Export Beagle Marker File per Chromosome

Author: Autumn Laughbaum, Golden Helix, Inc.

Overview

This script may be used to prepare the marker file required by BEAGLE when using multiple datasets in an imputation project. The script will output separate files for each chromosome with the following columns: marker name, position, 2+ alleles in space-delimited format. Each row is a marker and the output may have varying number of columns depending on how many alleles are present in the original spreadsheet's marker-column. Missing alleles are not included in the output. This script must be run from a marker-mapped spreadsheet. Prior to running the script, you should create a single spreadsheet containing the combined genotypes for all datasets that you are planning to use for imputation (reference data and all study datasets). Create a subset of this spreadsheet to the markers that you wish to use for imputation and apply a unified marker map. Then run this script to create the BEAGLE markers files. You may also use the "*Export Beagle By Chromosome*" script to create the genotype input files, but it is recommended that you run the genotype export separately for each dataset in the study.

Recommended Directory Location

Save the script to the following directory:
*..**\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\File**

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between versions. The easiest way to locate this directory on your computer is to open SVS and select the **Tools >Open Folder > User Scripts Folder** menu option. If saved to the proper folder, this script will be accessible from the spreadsheet **File** menu.

Using the Script

1. From a spreadsheet containing genotypic data, choose **File >Export Beagle Marker File per Chromosome**.

2. Select the directory to save the files to, the file base name, and the delimiter to use for the text files.

One text file will be created for each chromosome represented in the dataset. In the output, the markers are rows and the columns contain the marker information.