

USER'S GUIDE MANUAL

BGLR-R (Bayesian Generalized Linear Regression with R)

Francisco Rodríguez

Paulino Pérez

Gustavo de los Campos

Gregorio Alvarado

Angela Pacheco

Juan Burgueño

José Crossa

Contents

1	BGLR-R description	3
2	BGLR-R requirements and installation	4
3	Initial window	5
4	Opening input phenotypic files	7
5	Analyzing the data	9
6	Some details about the models	20
7	Problems with the data set	21
8	Examples of BGLR-R analyses.	22

1 BGLR-R description

BGLR is a software to simplify the selection of input files and parameters to perform Bayesian Generalized Linear Regression using R (statistical software). BGLR provides predictions, GWAS analysis and analysis of reaction norm model described in reference [2]. It allows the inclusion of markers information, relationship matrix (pedigree), environmental covariables, and other variables as fixed or random effects.

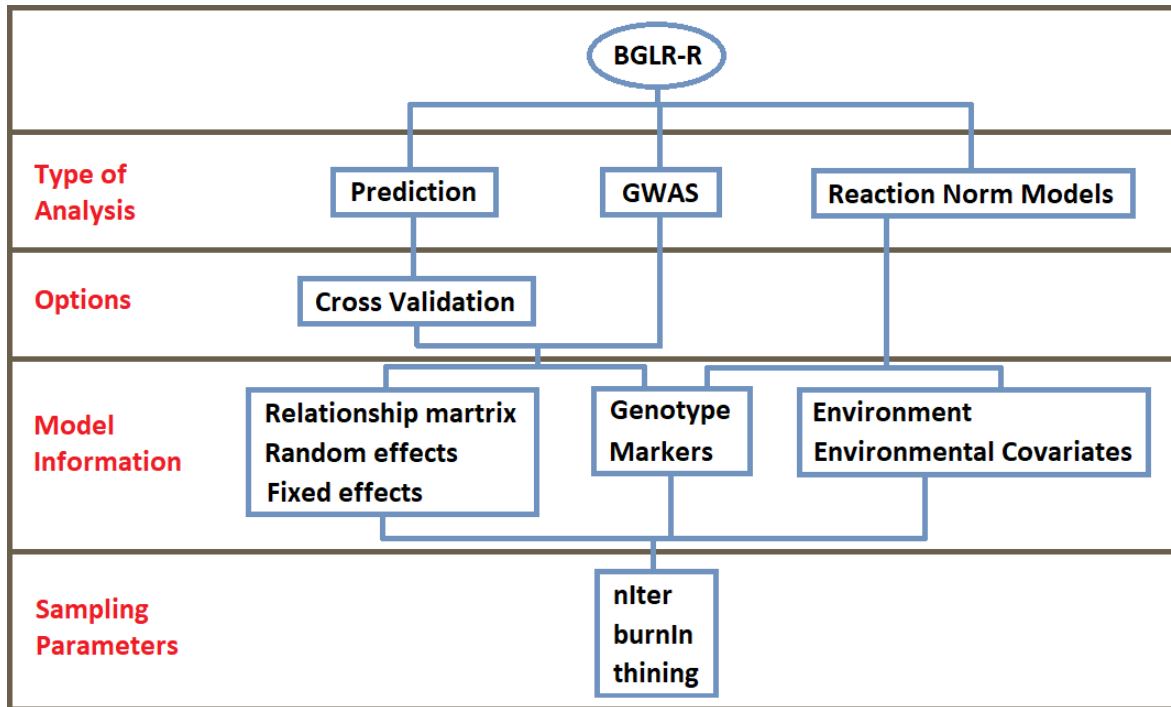


Figure 1: Flow diagram showing the logical process that BGLR-R follows for analysis

2 BGLR-R requirements and installation

The requirements to install BGLR-R in your computer are the following:

- Download BGLR-R from CIMMYT repository
- WINDOWS operating system (W7 or higher is necessary).
- Java Software. It can be downloaded from <https://www.java.com/>.

The installation is really easy, you just need to click twice the file BGLR-R_v2.0_setup.exe and follow the instruction in the setup installer.

3 Initial window

Once BGLR-R has been installed, you can start to analyze your own data sets. The first step is to open BGLR-R from the start menu or from the desktop if you have created a shortcut for BGLR-R. Figure 2 shows the initial window of BGLR-R. On this initial window there are the buttons “*Open phenotypic data*”, “*Analyze*” and “*Help*”.

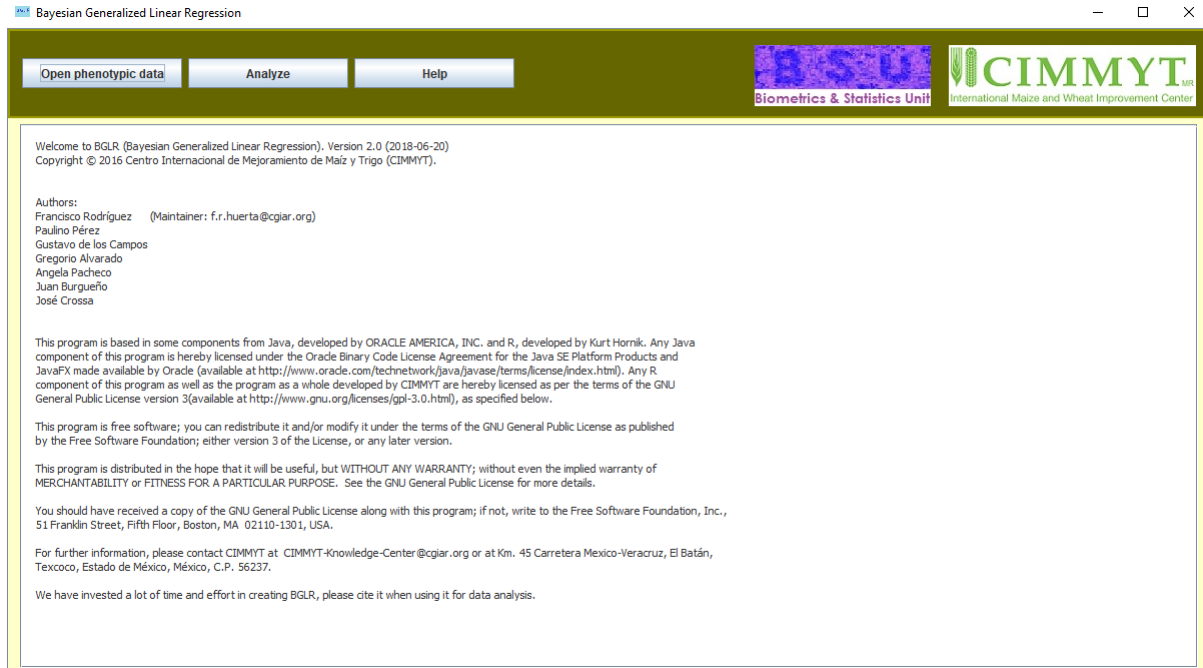


Figure 2: Initial window

- Button “*Open phenotypic file*” Opens a window to choose a phenotypic data set to analyze.
- Button “*Analyze*” Executes the analysis.
- Button “*Help*” displays a new window with three options:
 - *User’s Manual*: this option displays the manual with the description of BGLR-R.
 - *License*: this option displays the reference to cite BGLR-R and the license file. Figure 3 shows the window with this information.

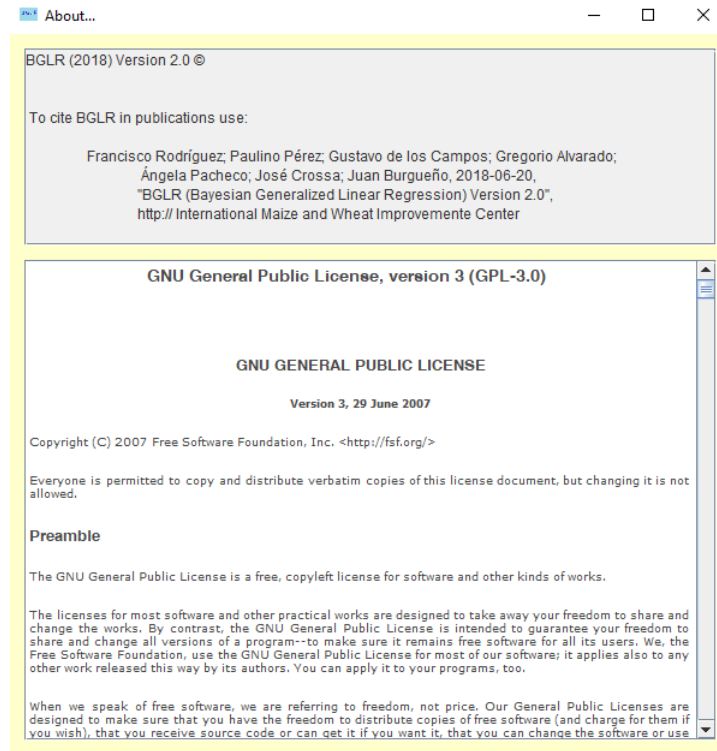


Figure 3: License window

- *Examples*: this option allows you to load the example data set. The example data sets are located in “C:\BGLR-R\Examples”.

Also, you can find in the initial window some information about the software, the authors and some other details.

4 Opening input phenotypic files

Input files can be opened by clicking on “Open phenotypic file” and choosing a file, three kind of files could be loaded: *comma separated file* (*.csv), *Excel file* (*.xlsx) first sheet, or *tab delimited Text file* (*.txt). The files can be stored in any folder. Missing values must be indicated by “.”, “-”, “NA” or blank (empty) spaces. Colum headers must be in the first row, as is shown in Figure 4.

	A	B	C	D	E	F	G	H	I	J	K
1	SUBJECT.NAME	PROJECT.NAME	PHENOTYPE.NAME	Obesity.BMI	Obesity.BodyLength	Date.Month	Date.Year	Date.Season	Date.StudyStartSeconds	Date.Hour	Date.StudyDay
2	A048005080	HS_mouse	Obesity	-0.520131667	8.2	5	2003	spring	9072000	0	106
3	A048006063	HS_mouse	Obesity	-0.401116443	8.2	3	2003	spring	5443200	0	64
4	A048006555	HS_mouse	Obesity	-0.526935426	8.1	4	2003	spring	6652800	0	78
5	A048007096	HS_mouse	Obesity	-0.41527421	7.6	3	2003	spring	3628800	0	43
6	A048010273	HS_mouse	Obesity	-0.546143102	7.8	1	2003	winter	-518400	0	-5
7	A048010371	HS_mouse	Obesity	-0.394471031	6.7	3	2003	spring	4838400	0	57
8	A048011040	HS_mouse	Obesity	-0.547702329	6	3	2003	spring	6048000	0	71
9	A048011287	HS_mouse	Obesity	-0.480725379	8.8	4	2003	spring	8467200	0	99
10	A048011567	HS_mouse	Obesity	-0.513760484	8.2	3	2003	spring	6048000	0	71
11	A048013559	HS_mouse	Obesity	-0.512076424	7.5	2	2003	winter	3024000	0	36
12	A048015047	HS_mouse	Obesity	-0.465510577	6.9	5	2003	spring	9072000	0	106
13	A048017615	HS_mouse	Obesity	-0.478513544	8.1	2	2003	winter	1209600	0	15
14	A048021023	HS_mouse	Obesity	-0.413322322	7.6	3	2003	spring	5443200	0	64

Figure 4: Example of csv data set

When you click the button “Open file”, the initial folder to look for a file is “Documents”. Figure 5 shows the window to look for and choose a dataset

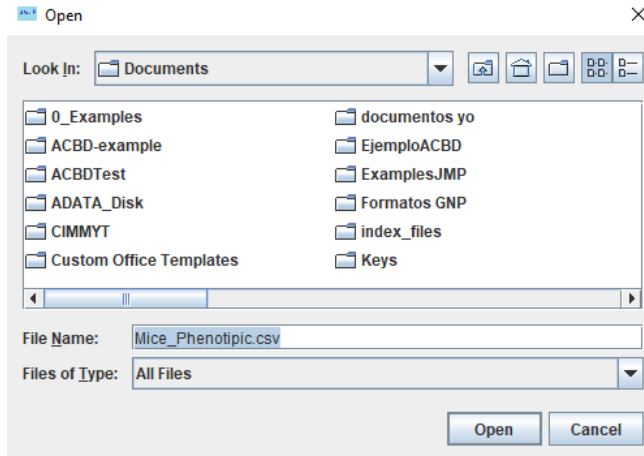


Figure 5: Open file

Once dataset has been chosen, BGLR-R will display the data set in a table at the bottom of the screen to allow the user to check whether the data were read correctly or whether an error was made. Also, some panels will be displays to select the options to analyze the data. Figure 6 shows the data set and panels in BGLR-R.

5 Analyzing the data

Analyzing a data set is easy and you just need to choose some options and set some parameter that are described in the following pharagraphs.

Analysis. In this menu, you can select among three different analyses:

- a) Prediction.
- b) GWAS.
- c) Models.

We will explain the different analysis options.

- a) **Prediction.** This option is used to get predictions for genotypes. To produce predictions you must define a column to identify the genotypes and add information like markers, relationship matrix (pedigree), fixed effects or random effects. Figure 6 shows the four options of additional information chosen. To include markers information you must click button “Load Markers” and a new window is opened to look for corresponding file in your documents. To include a valid markers file, you must go to apendix 1. Button “See Markers” allows you to see the first 50 rows and 50 columns of the original markers file. This markers matrix can be included using one of the following model: “GBLUP”, “LASSO”, “BayesA” or “BayesB”.

To include relationship matrix (pedigree) information you must click button “Load Relation” and a new window is opened to look for corresponding file in your documents. To include a valid relationship matrix file, you must go to apendix 2. Button “See Relation” allows you to see the first 50 rows and 50 columns of the original relationship matrix file.

Cross validation

The screenshot displays a software interface with five main panels, all of which have their respective options selected for a prediction analysis:

- Analysis type:** The 'Analysis' dropdown is set to 'Prediction'. The checkboxes for 'Markers', 'Relationship matrix', 'Random effects', and 'Fixed effects' are all checked. The 'Model for Markers' is set to 'Ridge'. The 'Plots Format' is set to 'PDF (*.pdf)'. The 'Output folder' is 'Analysis1'. Buttons for 'Load Markers', 'See Markers', 'Load Relation', and 'See Relation' are visible.
- Genotype selection:** The 'Genotype (L):' dropdown is set to 'Choose variable'. The 'Parameters' section shows 'nIter' at 5,000, 'burnIn' at 1,000, and 'thinning' at 10. The 'Cross Validation' checkbox is checked.
- Response Variables:** The 'Select one or more' list contains: SUBJECT.NAME, PROJECT.NAME, PHENOTYPE.NAME, Obesity.BMI, Obesity.BodyLength, Date.Month, and Date.Year. The 'Selected' list is empty. The 'Drop All' button is at the bottom.
- Random effects:** The 'Select one or more' list contains the same variables as the Response Variables panel. The 'Selected' list is empty. The 'Drop All' button is at the bottom.
- Fixed effects:** The 'Select one or more' list contains the same variables as the other panels. The 'Selected' list is empty. The 'Drop All' button is at the bottom.

Figure 7: All options selected for prediction

- b) **GWAS.** This option is used to perform an analysis of GWAS. To perform the GWAS analysis you must define a column to identify the genotypes and add information like relationship matrix (pedigree), fixed effects or random effects. In this case, markers information is a mandatory requirement and you can't unselect this option. Figure 7 shows the three options of additional information chosen.

To include markers information you must click button “Load Markers” and a new window is opened to

look for corresponding file in your documents. To include a valid markers file, you must go to appendix 1. Button “See Markers” allows you to see the first 50 rows and 50 columns of the original markers file. This markers matrix can be included using one of the following model: “GBLUP”, “LASSO”, “BayesA” or “BayesB”.

To include relationship matrix (pedigree) information you must click button “Load Relation” and a new window is opened to look for corresponding file in your documents. To include a valid relationship matrix file, you must go to appendix 2. Button “See Relation” allows you to see the first 50 rows and 50 columns of the original relationship matrix file.

Figure 8: Frame for covariates

- c) **Models.** This option allows you to try the reaction norm models used in reference [1]. There are eight defined models and according to the selected model, the options to set can change. For this models, the maximum options to set are shown in Figure 8 and they are “Sites” (environments), “Genotype”, markers and environmental covariables.

Figure 9: Frame for covariates

Load Markers

The file with markers information should be a .csv file. The first column of this matrix should contain the names of the genotypes that should match with the column of genotypes that you specify in phenotypic data. The following

columns must be the markers information and the first row must contain the names of the markers. Figure 16 shows an example of the structure of markers file.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		rs3683945_G	rs3707673_G	rs6269442_G	rs6336442_G	rs13475700_A	rs3658242_T	rs13475701_C	rs6198069_G	rs3659303_G	rs3674785_G	rs3677817_G	rs13475705_A
2	A048005080	1	1	1	1	0	1	0	1	1	1	1	0
3	A048006063	1	1	2	1	1	1	1	2	1	1	1	1
4	A048006555	2	0	2	2	0	2	0	2	2	0	0	0
5	A048007096	1	1	1	1	1	1	1	2	1	1	1	1
6	A048010273	2	0	2	2	0	2	0	2	2	0	0	0
7	A048010371	1	1	1	1	0	1	0	1	1	1	1	0
8	A048011040	1	1	1	1	0	1	0	1	1	1	1	0
9	A048011287	1	1	2	1	1	1	1	2	1	1	1	1
10	A048011567	1	1	2	1	1	1	1	2	1	1	1	1
11	A048013559	2	0	2	2	0	2	0	2	2	0	0	0
12	A048015047	2	0	2	2	0	2	0	2	2	0	0	0
13	A048017615	1	1	1	1	0	1	0	1	1	1	1	0
14	A048021023	1	1	1	1	0	1	0	1	1	1	1	0

Figure 10: Example of markers file

Also it is possible to consider other kind of marker codes, like examples in Figure 17, where letters are used separated by another character or simply both letters together. When recoding is done, you get a new folder called “Info_Markers” with the following two files:

- “Monomorphic_Markers.csv” which lists the markers that are monomorphic.
- “Recode_Markers_Info.csv” which lists all markers indicating the minor and major alleles (recoded as 0), and frequencies of minor allele (recoded as 2), NA (missing) and Heterozigous (recoded as 1).

	A	B	C	D	E	F	G
1	Entry	ZM001504	ZM001481	ZM001456	ZM001455	ZM001467	ZM001467
2	1	CG	CC	TT	GT	GG	TC
3	3	CG	CG	AA	GT	CC	CC
4	4	CC	??	TA	GT	GC	TC
5	5	GG	??	TT	TT	GC	TC
6	6	CC	CG	TA	GT	GC	TT
7	7	??	CG	TA	GT	CC	??
8	8	CG	CC	TA	GT	GC	CC
9	9	CG	CC	TA	GG	GC	TC
10	10	GG	CG	TT	TT	GG	TT
11	11	??	CC	TA	GG	GC	TC
12	12	CC	CG	TA	GG	CC	TC
13	13	CG	CG	TT	GT	CC	TC
14	14	GG	CG	AA	GT	GG	TC

	A	B	C	D	E	F	G
1	Entry	ZM001504	ZM001481	ZM001456	ZM001455	ZM001467	ZM001467
2	1	C_G	C_C	T_T	G_T	G_G	T_C
3	3	C_G	C_G	A_A	G_T	C_C	C_C
4	4	C_C	?_?	T_A	G_T	G_C	T_C
5	5	G_G	?_?	T_T	T_T	G_C	T_C
6	6	C_C	C_G	T_A	G_T	G_C	T_T
7	7	?_?	C_G	T_A	G_T	C_C	?_?
8	8	C_G	C_C	T_A	G_T	G_C	C_C
9	9	C_G	C_C	T_A	G_G	G_C	T_C
10	10	G_G	C_G	T_T	T_T	G_G	T_T
11	11	?_?	C_C	T_A	G_G	G_C	T_C
12	12	C_C	C_G	T_A	G_G	C_C	T_C
13	13	C_G	C_G	T_T	G_T	C_C	T_C
14	14	G_G	C_G	A_A	G_T	G_G	T_C

	A	B	C	D	E	F	G
1	Entry	ZM001504	ZM001481	ZM001456	ZM001455	ZM001467	ZM001467
2	1	C.G	C.C	T.T	G.T	G.G	T.C
3	3	C.G	C.G	A.A	G.T	C.C	C.C
4	4	C.C	?_?	T.A	G.T	G.C	T.C
5	5	G.G	?_?	T.T	T.T	G.C	T.C
6	6	C.C	C.G	T.A	G.T	G.C	T.T
7	7	?_?	C.G	T.A	G.T	C.C	?_?
8	8	C.G	C.C	T.A	G.T	G.C	C.C
9	9	C.G	C.C	T.A	G.G	G.C	T.C
10	10	G.G	C.G	T.T	T.T	G.G	T.T
11	11	?_?	C.C	T.A	G.G	G.C	T.C
12	12	C.C	C.G	T.A	G.G	C.C	T.C
13	13	C.G	C.G	T.T	G.T	C.C	T.C
14	14	G.G	C.G	A.A	G.T	G.G	T.C

Figure 11: Examples markers codification

Output of markers recode and imputing processes.

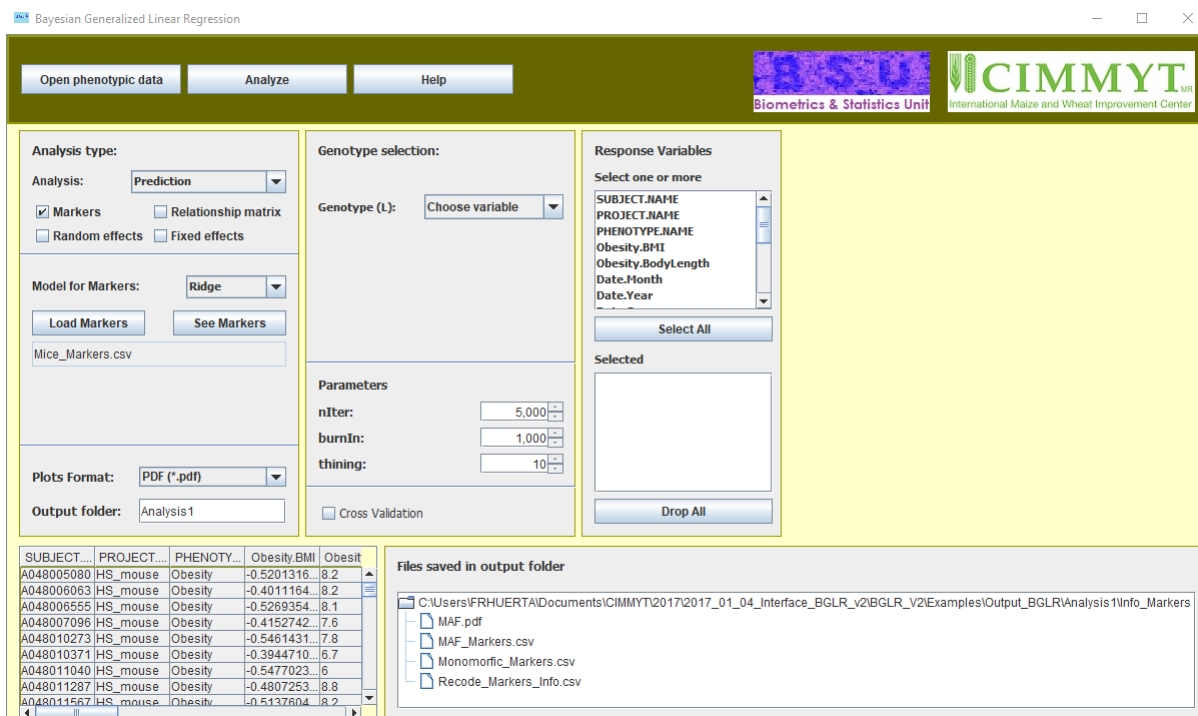


Figure 12: Outputs recode and impute

Markers window view

The 'Markers View' window displays a matrix of marker data. The columns represent different markers, and the rows represent individual samples. The data is binary, with values 0 and 1.

rs3683945...	rs3707673...	rs6269442...	rs6336442...	rs1347570...	rs3658242...	rs1347570...	rs6198069...	rs3659303...	rs3674785...	rs...
1	1	1	1	0	1	0	1	1	1	1
1	1	2	1	1	1	1	2	1	1	1
2	0	2	2	0	2	0	2	2	0	0
1	1	1	1	1	1	1	2	1	1	1
2	0	2	2	0	2	0	2	2	0	0
1	1	1	1	0	1	0	1	1	1	1
1	1	1	1	0	1	0	1	1	1	1
1	1	2	1	1	1	1	2	1	1	1
1	1	2	1	1	1	1	2	1	1	1
2	0	2	2	0	2	0	2	2	0	0
2	0	2	2	0	2	0	2	2	0	0
1	1	1	1	0	1	0	1	1	1	1
1	1	1	1	0	1	0	1	1	1	1
1	1	1	1	0	1	0	1	1	1	1
2	0	2	2	0	2	0	2	2	0	0
0	2	0	0	0	0	1	1	0	2	1
1	1	1	1	0	1	1	2	1	1	0
2	0	2	2	0	2	0	2	2	0	0
1	1	1	1	0	1	0	1	1	1	1
1	1	1	1	1	1	1	2	1	1	1

Figure 13: View of markers matrix

Load Relationship Matrix.

The file with relationship matrix information should be a .csv file. The first column and first row of this matrix should contain the names of the genotypes that should match with the column of genotypes that you specify in phenotypic data. Figure 18 shows an example of the structure of relationship matrix file.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		A048005080	A048006063	A048006555	A048007096	A048010273	A048010371	A048011040	A048011287	A048011567	A048013559	A048015047	A048017615	A048021023
2	A048005080	1	0	0	0	0	0	0	0	0	0	0	0	0
3	A048006063	0	1	0	0	0	0	0	0	0	0	0	0.5	0
4	A048006555	0	0	1	0	0	0	0	0	0	0	0	0	0
5	A048007096	0	0	0	1	0	0	0	0	0	0	0	0	0
6	A048010273	0	0	0	0	1	0	0	0	0	0	0	0	0
7	A048010371	0	0	0	0	0	1	0	0	0	0	0	0	0
8	A048011040	0	0	0	0	0	0	1	0	0	0	0	0	0
9	A048011287	0	0	0	0	0	0	0	1	0	0	0	0	0
10	A048011567	0	0	0	0	0	0	0	0	1	0	0	0	0
11	A048013559	0	0	0	0	0	0	0	0	0	1	0	0	0
12	A048015047	0	0	0	0	0	0	0	0	0	0	1	0	0
13	A048017615	0	0.5	0	0	0	0	0	0	0	0	0	1	0
14	A048021023	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 14: Example of relationship matrix

Relationship matrix view

A048005080	A048006063	A048006555	A048007096	A048010273	A048010371	A048011040	A048011287	A048011567	A048013559	A048015047	A048017615	A048021023
1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0
0	0.5	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0	0

Figure 15: View of relationship matrix

Load Environmental Covariables.

The file with environmental covariables should be a .csv file. Each column must be the an environmental covariable information and the firs row must contain the names of the environmental covariables. In this case, it is not necessary to include a first column with genotypes or locations information, this matrix should have the same number of rows like phenotypic information, and info of each row should match with info of phenotypic information. Figure 19 shows an example of the structure of environmental covariables file.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17
2	-0.10586	-0.13796	-1.29061	-0.90931	-0.39041	0.355177	-1.84527	-0.20853	-0.86982	0.542878	-0.37423	0.414712	0.520737	0.953359	-0.74662	0.130422	-0.22391
3	-0.10586	-0.79526	-1.56425	-1.23174	-0.62521	-0.14423	-1.84208	-0.57613	-0.86982	-0.13027	0.864394	0.693583	0.228146	0.97918	-0.61137	-0.23753	-0.22391
4	-0.10586	0.114852	-2.19304	-1.43461	-0.00085	0.178916	-1.67666	-0.11186	-0.86982	0.812428	-0.56014	0.338656	0.664843	1.001341	-0.92431	0.253191	-0.22391
5	-0.60319	-0.05743	-1.56425	-1.23174	-0.62521	-0.14423	-1.84208	-0.57613	-1.21844	0.364922	0.864394	0.693583	0.228146	0.97918	-0.61137	-0.23753	-0.27107
6	-0.41649	0.004989	-1.65158	-0.64484	0.420726	-0.76604	-1.79596	-0.39233	-1.02353	0.578159	-0.41114	0.508823	0.728076	0.484033	-0.60838	-0.1135	-0.27107
7	-0.69552	0.038073	-1.56425	-1.23174	-0.62521	-0.14423	-1.84208	-0.57613	-1.33388	0.496935	0.864394	0.693583	0.228146	0.97918	-0.61137	-0.23753	-0.27107
8	-0.69552	0.391382	-1.65158	-0.64484	0.420726	-0.76604	-1.79596	-0.39233	-1.33388	0.941146	-0.41114	0.508823	0.728076	0.484033	-0.60838	-0.1135	-0.27107
9	-0.41649	0.308985	-1.29061	-0.90931	-0.39041	0.355177	-1.84527	-0.20853	-1.02353	0.807097	-0.37423	0.414712	0.520737	0.953359	-0.74662	0.130422	-0.27107
10	0.075916	0.697874	-1.20328	-0.2717	0.559473	0.335592	-1.29491	0.352108	-0.71242	1.361901	-0.2279	0.462486	0.747514	1.057929	-0.92866	0.798784	-0.22391
11	-0.69552	0.695377	-1.29061	-0.90931	-0.39041	0.355177	-1.84527	-0.20853	-1.33388	1.170084	-0.37423	0.414712	0.520737	0.953359	-0.74662	0.130422	-0.27107
12	-0.10586	0.114852	-2.19304	-1.43461	-0.00085	0.178916	-1.67666	-0.11186	-0.86982	0.812428	-0.56014	0.338656	0.664843	1.001341	-0.92431	0.253191	-0.22391
13	-0.10586	-0.13796	-1.29061	-0.90931	-0.39041	0.355177	-1.84527	-0.20853	-0.86982	0.542878	-0.37423	0.414712	0.520737	0.953359	-0.74662	0.130422	-0.22391
14	-0.31473	-0.12859	-1.65158	-0.64484	0.420726	-0.76604	-1.79596	-0.39233	-0.94416	0.453027	-0.41114	0.508823	0.728076	0.484033	-0.60838	-0.1135	-0.27107
15	0.075916	1.204741	-1.46527	-0.67745	0.804949	0.369865	-0.74932	0.623961	-0.71242	1.760492	-0.12048	0.579233	0.778891	1.053378	-0.47429	1.087402	-0.22391

Figure 16: Example of covariables file

Environmental covariables matrix view

c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17
-0.1058638...	-0.1379575...	-1.2906077...	-0.9093072...	-0.3904098...	0.3551767...	-1.8452650...	-0.2085303...	-0.8698221...	0.5428776...	-0.37423...	0.414712...	0.520737...	0.953359...	-0.74662...	0.130422...	-0.22391...
-0.1058638...	-0.7952619...	-1.5642492...	-1.2317355...	-0.6252124...	-0.1442280...	-1.8420838...	-0.5761320...	-0.8698221...	-0.1302719...	0.864394...	0.693583...	0.228146...	0.97918...	-0.61137...	-0.23753...	-0.22391...
-0.1058638...	0.1148518...	-2.1930424...	-1.4346117...	-0.0008510...	0.1789162...	-1.6766587...	-0.1118578...	-0.8698221...	0.8124282...	-0.56014...	0.338656...	0.664843...	1.001341...	-0.92431...	0.253191...	-0.22391...
-0.6031926...	-0.0574330...	-1.5642492...	-1.2317355...	-0.6252124...	-0.1442280...	-1.8420838...	-0.5761320...	-1.2184360...	0.3649218...	0.364922...	0.864394...	0.693583...	0.228146...	0.97918...	-0.61137...	-0.23753...
-0.4164892...	0.0049890...	-1.6515816...	-0.6448436...	0.4207263...	-0.7660360...	-1.7959556...	-0.3923311...	-1.0235321...	0.5781586...	0.578159...	-0.41114...	0.508823...	0.728076...	0.484033...	-0.60838...	-0.1135...
-0.6955185...	0.0380727...	-1.5642492...	-1.2317355...	-0.6252124...	-0.1442280...	-1.8420838...	-0.5761320...	-1.3338803...	0.4969348...	0.496935...	0.864394...	0.693583...	0.228146...	0.97918...	-0.61137...	-0.23753...
-0.6955185...	0.3913816...	-1.6515816...	-0.6448436...	0.4207263...	-0.7660360...	-1.7959556...	-0.3923311...	-1.3338803...	0.9411457...	0.941146...	-0.41114...	0.508823...	0.728076...	0.484033...	-0.60838...	-0.1135...
-0.4164892...	0.3089845...	-1.2906077...	-0.9093072...	-0.3904098...	0.3551767...	-1.8452650...	-0.2085303...	-1.0235321...	0.8070973...	0.807097...	-0.37423...	0.414712...	0.520737...	0.953359...	-0.74662...	0.130422...
0.0759155...	0.6978740...	-1.2032753...	-0.2716963...	0.5594733...	0.3355922...	-1.2949085...	0.3521084...	-0.7124236...	1.3619006...	1.361900...	-0.2279...	0.462486...	0.747514...	1.057929...	-0.92866...	0.798784...
-0.6955185...	0.6953771...	-1.2906077...	-0.9093072...	-0.3904098...	0.3551767...	-1.8452650...	-0.2085303...	-1.3338803...	1.1700844...	1.170084...	-0.37423...	0.414712...	0.520737...	0.953359...	-0.74662...	0.130422...
-0.1058638...	0.1148518...	-2.1930424...	-1.4346117...	-0.0008510...	0.1789162...	-1.6766587...	-0.1118578...	-0.8698221...	0.8124282...	0.812428...	-0.56014...	0.338656...	0.664843...	1.001341...	-0.92431...	0.253191...
-0.1058638...	-0.1379575...	-1.2906077...	-0.9093072...	-0.3904098...	0.3551767...	-1.8452650...	-0.2085303...	-0.8698221...	0.5428776...	0.542877...	-0.37423...	0.414712...	0.520737...	0.953359...	-0.74662...	0.130422...
-0.3147255...	-0.1285942...	-1.6515816...	-0.6448436...	0.4207263...	-0.7660360...	-1.7959556...	-0.3923311...	-0.9441571...	0.4530274...	0.453027...	-0.41114...	0.508823...	0.728076...	0.484033...	-0.60838...	-0.1135...
0.0759155...	1.2047412...	-1.4652725...	-0.6774487...	0.8049487...	0.3698651...	-0.7493238...	0.6239613...	-0.7124236...	1.7604917...	1.760491...	-0.12048...	0.579233...	0.778891...	1.053378...	-0.47429...	1.087402...
-0.2104998...	1.1679122...	-1.2032753...	-0.2716963...	0.5594733...	0.3355922...	-1.2949085...	0.3521084...	-0.9023155...	1.5440564...	1.544056...	-0.2279...	0.462486...	0.747514...	1.057929...	-0.92866...	0.798784...
-0.1058638...	-0.4419530...	-1.6515816...	-0.6448436...	0.4207263...	-0.7660360...	-1.7959556...	-0.3923311...	-0.8698221...	0.3139389...	0.313938...	-0.41114...	0.508823...	0.728076...	0.484033...	-0.60838...	-0.1135...
-0.4164892...	0.3089845...	-1.2906077...	-0.9093072...	-0.3904098...	0.3551767...	-1.8452650...	-0.2085303...	-1.0235321...	0.8070973...	0.807097...	-0.37423...	0.414712...	0.520737...	0.953359...	-0.74662...	0.130422...
0.0759155...	0.6978740...	-1.2032753...	-0.2716963...	0.5594733...	0.3355922...	-1.2949085...	0.3521084...	-0.7124236...	1.3619006...	1.361900...	-0.2279...	0.462486...	0.747514...	1.057929...	-0.92866...	0.798784...
-0.6031926...	-0.0574330...	-1.5642492...	-1.2317355...	-0.6252124...	-0.1442280...	-1.8420838...	-0.5761320...	-1.2184360...	0.3649218...	0.364921...	-0.41114...	0.508823...	0.728076...	0.484033...	-0.60838...	-0.1135...
-0.3147255...	1.3227190...	-1.2032753...	-0.2716963...	0.5594733...	0.3355922...	-1.2949085...	0.3521084...	-0.9441571...	1.6351344...	1.635134...	-0.12048...	0.579233...	0.778891...	1.053378...	-0.47429...	1.087402...
-0.1058638...	0.1148518...	-2.1930424...	-1.4346117...	-0.0008510...	0.1789162...	-1.6766587...	-0.1118578...	-0.8698221...	0.8124282...	0.812428...	-0.56014...	0.338656...	0.664843...	1.001341...	-0.92431...	0.253191...

Figure 17: View of relationship matrix

Response variables. The panel *Response Variables* is used to choose the response variables that you want to analyze. Response variables must be numeric. There is an option to standardize the response variables, it can be used when the interest is to select variables and the best option is to use all variables in the same scale. Figure 18 shows the check box used to standardize the response variables.

The parameters *nIter*, *burnIn* and *thining* are common for the three type of analysis, and they must be set to get the best results. A small description of these parameters is as follows:

- *nIter*: the number of iterations of the sampler.
- *burnIn*: the number of samples discarded.
- *thining*: the thinning used to compute posterior means

Figure 9 shows the parameters *nIter*, *burnIn* and *thining* with they default values.

Genotype selection:

Genotype (L): Choose variable ▼

☒ Standardize Response Variables

Parameters

nIter: 5,000

burnIn: 1,000

thining: 10

☐ Cross Validation

Figure 18: Option to standardize response variables

Genotype selection:

Genotype (L): Choose variable ▼

Parameters

nIter: 5,000

burnIn: 1,000

thining: 10

☐ Cross Validation

Figure 19: Parameters

In the field **Plots Format**, you can choose between three types of format for plot. By default “PDF (*.pdf)” is selected, but you can change by “PNG (*.png)” or “Metafile (*.wmf)” which are formats that can be modified and incorporated to different editors when you write reports.

In the field **Output folder**, you must type the name of the output folder where results will be saved; it will be created inside *Output_BGLR-R* folder which is inside the folder where you choose the file to analyze. You can change the name to separate outputs of different data sets. If you don’t change the name and you run the same analyses again, the previous analyses will be replaced by the new ones.

Genotype selection:

Genotype (L): Choose variable ▼

Parameters

nIter: 5,000

burnIn: 1,000

thinning: 10

☒ Cross Validation CV Module

Figure 21: Selection of Cross Validation module

Analysis type:

Analysis: Prediction ▼

☒ Markers ☐ Relationship matrix

☐ Random effects ☐ Fixed effects

Model for Markers: Ridge ▼

Load Markers See Markers

Plots Format: PDF (*.pdf) ▼

Output folder: Analysis1

Figure 20: Options for plots and output folder

Cross Validation module

It shows

When CV is selected, the button “CV Module” appears, and it allows to set the type of CV. next figures show the options

Cross validation using folds

Cross validation using Training-testing sets

Cross validation using defined sets

After setting all options for the analysis that you want to perform, the analysis starts by clicking the “*Analyze*” button.

If there are any specification error or warning because something is missing or it was incorrectly specified, you will get a message like in Figure 11:

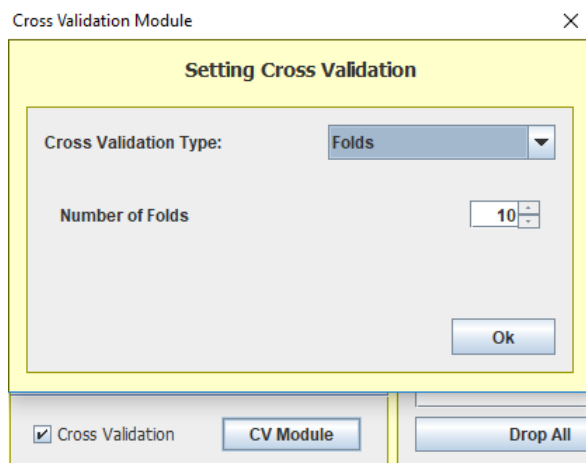


Figure 22: CV folds

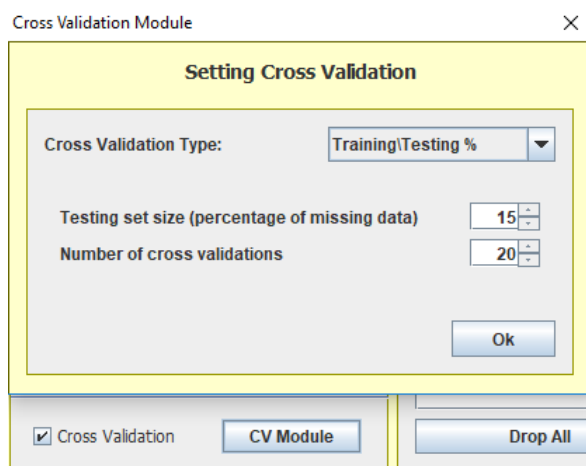


Figure 23: CV training-testing

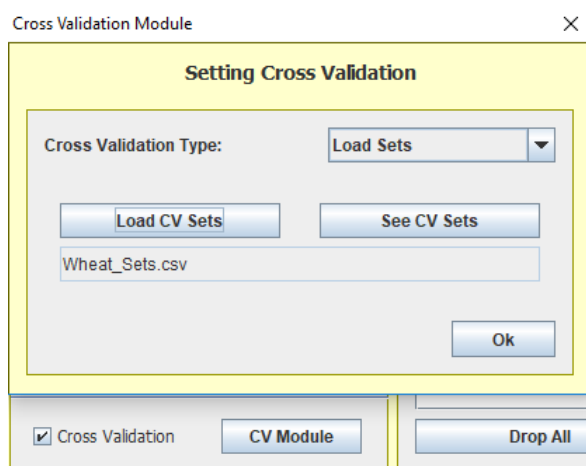


Figure 24: CV defined sets

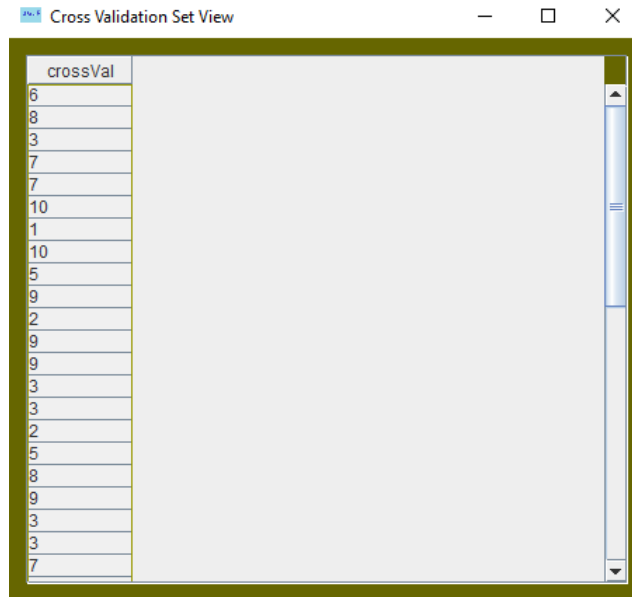


Figure 25: View of sets

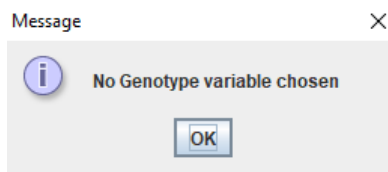


Figure 26: Warning message example

When everything was correctly set, the analysis will start. You can realize BGLR-R is working when the waiting bar at the bottom of screen appears.



Figure 27: Analyzing bar

When the analysis is done, you will receive a message similar to the message in Figure 17.

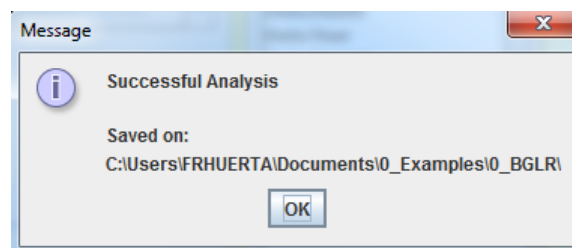


Figure 28: Successful message

Then you will receive a box similar to Figure 18, where you can see all output files that were created and stored in your *output* folder. To open, double-click on them.

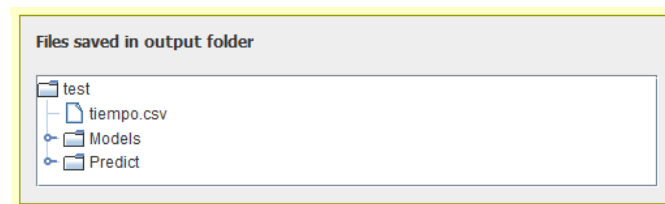


Figure 29: Output Folder

Remember that all outputs are in the same folder where your phenotypic data set was load.

6 Some details about the models

The BGLR function needs the definition of ETA parameter in which all effects are defined.

Fixed effects are defined in ETA using the parameter “model” with the value “FIXED”. Random effects are defined in ETA using the parameter “model” with the value “BRR”. The relationship matrix is defined in ETA using the parameter “model” with the value “BRR”.

7 Problems with the data set

Cases will be excluded from the analysis if BGLR-R detects that:

- Missing values only.
- No variation (they are constant).

If these problems occur, BGLR-R will let you know through the message screen (See some examples in the figure above).

8 Examples of BGLR-R analyses.

Example 1. Lattice data set, drought tolerance data

References

- [1] Paulino Pérez, Gustavo de los Campos. BGLR: A Statistical Package for Whole Genome Regression and Prediction
- [2] Diego Jarquín, José Crossa, et al (2013) A reaction norm model for genomic selection using high dimensional genomic and environmental data. *Theor Appl Genet* (2014) 127:595–607