Although ArcGIS can handle many types of data provided they also contain a set of coordinates, the crux of *geneGIS* is its ability to handle spatially referenced genetic data (SRGD). To ensure *geneGIS* functions as intended, we developed a standardized input file format. The SRGD input file is a comma separated value (CSV) file and specifies the *minimum* data requirements needed to use *geneGIS* (Table 1).

Table 1. Minimum requirements for the SRGD.csv input file. Field names must be identical to those listed in the table below. Additional fields of interest can be included.

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Required?** | **Expected Data Input Examples** |
| Sample\_ID | X | A unique identifier for each record in the table or dataset |
| Individual\_ID |  | A unique identifier for each known individual in the dataset |
| Source\_ID |  | An identifier used to indicate how individual identity was determined (photo, genetic, photo/genetic etc.) |
| Latitude | X | 12.8006 |
| Longitude | X | -122.8570 |
| Date\_Time | X | Recommend ISO 8601 format (2004/06/12T14:12:50)  Other Excel friendly formats supported, see below |
| Region |  | Mexico, Oregon, etc.  OR  Plot 1, Plot 2, Plot 3, etc. |
| Sex\* |  | M  F  U |
| Haplotype\* |  | A+  F2 |
| L\_YourLocusName\* |  | 211 |
| L\_YourLocusName\* |  | 220 |

\* NOTE: These fields are considered optional because any combination of Sex, Haplotype and Loci data are possible. However, ***at least one is required*** to make use of the intended *geneGIS* functionality.

**Sample\_ID**

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).

**Source\_ID**

An identifier used to indicate how individual identity was determined. For example, was an individual’s identity determined from a photograph, analysis of a genetic sample, or both? This field is critical when data is extended to encounters that would not typically have this information. For example, an encounter in which an individual is identified with a photograph will not contain genetic data unless the data has been extended.

**Latitude**

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.

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

**Longitude**

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.

**Date\_Time**

The date and time for when the sample or photograph was collected. Recommended format ISO 8601 YYYY/MM/DDThh:mm:ss.

Other supported formats for date include: MM/DD/YYYY and DD-Mon-YYYY

Other supported formats for time include: HH:MM:SS AM/PM and HH:MM:SS (24 hour)

Note: date is required, time is optional.

**Region**

The general region from where the sample or photograph was collected. This can be based on political or geographic boundaries (e.g. oceans, countries, states, etc.) or the study design (e.g., a specific quadrat or plot).

**Sex\***

The genetic sex of the sample or individual, if known.

**Haplotype\***

The mitochondrial haplotype of the sample or individual, if known.

**L\_YourLocusName\***

The name of a specific locus. For biallelic data (microsatellites or SNPs), there should be 2 columns with the same locus name. Each column should hold a value for each of the alleles for that locus.

Include as many loci and allelic values as necessary for your data.

The L\_ prefix will allow the Classified Import Tool to correctly recognize Loci.

ArcGIS does not allow more than one field with the same name. Upon import into AcGIS, the name for each locus will be appended as follows: L\_YourLocusName\_1, L\_YourLocusName\_2.

**Example SRGD.csv file with additional loci data fields:**

