Version 1.0 (Feb, 2012)

BIASLESS User Guide

Hsin-Chou Yang[†], Pei-Li Wang and Chien-Wei Lin

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

[†] hsinchou@stat.sinica.edu.tw

Table of Contents:

1. BIASLESS LICENSE

2. INTRODUCTION

BIASLESS (**B**iomarkers **I**dentification **a**nd **S**amp**les** **S**ubdivision) written in R and R-GUI is a user-friendly tool for the identification of key predictive markers to discriminate samples from different populations/groups based on high-dimensional genomic and transcriptomic marker data. **BIASLESS** integrates forward variable selection and cross-validation procedures with flexible discriminant analysis to identify key single-nucleotide polymorphism (SNP) and/or gene expression (GE) markers. The identified informative markers can be used to classify samples into populations/groups with the highest cross-validation prediction accuracy.

3. SOFTWARE DOWNLOAD AND INSTALLATION

Execution of **BIASLESS** requires the installation of **BIASLESS** program and R program. Procedures for downloading and installing the two programs are described as follows:

1. **BIASLESS**:

   BIASLESS program is available at the **BIASLESS** website at http://www.stat.sinica.edu.tw/hsinchou/genetics/prediction/BIASLESS.htm. The zipped file "BIASLESS_v1.0.zip" can be downloaded and then unzipped to obtain a directory "BIASLESS" containing the program codes of **BIASLESS** and two illustrated examples.

2. R:

   Users can download R program "R-2.15.0-win.exe" from the **BIASLESS** website. Or users can download R from the website of "The R Project for Statistical Computing" at http://www.r-project.org/. Users click "CRAN" (Comprehensive R Archive Network) in the left of the page and then select a suitable mirror site to download R. Select a platform (Linux, MacOS X, or Windows) for R execution in your end. Click the hyperlink "base" and select "R-2.15.0-win.exe". Then execute the file to install R to "C:\Program Files\R\R-2.15.0". After finishing the installation of R, doubly click the icon "R i386 2.15.0" or "R x64 2.15.0" to initialize R in a 32-bit or 64-bit system, respectively. A window "RGui" with a sub-window "R Console" jumps up await for the subsequent analysis action. Users are suggested to update packages in R. They can select "Packages" in the tool bar, click "Update packages" and then select a suitable mirror site to update packages. A window "CRAN mirror" jumps up and the icon "OK" is clicked to update packages. Note that the analyses provided by **BIASLESS** require two additional R packages: mda and gtools. These packages will be automatically downloaded if users use a latest version of R, e.g., R-2.15.0. **Note that users are suggested to use program R-2.15.0 or a version of R program newer than program R-2.15.0 for execution of BIASLESS.**

4. BIASLESS INITIALIZATION

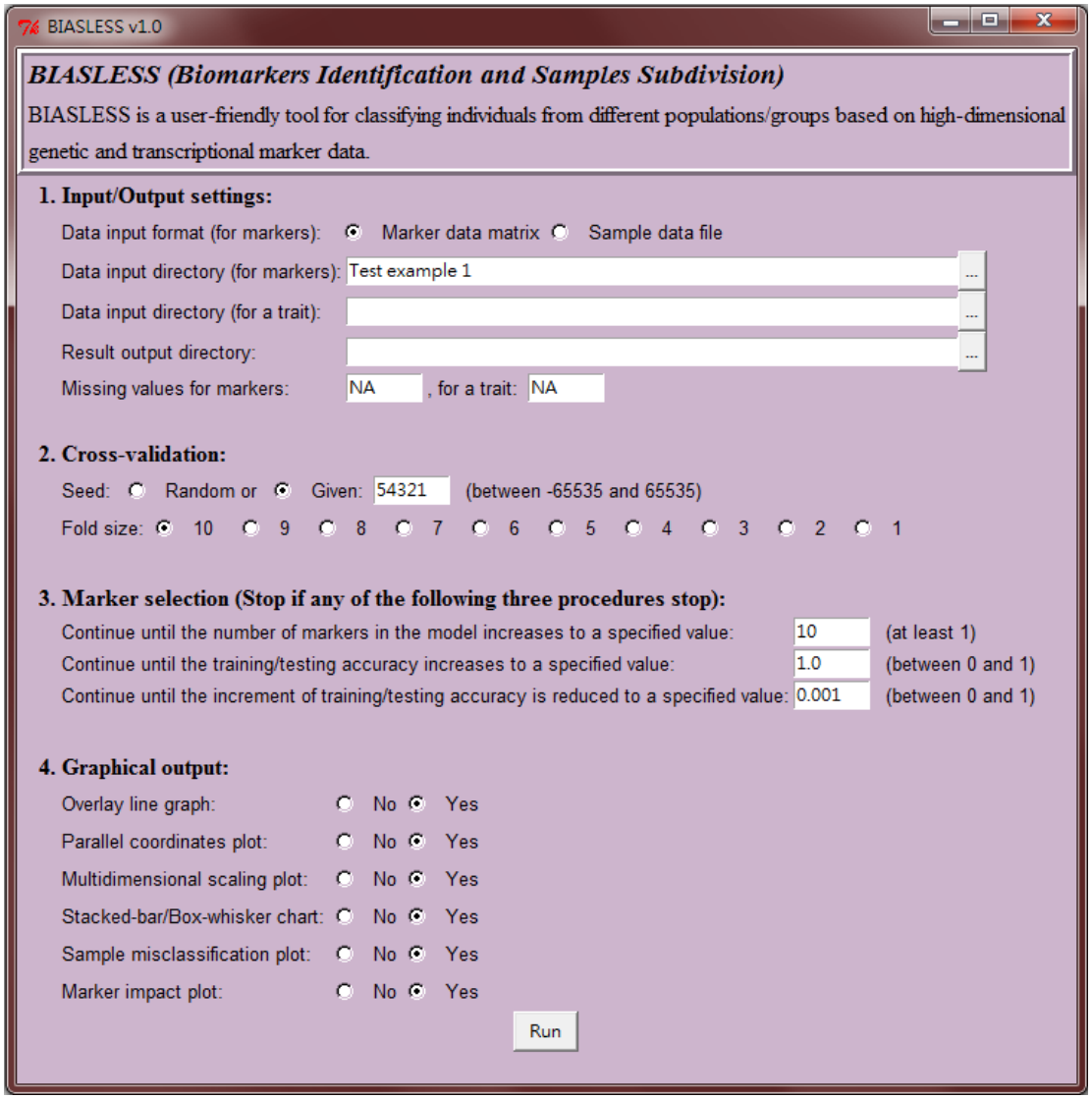
After an installation of **BIASLESS** and R, **BIASLESS** can be initialized by the following procedures. In this user guide, we suppose that users have installed the **BIASLESS** in the destination directory "C:\BIASLESS".

1. Initialize R by doubly clicking the icon "R i386 2.15.0" or "R x64 2.15.0" in a 32-bit or 64-bit system, respectively.

2. Key in the command **BIASLESSgui=paste("C:/BIASLESS/BIASLESS_interface.r",sep="")** in the command line in the window "R Console" and press the Enter key.
3. Type the command, **source(BIASLESSgui)**, in the command line and press the Enter key to initialize **BIASLESS**. The interface of **BIASLESS** jumps up and waits for the data entry after pressing the Enter key (**Figure 1**).

The commands are also provided in a file "BIASLESS_path.txt" in **BIASLESS**. Users can copy the commands to the command line of R and execute the commands to initialize **BIASLESS**.

Figure 1. Interface of BIASLESS



5. DESCRIPTION OF WORKING DIRECTORIES

BIASLESS is easy-to-use software written in R and provides a user-friendly interface written in R-GUI (**Figure 1**). The interface contains a short preface to introduce **BIASLESS** briefly. Two test examples are provided in **BIASLESS** to demonstrate the data input format, result output format and the execution of **BIASLESS**. Structure of directories and files in **BIASLESS** are shown in a Free Mind map (see **Figure 2**). Four main item questions are designed for providing required/optional information for data analysis by using **BIASLESS**.

Item 1: Input/Output settings:

- Data input format (for markers) – **BIASLESS** accepts two input formats of marker data: "Marker data matrix" and "Sample data file". **Test Example 1** (**Section 7.1**) is prepared complying with the format of a marker data matrix. **Test Example 2** (**Section 7.2**) is prepared complying with the format of a sample data file.
- Data input directory (for markers) – Users can press the browse button to specify the working directory where their marker data are saved.
- Data input directory (for a trait) – Users can press the browse button to specify the working directory where their trait (phenotype) data are saved. A study trait can be two groups (e.g., case group and control group) or more than two groups (e.g., African population, European population, and Asian population).
- Result output directory – Users can press the browse button to specify the working directory where their results should be saved. Note that the output directory must exist before executing **BIASLESS**.
- Missing values (for markers and a trait) – Users should fill in a notation or code to indicate a missing value in their marker data and trait data. The default is NA.

Item 2: Cross-validation:

- Seed – R generates a "Random" seed or users can "Give" a fixed seed (a real value between -65535 and 65535) for partitioning study samples into *x*-fold cross-validation samples (*x*-1 folds for training samples and the remaining fold for testing samples). Note that the results of two analyses for

the same data may be different if a random seed is used. The default is to use a fixed seed, 54321.

- Fold size – Specify the number of folds in a cross-validation procedure (from 10 to 1). For example, if a fold number is 2, all samples are partitioned into two separated parts. In a first cross-validation run, the first part of samples is a training data set and the second part is a testing data set. In a second cross-validation run, the second part of samples becomes a training data set and the first part becomes a testing data set. If a fold number is 1, all samples are training samples and testing accuracy will be calculated by using a leave-one-out procedure. The default is to perform a 10-fold cross-validation.

Item 3: Marker selection (Stop if any of the following three procedures stop):

- Continue until the number of markers in the model increases to a specified value (the default is 10).
- Continue until the training/testing accuracy increases to a specified value (the default is 1).
- Continue until the increment of training/testing accuracy is reduced to a specified value (the default is 0.001).

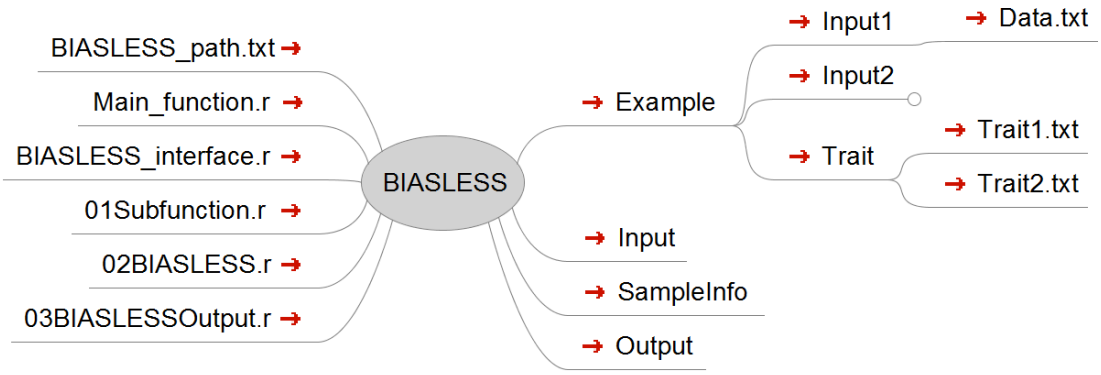
Item 4: Graphical output:

- Overlay line graph – Users can choose to draw the plot or not.
- Parallel coordinates plot – Users can choose to draw the plot or not.
- Multidimensional scaling (MDS) plot – Users can choose to draw the plot or not.
- Stacked-bar/Box-whisker (SBBW) plot – Users can choose to draw the plot or not.
- Sample misclassification plot – Users can choose to draw the plot or not.
- Marker impact plot – Users can choose to draw the plot or not.

Once all item questions are answered, icon “Run” is clicked to submit computational job. Then, **BIASLESS** will check the inputted data information and data files. If the inputted information is invalid or the data files are ill-format, **BIASLESS** will show warning message(s) or error message(s), which provides users to make corrections. If the inputted data pass the examination, **BIASLESS** starts to perform analysis. Then **BIASLESS** starts to perform analysis and a message “Please wait a while, **BIASLESS** is running…”, and a window of *x*-fold cross-validation will

jump up to show the progress of the cross-validation analysis. Please wait until a new window with the message "Computation of **BIASLESS** is finished." jumps up to acknowledge users the completion of **BIASLESS** computation. **Note that users can interrupt the execution of BIASLESS anytime by clicking ESC in the window "R Console".** Once the execution of **BIASLESS** is finished, the numerical results and graphic outputs will be automatically saved in the output directory that users specified. **Note that if users specify an existing directory containing outputs from a previous analysis, all the files and subdirectories in the directory will be removed automatically after executing a new analysis.**

Figure 2. Structure of directories and files in BIASLESS



6. DATA INPUT FORMAT

This section introduces the input format of marker and trait data in **BIASLESS**. **BIASLESS** accepts two input formats of marker data: "Marker data matrix" or "Sample data file". **Test Example 1** is prepared to illustrate the format of a marker data matrix (**Section 7.1**) and **Test Example 2** is prepared to illustrate the format of a sample data file (**Section 7.2**).

6.1 Marker data

- *Marker data matrix:*

As mentioned in Item 1 in **Section 5**, users can specify a working directory where users' marker data matrix is saved, e.g., "C:\BIASLESS\Work". **Note that the directory names are case sensitive.** Marker data of all samples can be arranged in a single data file or multiple data files. Suppose that there are M markers (M_1 SNP

markers and M-M_1 GE markers) and N samples in users' data set.

- ✓ Single data file: If all data are arranged in a single file, then there are M+1 rows and N+1 columns in the file. The first row is a header row to indicate sample ID. The first column is a header column to indicate the marker ID. Marker data of a first sample are arranged in the second column from the second row to the (M+1)th row, marker data of a second sample are arranged in the third column from the second row to (M+1)th row, and so on. Please refer to **Test Example 1** (**Section 7.1**).
- ✓ Multiple data files: If data are arranged in more than one data file, **note that the number of samples (columns) and the column order in different marker data matrices MUST be identical.** For example, genotype data of M_1 SNPs and N samples are arranged in the first marker data matrix, and expression data of M-M_1 GE markers and N samples are arranged in the second marker data matrix. Then the first data matrix has M_1+1 rows and N+1 columns and the second data matrix has M-M_1+1 rows and N+1 columns, where the first row in the two data matrices is a header row for sample ID and the first column in the two data matrices is a header column for marker name.

- *Sample data file:*

As mentioned in Item 1 in **Section 5**, users can specify a working directory where users' sample data files are saved, e.g., "C:\BIASLESS\Work". **Note that the directory names are case sensitive.** Marker data of each sample are saved in an individual data file. There are N data files in total. Each data file contains two columns and M+1 rows. The first column provides ID of markers with a header "Marker ID". The second column provides values of markers with a header "Value". Please refer to **Test Example 2** (**Section 7.2**). **BIASLESS** combines marker value data of all samples as a data matrix having M+1 rows and N+1 columns including a column header and a row header automatically. Partitioned subsets of the data matrix are saved in the "Temp" subdirectory under the BIASLESS directory, and the directory will be removed automatically in the beginning of a next run. **Note that users can repeat the same data analysis with a same or different seed by the proceeding steps: (1) copy the sub-matrices in the Temp subdirectory to a new directory, (2) choose "Marker data matrix" for their input format, and (3) specify the directory as an input directory in a new run.**

6.2 Trait data

Trait data of all samples are arranged in a single data file. The first column indicates IDs of samples with a header "Sample_ID". The second column indicates IDs of groups which the samples belong to (e.g., YRI or CEU in **Test Example 1**, **Section 7.1**) with a header "Group_ID". The third column indicates the ID of cross-validation fold where the sample was assigned as a testing sample with a header "Fold_ID". Please refer to **Test Example 1** and **Test Example 2** for details. **Note that users can specify samples (rather than a random assign by program R) belonging to training samples or testing samples in a cross-validation by giving Fold_ID to samples.**

7. EXAMPLES

In this section, we illustrated the execution of **BIASLESS** by using two real examples.

Test Example 1: Ancestry informative markers for discriminating samples from African and European populations (Data input format – Marker data matrix)

This example is the default example of **BIASLESS** and can be run easily by pressing the "Run" button (keying in **Test example 1** in the directory of data input in Item 1). **Note that the commands, filenames, directory names are case sensitive.** Then **BIASLESS** starts to perform analysis and a message "Please wait a while, **BIASLESS** is running…" will be shown in the command line. The percentage of the analysis completed will be shown. When the computation is finished, a message "Computation of **BIASLESS** is finished." shown to acknowledge users the completion of **BIASLESS** computation. The computational procedure will take about 1 minute and 30 seconds using a machine with a CPU of Intel Core2 Duo E8400 3.00GHz and RAM of DDR2 3.25G. Results of the analysis will be automatically saved in the output directory "C:\BIASLESS\Output\Example1".

In this example, we studied ancestry informative markers for discerning 30 African marriage pairs from Yoruba in Ibadan (YRI) and 30 Caucasian marriage pairs of European descent resided in Utah (CEU) based on 527 SNP markers (on chromosome 19 of Affymetrix 100K SNP array). The conditions set for this example

are shown in the **BIASLESS** interface (**Figure 3**). Marker data were prepared in a marker data matrix format. Namely, except for the header (sample name) and the first column (SNP ID), genotype data of 527 SNPs of 120 samples were arranged as a 527 by 120 data matrix. Each SNP was recoded as 1, 0.5 and 0 for *AA*, *AB* and *BB*. Missing genotypes and trait were indicated by NA. A fixed seed 54321 was used and a 10-fold cross-validation was performed. Marker selection stopped if any one of the following three procedures stopped: (1) marker selection continues until the number of markers in the model increased to 10, (2) marker selection continues until the training/testing accuracy increased to 1, and (3) marker selection continues until the increment of training/testing accuracy was reduced to 0.001.

Numerical output files contain "Description.txt", "Table.csv", and "TraitTable.txt". Description of the conditions and data used for the analysis of **Test Example 1** is provided in file "Description.txt" (**Table 1**). File "Table.csv" summarizes the results of cross-validation analysis. In this example, all 10-fold cross-validation analyses selected either 3 or 4 SNPs and the leave-one-out testing accuracies were greater than 0.96. The sample size used for calculating leave-one-out testing accuracies is listed in the final column. The analyses in folds 3, 6 and 7 selected the same SNP markers (CV accuracy = 3) and their corresponding testing accuracies attained 1 (perfect classification). The analyses in folds 1 and 2 also attained a testing accuracy of 1. Detailed results for each step of marker selection in every fold were also provided. In file "TraitTable.txt" the last column provided the information about in which fold that the study sample was assigned to a testing sample.

In addition to numerical results, graphical results were also provided by **BIASLESS**, including "Overlay line graph.pdf", "Parallel coordinates plot.pdf", MDS plot file(s), SBBW plot files(s), "Sample misclassification plot.pdf", and "Marker impact plot.pdf" where multiple MDS plots and multiple SBBW plots will be provided if multiple folds attained the same maximum testing accuracy. File "Overlay line graph.pdf" (**Figure 4**) shows the profiles of training accuracy (black solid line) and testing accuracy (red dash line) of all 10-fold cross-validation analyses. The results of folds 1, 2, 3, 6, and 7 were framed in blue color, indicating they were the best classification models. "Parallel coordinates plot.pdf" (**Figure 5**) provides the information of fold index, number of markers selected, training accuracy, and testing accuracy in a 10-fold cross-validation analysis. In this example, all 10-fold

cross-validation analyses selected either 3 or 4 SNPs (3 SNPs in folds 2, 4, 5, 8, 9, and 10, and 4 SNPs in folds, 1, 3, 6, and 7). All models in the 10-fold cross-validation analysis attained a training accuracy of 1 but only the model in folds 1, 2, 3, 6 and 7 had a testing accuracy of 1. File "MDS_Fold_3.pdf" (**Figure 6**) displayed a 2-dimensional configuration of HapMap YRI samples (African population, blue color) and HapMap CEU sample (European samples, red color). Samples from African and European populations can be separated clearly. File "SBBW_Fold_3.pdf" (**Figure 7**) displays genotypic distributions (*AA* call – blue color, *AB* call – green color, and *BB* call – brown color) of SNPs and expression distributions (pink color) of genes selected in the best prediction model. In this example, the selected markers presented quite different genetic distributions in HapMap YRI samples and HapMap CEU samples, illustrating that these markers were important ancestry informative markers for discerning samples from African and Caucasian populations.

File "Sample misclassification plot.pdf" (**Figure 8**) displays states of correct classification or misclassification in training and testing samples in each of 10-fold cross-validations. Misclassification proportion across all cross-validations was shown for each individual in training and testing samples. Misclassification proportions across all training samples and all testing samples also were shown respectively for each step of marker selection in each of 10-fold cross-validations. An African sample NA19201 (Obs 55 in TraitTable.txt) had the highest training misclassification rate (34 %) and the highest testing misclassification rate (100%). File "Marker impact plot.pdf" (**Figure 9**) displays states of markers selected (red color) or unselected (blue color) in training samples in each of 10-fold cross-validations. The selection times of markers across all cross-validations was shown in a horizontal bar chart in the right-hand side of this figure. Moreover, this figure also provided the information about which step a marker was selected. For example, SNP_A-1662079 was only selected once in 10-fold cross-validations; this marker entered the model in the third step in the 5^{th} cross-validation.

Figure 3. Conditions used for the analysis of **Test Example 1**.

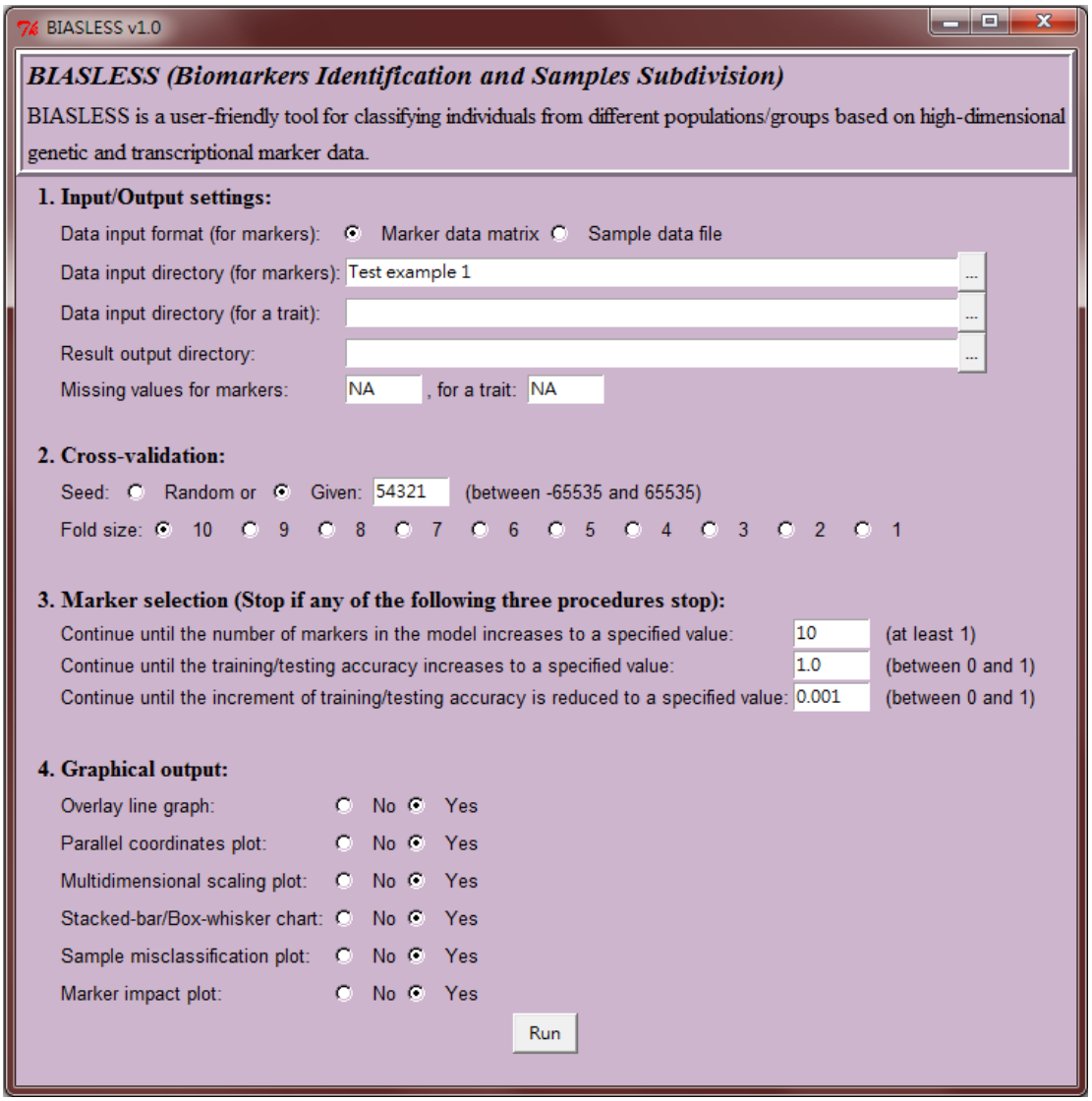


Table 1. Summary of the conditions and data used in the analysis of **Test Example 1**.

==========
========== Welcome to use BIASLESS software ==========
==========

1. Input/Output settings -
   (a) Input data format (for markers): Marker data matrix
   (b) Input data directory name (for markers): C:/BIASLESS/Example/Input1/
   (c) Input data file name (for a trait): C:/BIASLESS/Example/Trait/Trait1.txt
   (d) Result output directory name: C:/BIASLESS/Output/Example1/
   (e) Missing data (for markers and a trait): NA for markers and NA for a trait
   (f) The number of individuals: 120
   (g) The number of markers (SNPs/GE): 527

2. Cross-validation -
   (a) Seed for cross-validations: 54321
   (b) Fold size: 10 folds

3. Marker selection (Stop if any of the following three procedures stop) -
   (a) Continue until the number of markers in the model increases to a specified value: 10.
   (b) Continue until the training/testing accuracy increases to a specified value: 1.
   (c) Continue until the increment of training/testing accuracy is reduced to a specified value: 0.001.

4. Graphical output -
   (a) Overlay line graph: Y
   (b) Parallel coordinates plot: Y

(c) Multidimensional scaling plot: Y
(d) Stacked-bar/Box-whisker chart: Y
(e) Sample misclassification plot: Y
(f) Marker impact plot: Y

Elapsed time: 0-H, 1-M, 35-S

Figure 4. Overlay line graph of a 10-fold cross-validation analysis in **Test Example 1**.

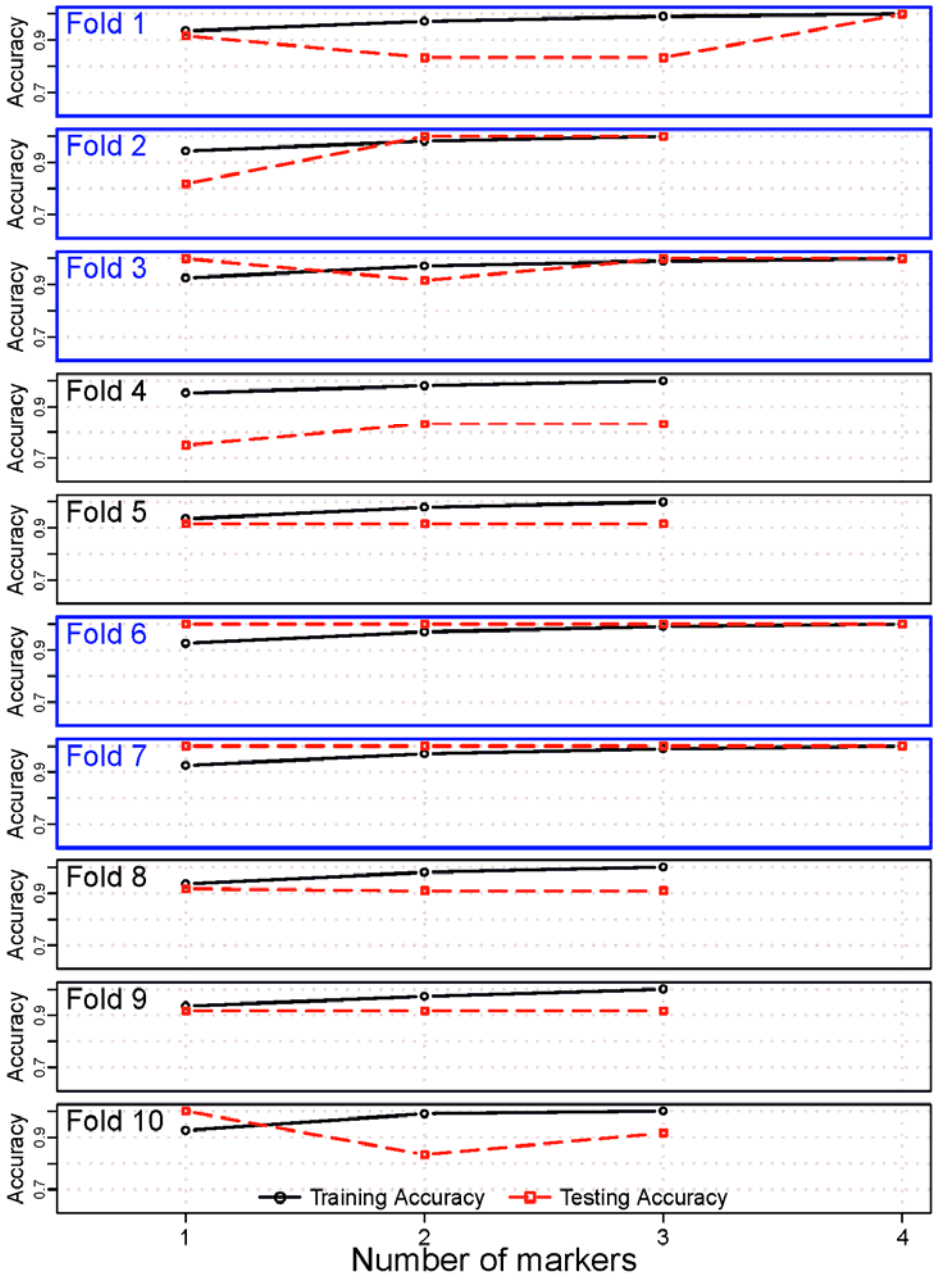


Figure 5. Parallel coordinates plot of a 10-fold cross-validation analysis in **Test Example 1**.

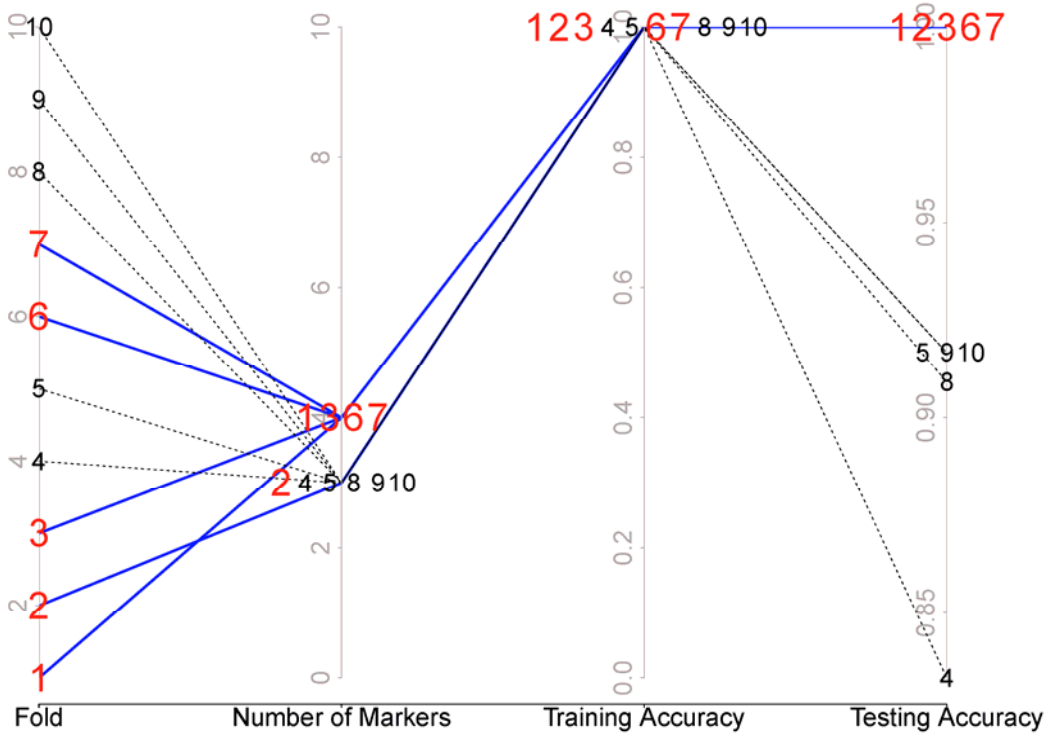


Figure 6. Multidimensional scaling plot of a 10-fold cross-validation analysis in **Test Example 1**.

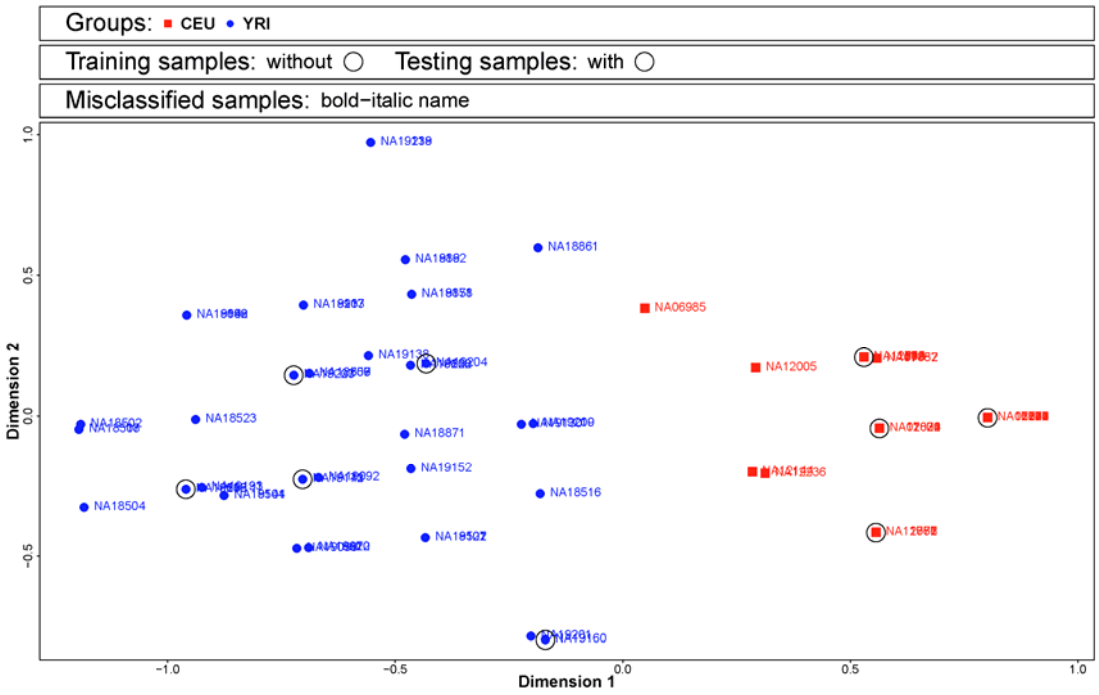


Figure 7. Stacked-bar/Box-whisker plot of a 10-fold cross-validation analysis in **Test Example 1**.

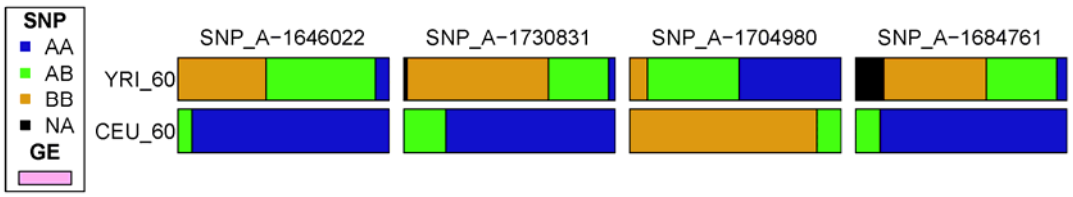


Figure 8. Sample misclassification plot of a 10-fold cross-validation analysis in **Test Example 1**.

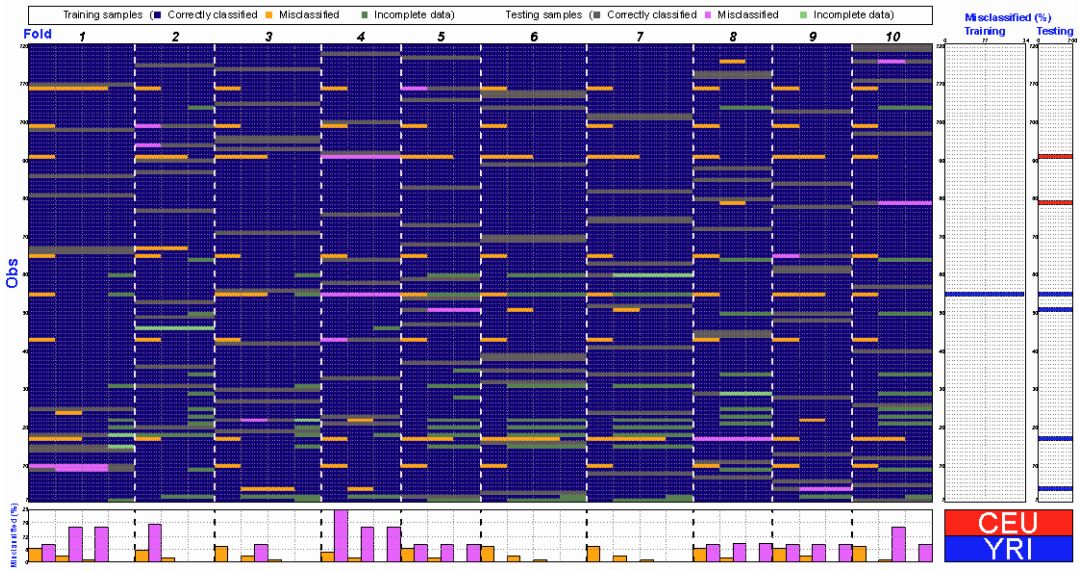
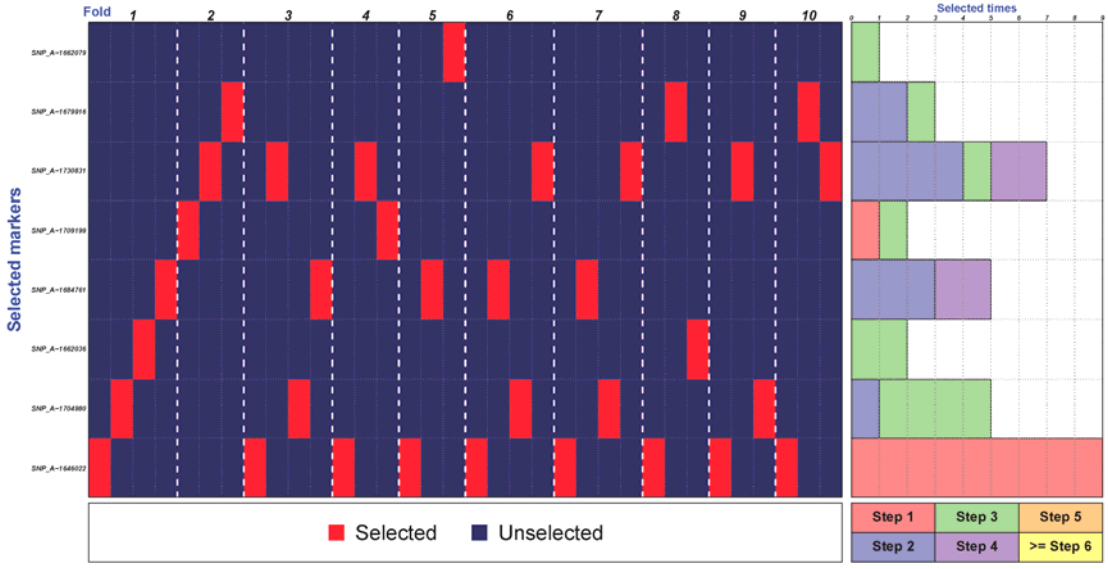


Figure 9. Marker impact plot of a 10-fold cross-validation analysis in **Test Example 1**.



Test Example 2: Ancestry informative markers for discriminating samples from

two closely related ethnic populations (Data input format – Sample data file)

This example is also provided in **BIASLESS** and can be run by pressing the "Run" button easily (keying in **Test example 2** in the directory of data input in Item 1). **Note that the commands, filenames, directory names are case sensitive.** Then **BIASLESS** starts to perform analysis and a message "Please wait a while, **BIASLESS** is running…" will be shown in the command line. The percentage of the analysis completed will be shown. When the computation is finished, a message "Computation of **BIASLESS** is finished." shown to acknowledge users the completion of **BIASLESS** computation. The computational procedure will take about 15 minutes using a machine with a CPU of Intel Core2 Duo E8400 3.00GHz and RAM of DDR2 3.25G. Results of the analysis will be automatically saved in the output directory "C:\BIASLESS\Output\Example2".

In this example, we studied ancestry informative markers for distinguishing 45 Han Chinese persons in Beijing (CHB) and 45 Japanese persons in Tokyo (JPT) based on 5,421 SNP markers and 1,312 GE markers on chromosome 19. The conditions set for this example are shown in the **BIASLESS** interface (**Figure 10**). Marker data of each sample were saved in an individual data file. There were 90 data files in total and each data file contained two columns and 6,734 rows. The first column indicated marker ID and the second column indicates marker value of study sample. Each SNP was recoded as 1, 0.5 and 0 for *AA*, *AB* and *BB*. Missing genotypes and trait were indicated by NA. A fixed seed 54321 was used and a 10-fold cross-validation was performed. Marker selection stopped if any one of the following three procedures stopped: (1) marker selection continues until the number of markers in the model increased to 10, (2) marker selection continues until the training/testing accuracy increased to 1, and (3) marker selection continues until the increment of training/testing accuracy was reduced to 0.001.

Numerical output files contain "Description.txt", "Table.csv", and "TraitTable.txt". Description of the conditions and data used for the analysis of **Test Example 2** is provided in file "Description.txt" (**Table 2**). File "Table.csv" summarizes the results of cross-validation analysis. In this example, all 10-fold cross-validation analyses selected either 2 or 3 SNP/GE markers. Models in folds 1, 4, 6, 7, 8, and 9 attained a testing accuracy of 1 (perfect classification). The leave-one-out testing accuracies in all folds were greater than 0.96. The sample size used for calculating leave-one-out testing accuracies was listed in the final column.

The analyses in folds 1, 4, 6 and 7 identified the same SNP and GE markers (CV accuracy = 4) and their corresponding testing accuracies attained 1. Detailed results for each step of marker selection in every fold were also provided. In file "TraitTable.txt" the last column provides the information about in which fold that the study sample was assigned to a testing sample.

In addition to numerical results, graphical results are also provided by **BIASLESS**, including "Overlay line graph.pdf", "Parallel coordinates plot.pdf", MDS plot file(s), SBBW plot files(s), "Sample misclassification plot.pdf", and "Marker impact plot.pdf" where multiple MDS plots and multiple SBBW plots will be provided if multiple folds attain the same maximum testing accuracy. File "Overlay line graph.pdf" (**Figure 11**) shows the profiles of training accuracy (black solid line) and testing accuracy (red dash line) of all 10-fold cross-validation analyses. The results of folds 1, 4, 6, 7, 8, and 9 were framed in blue color, indicating they were the best classification models. "Parallel coordinates plot.pdf" (**Figure 12**) provides the information of fold index, number of markers selected, training accuracy, and testing accuracy in a 10-fold cross-validation analysis. In this example, all 10-fold cross-validation analyses selected either 2 or 3 SNP and GE markers. All models in the 10-fold cross-validation analysis attained a training accuracy of 1 but only the model in folds 1, 4, 6, 7, 8, and 9 had a testing accuracy of 1. File "MDS_Fold_1.pdf" (**Figure 13**) displayed a 2-dimensional configuration of HapMap CHB samples (Chinese population, red color) and HapMap JPT sample (Japanese population, blue color). Samples from Chinese and Japanese populations can be separated clearly. File "SBBW_Fold_1.pdf" (**Figure 14**) displays genotypic distributions (*AA* call – blue color, *AB* call – green color, and *BB* call – brown color) of SNPs and expression distributions (pink color) of genes selected in the best prediction model. In this example, the selected markers presented quite different genetic distributions in HapMap CHB samples and HapMap JPT samples, illustrating that these markers were important ancestry informative markers for discriminate samples from two closely related ethnic populations. File "Sample misclassification plot.pdf" (**Figure 15**) displays states of correct classification or misclassification in training and testing samples in each of 10-fold cross-validations. Misclassification proportion across all cross-validations was shown for each individual in training and testing samples. Misclassification proportions across all training samples and all testing samples also were shown respectively for each step of marker selection in each of 10-fold

cross-validations. A Japanese sample NA18960 (Obs = 59) was misclassified in all 10-fold cross-validations. File “Marker impact plot.pdf” (**Figure 16**) displays states of markers selected (red color) or unselected (blue color) in training samples in each of 10-fold cross-validations. The selection times of markers across all cross-validations was shown in a horizontal bar chart in the right-hand side of this figure. Moreover, this figure also provided the information about which step a marker was selected. For example, GI_4506928-S was selected in all 10-fold cross-validations and this marker always entered the model in the first, implicating the importance of this GE in differentiating samples from CHB and JPT populations.

Figure 10. Conditions used for the analysis of **Test Example 2**.

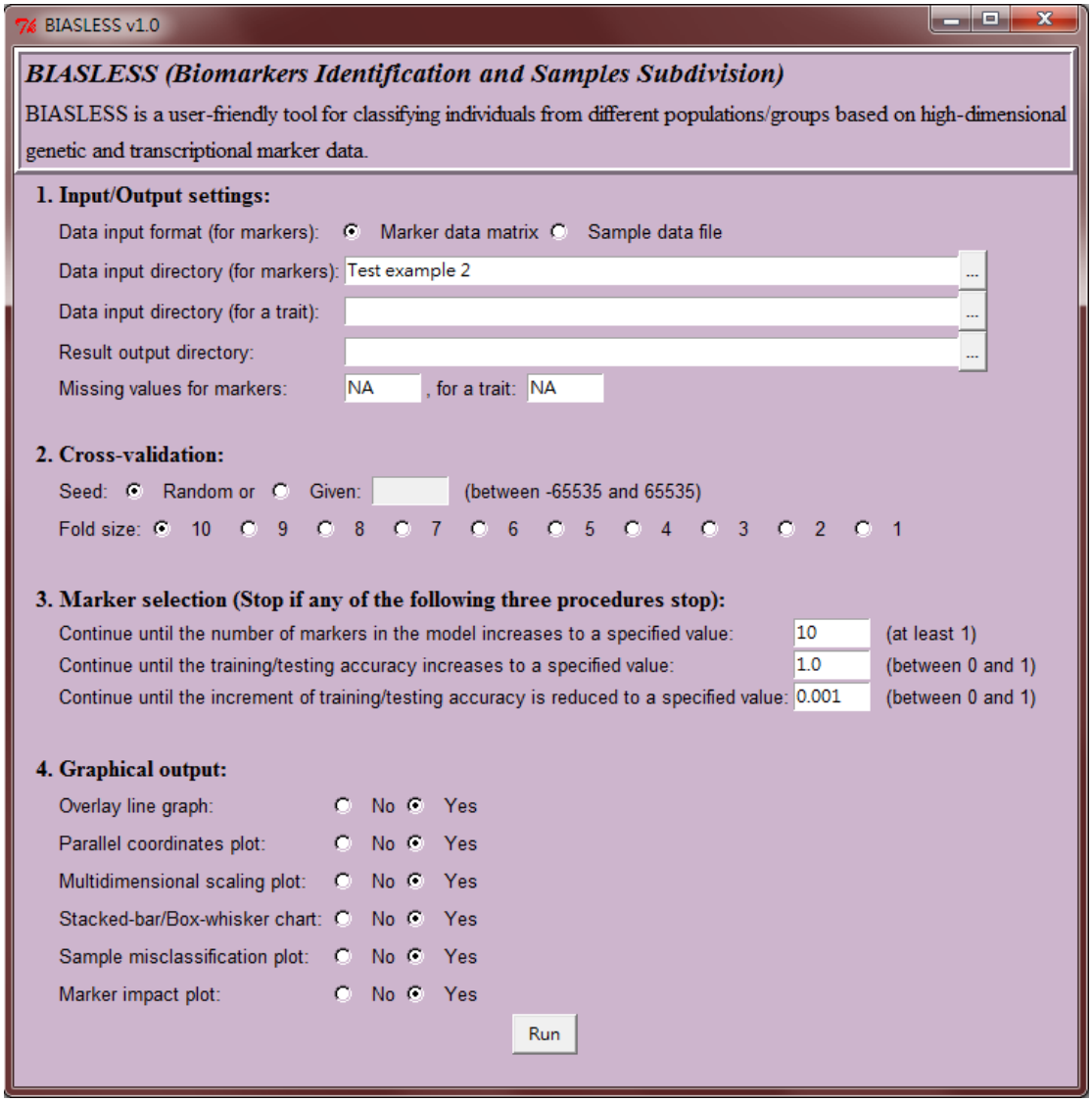


Table 2. Summary of the conditions and data used in the analysis of **Test Example 2**.

```
================================================================
================       Welcome to use BIASLESS software       ================
================================================================
```

1. Input/Output settings -
   (a) Input data format (for markers): Sample data file
   (b) Input data directory name (for markers): C:/BIASLESS/Example/Input2/
   (c) Input data file name (for a trait): C:/BIASLESS/Example/Trait/Trait2.txt
   (d) Result output directory name: C:/BIASLESS/Output/Example2/
   (e) Missing data (for markers and a trait): NA for marker and NA for trait
   (f) The number of individuals: 90
   (g) The number of markers (SNPs/GE): 6733

2. Cross-validation -
   (a) Seed for cross-validations: random
   (b) Fold size: 10 folds

3. Marker selection (Stop if any of the following three procedures stop) -
   (a) Continue until the number of markers in the model increases to a specified value: 10.
   (b) Continue until the training/testing accuracy increases to a specified value: 1.
   (c) Continue until the increment of training/testing accuracy is reduced to a specified value: 0.001.

4. Graphical output -
   (a) Overlay line graph: Y
   (b) Parallel coordinates plot: Y
   (c) Multidimensional scaling plot: Y
   (d) Stacked-bar/Box-whisker chart: Y
   (e) Sample misclassification plot: Y
   (f) Marker impact plot: Y

Elapsed time: 0-H, 14-M, 18-S

Figure 11. Overlay line graph of a 10-fold cross-validation analysis in **Test Example 2**.

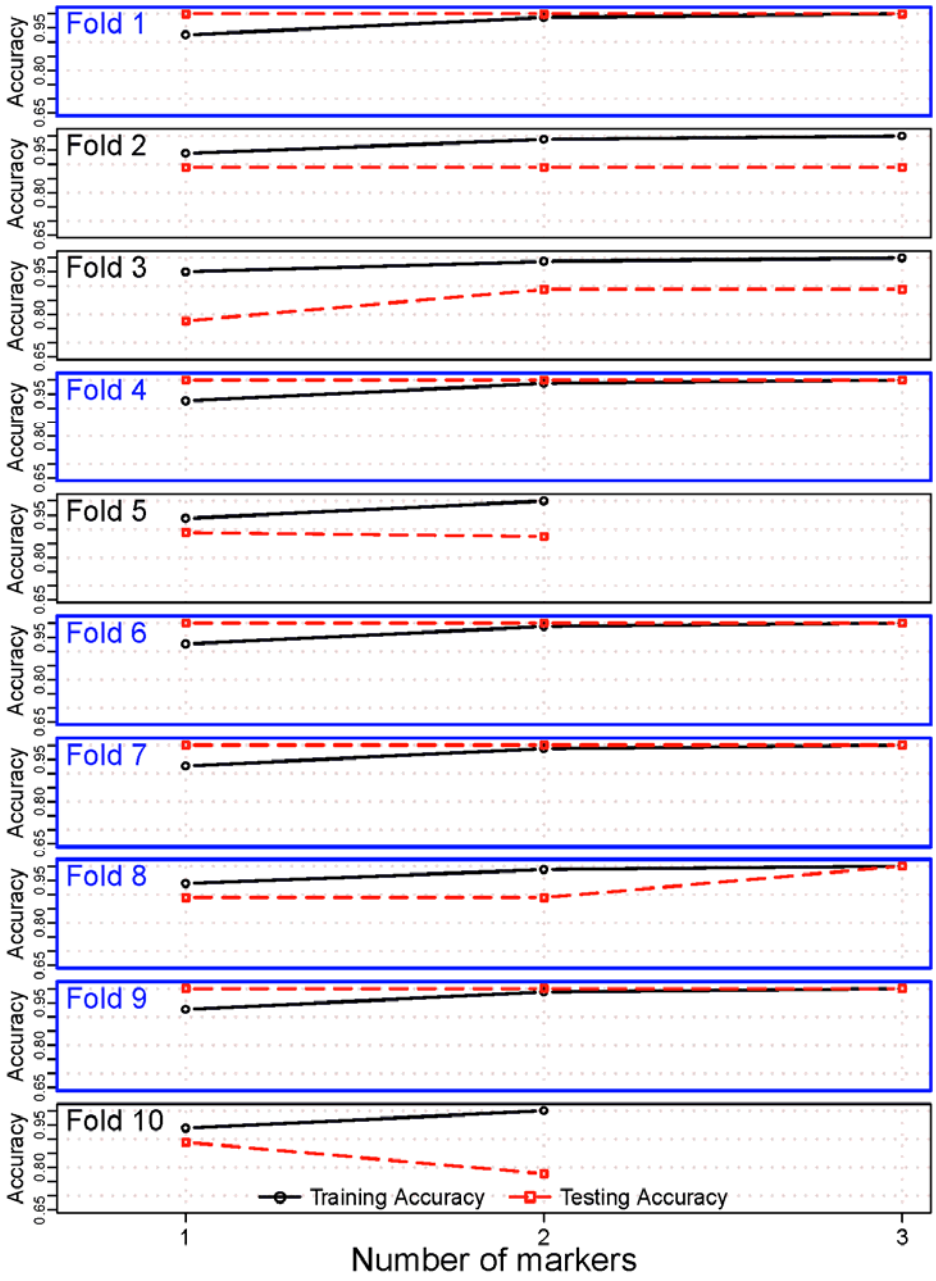


Figure 12. Parallel coordinates plot of a 10-fold cross-validation analysis in **Test Example 2**.

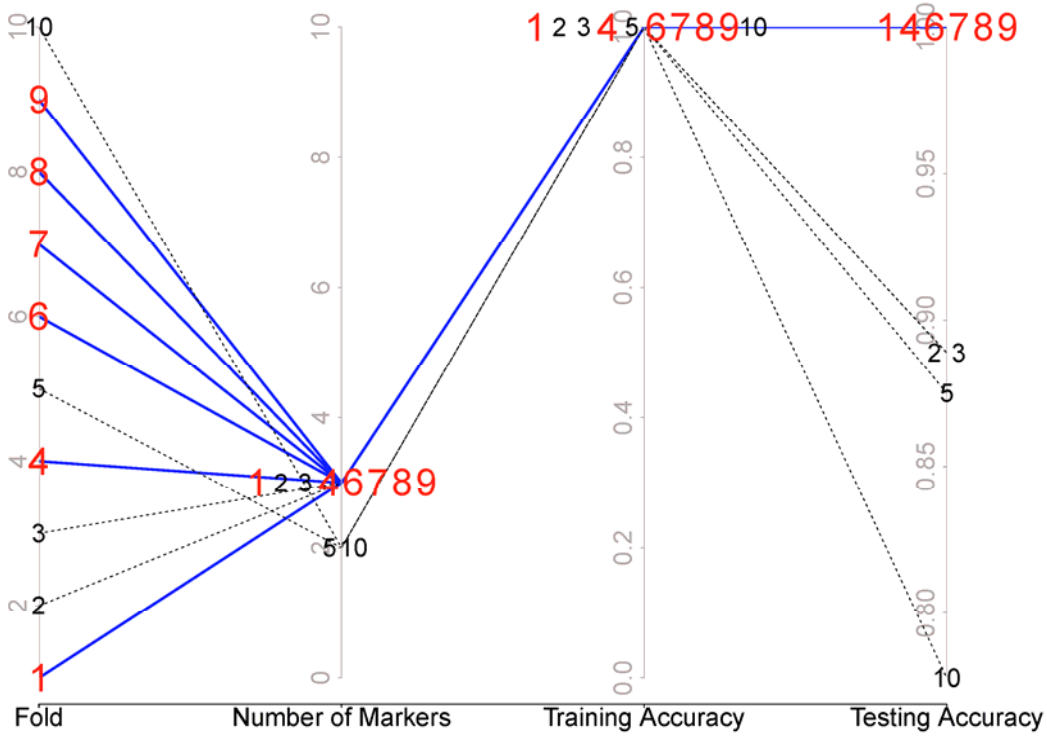


Figure 13. Multidimensional scaling plot of a 10-fold cross-validation analysis in **Test Example 2**.

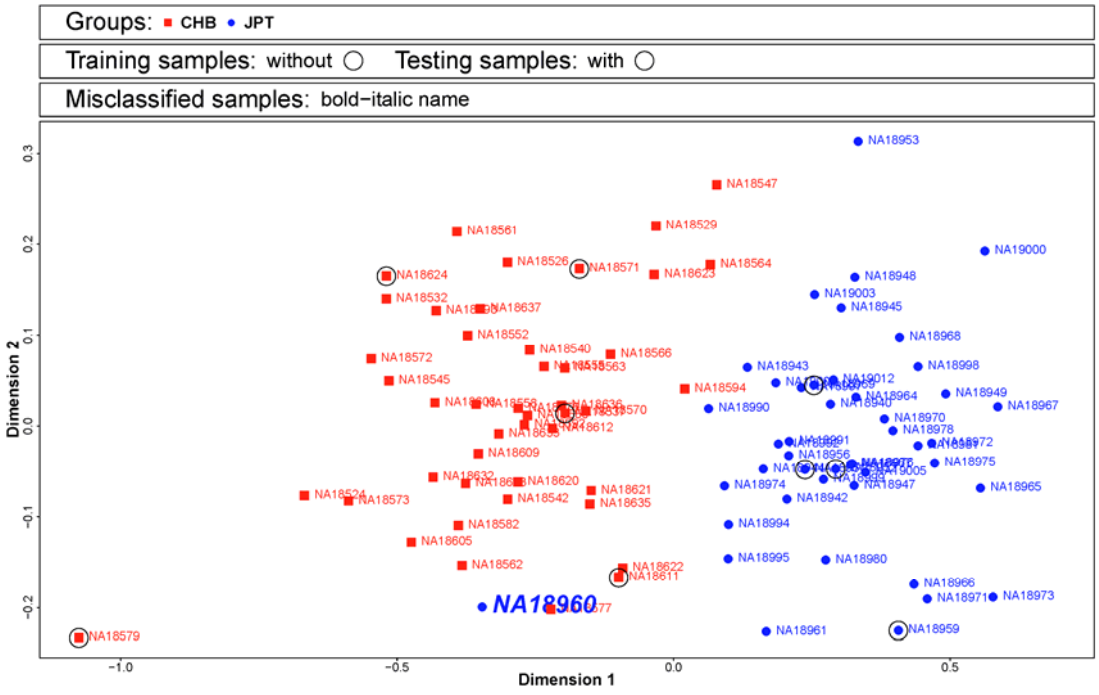


Figure 14. Stacked-bar/Box-whisker plot of a 10-fold cross-validation analysis in **Test Example 2**.

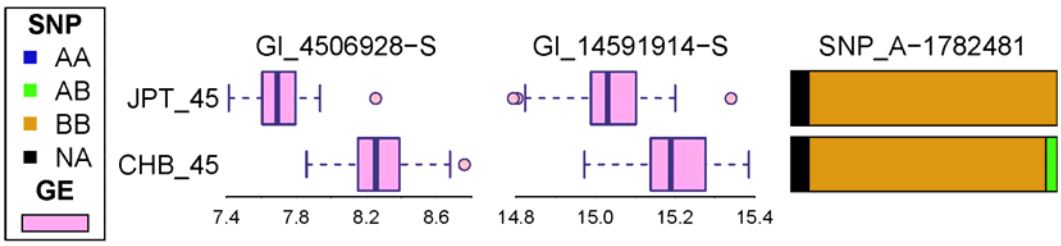


Figure 15. Sample misclassification plot of a 10-fold cross-validation analysis in **Test Example 2**.

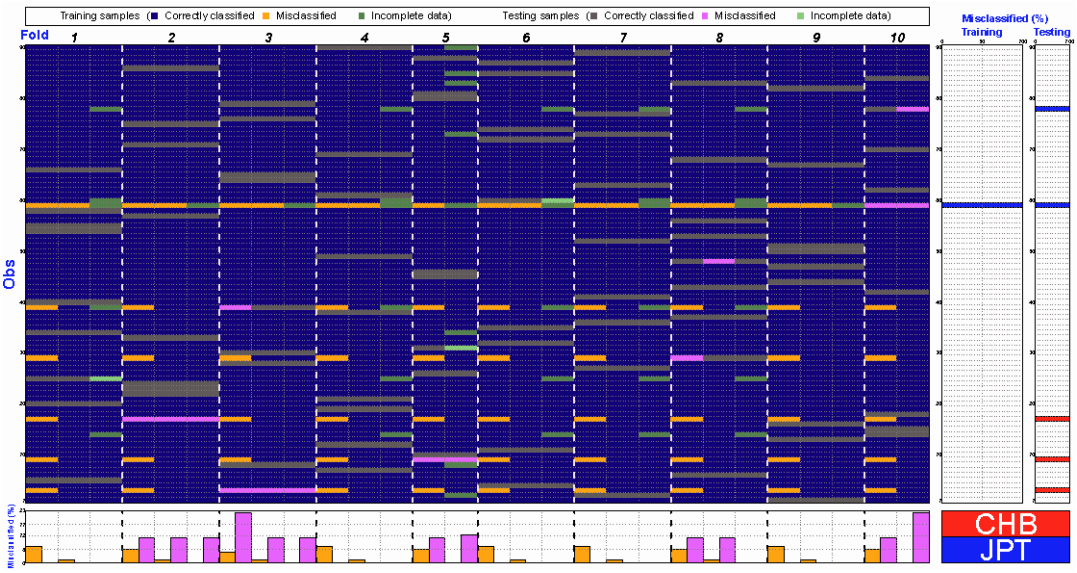


Figure 16. Marker impact plot of a 10-fold cross-validation analysis in **Test Example 2**.