Import Concatenated Genotype String File

Author: Greta Peterson, Golden Helix, Inc.

Overview

This import script is designed to import genotypic data that is stored in a concatenated string format. The user will specify the variant name column, sample name column, and data column(s) as well as the genotype encoding.

Recommended Directory Location

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

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Windows 7. The easiest way to locate this directory on your computer is to open SVS and select **Tools > Open Folder > UserScripts Folder**. If saved to the proper folder, this script will be accessible from the project navigator's **Import** menu.

Using the Script

1. From the project navigator, choose **Import > Import Concatenated Genotype String File**.

2. Select the appropriate genotype file, the file delimiter, whether or not the file has a row indicating the column/field names and whether or not there are header lines that need to be skipped in the genotype file. If so, the number of lines to skip must be specified.

3. Select the appropriate marker map file, the file delimiter, whether or not the file has a row indicating the column/field names and whether or not there are header lines that need to be skipped in the marker map file. If so, the number of lines to skip must be specified.

4. There are two advanced options that can be specified, whether or not there is a marker map field that defines the A/B allele encoding, (i.e. Allele A = C and Allele B = G) and whether or not the intermediate spreadsheets should be deleted.

5. Once these parameters have been specified, click **Next >**. See **Figure 1**.
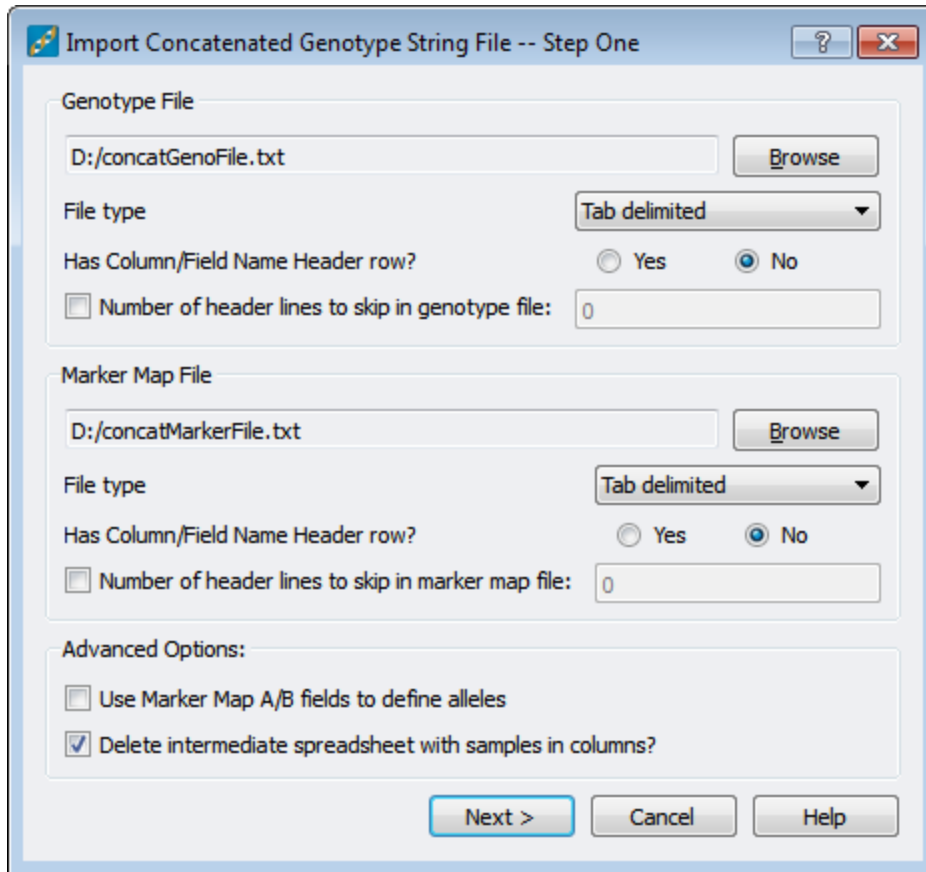
Figure 1: Step One -- Files and file parameters chosen

6. In the second dialog, select the field containing the column headers (marker names) from the map file, the field containing the sample names and the field containing the concatenated string from the genotype file. If the genotype recoding needs to be adjusted from the default this can be specified at this time. The default is A_A=0, A_B = 1, B_B = 2 and ?_? = 5. If the option to use a marker map field to specify the A and B alleles was selected on the first dialog additional prompts will be available to make the choices for the A and B allele fields. Once all of the parameters are correctly specified click **OK**. See **Figure 2**.
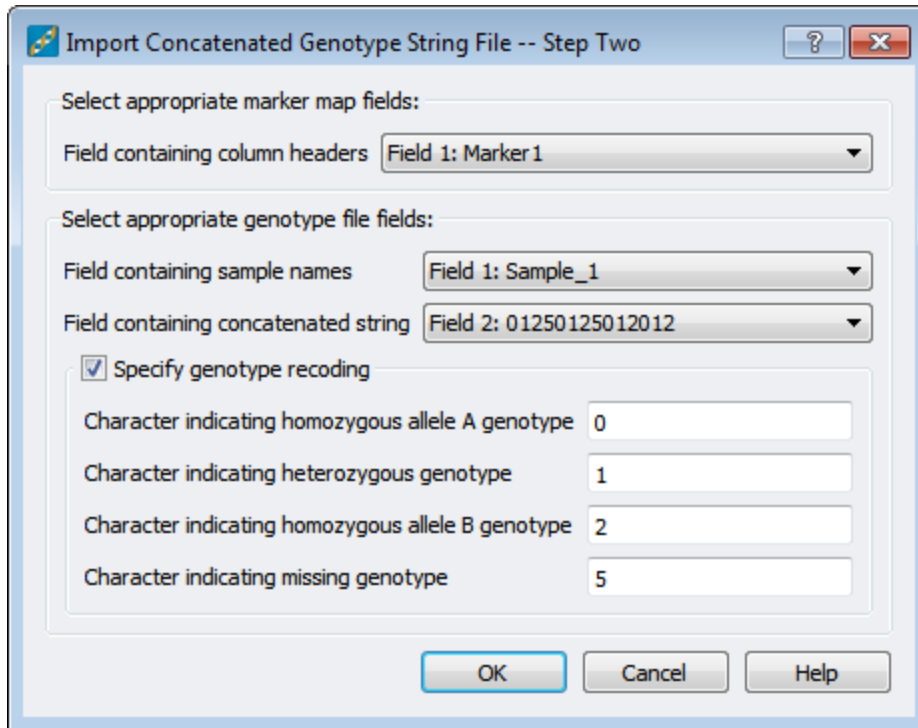
Figure 2: Step Two -- Setting specific field and recoding parameters

7. The resulting spreadsheet will be a genotype spreadsheet where the samples are in rows and the markers are in columns.

Figure 3: Resulting spreadsheet after import