Chapter 8

Rindsel: An R Package for Phenotypic and Molecular Selection Indices Used in Plant Breeding

Sergío Perez-Elizalde, Jesús J. Cerón-Rojas, José Crossa, Delphine Fleury, and Gregorio Alvarado

Abstract

Selection indices are estimates of the net genetic merit of the individual candidates for selection and are calculated based on phenotyping and molecular marker information collected on plants under selection in a breeding program. They reflect the breeding value of the plants and help breeders to choose the best ones for next generation. *Rindsel* is an R package that calculates phenotypic and molecular selection indices.

Key words Smith selection index, Eigen selection index method, Lande and Thompson index, Molecular Eigen selection index method

1 Introduction

Marker-assisted selection uses phenotypic and molecular markers information to select best performer lines at each step of a breeding program. Markers target either major loci controlling desirable traits such as flowering date or quantitative trait loci (QTL) for complex trait such as yield. While it is relatively easy to select the lines carrying the best alleles for 1–3 traits, identifying the best combination of alleles becomes more complicated when multiple loci are tracked simultaneously. Moreover phenotype data of lines under selection remain important information that should be combined with the molecular value to predict the net genetic merit of the individual candidates for selection. Phenotypic and molecular information can be combined in a single value called "selection index."

There are several ways of calculating selection indices using either phenotype data and/or molecular markers information. *Rindsel* is an R package that calculates both types of selection indices. *Rindsel* calculates: the Smith selection index [1], the Eigen selection index method (ESIM) [2, 3], the Restrictive Kempthorne

Delphine Fleury and Ryan Whitford (eds.), *Crop Breeding: Methods and Protocols*, Methods in Molecular Biology, vol. 1145, DOI 10.1007/978-1-4939-0446-4_8,

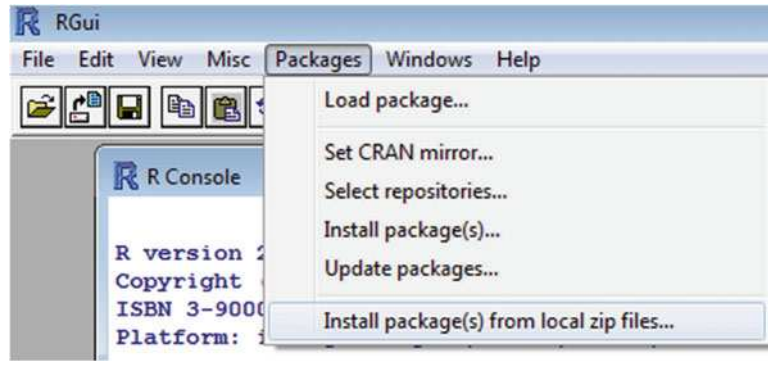

Fig. 1 R console showing the installation of Rindsel package

and Nordskog selection index [4], and the Restrictive Eigen selection index method (RESIM) [5]. These indices are defined as linear combinations of the observed mean phenotypic values of the traits of interest with the traits economic weights previously defined by the breeder. *Rindsel* also calculates molecular selection indices using Lande and Thompson method [6] and Molecular ESIM method (MESIM) [3]. A molecular selection index uses molecular marker or QTL information, i.e., the estimated effect of the QTL linked to the molecular marker (MQTL effect), that is, a combination of the MQTL effects and phenotypic information. The method aims to improve one trait by incorporating information on MQTL effects by means of the molecular selection indices. The final output incorporates all MQTL effects in one index.

2 Software Installation and Help

2.1 Installation

1. Install R from The Comprehensive R Archive Network (CRAN): http://cran.r-project.org/.
2. Save the R zip files of lme4 and Hmisc packages from CRAN.
3. Save the Rindsel zip file from the Integrated Breeding Platform tools repository https://www.integratedbreeding.net/.
4. Install the R packages lme4, Hmisc and Rindsel:
   From the menu, select Packages, then select Install package(s) from local zip file (Fig. 1). Select the zip file from the directory where you saved it.

2.2 Help

1. From the menu, select Help, then Html help (Fig.2). It will open the help browser.
2. Select the link packages and search for *Rindsel*.
   An alternative is to type help.search("*Rindsel*") in the R command prompt.

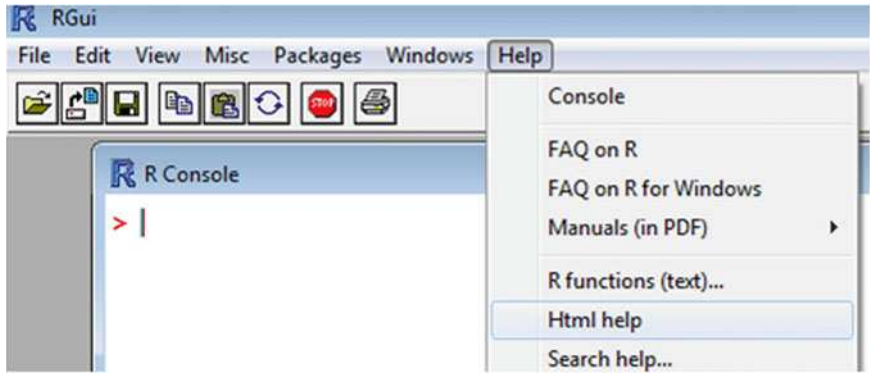

Fig. 2 R console showing how to call the Help menu

	A	B	C	D	E
1	REP	Block	ENTRY	MFL1	FFL1
2	1	12	1	103	101
3	1	12	2	106.86	106.67
4	1	24	3	97.75	101.67
5	1	10	4	103.33	105.14
6	1	19	5	95.3	99.33
7	1	22	6	102.25	104.33
8	1	19	8	103.25	104
9	1	14	9	100	100.6

Fig. 3 Organization of phenotypic data in an input Excel file (example in Rindsel\data\traits.csv)

3 Run *Rindsel*

3.1 Prepare the Input Files

1. Prepare the QTL file.
   Phenotypic data include field design and the trait data per entry (lines or plants). Create and Excel sheet. Data are organized as in Figure with replicates in first column, block in 2d column, entry number in the 3d column; then each trait is organized in the next columns (Fig. 3). The column headings are in the first line.
2. Prepare the markers file.
   Create an Excel spreadsheet. List the traits names in the first column, the indicator variable of the selected traits (0 absence, 1 presence) in the 2d column named Ind, the economic weights (−1 negative, 1 positive, necessary for Lande and Thompson index) in 3d column named eco, and the desired effect of selection (MESIMIndex) in the fourth column named sgn (−1 negative, 1 positive) (Fig. 4).

	A	B	C	D
1	TRAIT	Ind	eco	sgn
2	MFL1	1	-1	-1
3	FFL1	1	-1	-1
4	EHT1	1	-1	-1
5	PHT1	1	-1	-1
6	GY1	1	1	1
7	MFL2	0	-1	-1
8	FFL2	0	-1	-1
9	EHT2	0	-1	-1
10	PHT2	0	-1	-1
11	GY2	0	1	1
12	MFL3	0	-1	-1
13	FFL3	0	-1	-1
14	EHT3	0	-1	-1
15	PHT3	0	-1	-1
16	GY3	0	1	1

Fig. 4 Organization of weight data in an input Excel file (example in Rindsel\data\weights.csv). If known, heritability values can be used as traits' economic weights. In this example, the breeder wants to select plants of early flowering time in environment 1, which is defined by low values of male MFL1 and female FML1 plants. The desired effect for both traits is then set to −1. By contrast, lines of high grain yield in environment1 (GY1) will be selected, therefore the desired effect of GY1 is set to 1

	A	B	C	D
1	V	M1	M2	M3
2	1	1	1	1
3	2	-1	0	1
4	3	-1	0	0
5	4	1	0	1
6	5	0	1	0
7	6	0	-1	0
8	7	-1	-1	0
9	8	-1	-1	0

Fig. 5 Organization of markers data in the input Excel file (example in Rindsel\data\markers.csv). Markers names are headed in the first line

3. Prepare the weight file.
   Create an Excel spreadsheet. Alleles will be listed for each marker in columns. The alleles are noted 1 for AA, 0 for Aa, and −1 for aa (Fig. 5).
4. Prepare the phenotypic data file:
   Create an Excel spreadsheet. For each trait, "Score" is the QTL effect of the marker and reported in the odd columns (Fig. 6). Corresponding marker numbers are reported in the even column headed "Marker." Markers names are headed in the first line.

	A	B	C	D	E	F
1	Score	Marker	Score	Marker	Score	Marker
2	MFL1	MFL1	FFL1	FFL1	EHT1	EHT1
3	7.28	1	-0.42	1	2.65	11
4	-0.63	14	-0.84	10	1.36	14
5	6.65	2	0.34	11	-3.2	8
6	6.71	4	-1.49	13	2.39	9
7	0.11	4	0.48	17	2.82	4
8	-9.47	4	-0.59	3	-8.25	4
9	-6.66	5	0.53	4	-0.84	5
10	-1.09	6	-1.33	6	-8.18	6
11	1.53	7	-0.13	7	0.19	7

Fig. 6 Organization of QTL data in the input Excel file (example in Rindsel\data\qtl.csv)

Fig. 7 R console showing how to load the Rindsel package

3.2 Load Rindsel

1. From the Packages menu select Load package (Fig. 7).
2. The available packages are displayed (Fig. 8). Select *Rindsel*.

3.3 Calculate Lande and Thompson Index

1. On the command prompt, write IndexName() to display the main menu (Fig. 9). Then type the index number you would like to calculate (in the example, 5 for Lande and Thompson index) and press Enter.
2. A window automatically opens requesting the phenotypic data file (Fig. 10). Browse and select the file.
3. A new window opens requesting the weight data file (Fig. 11). Browse and select the file.
4. The R routine starts by calculating the genetic and phenotypic covariance matrices. When the calculation is done, a new window opens requesting the markers file (Fig. 12). Browse and select the marker file.
5. A new window opens requesting the QTL data file (Fig. 13). Browse and select the file.
6. The output file is displayed in the R console (Fig. 14). Genetic, phenotypic, and molecular covariance matrices are displayed. A percentage of the individuals are automatically selected, by default 5 %, and sorted in LT index from highest to lowest.

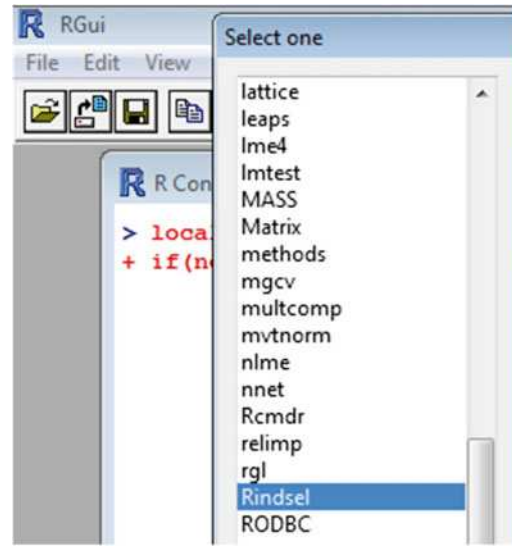

Fig. 8 R console showing how to start Rindsel

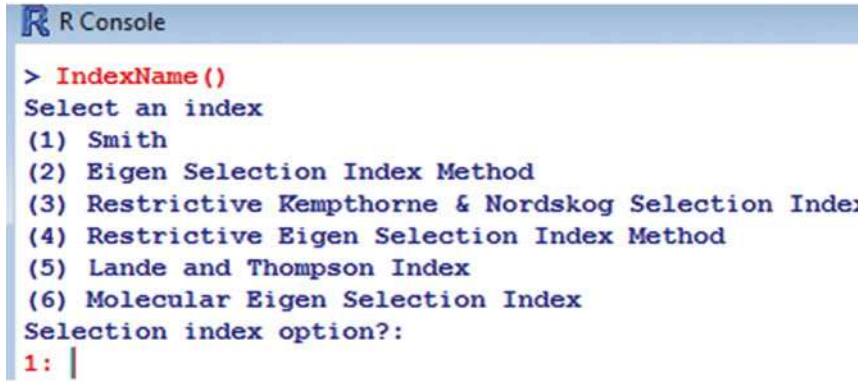

Fig. 9 R console showing how to run an index function

The values per trait and the LT index are also provided for all plants of the population.

3.4 Calculate Other Selection Index

1. For the other selection index methods, we proceed the same way by typing the number corresponding of the method, for example 6 for Molecular Eigen Selection Index.
2. It is possible to modify the default parameters for each method.

   For example, by typing MESIMIndex(selval = 10, rawdata = FALSE), selval = 10 means that we select the 10 % of traits with the highest values of the MESIM index, rawdat = FALSE means that we use covariance matrices that are already calculated. In the example above, we can use the covariance matrices obtained by calculating the LT index to calculate the MESIMIndex.

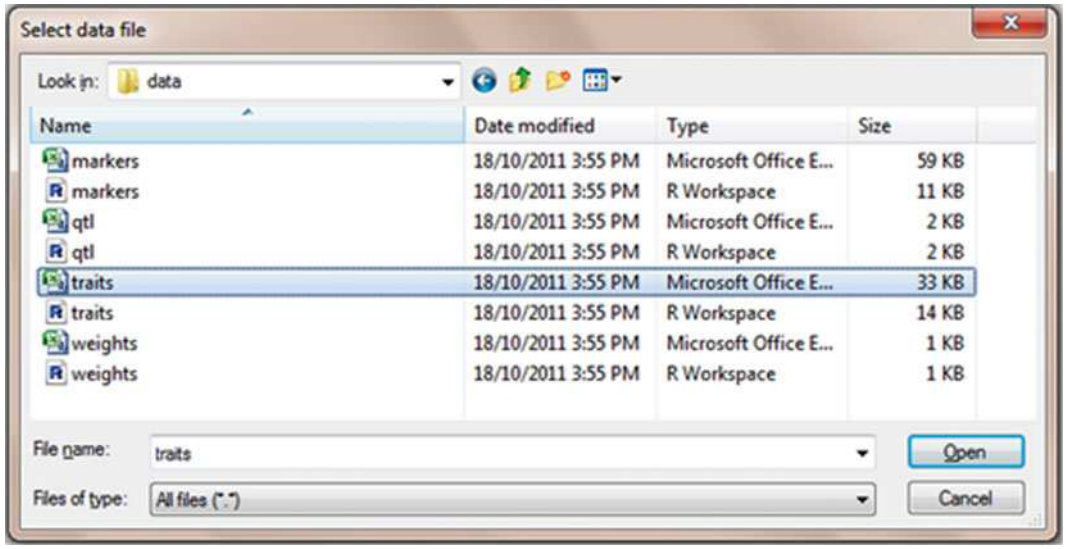

Fig. 10 R console showing how to load the phenotypic data file

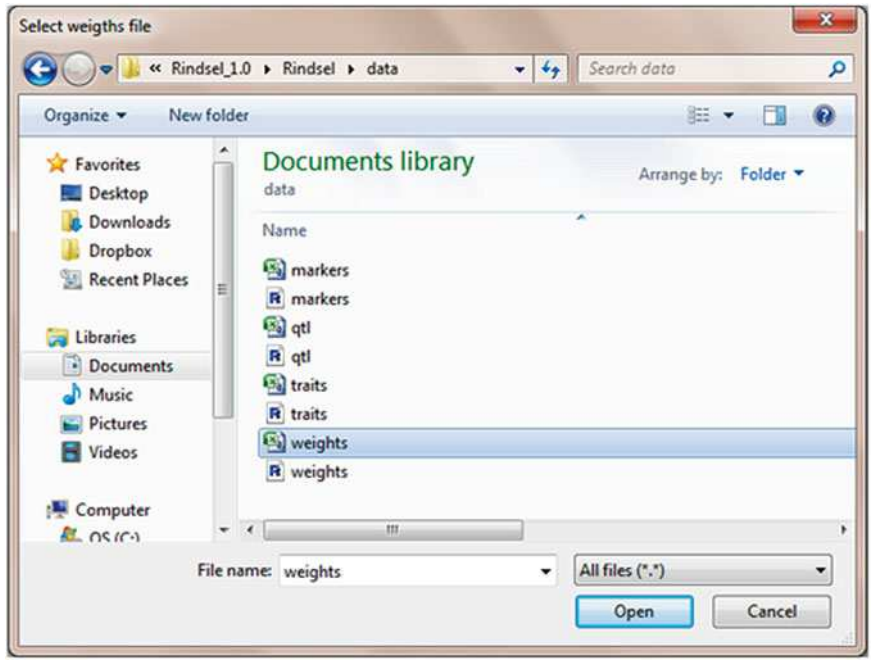

Fig. 11 R console showing how to load the weight data file

4 Notes

Simulation was used to compare the efficiency of selection indices (3). The simulation program used original data from actual doubled-haploid maize mapping population of 236 genotypes with five traits and their respective QTL. The five traits were male flowering

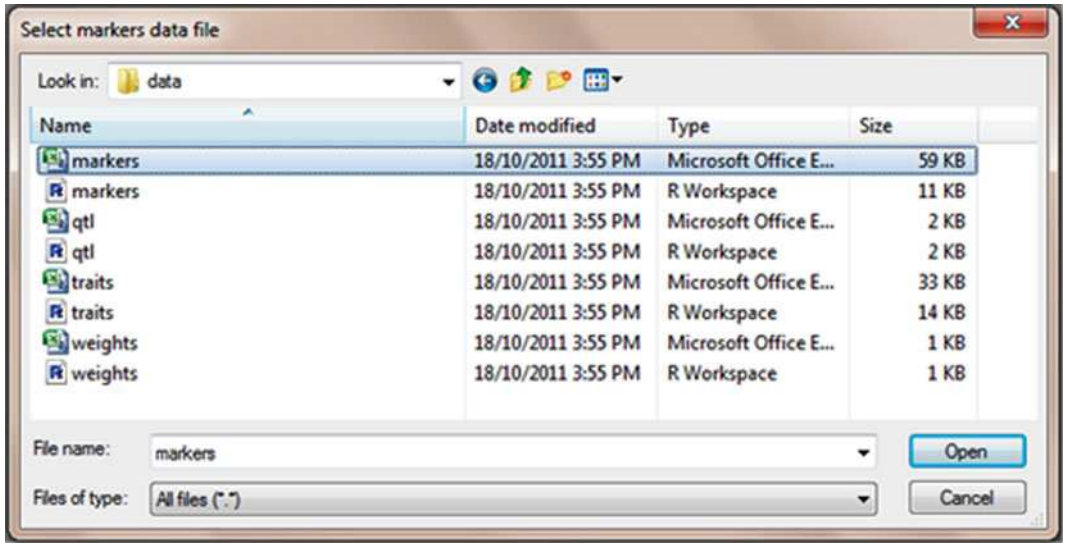

Fig. 12 R console showing how to load the marker data file

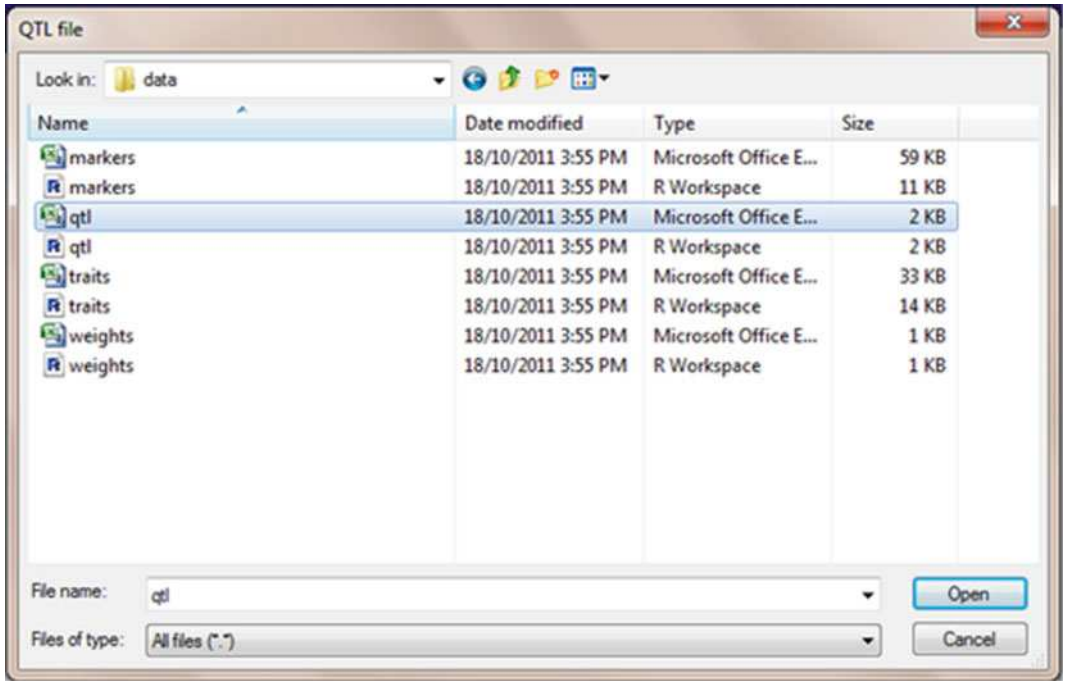

Fig. 13 R console showing how to load the QTL data file

time (MFL) (days), female flowering time (FFL) (days), plant height (PHT) (cm), ear height (EHT) (cm), and 100-kernel weight (HKF) (g). The data are provided with *Rindsel* package (under Rindsel/data folder) as an example dataset for training purposes.

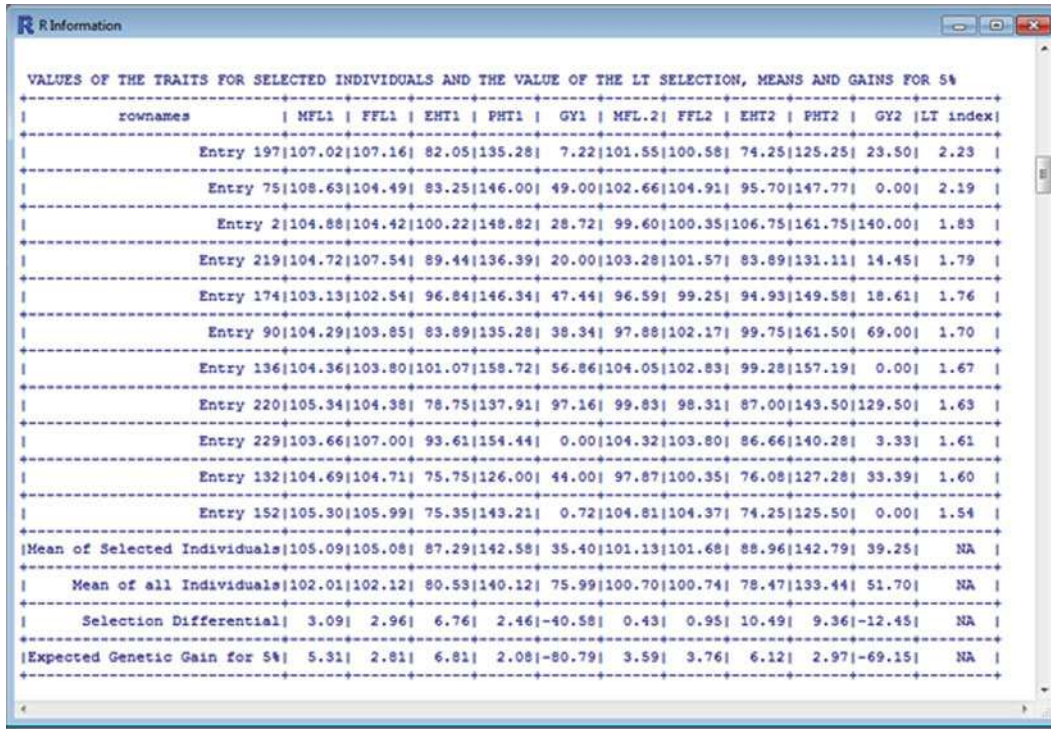

Fig. 14 Output results displayed in the R console showing the Lande and Thompson index per trait and per entry line (LT)

Simulation results show that when genotypes are selected on the basis of individual traits, MESIM increased the response to selection over the Lande and Thompson index. When several traits are selected simultaneously, MESIM outperformed Lande and Thompson for traits with low heritability. For traits with high heritability, ESIM performed very well. MESIM can be considered a generalization of ESIM (2) when information on QTL is incorporated through molecular markers. Lande and Thompson index requires economic weights, which can be chosen as –1, –1, –1, –1, and 1 for each trait in the maize example above. Heritability of the traits can also be used as weights. However it can be difficult to choose weights when no estimates of economic weights are available. MESIM has the advantage over Lande and Thompson to do not require economic weight.

Acknowledgments

This work is supported by the Generation Challenge Programme through the Integrated Breeding Platform project and CIMMYT.

References

1. Smith HF (1936) A discriminant function for plant selection. In: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, NC, pp 466–476
2. Cerón-Rojas JJ, Crossa J, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A (2006) A selection index method based on Eigen analysis. Crop Sci 46:1711–1721
3. Cerón-Rojas J, Castillo-González F, Sahagún-Castellanos J, Santacruz-Varela A, Benítez-Riquelme I, Crossa J (2008) A molecular selection index method based on Eigen analysis. Genetics 180:547–557
4. Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19
5. Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008) A restricted selection index method based on Eigen analysis. J Agric Biol Environ Stat 13:421–438
6. Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756