

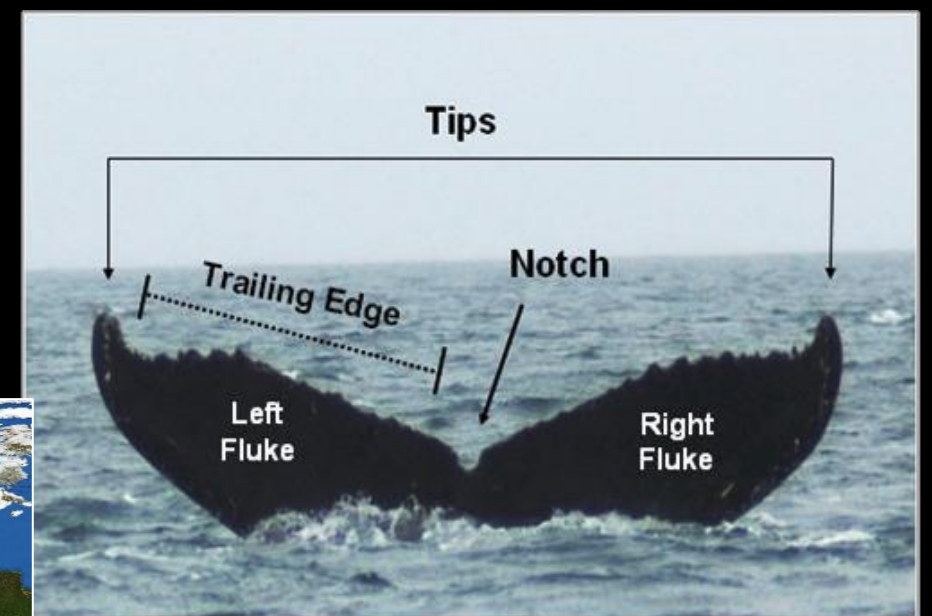
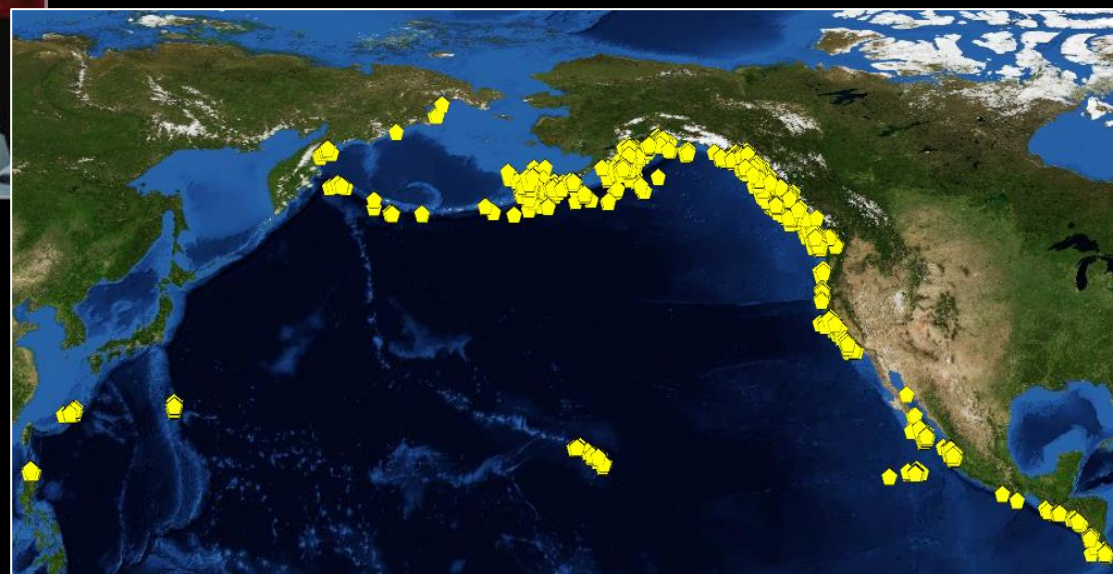
geneGIS:

Computational tools for spatial analyses of DNA profiles
with associated photo-identification records

Dori Dick, Shaun Walbridge, Dawn Wright, John Calambokidis,
Erin Falcone, Debbie Steel, Tomas Follett,
Jason Holmberg, C. Scott Baker



<http://www.nationalgeographicstock.com/ngsimages/explore/explorecomp.jsf?xsys=SE&id=1231238>



<http://www.afsc.noaa.gov/ABL/Humpback/WhyPhoto.htm>

The Problem...

Cetacean research:

- Individual-based studies using photo-identification and genetics are becoming more common

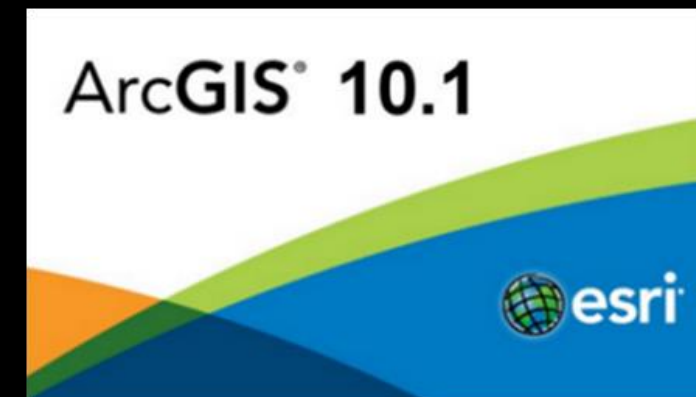


- Integration of these databases is rare
- Few tools exist to handle, explore or visualize the spatial patterns of such data

The Solution...

Our Approach:

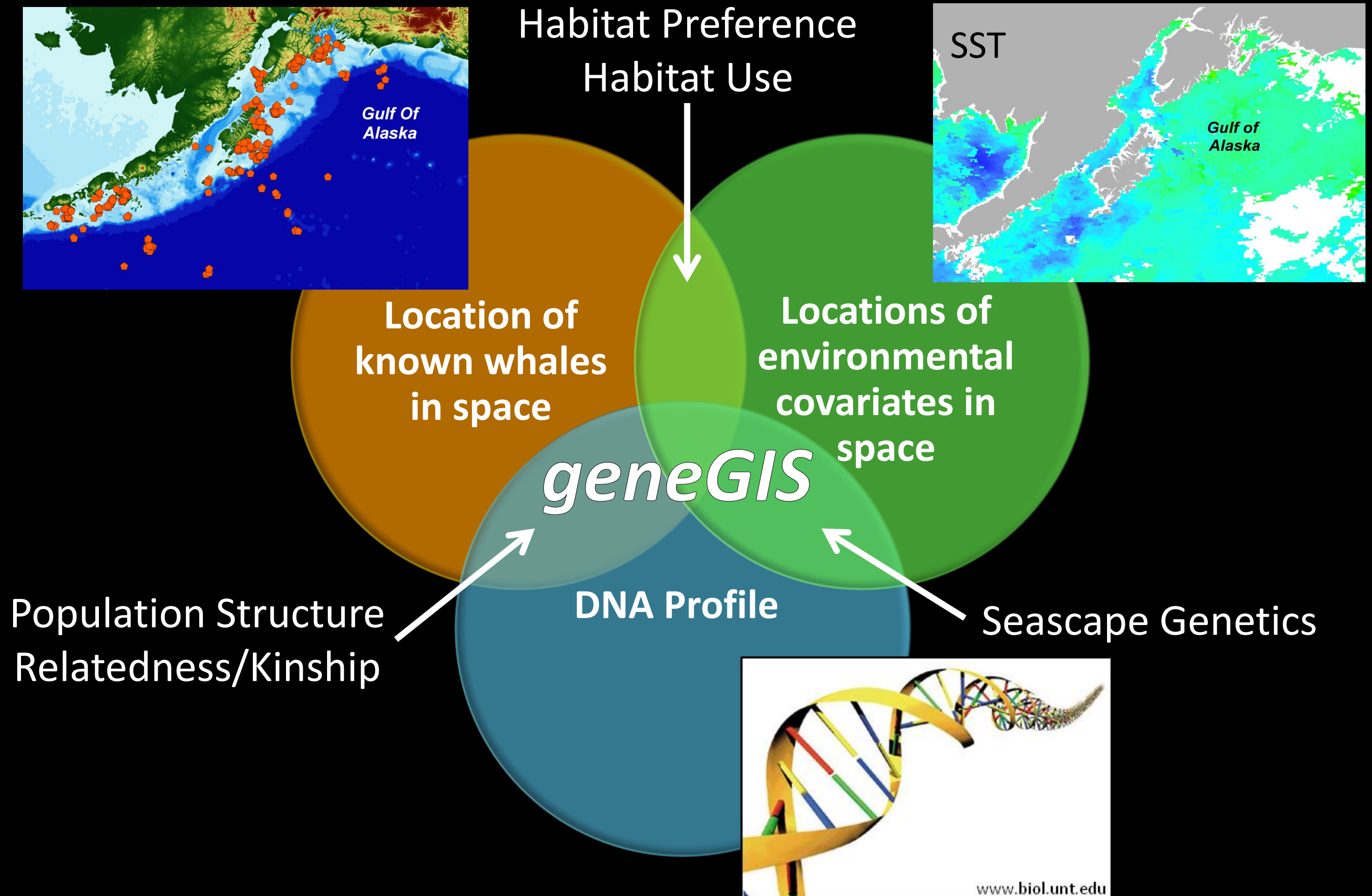
- Provide suite of ArcGIS tools for use with integrated individual-based data



- Easily accessible for non-GIS users
- Tools and geoprocessing scripts open source

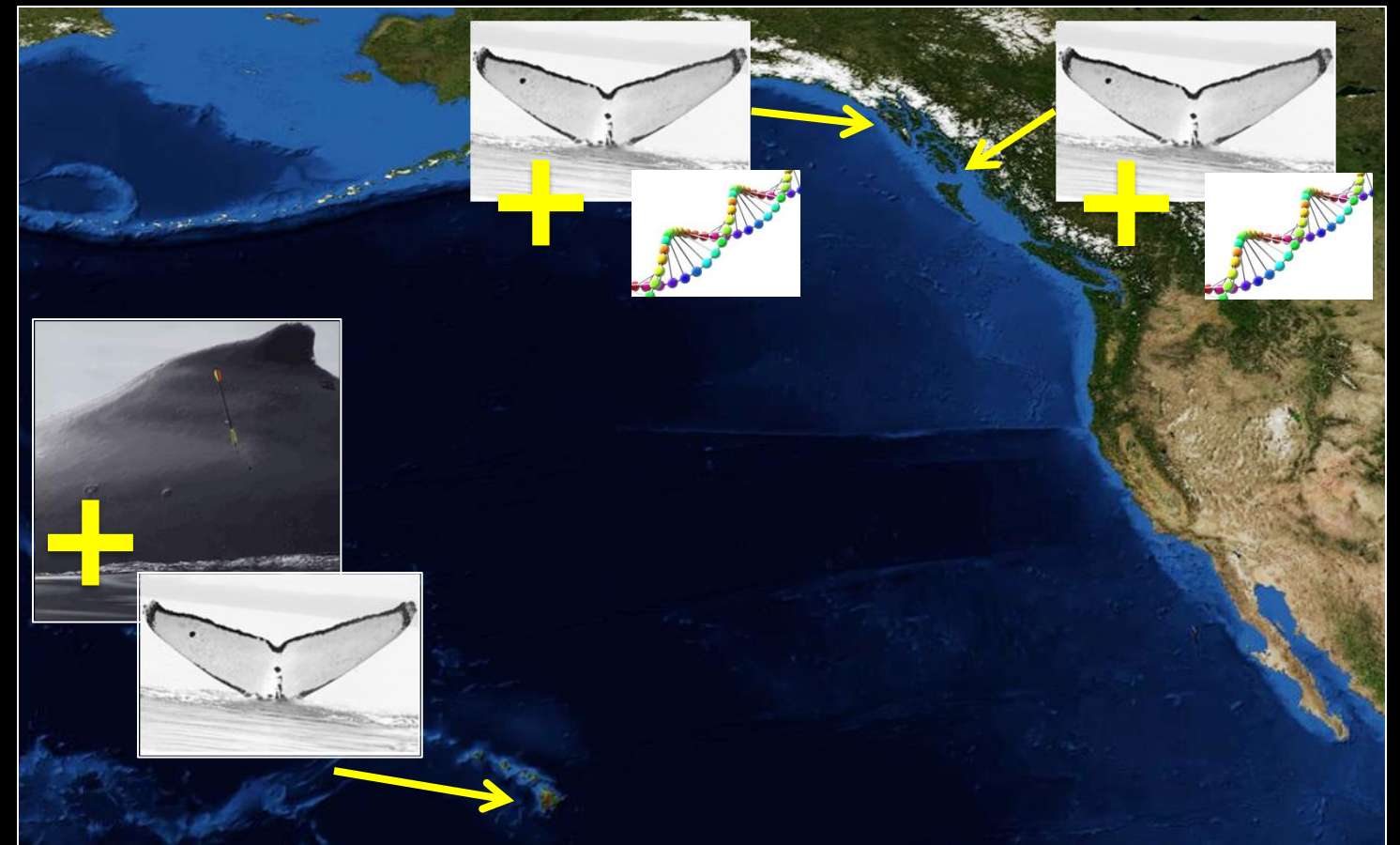


geneGIS Framework



Our database...

- *geneGIS* developed and implemented using SPLASH*, a collated & extended database of N. Pacific humpback whale records
 - 8,000+ naturally marked individuals
 - 2,700 associated DNA profiles (10 microsatellite loci, mtDNA, sex)
- 2 different development environments
 1. Java-based, web interface data management system supported by Wildbook
<http://www.wildme.org/wildbook/doku.php>
 2. ArcGIS



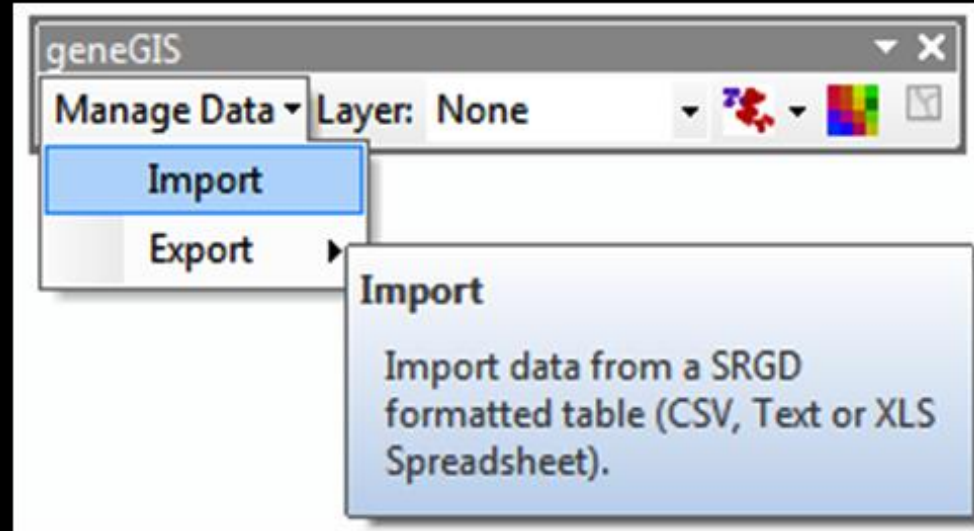
Working within *geneGIS*

- To provide a suite of ArcGIS tools to researchers who want to:
 1. **Visualize** data on a map
 2. Spatially **explore, display, and select** data
 3. **Export** data to formats required by other genetic analyses software
 4. **Extract** data from environmental layers
 5. Conduct **basic spatial analyses**

Working within *geneGIS*

To provide *quantitative* approach to *conservation priorities* by mapping spatial data and attributes of individuals to improve ability to *study influence of seascape on population structure*

1. Data Visualization



1. Data Visualization

Import Data

SRGD Input File

SRGD Input File
C:\Dori\Dissertation\geneGISChapter\SPLASH_Data_20:

Spatial Reference (optional)
GCS_WGS_1984

File Geodatabase Location
C:\Dori\Dissertation\geneGISChapter

File Geodatabase Name
Encounters

Genetic Columns

- ☒ Sex
- ☒ Haplotype
- ☒ L_GATA417_1
- ☒ L_GATA417_2
- ☒ L_Ev37_1
- ☒ L_Ev37_2
- ☒ L_Ev96_1
- ☒ L_Ev96_2
- ☒ L_rw4_10_1

Select All Unselect All Add Value

Identification Columns

- ☒ Sample_ID
- ☒ Individual_ID

Select All Unselect All Add Value

Location Columns

- ☒ Longitude
- ☒ Latitude

SRGD Input File

The following data fields for the SRGD input file **MUST** be formatted as follows:

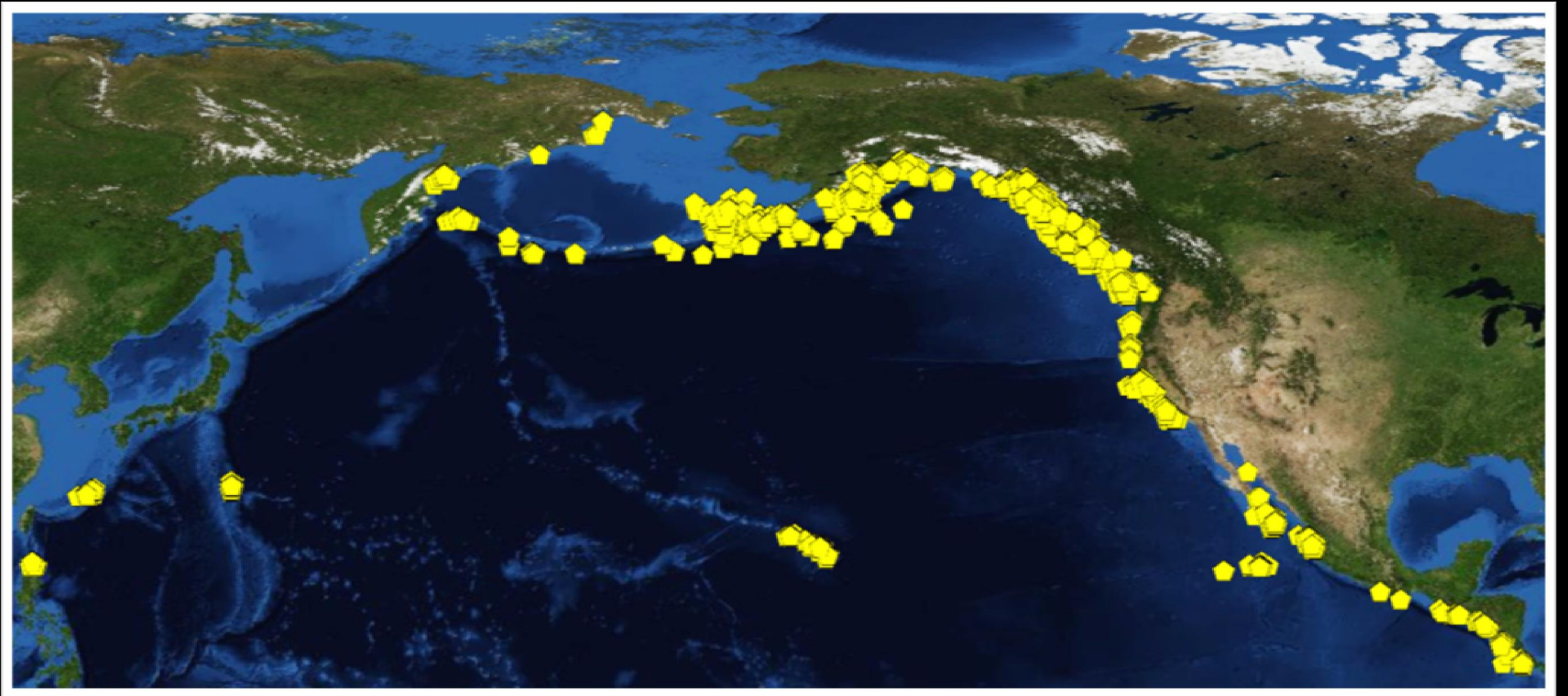
- **Sample_ID** (required)
A unique identifier for each data record or encounter with an animal that has an associated spatial (latitude, longitude) and temporal element (date, time) resulting in the collection of a sample (e.g., biopsy, skin, ear tag) and/or a photograph, tag or band that will allow the same individual to be resighted or recaptured. This information may have been used to determine an individual's identity.
- **Individual_ID**
A unique identifier for each known individual in the dataset. This information may have been derived from some previous type of analysis (e.g., determination of identity using genotyping or photo-identification methods).
- **Latitude** (required)
The latitude, in decimal degrees, where the sample or photograph was collected. Latitudes are positive north of the equator and negative south of the equator.
- **Longitude** (required)
The longitude, in decimal degrees, where the sample or photograph was collected. Longitudes are positive east of the Prime Meridian and negative west of the Prime Meridian.

NOTE: the default spatial reference, unless indicated by the user during data import, is Geographic (WGS84).

Alternatively, data can specify X and Y columns for projected datasets, such as coordinates stored in UTM.
- **Date_Time**(Date required; Time optional)

OK Cancel Environments... << Hide Help Tool Help

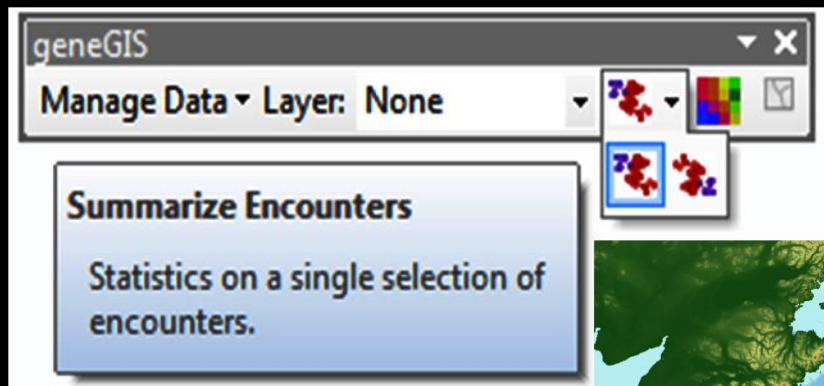
1. Data Visualization



All encounters of known individuals from the SPLASH Program

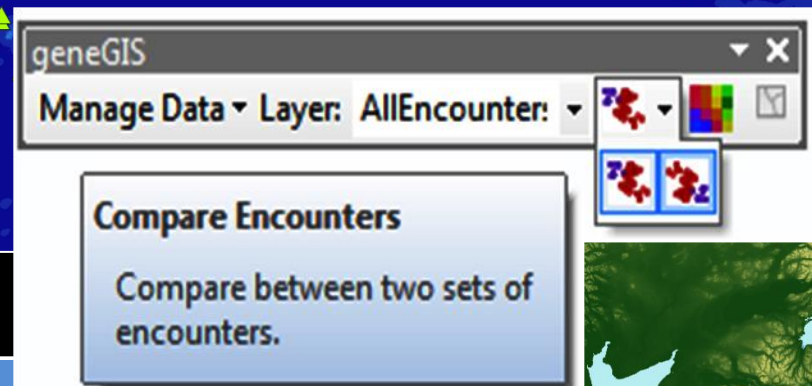
2. Spatially Explore, Display and Select

Are humpback whale populations in the Western Gulf of Alaska and Southeast Alaska genetically differentiated?



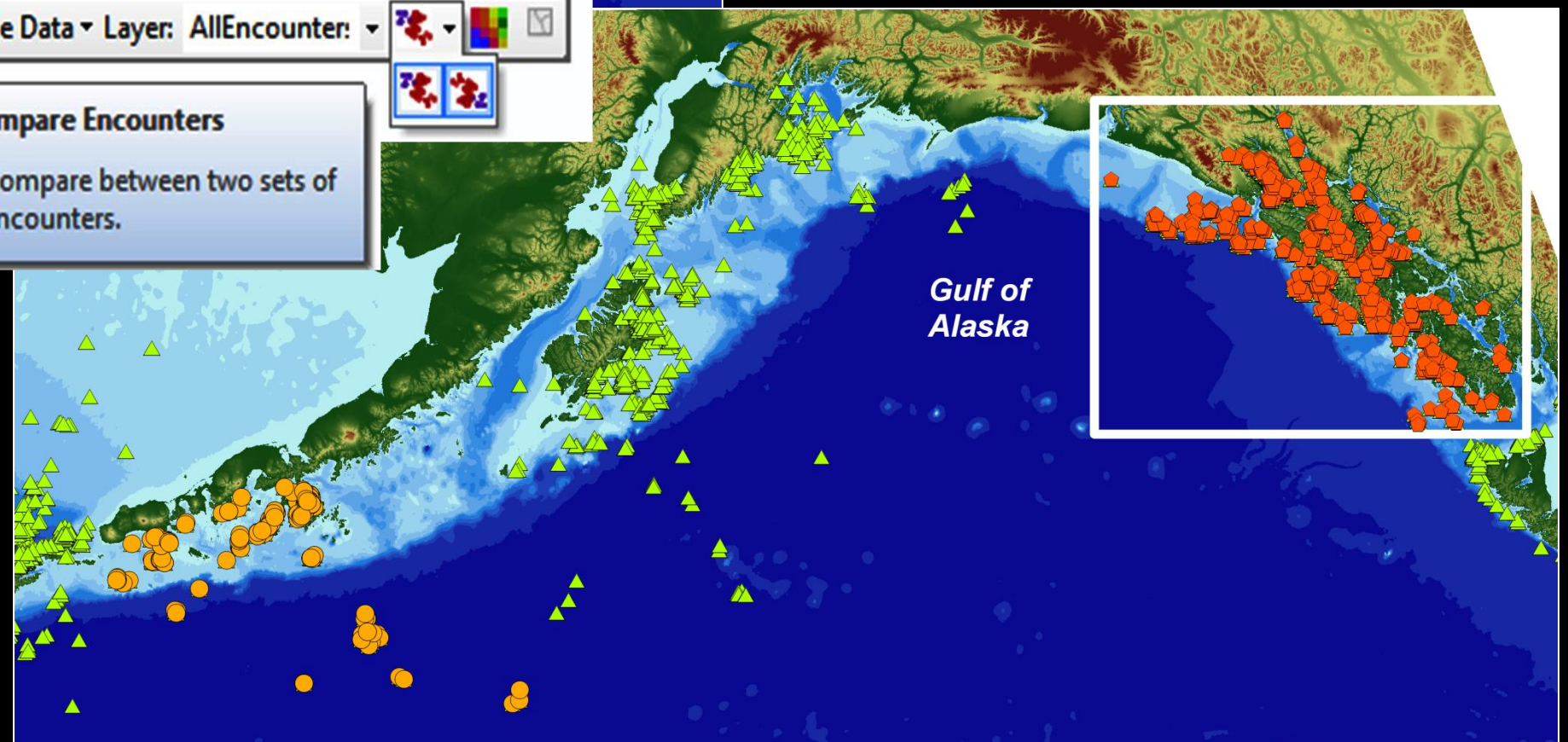
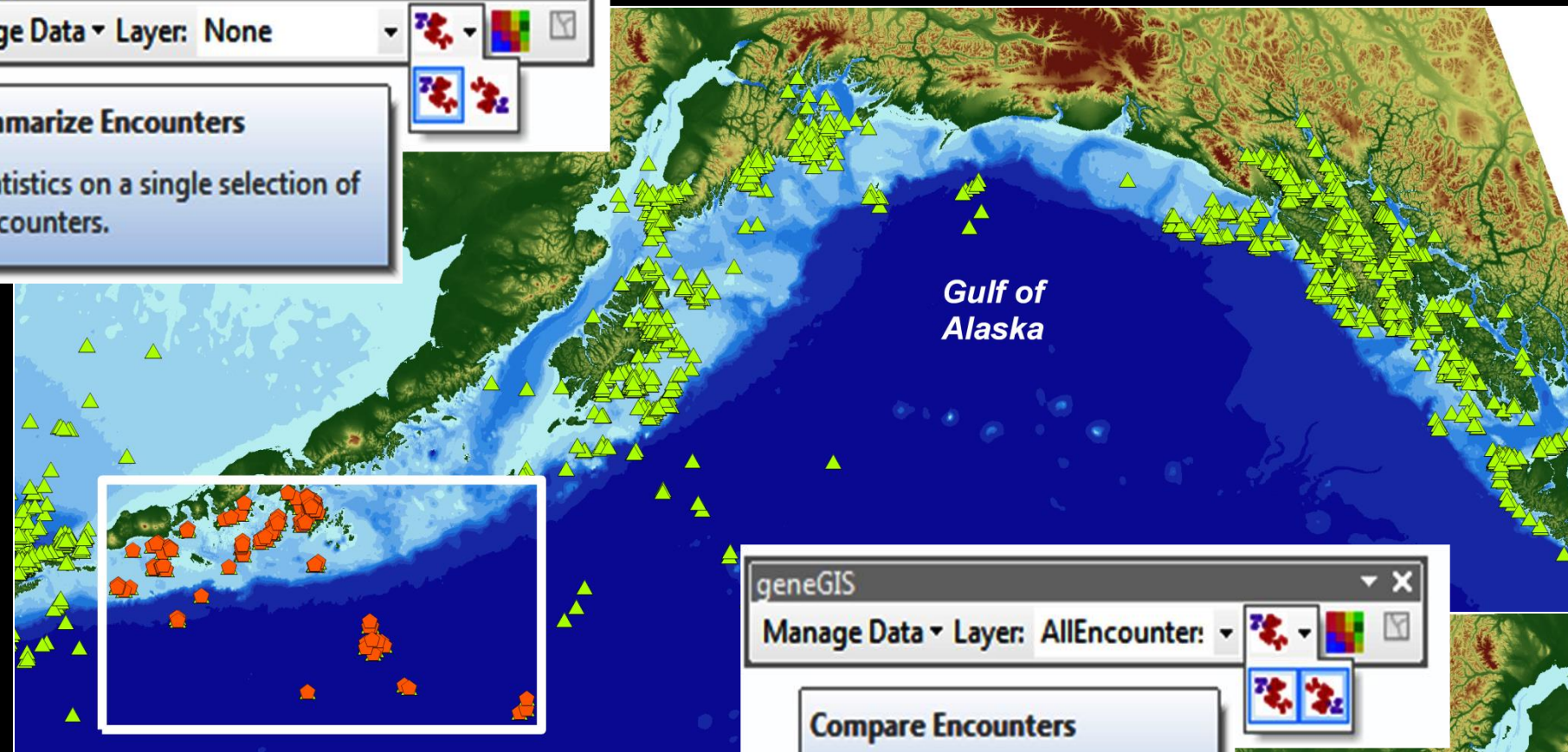
Encounters in selection

Encounters: 586
Individuals: 339

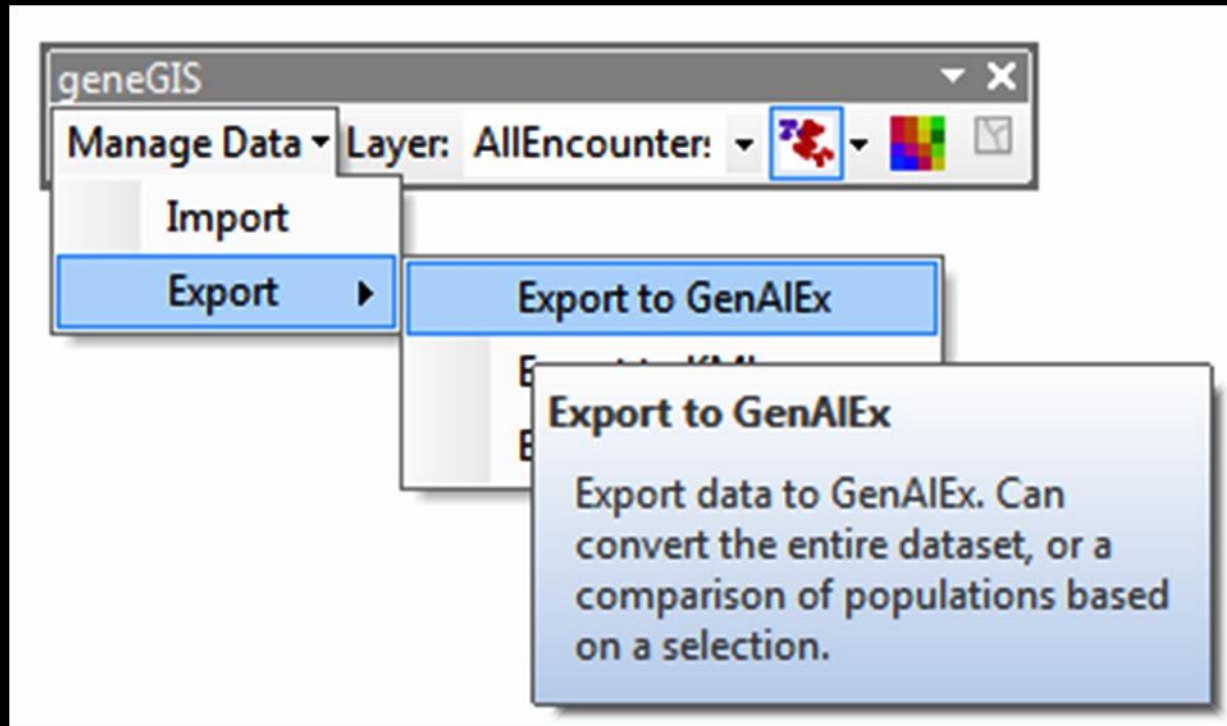


Comparison Results

First Set: 586 encounters,
339 individuals
Second Set: 2770 encounters,
1172 individuals
Common to both: 2

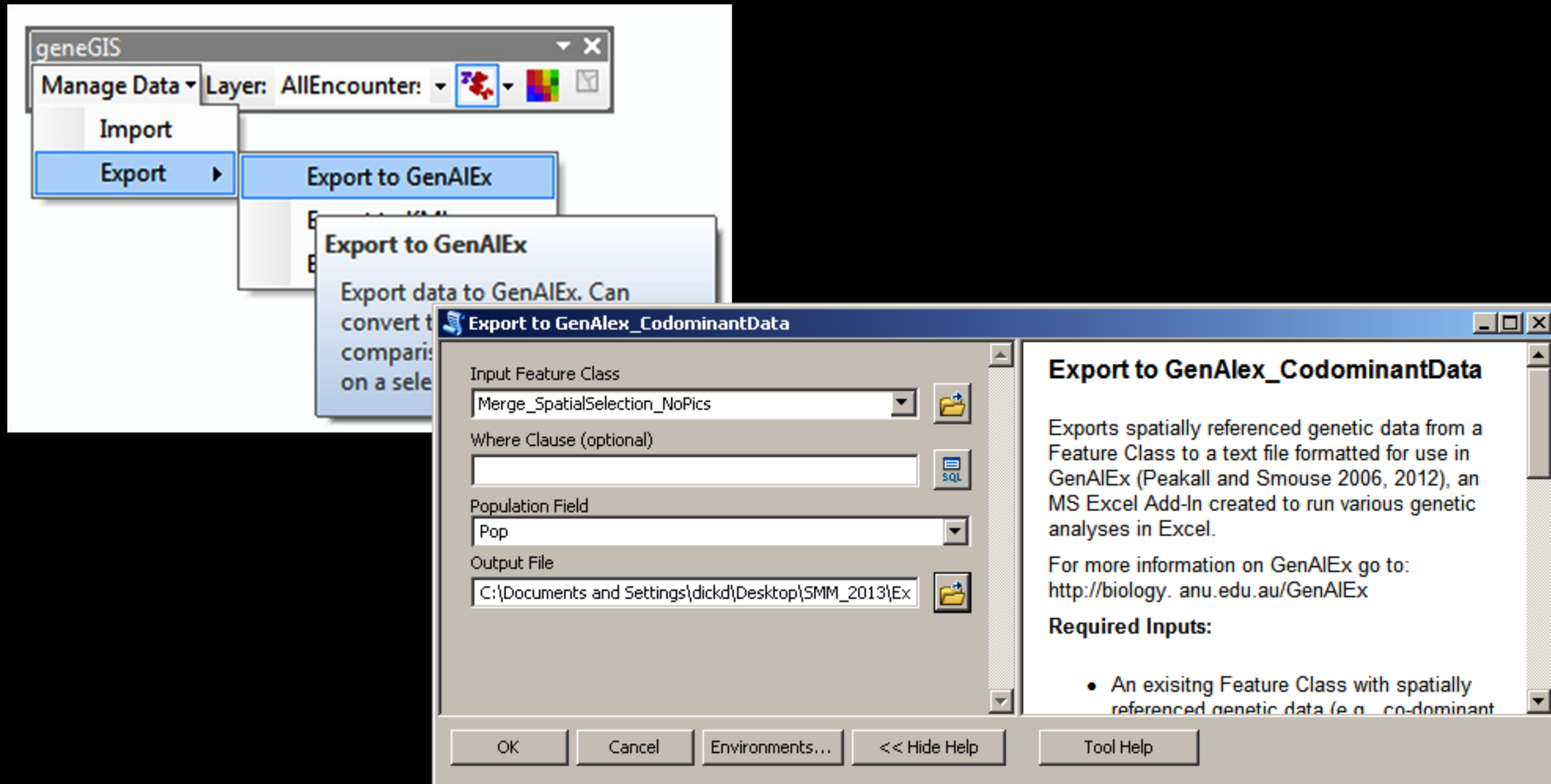


3. Data Export & Analysis Outside ArcGIS



Other Export Options:
Alleles in Space
GenAlEx
Genepop
KML files

3. Data Export & Analysis Outside ArcGIS



3. Data Export & Analysis Outside ArcGIS

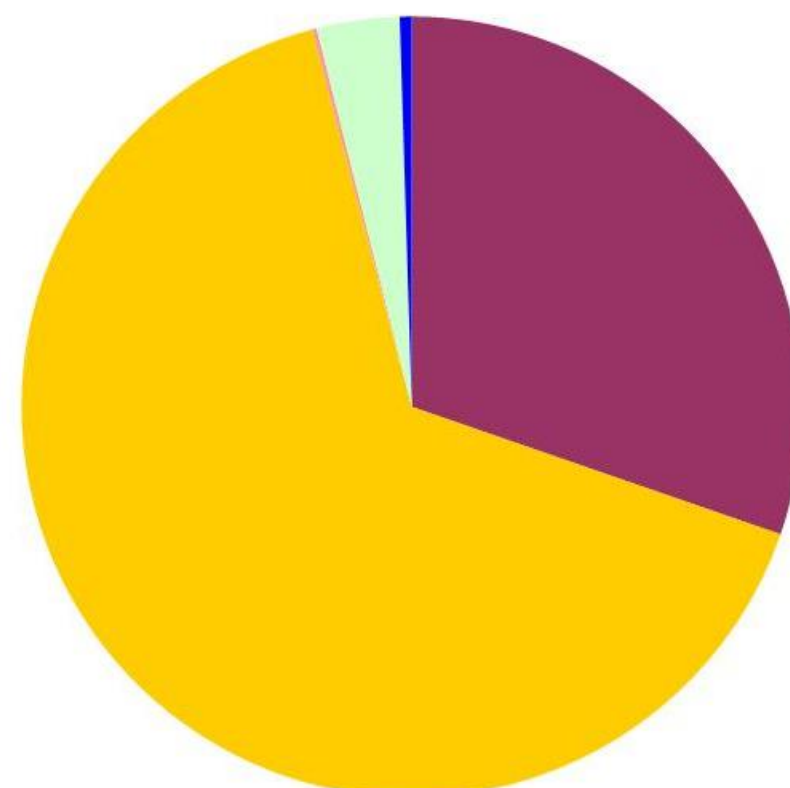
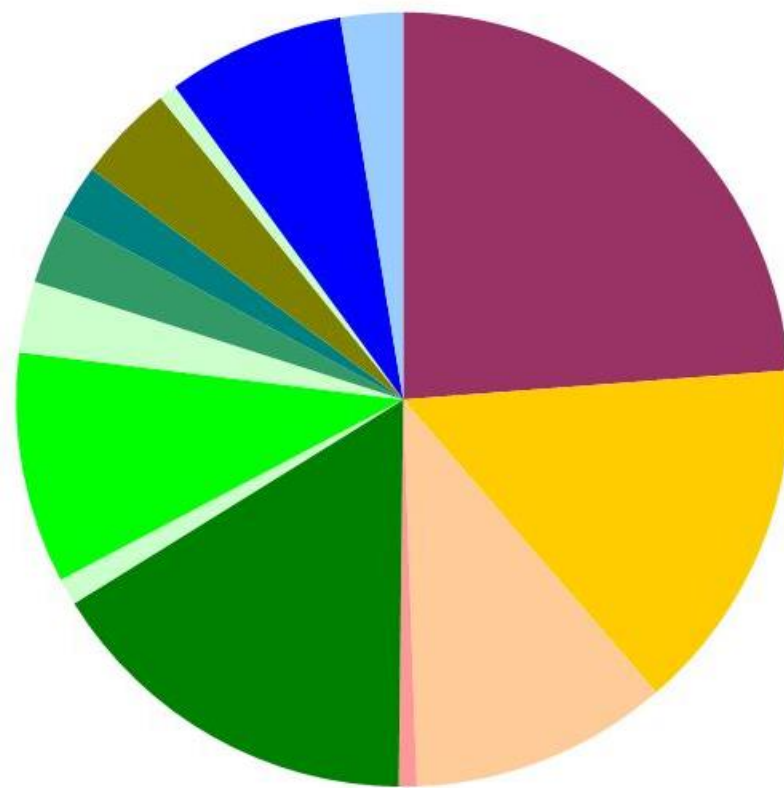
GenAlEx: ($F_{ST} = 0.197$, $p < 0.01$)

Populations are significantly genetically differentiated based on mtDNA haplotypes

mt DNA haplotype frequencies for:

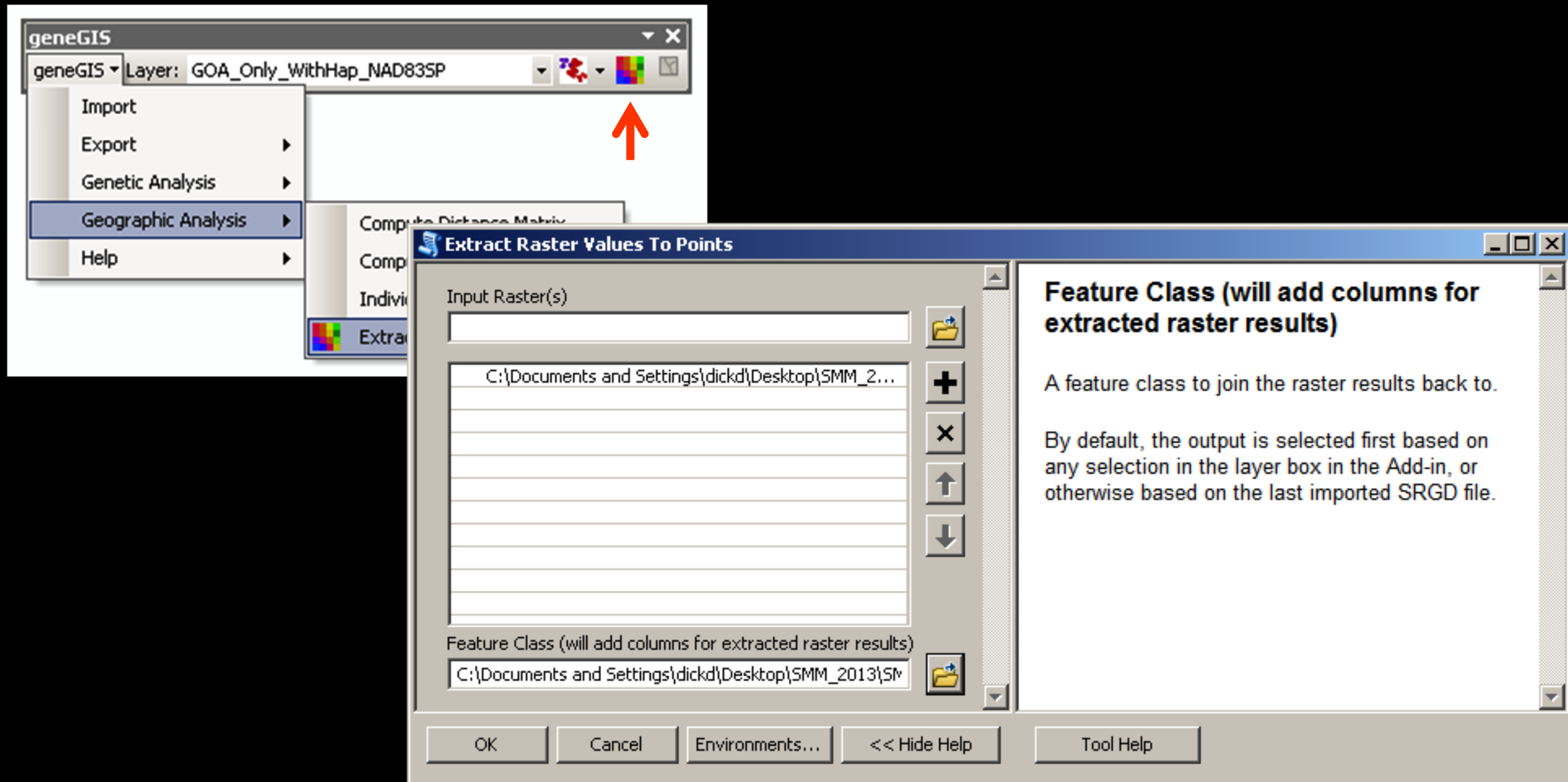
A) Western Gulf of Alaska

B) Southeast Alaska



4. Data Extraction

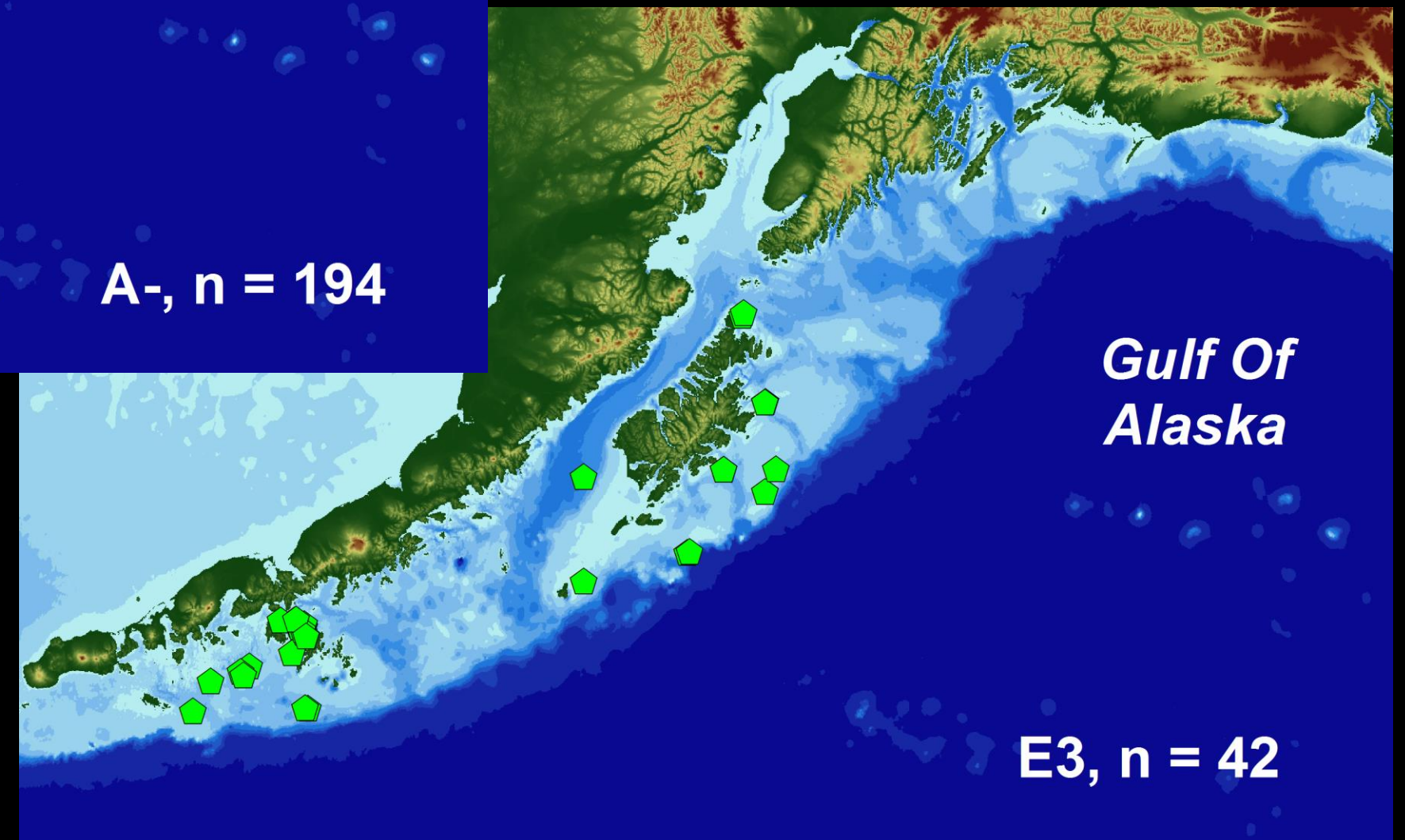
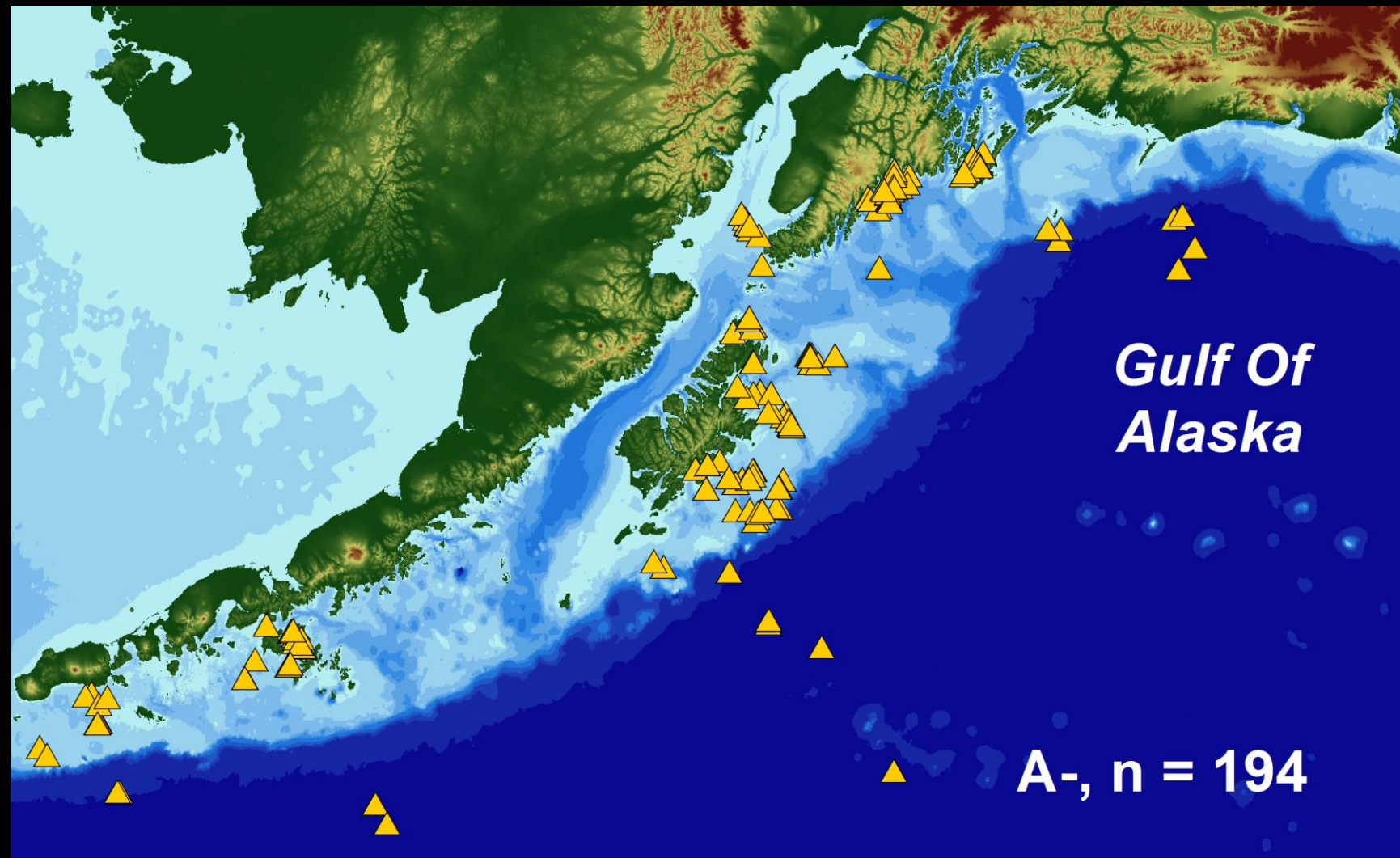
Within a set of whales of known mtDNA haplotype is there any evidence of preference for particular depths?



Cell values from a raster surface (bathymetry, SST, Chl a, etc.) are added to the input feature class.

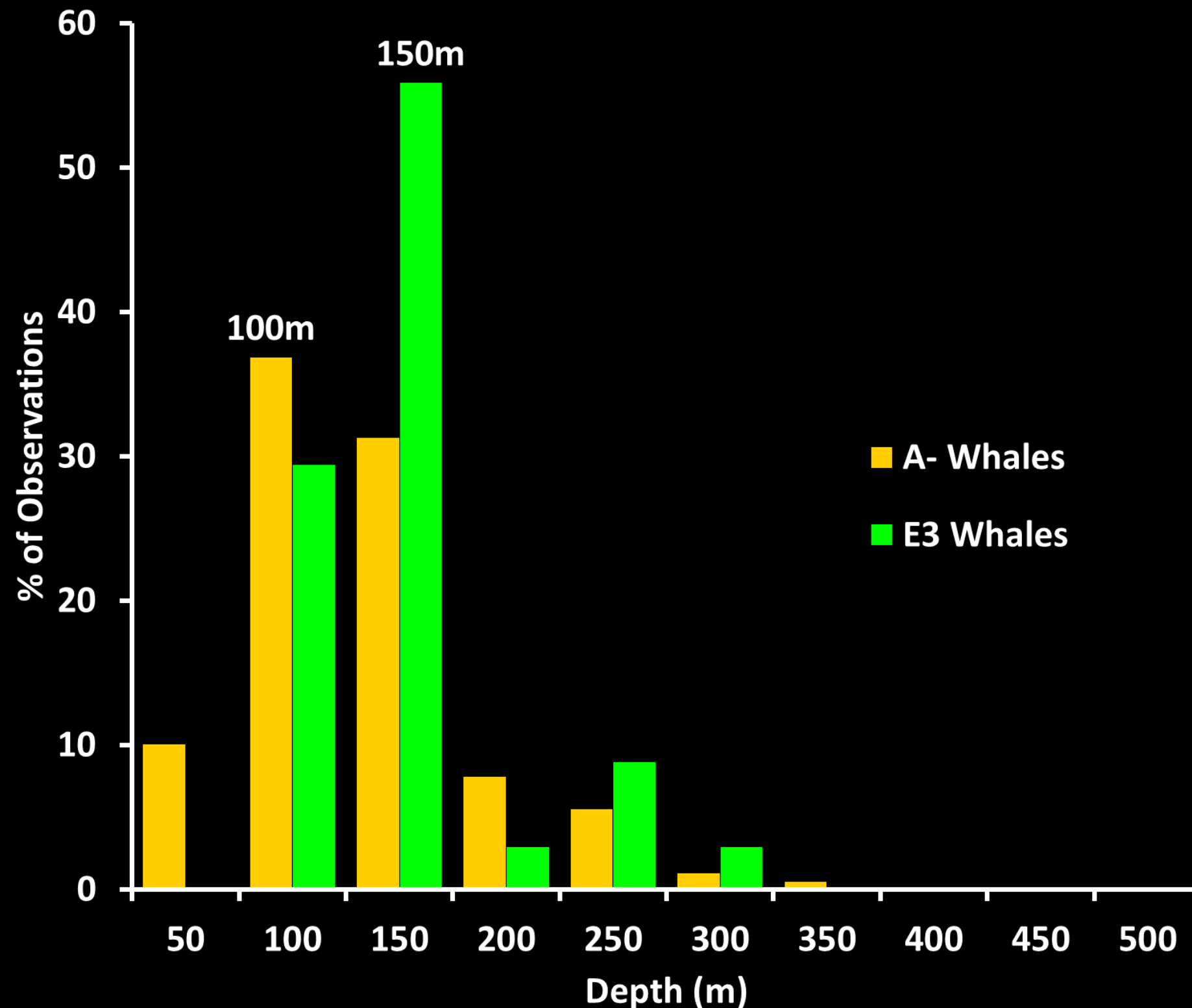
4. Data Extraction

Within a set of whales of known mtDNA haplotype is there any evidence of preference for particular depths?



4. Data Extraction

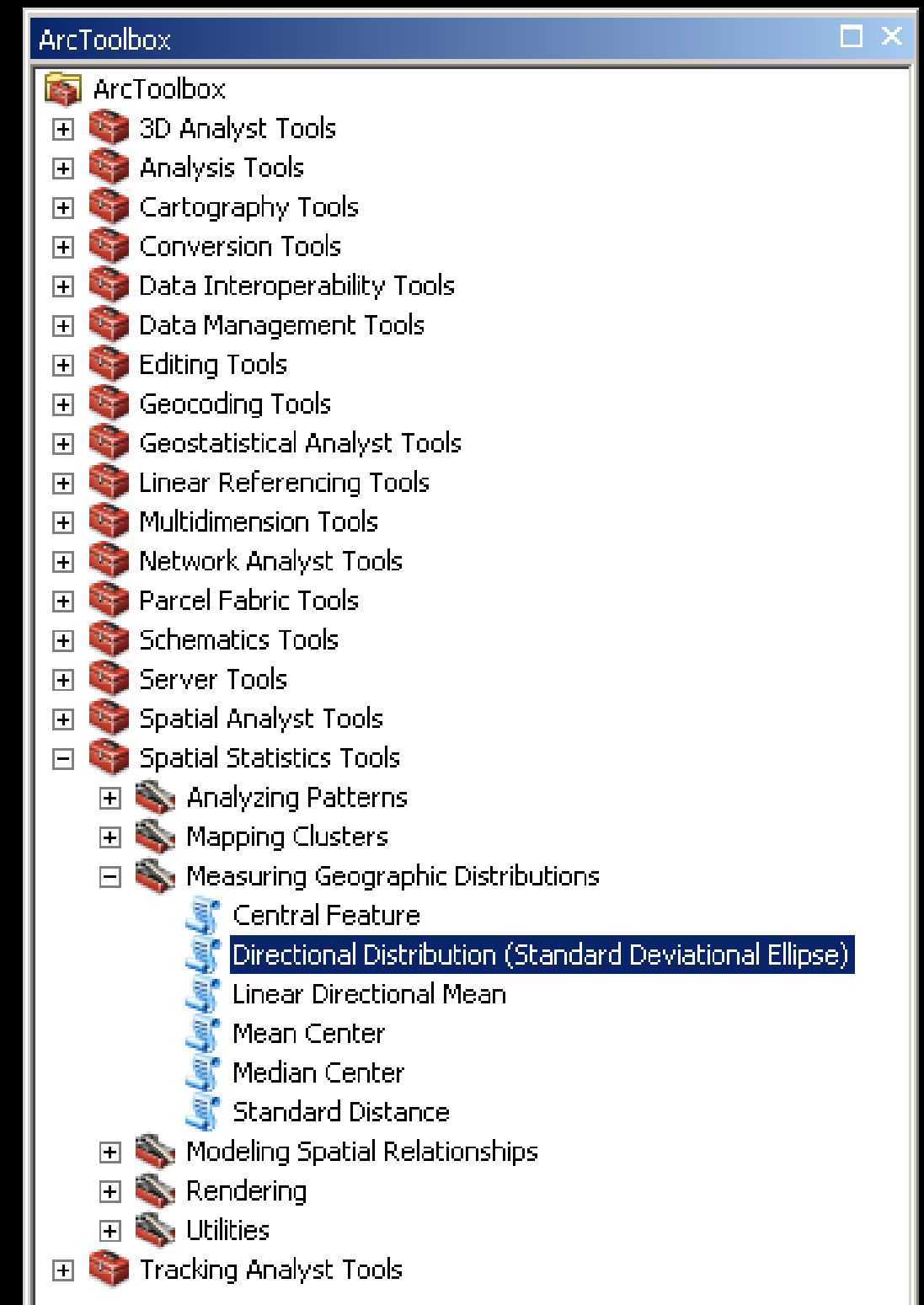
Within a set of whales of known mtDNA haplotype is there any evidence of preference for particular depths?



5. Spatial Analyses

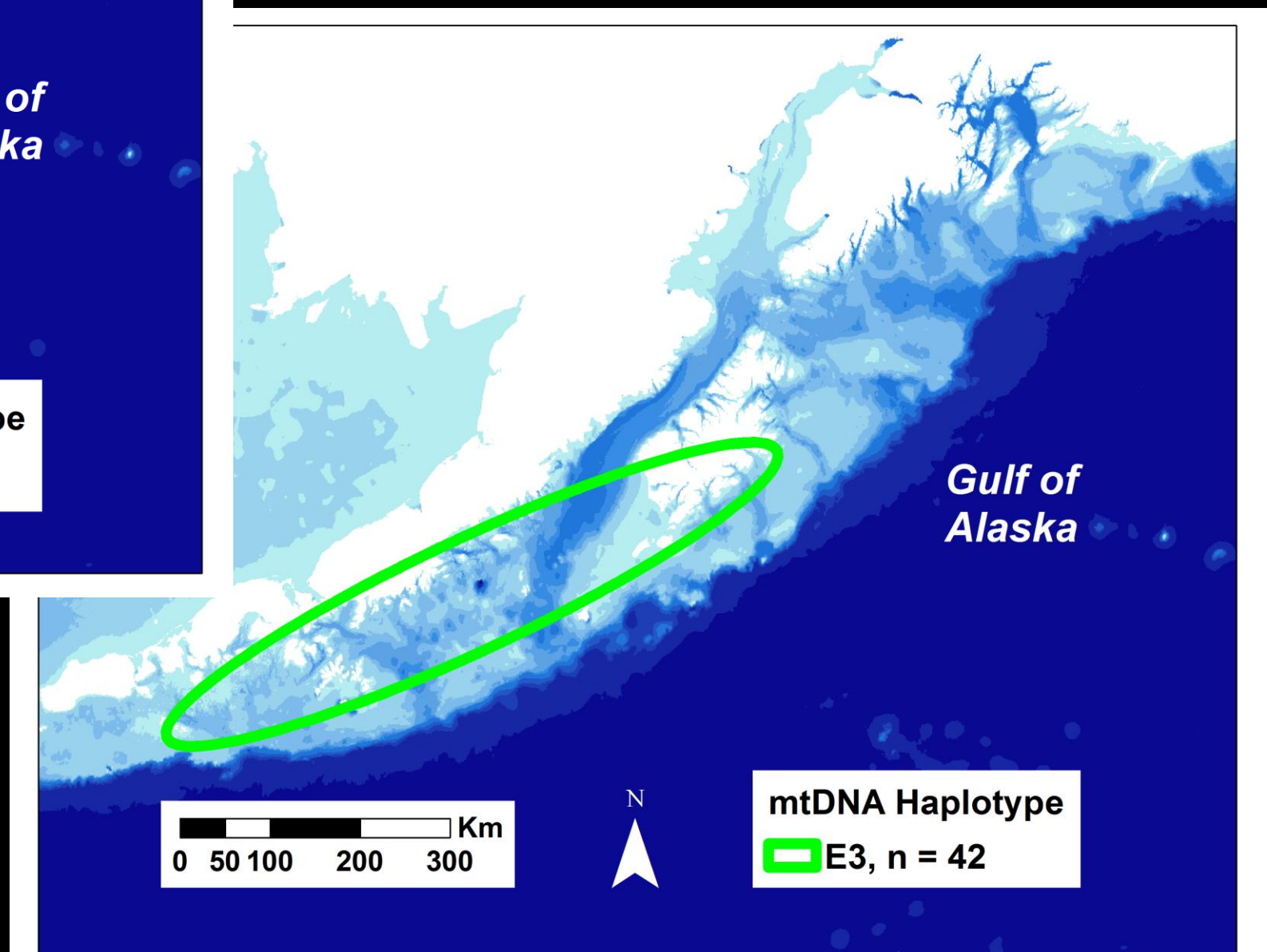
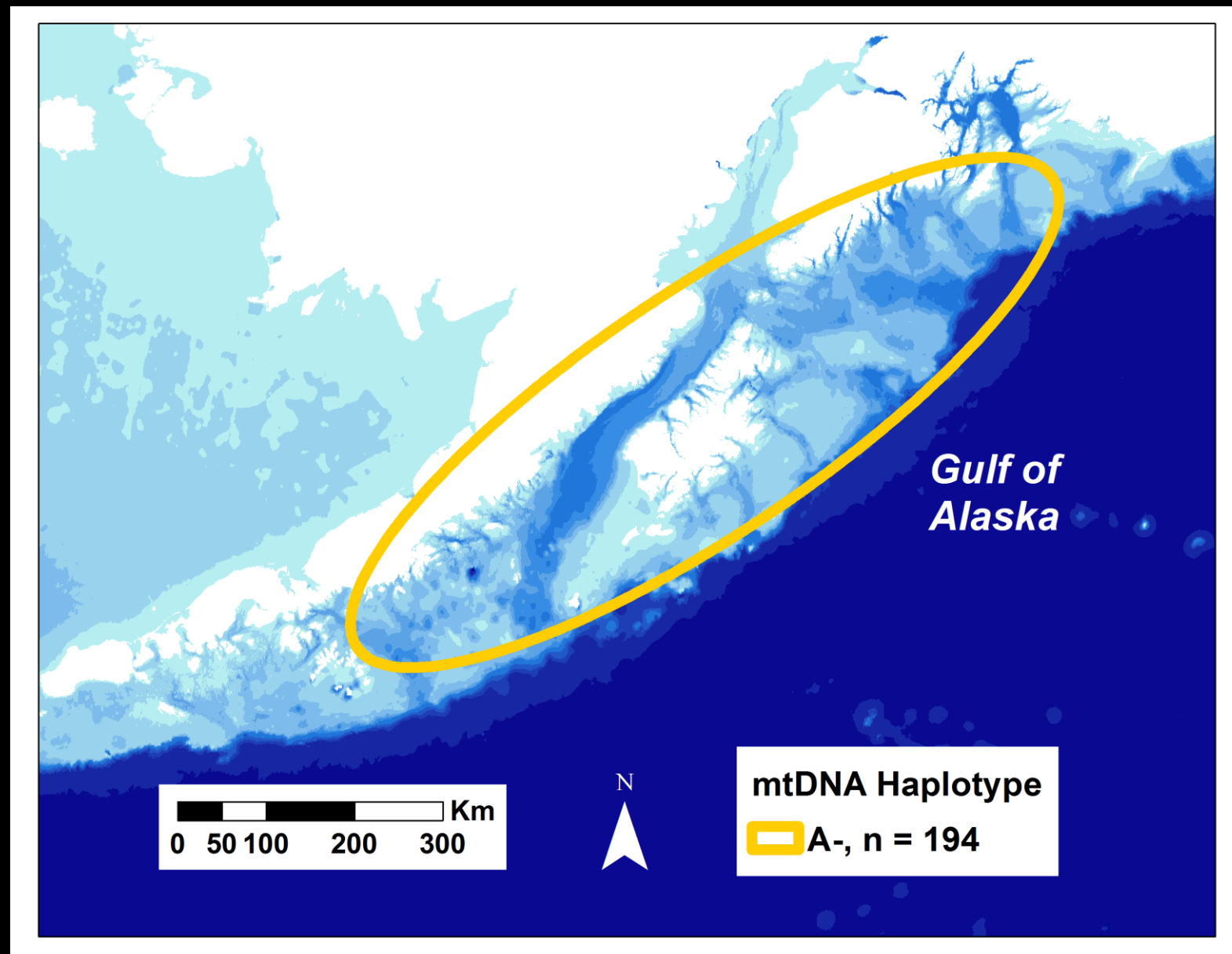
How do the spatial distributions of humpback whale haplotypes vary?

- Turn to the Standard ArcGIS ArcToolbox options
- Directional Distribution
- Conduct standard deviation ellipse analysis
- Summarizes central tendency, dispersion and directional trends in X and Y directions



5. Spatial Analyses

How do the spatial distributions of humpback whale haplotypes vary?



Welcome to geneGIS!
geneGIS: spatial analysis of individually linked sighting records and genetic samples. It's an extension for ArcGIS, and integrates with the Shepard Project to get your data online.

Sample_ID	Individual_ID	Date_Time	Region	Latitude	Longitude	Sex	Haplotype	L_GATA	L_GATA
1	100	1/14/2004 9:30	Cent Am	13.9038	-90.7658			417	417
2	100	1/17/2005 12:12	MX-ML	20.6	-105.6				
3	100	1/31/2005 10:37	MX-ML	20.7206	-105.526				
4	192	2/13/2006 14:19	Asia-OK	26.7534	127.6401	M	E1	206	218
5	192	2/13/2006 10:53	Asia-OK	26.6853	127.6922	M	E1	206	218
6	192	2/13/2006 15:16	Asia-OK	26.7308	127.6091	M	E1	206	218

PhotoID Only DNA Profile Only
Tabular genetic data
Combining genetic records with spatial information.

Spatial context

Not convinced? Check out our screencast: *coming soon*

DEPARTMENT OF THE NAVY
ONR
Science & Technology

OSU
Oregon State University

esri

WILDBOOK

Mack-recapture
Animal Biometrics
Molecular Ecology
Wildlife Toxicology
Social Ecology
Behavioral Ecology

- Installation File with instructions
- Basic background information & documentation including tool help
- Link to GitHub repository

Summary

- Introduced and demonstrated some geneGIS functionality to explore spatially explicit individual-based cetacean data
- Shown geneGIS can be combined with standard ArcGIS functions → further exploration on the influence of seascape on humpback whale distribution and population structure
- Provided an example of basic spatial analysis one can do without a strong GIS background → develop spatially explicit hypotheses



Thanks to:

All the researchers involved with SPLASH

Cascadia Research Collective for database maintenance

OSU Cetacean Conservation Genetics Lab for genetic analyses

This research is funded by ONR contract N0270A to Oregon State University.

Travel funding provided by OSU Graduate Student Travel Award,
American Cetacean Society Puget Sound Student Travel Grant



Contact: dickd@geo.oregonstate.edu