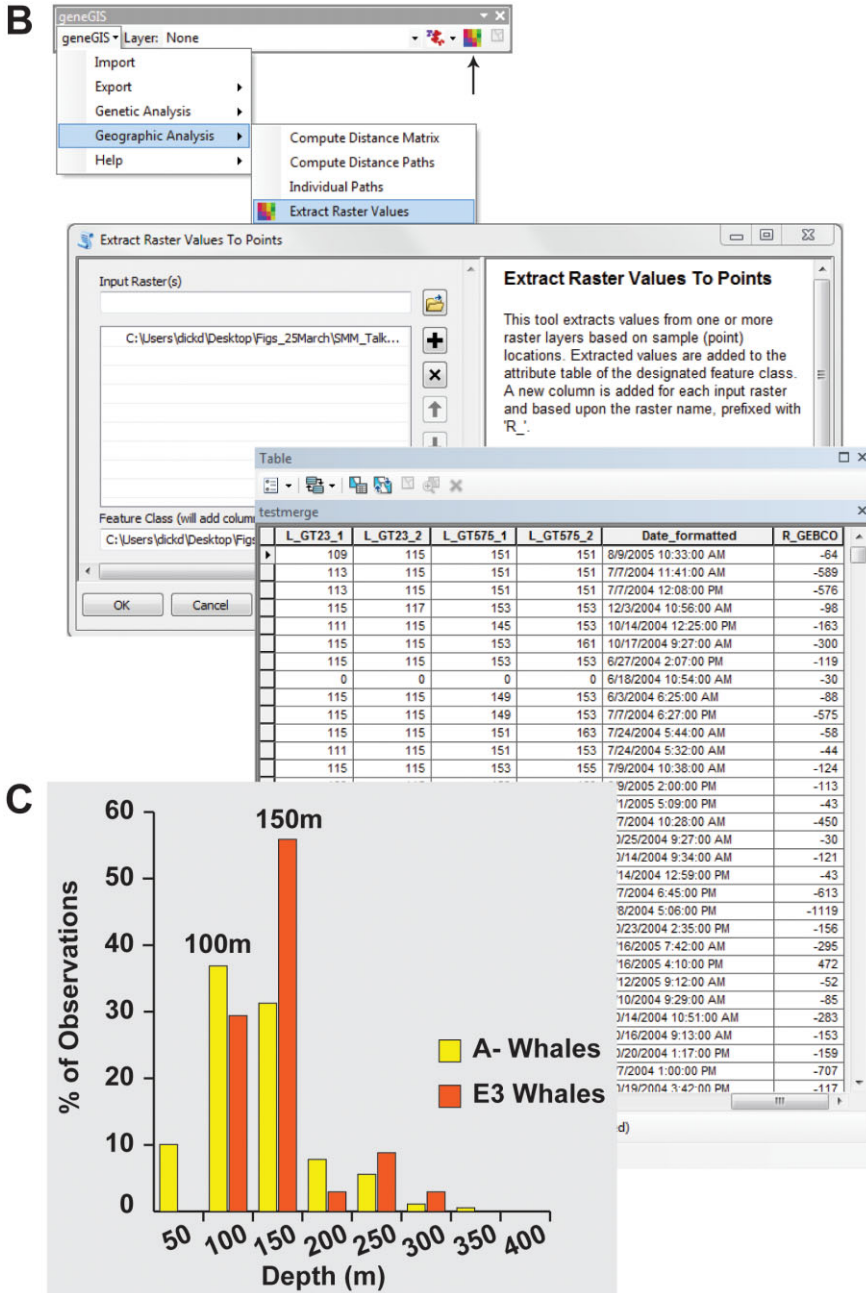


Figure 6B–C An example workflow in *geneGIS* using the Extract Raster Values tool with the SPLASH data. (B) The Extract Raster Values tool is used to extract cell values from the raster (top and middle) and record them into a new field named after the raster, R_GEBCO (bottom). (C) The data are exported using ArcGIS's table export option to enable further analysis in Excel or some other graphing software package to create a frequency histogram. In this example, the results from the data extraction suggest that the A- and E3 mtDNA haplotypes occur more frequently at different modal depths, 110 m and 150 m, respectively

3.4.4 Basic Spatial Analysis

Basic spatial analysis can assist with answering research questions such as: “How do the spatial distributions of humpback whales with different mtDNA haplotypes vary within a region?” The loading of genetic data into ArcGIS via *geneGIS* now provides the user with additional opportunities to conduct further spatial analyses using the standard default tools within ArcGIS Toolbox. For example, the Directional Distribution tool within the Spatial Statistics toolbox summarizes the central tendency, dispersion and directional trends in both the X and Y direction to visualize differences in the spatial distributions of the variable of choice (Mitchell 2005). Using one standard deviation and the same mtDNA haplotypes from above (A- and E3), the output polygons represent the location where 68% of whale encounters occurred (Figure 7) and quickly enables the visualization of different haplotype distributions.

4 Arc Marine Customization

4.1 Brief Background

Wright and Goodchild (1997) challenged the predominantly terrestrial-based GIS community to expand the capabilities of GIS to include the marine environment and the unique properties of ocean data. Released a decade later, the Arc Marine data model provided a GIS framework developed specifically for managing and mapping typical marine data types and conducting complex spatial analyses in the oceans (Wright et al. 2007). The data model produces a geodatabase resulting from the ability of the user to build validation rules, apply real-world behavior to features, and combine or link them to tables using relationship classes (Wright et al. 2007). Arc Marine is used worldwide by hundreds of researchers in marine ecology, marine geology, and marine physics (Isenor and Spears 2013). In addition, because marine research is so widely varied in the types of research conducted and the data required for that research, Arc Marine provides a common structural template that researchers can customize for their needs.

Lord-Castillo et al. (2009) provided one of the earliest Arc Marine customizations developed to map the movement and distribution of endangered whale species from satellite telemetry data. Keeping the core of the data model, this customization relies on three Arc Marine base classes: (1) the Vehicle object class to model a moving instrument carrying platform represented by the tagged animal; (2) the InstantaneousPoint feature class, subtype Location Series to hold the spatial and temporal sequence of the Argos satellite locations; and (3) the MarineEvent object class to enable dynamic sequencing of the time stamped animal movement paths to create spatial locations (Lord-Castillo et al. 2009).

We use the Lord-Castillo et al. (2009) customization as the starting point for our customization for two reasons. First, the Lord-Castillo et al. (2009) structure already considers the concept of an “individual”. Although the identity of the whale might be unknown relative to the population, it can be used as a means to recognize the one-to-one relationship between a whale and a satellite tag. Second, it provides the flexibility of merging the two customizations together at some point in the future if satellite telemetry data are added to the reconciled photo-identification and genetic databases.

4.2 Customization Specifics

To include individual-based genetic and photographic data within the Arc Marine framework, we retain nine key Arc Marine classes and populate them as illustrated in Figure 8. The Cruise

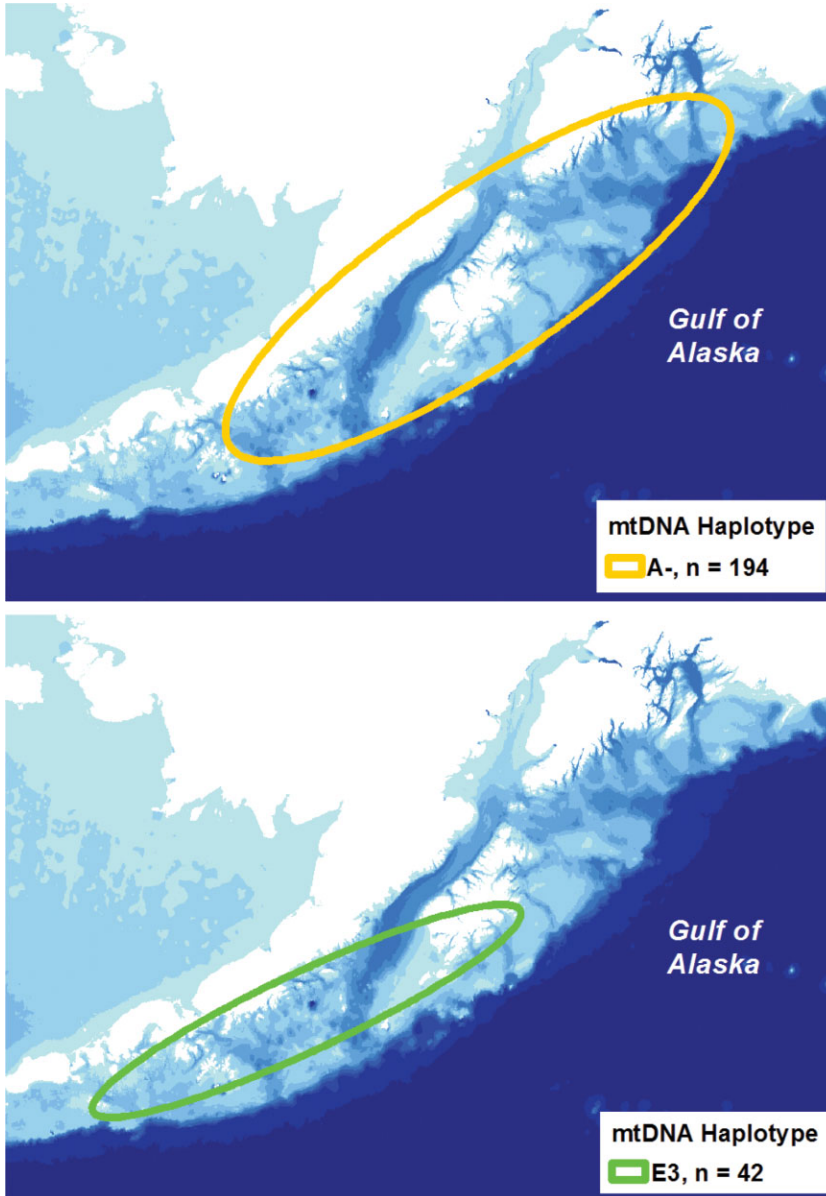


Figure 7 An example of a basic spatial analysis available as a result of loading genetic data into ArcGIS via *geneGIS*. Output from the Directional Distribution tool, a standard default tool from ArcGIS within the Spatial Statistics toolbox, displays the spatial distribution trends for the A- (top) and E3 (bottom) mtDNA haplotypes using one standard deviation ellipse

and SurveyInfo classes are preserved, containing information related to the specific cruise and survey, while the MarineEvent class is used to record the Occurrence, a point in time and space when one or more whales are observed (Figure 8A). Similar to Lord-Castillo et al. (2009), the Vehicle class represents the animal, but is further specified as a known individual with an

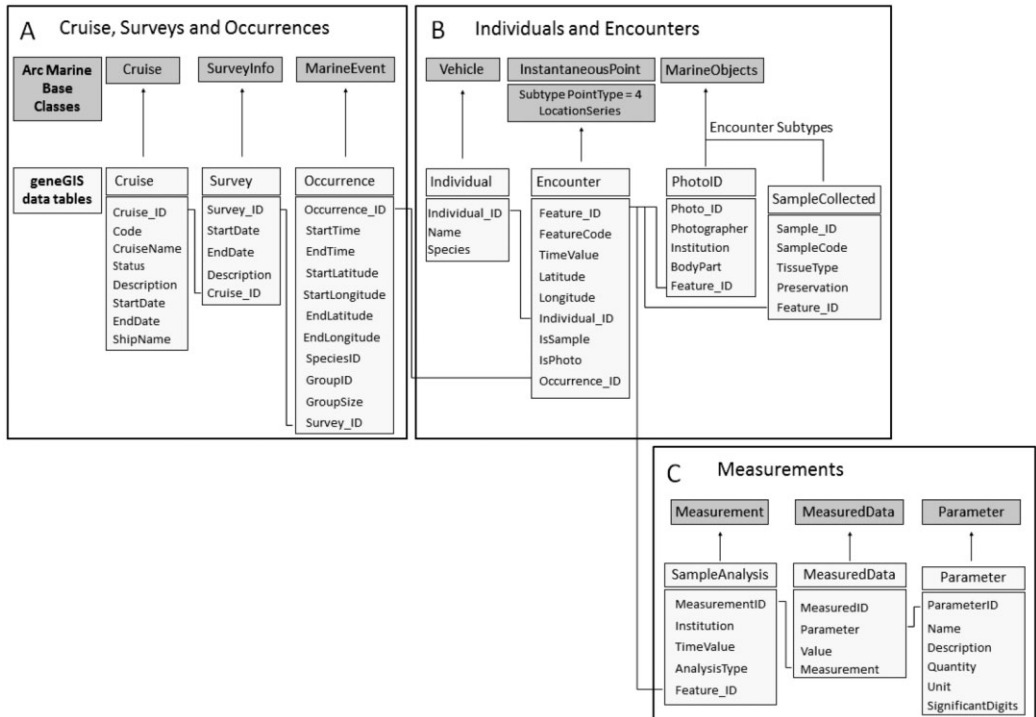


Figure 8 Generalized diagram of an Arc Marine customization for individual-based genetic and photo-identification data from cetaceans. Illustrations of how *geneGIS* data tables fit into Arc Marine base classes for: (A) Cruise, Surveys and Occurrences, (B) Individuals and Encounters (PhotoID or Samples) and (C) Measurements (SampleAnalysis, MeasuredData and Parameter). This customization developed as part of the *geneGIS* initiative

assigned identity. The InstantaneousPoint feature class subtype LocationSeries represents the Encounter, a discrete point in time and space at which an individual is sampled, while the MarineObjects class is used to define two Encounter Subtypes – PhotoID and SampleCollected (Figure 8B). Depending on the Encounter subtype, the Measurement class is used to define the type of SampleAnalysis conducted, while the MeasuredData and Parameter classes hold the information related to the analyses outputs (e.g. sex, mtDNA haplotype, microsatellite alleles) (Figure 8C).

4.3 SPLASH Implementation

The application of the Arc Marine customization for reconciled genetic and photo-identification SPLASH data is shown in Figure 9 and described here. During the three-year SPLASH program, there were multiple research cruises. Each Cruise is given a unique identifier and the table is populated with all relevant cruise information. During a single cruise, there are multiple daily Surveys, each with its own identifier, and within each survey, whenever a group of whales is sighted, there is an Occurrence and information about the group is recorded (Figure 9A). An Occurrence may lead to a related Encounter when either of the two

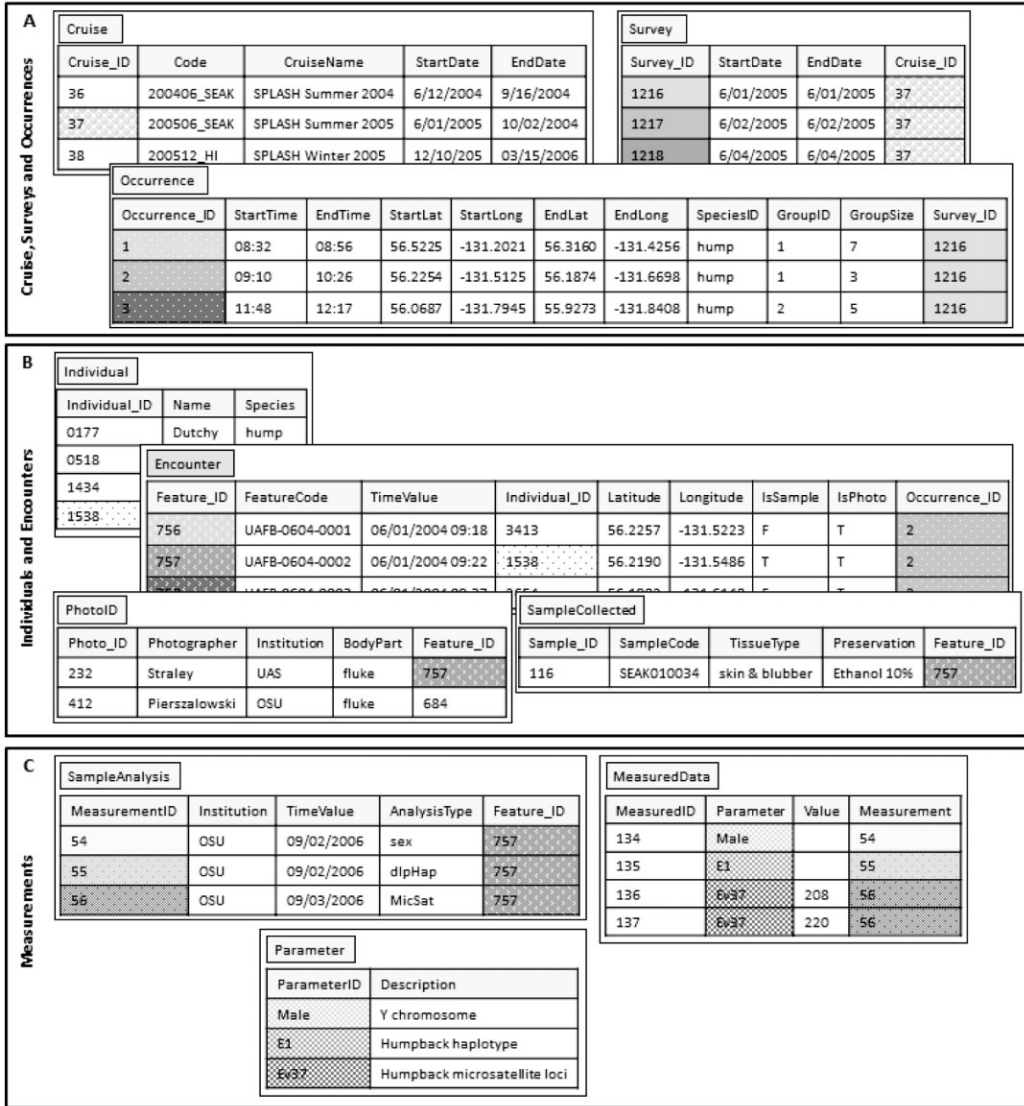


Figure 9 A diagram of how individual-based genetic and photo-identification SPLASH data fit in the customized *geneGIS* Arc Marine data model for (A) Cruise, Surveys and Occurrences; (B) Individuals and Encounters; and (C) Measurements. Tables are related according to similar grey scale colored data fields

Encounter Subtypes, PhotoID or SampleCollected, takes place (Figure 9B). The type of analyses (SampleAnalysis) and the subsequent results (MeasuredData and Parameter) of the data collected from the Encounter Subtypes (Figure 9C) provides the information necessary to assign a unique Individual identity (Figure 9B). Figure 10A shows the data loaded into the ArcMarine customization using the classes outlined above. Note that there is a single feature

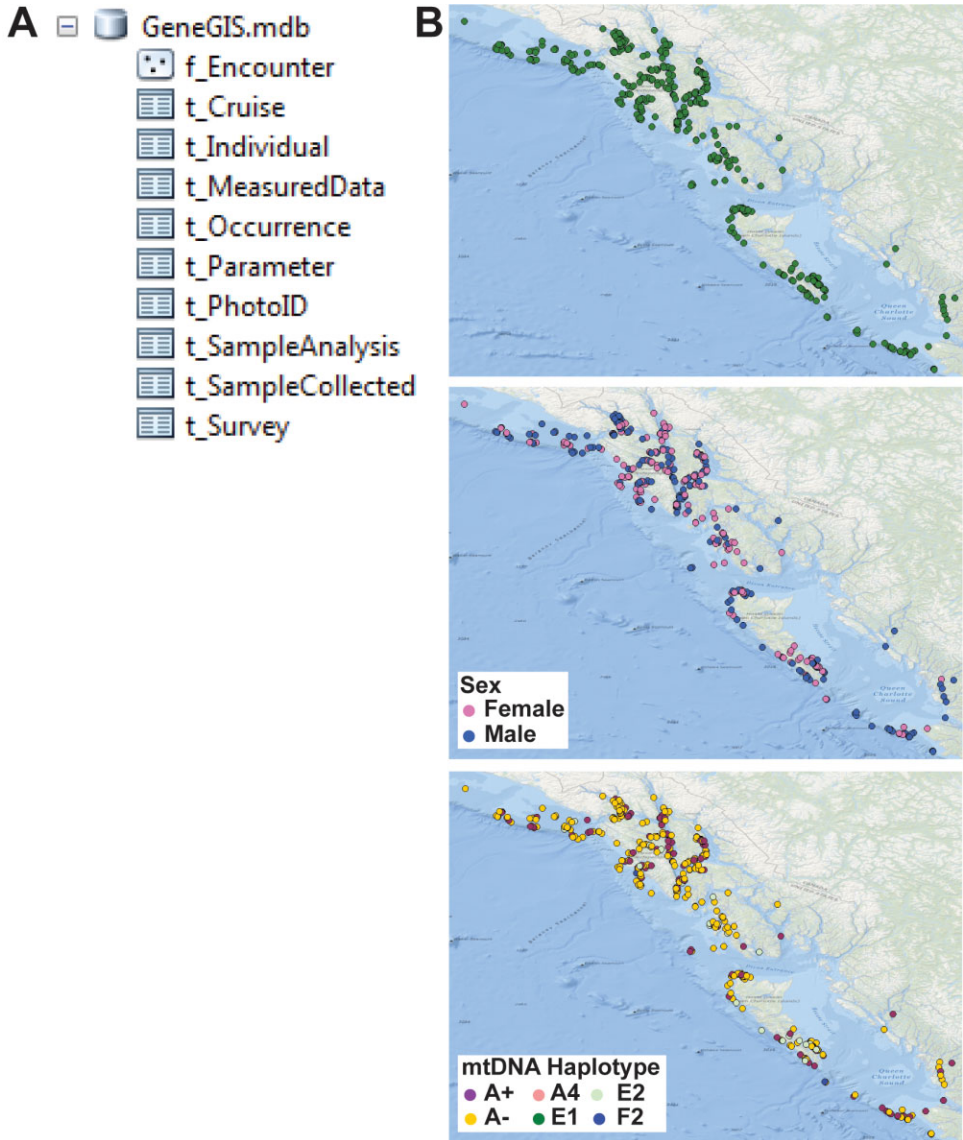


Figure 10 (A) SPLASH data loaded into the customized Arc Marine data model, showing one point feature class (f_Encounter) and nine tables tailored to handle reconciled photo-identification and genetic data associated with long-term cetacean studies. (B) ArcMap screen shots of humpback whale encounters from SE Alaska and Northern British Columbia loaded into the geodatabase by spatial location (top), by sex (middle) and by mtDNA haplotype (bottom). Basemap courtesy of Esri, <http://esriurl.com/obm>, and its partners

class (f_Encounter) while there are nine tables (denoted with a t_). Figure 10B shows the point locations of humpback whales off Southeast Alaska and Northern British Columbia mapped by spatial location (top), by sex (middle) and by mitochondrial haplotype (bottom) from the geodatabase.

5 Visualizing the Spatial Distribution of Humpback Whale mtDNA Haplotypes

Genetic analyses using both mtDNA haplotypes and microsatellite loci reveal North Pacific humpback whales have a complex population structure (Baker et al. 1998, 2013). Significant genetic differentiation occurs among breeding and feeding grounds. Further, although humpback whales show strong site fidelity to both breeding and feeding grounds, there is greater mtDNA haplotype diversity on some feeding grounds, suggesting that a very different population structure occurs while whales are feeding compared to breeding (Baker et al. 2013). Such findings have a number of important conservation implications, including the recognition that protective measures based solely on the breeding grounds will not successfully capture the species' genetic diversity.

In the Gulf of Alaska, for example, mtDNA diversity is high and population boundaries are not obvious, confounding conventional molecular ecology methods that require researchers to define spatial strata *a priori* (Beebee and Rowe 2008; Baker et al. 2013). Consideration of the spatial component, such as the distribution of individual animals across space using explicit geographic coordinates, is relatively rare. By incorporating spatially-explicit genetic data using *geneGIS*, the missing spatial component can be included. In addition, it allows for further analyses that incorporate environmental data (e.g. sea surface temperature, bathymetry, etc.) to explore the relationship between population divisions and the seascape.

Using *geneGIS* we build upon Section 3.4.3 to present one possible method using spatially-explicit encounters of known individuals to explore the spatial distribution of mtDNA haplotypes. SPLASH data following the SRGD.csv format are imported into ArcGIS using the Import tool from the *geneGIS* toolbar. Although this data format includes the field 'Region' as a means to provide some locational information, this represents researcher-defined strata and we purposely choose not to use it. Instead, we use the Summarize Encounter button to spatially select the points located in the region of interest – Northern and Western Gulf of Alaska. Whale encounters are mapped by haplotype to visually demonstrate the high diversity of mtDNA haplotypes (Figure 11A). The Directional Distribution tool within the Spatial Statistics toolbox is used to measure the orientation and direction of the haplotype distributions. Of the 18 haplotypes recorded in this area, ellipses using one standard deviation for nine ($n \geq 10$) haplotypes are calculated (Figure 11B). Although the visual interpretation of plotting the encounters by haplotype (Figure 11A) may provide a sense of orientation, the standard deviation ellipse analysis makes the trend in haplotype distribution clear while also using statistical calculation (Figure 11B) (Mitchell 2005). As a next step this information can be combined with various environmental variables deemed important to humpback whales on their foraging grounds to begin to answer spatially-explicit ecological questions related to pattern and process.

6 Conclusions

geneGIS is the first suite of ArcGIS tools and the first customized Arc Marine data model to incorporate and analyze individual-based genetic data in a seascape context. The suite of tools in *geneGIS* provide novel methods of data visualization, spatial selection, data extraction and spatial analyses to the field of molecular ecology, while the customization of the data model to include these data types will provide the opportunity to link with other data sets and tools created by the broader marine GIS community. The inclusion of the spatial component moves the visualization and analyses of population structure data beyond traditional descriptive text

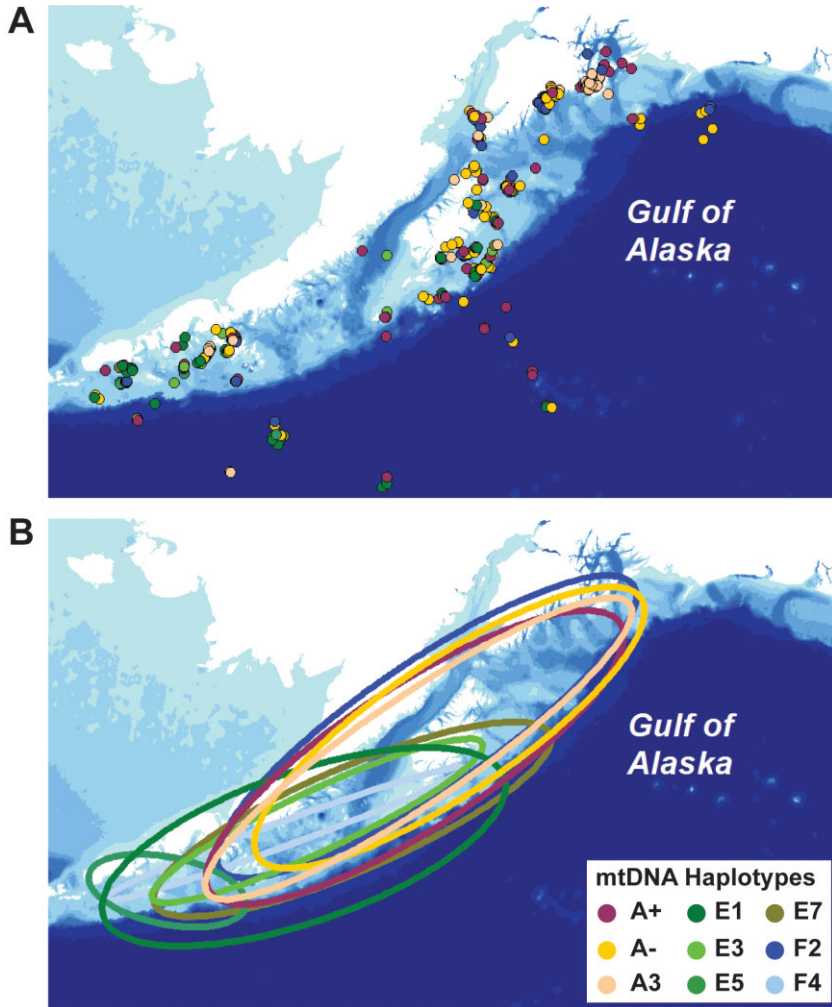


Figure 11 Screenshots of spatially selected humpback whale encounters, mapped by (A) mtDNA haplotype and (B) calculated ellipses with one standard deviation showing the spatial distribution of humpback whale haplotypes in the Northern and Western Gulf of Alaska

and pie charts of haplotype/allele frequencies based on researcher-defined boundaries. In this way, researchers are now better equipped to pose and answer questions using environmental information relevant to the study species in geographic space, which will be increasingly important as the marine environment continues to change due to anthropogenic modifications and global climate change.

This process revealed three primary directions to help guide future development of GIS tools for molecular ecology research. First, although we did not want to duplicate the efforts of already existing analytical packages (e.g. GenAlEx, Genepop), providing the option to calculate some of the more common genetic analyses (e.g. F-statistics for both microsatellite and mitochondrial DNA, Mantel tests) directly within *geneGIS* reduces the need to move back and forth between software packages and the need to learn additional applications. Second,

there is a need to build upon the strengths of GIS and improve the accessibility of data visualization methods and spatial analyses available to non-GIS users. This could include developing methods to calculate a continuous raster surface of relatedness/kinship (e.g. heat maps or kernel distributions) across a landscape. Finally, although the user base we target is likely to know what environmental variables to include, they may be unaware of how to acquire them. Thus, it would be useful to include a method for improved access to relevant environmental layers. One possibility would be to develop a button on the *geneGIS* toolbar that links the user to some of the common websites where such data are available for download. Perhaps even more useful might be to provide a direct link to the ArcGIS compatible Marine Geospatial Ecology Tools (MGET) developed by Roberts et al. (2010) that provide easy-to-use, open-source geoprocessing tools to access oceanographic data in addition to other forms of spatial environmental analyses.

The customization of Arc Marine also revealed two areas worth considering for further development of both this customization and the Arc Marine data model. First, because the customization of Arc Marine brings a new user group to the GIS table, it would be useful to develop a tool to enable easy data import into the *geneGIS* Arc Marine data model. Second, the Arc Marine data model is based on an older data structure, the personal geodatabase, for which a number of issues including file size limitations and stability have been identified. Revising the Arc Marine data model to the now Esri recommended standard, a file geodatabase, would resolve these issues and allow for a smoother interface between the Arc Marine customization for genetic data and the *geneGIS* tools.

7 Availability

geneGIS is freely available for download from the website <http://genegis.org>. Source code is hosted on Github (<https://github.com/genegis/genegis>). The *geneGIS* website contains an online manual, installation instructions, a tutorial with sample data, conference presentations, and a list of relevant literature.

References

- Baird R W and Whitehead H 2000 Social organization of mammal-eating killer whales: Group stability and dispersal patterns. *Canadian Journal of Zoology* 78: 2096–105
- Baker C S, Herman L, Perry A, Lawton W, Straley J, Wolman A, Winn H, Hall J, Kaufman G, Reinke J, and Ostman J 1986 The migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the central and eastern North Pacific. *Marine Ecology Progress Series* 31: 105–19
- Baker C S, Medrano-Gonzalez L, Calambokidis J, Perry A, Pichler F, Rosenbaum H, Straley J M, Urbán-Ramirez J, Yamaguchi M, and von Ziegeler O 1998 Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Molecular Ecology* 6: 695–707
- Baker C S, Steel D, Calambokidis J, Falcone E, González-Peral U, Barlow J, Durdin A M, Clapham P J, Ford J K B, Gabriele C M, Mattila D, Rojas-Bracho L, Straley J M, Taylor B L, Urbán J R, Wade P R, Weller D, Witteveen B H, and Yamaguchi M 2013 Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. *Marine Ecology Progress Series* 494: 291–306
- Barlow J, Calambokidis J, Falcone E A, Baker C S, Burdin A M, Clapham P J, Ford J K B, Gabriele C M, Leduc R, Mattila D K, Quinn T J, Rojas-Bracho L, Straley J M, Taylor B L, Urbán R J, Wade P, Weller D, Witteveen B H, and Yamaguchi M 2011 Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies. *Marine Mammal Science* 27: 793–818

- Beebee T J C and Rowe G 2008 *An Introduction to Molecular Ecology*. Oxford, UK, Oxford University Press
- Best P B 1993 Increase rates in severely depleted stocks of baleen whales. *ICES Journal of Marine Science* 50: 169–86
- Calambokidis J, Falcone E A, Quinn T J, Burdin A M, Clapham P J, Ford J K B, Gabriele C M, LeDuc R, Mattila D, Rojas-Bracho L, Straley J M, Taylor B L, Urbán J, Weller D, Witteveen B H, Yamaguchi M, Bendlin A, Camacho D, Flynn K, Havron A, Huggins J, Maloney N, Barlow J, and Wade P R 2008 SPLASH: Structure of Populations, Levels of Abundance and Status of Humpback Whales in the North Pacific. Washington, D.C., U.S. Department of Commerce Report AB133F-03-RP-00078
- Clapham P J 2009 Humpback whale (*Megaptera novaeangliae*). In Perrin W F, Würsig B, and Thewissen J G M (eds) *Encyclopedia of Marine Mammals* (Second Edition). Amsterdam, The Netherlands, Academic Press: 582–85
- CBD 2012 Decisions Adopted by the Conference of the Parties to the Convention on Biological Diversity. WWW document, <http://www.cbd.int/decisions/cop/?m=cop-11>
- Dalla Rosa L, Ford J K, and Trites A W 2012 Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Continental Shelf Research* 36: 89–104
- Etherington T R 2011 Python based GIS tools for landscape genetics: Visualizing genetic relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution* 2: 52–5
- Ford J K B, Ellis G M, and Balcomb K C 2000 *Killer Whales: The Natural History and Genealogy of Orcinus orca in British Columbia and Washington*. Vancouver, University of British Columbia Press
- Hammond P S, Mizroch S A, and Donovan G P 1990 Report of the workshop on individual recognition and the estimation of cetacean population parameters. In Hammond P S, Mizroch S A, and Donovan G P (eds) *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters*. Cambridge, Report of the International Whaling Commission Special Issue 12: 3–17
- Hardy O J and Vekemans X 2002 SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–20
- Isenor A W and Spears T W 2013 Combining the Arc Marine framework with geographic metadata to support ocean acoustic modeling. *Transactions in GIS* 18: 183–200
- Johnson J H and Wolman A A 1984 The humpback whale. *Marine Fisheries Review* 46: 30–7
- Lord-Castillo B, Wright D J, Mate B R, and Follett T 2009 A customization of the Arc Marine data model to support whale tracking via satellite telemetry. *Transactions in GIS* 13(s1): 63–83
- Manel S, Schwartz M K, Luikart G, and Taberlet P 2003 Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18: 189–97
- Markowitz T M, Harlin A D, and Würsig B 2003 Digital photography improves efficiency of individual dolphin identification. *Marine Mammal Science* 19: 217–23
- Miller M 2005 Alleles In Space: Computer software for the joint analysis of inter-individual spatial and genetic information. *Journal of Heredity* 96: 722–23
- Mitchell A 2005 *The ESRI Guide to GIS Analysis* (Volume 2). Redlands, CA, Esri Press
- Noren D P and Mocklin J A 2012 Review of cetacean biopsy techniques: Factors contributing to successful sample collection and physiological and behavioral impacts. *Marine Mammal Science* 28: 154–99
- Parks D H, Mankowski T, Zangooei S, Porter M S, Armanini D G, Baird D J, Langille M G I, and Beiko R G 2013 GenGIS 2: Geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. *PLOS One* 8: 1–10
- Peakall R and Smouse P E 2006 GENALEX 6: Genetic analysis in Excel, Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–95
- Peakall R and Smouse P E 2012 GenAlEx 6.5: Genetic analysis in Excel, Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537–39
- Primack R B 2010 *Essentials of Conservation Biology*. Sunderland, UK, Sinauer Associates
- Rasmussen K, Palacios D M, Calambokidis J, Saborío M T, Dalla Rosa L, Secchi E R, Steiger G H, Allen J M, and Stone G S 2007 Southern hemisphere humpback whales wintering off Central America: Insights from water temperature into the longest mammalian migration. *Biology Letters* 3: 302–5
- Reeves R R, Smith B D, Crespo E A, and di Sciara D N (compilers) 2003 *Dolphins, Whales and Porpoises: 2002–2010 Conservation Action Plan for the World's Cetaceans*. Gland, Switzerland, IUCN/SSC Cetacean Specialist Group
- Reilly S B, Bannister J L, Best P B, Brown M, Brownell Jr R L, Butterworth D S, Clapham P J, Cooke J, Donovan G P, Urbán J and Zerbini A N 2008 *Megaptera novaeangliae*. WWW document, <http://www.iucnredlist.org>
- Roberts J J, Best B D, Dunn D C, Trembl E A, and Halpin P N 2010 Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling and Software* 25: 1197–207

- Raymond M and Rousset F 1995 GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–49
- Rousset F 2008 Genepop'007: A complete reimplementaion of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–06
- Stevick P T, Allen J, Clapham P J, Friday N, Katona S K, Larsen F, Lein J, Mattila D K, Palsbøll P J, Sigurjónsson J, Smith T D, Øien N, and Hammond P S 2003 North Atlantic humpback whale abundance and rate of increase four decades after protection from whaling. *Marine Ecology Progress Series* 258: 263–73
- Storfer A, Murphy M A, Evans J S, Goldberg C S, Robinson S, Spear S F, Dezzani R, Delmelle E, Vierling L, and Waits L P 2007 Putting the 'landscape' in landscape genetics. *Heredity* 98: 128–42
- Vandergast A G, Perry W M, Lugo R V, and Hathaway S A 2011 Genetic landscapes GIS toolbox: Tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* 11: 158–61
- Wagner H H, Murphy M A, Holderegger R, and Waits L 2012 Developing an interdisciplinary, distributed graduate course for twenty-first century scientists. *Bioscience* 62: 182–8
- Wright D J and Goodchild M F 1997 Data from the deep: Implications for the GIS community. *International Journal of Geographical Information Science* 11: 523–28
- Wright D J, Blongwicz M J, Halpin P N, and Breman J 2007 *Arc Marine GIS for a Blue Planet*. Redlands, CA, Esri Press
- Würsig B, Perrin W F, and Thewissen J G M 2009 History of marine mammal research. In Perrin W F, Würsig B, and Thewissen J G M (eds) *Encyclopedia of Marine Mammals* (Second Edition). Amsterdam, The Netherlands, Academic Press: 565–69