User Manual for mrMLM

1. Introduction

mrMLM, a R package, aims to provide a user-friendly interface to conduct genome-wide association study (GWAS) via a multi-locus random-SNP-effect mixed linear model (mrMLM) methodology and to visualize its results. It works on the platforms of Windows, Linux and MacOS. The GUI is based on available add-on package RGtk2, via the aid of another package gWidgetsRGtk2. The visualization of results is based on package qqman, such as Manhattan and QQ plots.

2. Installation

2.1 Install GTK+

You may need to install GTK+ before installing RGtk2, because RGtk2 depends on GTK+.

For **Windows** user, you do as below:
Download GTK+ here
(http://sourceforge.net/projects/gladewin32/files/gtk%2B-win32-runtime/2.10.11/gtk-2.10.11-win32-1.exe).
Run the resulting file (gtk-2.10.11-win32-1.exe), which is an automated installer that will help you complete the installation of Gtk2 libraries.

For **Mac OS** users, you do as below:
Download GTK+ here (http://sourceforge.net/projects/gtk-osx/files/latest/download).
Extract and run the resulting file (gtk-osx-docbook-1.2.tar.gz).

For **Linux** users, you do as below:
You may or may not upgrade the GTK libraries depending on your distribution.
There are more details on RGtk2 at RGtk2's home page (http://www.ggobi.org/rgtk2/).

2.2 Install R

Download R from CRAN (https://cran.r-project.org/) and install it by running the downloaded file.

2.3 Install the R packages

The following R packages are needed: RGtk2, cairoDevice, gWidgets, gWidgetsRGtk2, RGtk2Extras and qqman, which can be downloaded from CRAN (https://cran.r-project.org/). Install them in order, as some depend on others. Within R environment, these packages can be installed directly using the below command:

```
install.packages(pkgs=c("RGtk2","cairoDevice","gWidgets","gWidgetsRGtk2","RGtk2Extras","qqman"))
```

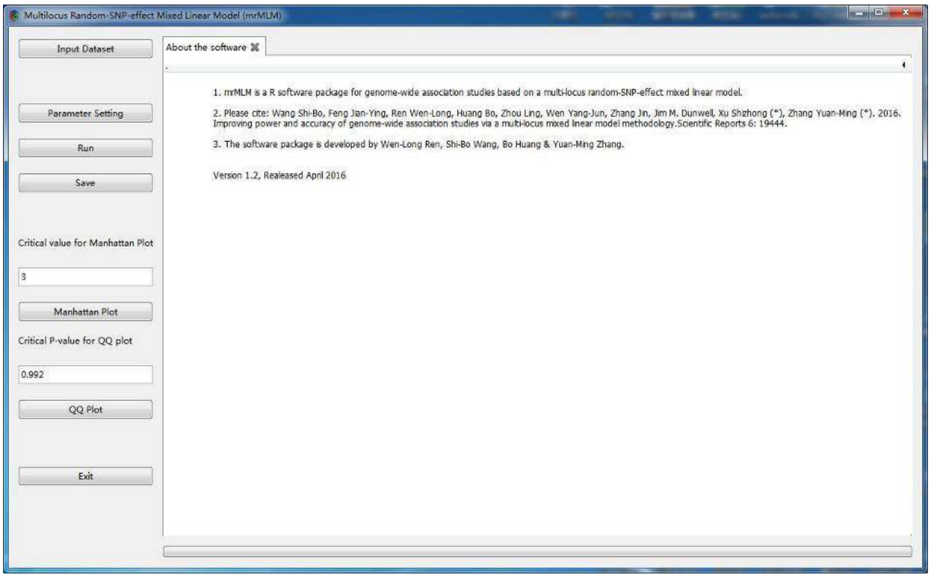
2.4 Install mrMLM

The mrMLM package is freely available at the CRAN (https://cran.r-project.org/web/packages/mrMLM/index.html or soyzhang@mail.hzau.edu.cn or soyzhang@hotmail.com), you can download or request this R software. Within R environment, the mrMLM software can be installed directly using the below command:

```
install.packages(pkgs="mrMLM")
```

3. Running

The **RUN** steps are described as below. Within R environment, launch the mrMLM by command: library(mrMLM), then the following dialog will appear.



1. mrMLM is a R software package for genome-wide association studies based on a multi-locus random-SNP-effect mixed linear model.

2. Please cite: Wang Shi-Bo, Feng Jian-Ying, Ren Wen-Long, Huang Bo, Zhou Ling, Wen Yang-Jun, Zhang Jin, Jim M. Dunwell, Xu Shizhong (*), Zhang Yuan-Ming (*). 2016. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology.Scientific Reports 6: 19444.

3. The software package is developed by Wen-Long Ren, Shi-Bo Wang, Bo Huang & Yuan-Ming Zhang.

Version 1.2, Realeased April 2016

Critical value for Manhattan Plot: 3

Critical P-value for QQ plot: 0.992

Figure 1. The GUI for the mrMLM

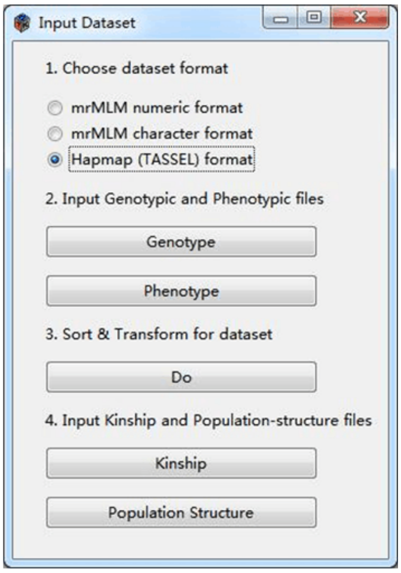
To restart the GUI, the command mrMLM() can be issued.

3.1 Input Dataset

Use the **Input Dataset** button to input dataset files, and then a dialog box will be appeared. In the dialog box, there are four steps. First, users select the dataset formats, which include **mrMLM numeric** format, **mrMLM character** format and **hapmap** format used in the TASSEL software. Then, use the **Genotype** and **Phenotype** buttons to input the genotypic and phenotypic datasets, respectively. Once one file is successfully uploaded, one tabbed page is added to the software notebook. Third, two things will be implemented in this step. One is to sort the individuals between the genotypic and phenotypic files and all the common individuals between the two files are selected to be analyzed in the further analyses. Another is to transfer the character genotypes into the numeric genotypes if the genotypes are character. Once users press the **DO** button, the two things will be conducted. Once the two files will be successfully uploaded, two tabbed pages (Genotype and Phenotype) will be added to the software notebook. Finally, use the **Kinship** and

Population Structure buttons to input the kinship and population structure matrices, respectively. If one file is successfully uploaded, the corresponding data page will be added to the notebook. Note that the **Kinship** and **Population Structure** buttons have two options. For the **kinship** button, one is to directly upload the kinship matrix and another is to calculate the kinship matrix in this software. For the **Population Structure** button, the population structure matrix may be not included in the mixed linear model of the GWAS if it has no effect on GWAS. If not, it should be included in the mixed model.

Note: About the **input file formats** in details, please see **Direction 1** in the end of the manual.



Input Dataset

1. Choose dataset format

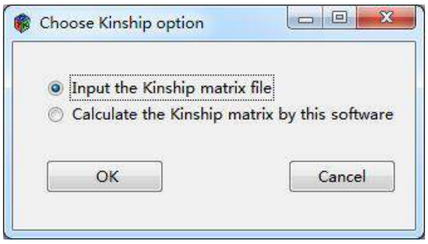
○ mrMLM numeric format
○ mrMLM character format
● Hapmap (TASSEL) format

2. Input Genotypic and Phenotypic files

3. Sort & Transform for dataset

4. Input Kinship and Population-structure files

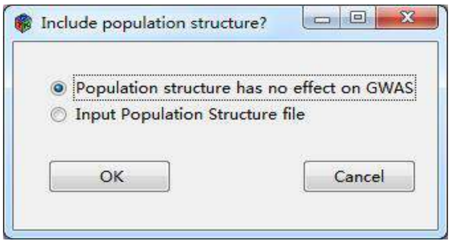
Figure 2. The **Input Dataset** dialog



Choose Kinship option

● Input the Kinship matrix file
○ Calculate the Kinship matrix by this software

Figure 3. The **Kinship** dialog



Include population structure?

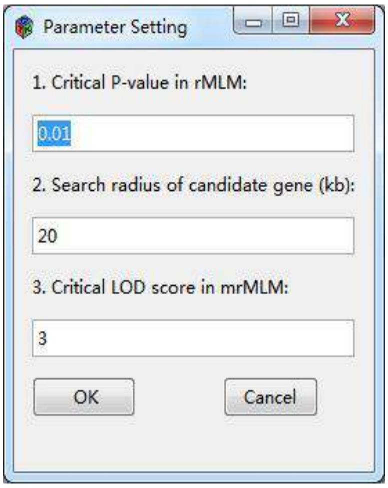
● Population structure has no effect on GWAS
○ Input Population Structure file

Figure 4. The **population structure** dialog

3.2 Run Program

Use the **Parameter Setting** button to set parameters before run the program. In the rMLM, "Critical P-value in Manhattan plot" is set at $0.05/m_e$, where m_e is the effective number of markers (please see Wang et al. Scientific Reports 2016, 6: 19444). All the SNPs significantly associated with the trait are marked by the value of $0.05/m_e$ in the last column of the **Result1** file. If users want to change the **Critical P-value** in Manhattan plot, a new $-\log_{10}$(P-value) may be input in the box above the **Manhattan Plot** button.

"Search radius of candidate gene (kb)" means to keep the one marker with the least P-value, and to delete all the other markers within the radius of the associated marker with the least P-value. Use the **Run** button to execute the software. If the program runs, a progress bar with the "**Please be patient...**" words will appear in the bottom of the interface. If the program finished, a bar with the "**All done.**" will appear.

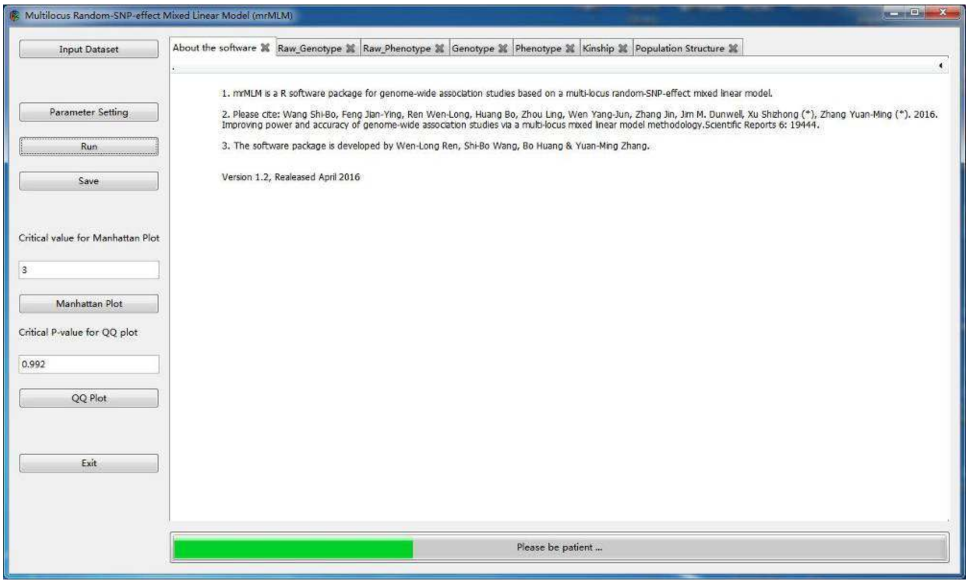


1. Critical P-value in rMLM: 0.01

2. Search radius of candidate gene (kb): 20

3. Critical LOD score in mrMLM: 3

Figure 5. The **Parameter Setting** dialog



Critical value for Manhattan Plot: 3

Critical P-value for QQ plot: 0.992

1. mrMLM is a R software package for genome-wide association studies based on a multi-locus random-SNP-effect mixed linear model.

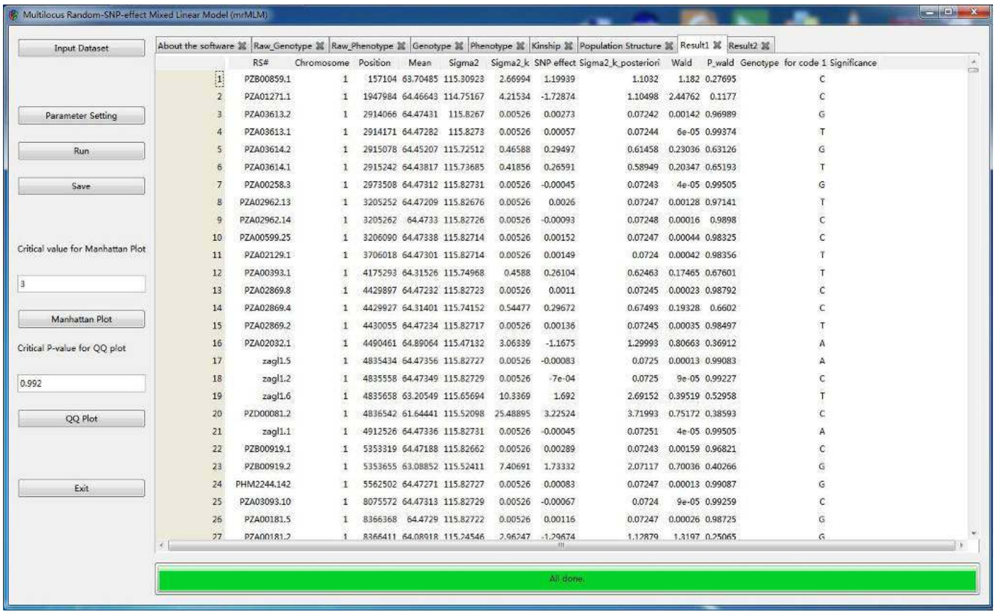
2. Please cite: Wang Shi-Bo, Feng Jian-Ying, Ren Wen-Long, Huang Bo, Zhou Ling, Wen Yang-Jun, Zhang Jin, Jim M. Dunwell, Xu Shizhong (*), Zhang Yuan-Ming (*). 2016. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology.Scientific Reports 6: 19444.

3. The software package is developed by Wen-Long Ren, Shi-Bo Wang, Bo Huang & Yuan-Ming Zhang.

Version 1.2, Realeased April 2016

Please be patient ...

Figure 6. A **running** program interface



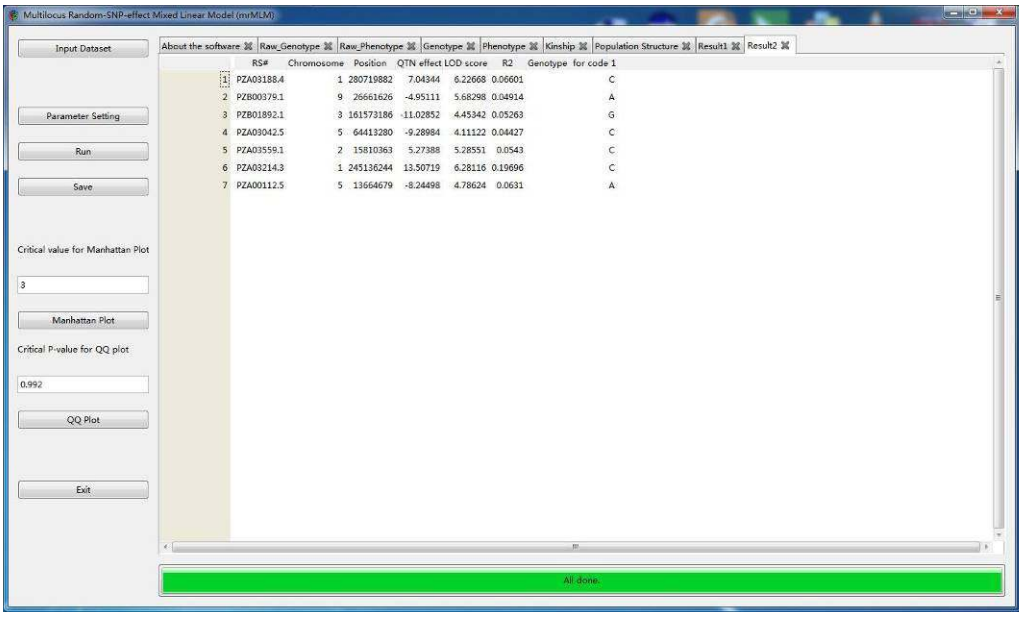
Critical value for Manhattan Plot: 3

Critical P-value for QQ plot: 0.992

	RS#	Chromosome	Position	Mean	Sigma2	Sigma2_k	SNP effect	Sigma2_k_posteriori	Wald	P_wald	Genotype for code 1	Significance
1	PZB00859.1	1	157104	63.70485	115.30923	2.66994	1.19939	1.1032	1.182	0.27695	C	
2	PZA01271.1	1	1947984	64.46643	114.75167	4.21534	-1.72874	1.10498	2.44762	0.1177	C	
3	PZA03613.2	1	2914066	64.47431	115.8267	0.00526	0.00273	0.07242	0.00142	0.96989	G	
4	PZA03613.1	1	2914171	64.47282	115.8273	0.00526	0.00057	0.07244	6e-05	0.99374	T	
5	PZA03614.2	1	2915078	64.45207	115.72512	0.46588	0.29497	0.61458	0.23036	0.63126	G	
6	PZA03614.1	1	2915242	64.43817	115.73685	0.41856	0.26591	0.58949	0.20347	0.65193	T	
7	PZA00258.3	1	2973508	64.47312	115.82731	0.00526	-0.00045	0.07243	4e-05	0.99505	G	
8	PZA02962.13	1	3205252	64.47209	115.82676	0.00526	0.0026	0.07247	0.00128	0.97141	T	
9	PZA02962.14	1	3205262	64.4733	115.82726	0.00526	-0.00093	0.07248	0.00016	0.9898	C	
10	PZA00599.25	1	3206090	64.47338	115.82714	0.00526	0.00152	0.07247	0.00044	0.98325	C	
11	PZA02129.1	1	3706018	64.47301	115.82714	0.00526	0.00149	0.0724	0.00042	0.98356	T	
12	PZA00393.1	1	4175293	64.31526	115.74968	0.4588	0.26104	0.62463	0.17465	0.67601	T	
13	PZA02869.8	1	4429897	64.47232	115.82723	0.00526	0.0011	0.07245	0.00023	0.98792	C	
14	PZA02869.4	1	4429927	64.31401	115.74152	0.54477	0.29672	0.67493	0.19328	0.6602	C	
15	PZA02869.2	1	4430055	64.47234	115.82717	0.00526	0.00136	0.07245	0.00035	0.98497	T	
16	PZA02032.1	1	4490461	64.89064	115.47132	3.06339	-1.1675	1.29993	0.80663	0.36912	A	
17	zagl1.5	1	4835434	64.47356	115.82727	0.00526	-0.00083	0.0725	0.00013	0.99083	A	
18	zagl1.2	1	4835558	64.47349	115.82729	0.00526	-7e-04	0.0725	9e-05	0.99227	C	
19	zagl1.6	1	4835658	63.20549	115.65694	10.3369	1.692	2.69152	0.39519	0.52958	T	
20	PZD00081.2	1	4836542	61.64441	115.52098	25.48895	3.22524	3.71993	0.75172	0.38593	C	
21	zagl1.1	1	4912526	64.47336	115.82731	0.00526	-0.00045	0.07251	4e-05	0.99505	A	
22	PZB00919.1	1	5353319	64.47188	115.82662	0.00526	0.00289	0.07243	0.00159	0.96821	C	
23	PZB00919.2	1	5353655	63.08852	115.52411	7.40691	1.73332	2.07117	0.70036	0.40266	G	
24	PHM2244.142	1	5562502	64.47271	115.82727	0.00526	0.00083	0.07247	0.00013	0.99087	G	
25	PZA03093.10	1	8075572	64.47313	115.82729	0.00526	-0.00067	0.0724	9e-05	0.99259	C	
26	PZA00181.5	1	8366368	64.4729	115.82722	0.00526	0.00116	0.07247	0.00026	0.98725	G	
27	PZA00181.2	1	8366411	64.08918	115.24546	2.96247	-1.29674	1.12879	1.3197	0.25065	G	

All done.

Figure 7. A finished program interface (the **rMLM Results: Result1**)



Critical value for Manhattan Plot: 3

Critical P-value for QQ plot: 0.992

	RS#	Chromosome	Position	QTN effect	LOD score	R2	Genotype for code 1
1	PZA03188.4	1	280719882	7.04344	6.22668	0.06601	C
2	PZB00379.1	9	26661626	-4.95111	5.68298	0.04914	A
3	PZB01892.1	3	161573186	-11.02852	4.45342	0.05263	G
4	PZA03042.5	5	64413280	-9.28984	4.11122	0.04427	C
5	PZA03559.1	2	15810363	5.27388	5.28551	0.0543	C
6	PZA03214.3	1	245136244	13.50719	6.28116	0.19696	C
7	PZA00112.5	5	13664679	-8.24498	4.78624	0.0631	A

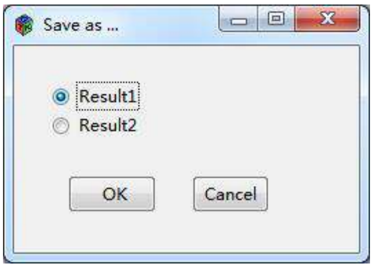
All done.

Figure 8. A finished program interface (the **mrMLM Results: Result2**)

3.3 Save results

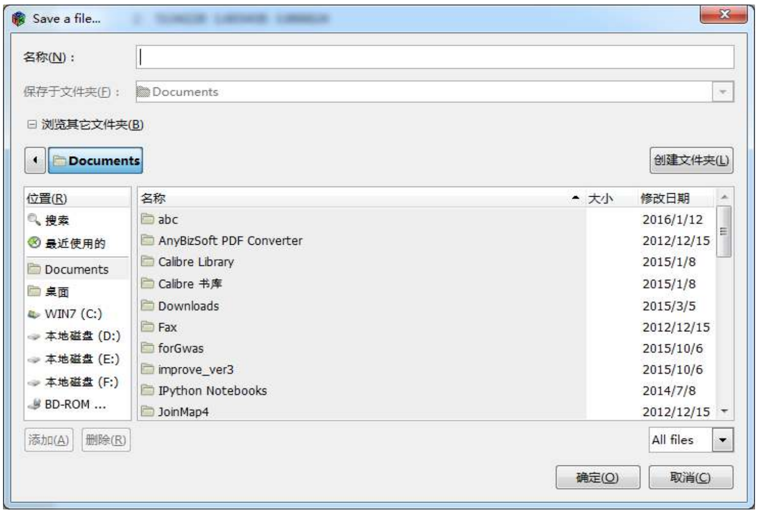
Use **Save** button to save the results as ***.csv** files. The **Results in the rMLM** are saved as **Result1.csv** and the **Results in the mrMLM** are saved as **Result2.csv**. If click **OK** button (after the **Save** button), a dialog is used to choose the pathway and the file name for the saving files.

Note: About the **explanation of Result1 and Result2** in details, please see **Direction 2** in the end of the manual.



● Result1
○ Result2

Figure 9. The result **Save** dialog for the rMLM and mrMLM methods



名称(N)：

保存于文件夹(F)： Documents

名称	大小	修改日期
abc		2016/1/12
AnyBizSoft PDF Converter		2012/12/15
Calibre Library		2015/1/8
Calibre 书库		2015/1/8
Downloads		2015/3/5
Fax		2012/12/15
forGwas		2015/10/6
improve_ver3		2015/10/6
IPython Notebooks		2014/7/8
JoinMap4		2012/12/15

All files

Figure 10. The **Save** dialog

Warning: It is better not to include other languages except English in the pathway and file name. Otherwise, there may be something wrong.

4. Visualization of Results

If the program have finished, you may have the result visualization. Before use the **Manhattan Plot** button, please set the critical value for –$\log_{10}$(P), which is defaulted the value of –log(0.05/m_e), where m_e is the effective number of markers. Of course, users may change this value. Before press the **QQ Plot** button, please set the critical P-value for QQ plot, which is defaulted the value of 0.992. This is because the P-values are a mixture of a χ^2 distribution with one degree of freedom and a point mass at one. Note that users may also change this 0.992 based on yourself results.

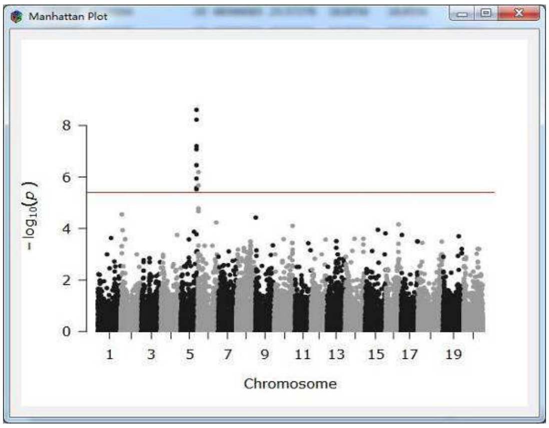


Figure 11. The **Manhattan Plot**

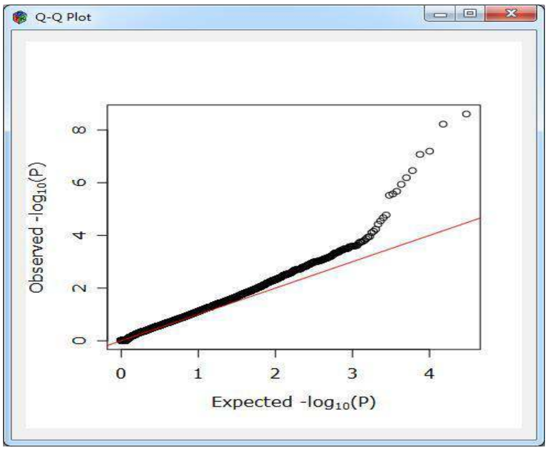


Figure 12. The **QQ Plot**

Directions

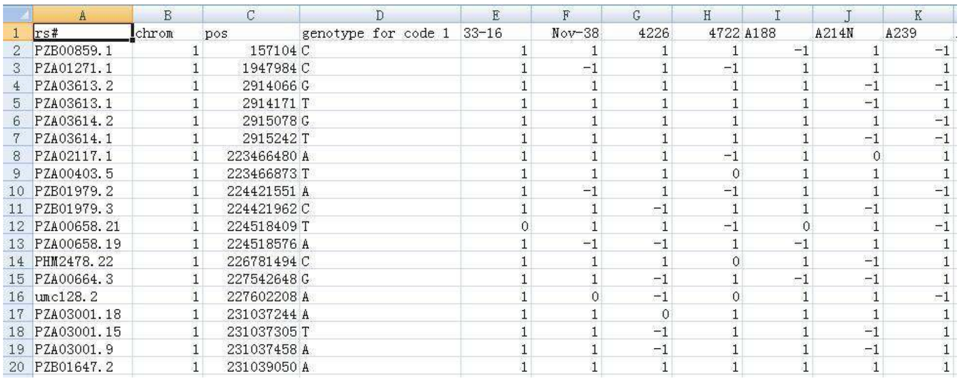
Direction 1: Explanation of input files in details

D1.1 The Genotypic file

The **Genotypic** file should be a ***.csv** format file.

D1.1.1 mrMLM numeric format

The first column, named "**rs#**" in the first row, stands for marker ID. The second column, named "**chrom**" in the first row, stands for chromosome. The third column, named "**pos**" in the first row, stands for the position (bp) of SNP in the chromosome. The fourth column, named "**genotype for code 1**" in the first row, stands for reference bases. And each of the remaining columns stands for one individual. In their first rows, the individual names are appeared. For each marker, homozygous genotypes are expressed by 1 and -1, respectively, and the heterozygous and missing genotypes are indicated by zero. Note that the genotypes with code **1** will be listed in the **Result** files.

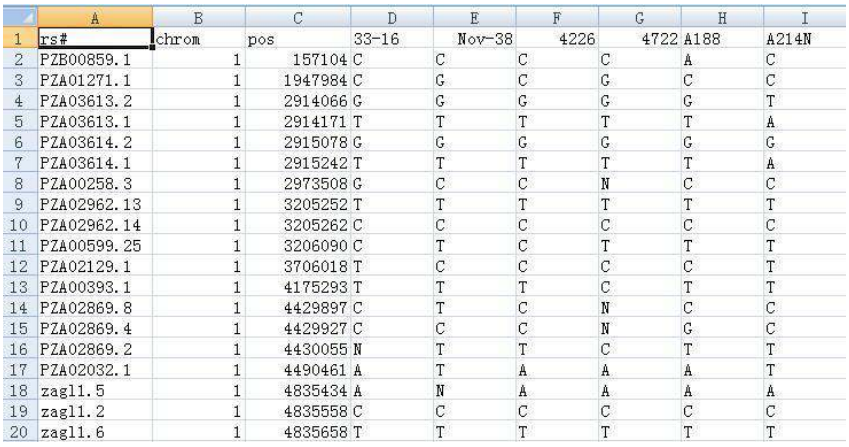


	A	B	C	D	E	F	G	H	I	J	K
1	rs#	chrom	pos	genotype for code 1	33-16	Nov-38	4226	4722	A188	A214N	A239
2	PZB00859.1	1	157104	C	1	1	1	1	-1	1	-1
3	PZA01271.1	1	1947984	C	1	-1	1	-1	1	1	1
4	PZA03613.2	1	2914066	G	1	1	1	1	1	-1	-1
5	PZA03613.1	1	2914171	T	1	1	1	1	1	-1	1
6	PZA03614.2	1	2915078	G	1	1	1	1	1	1	-1
7	PZA03614.1	1	2915242	T	1	1	1	1	1	-1	-1
8	PZA02117.1	1	223466480	A	1	1	1	-1	1	0	1
9	PZA00403.5	1	223466873	T	1	1	1	0	1	1	1
10	PZB01979.2	1	224421551	A	1	-1	1	-1	1	1	-1
11	PZB01979.3	1	224421962	C	1	1	-1	1	1	-1	1
12	PZA00658.21	1	224518409	T	0	1	1	-1	0	1	-1
13	PZA00658.19	1	224518576	A	1	-1	-1	1	-1	1	1
14	PHM2478.22	1	226781494	C	1	1	1	0	1	-1	1
15	PZA00664.3	1	227542648	G	1	1	-1	1	-1	-1	1
16	umc128.2	1	227602208	A	1	0	-1	0	1	1	-1
17	PZA03001.18	1	231037244	A	1	1	0	1	1	1	1
18	PZA03001.15	1	231037305	T	1	1	-1	1	1	-1	1
19	PZA03001.9	1	231037458	A	1	1	-1	1	1	-1	1
20	PZB01647.2	1	231039050	A	1	1	1	1	1	1	1

Figure D1.1.1 The **genotypic** file with **mrMLM numeric** format

D1.1.2 mrMLM character format

The first three columns are same as those in the "**D1.1.1 mrMLM numeric format**". The marker values are character, such as **A, T, C, G** and **N**, and the other notations are heterozygous genotypes. The "**N**" indicates missing. The first rows from the fourth to last columns are individual code.

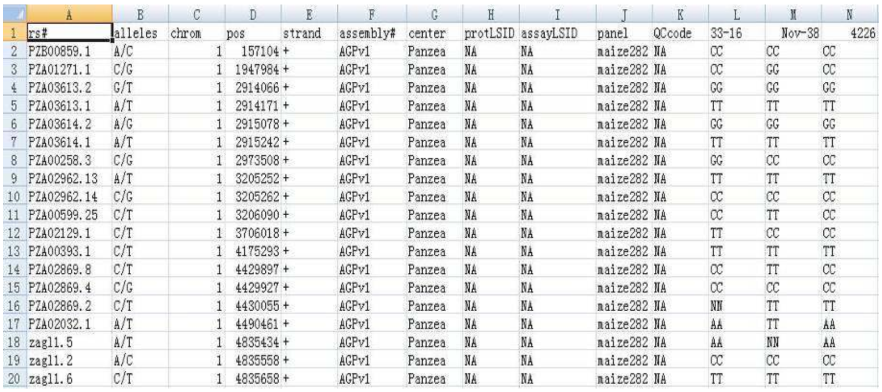


	A	B	C	D	E	F	G	H	I
1	rs#	chrom	pos	33-16	Nov-38	4226	4722	A188	A214N
2	PZB00859.1	1	157104	C	C	C	C	A	C
3	PZA01271.1	1	1947984	C	G	C	G	C	C
4	PZA03613.2	1	2914066	G	G	G	G	G	T
5	PZA03613.1	1	2914171	T	T	T	T	T	A
6	PZA03614.2	1	2915078	G	G	G	G	G	G
7	PZA03614.1	1	2915242	T	T	T	T	T	A
8	PZA00258.3	1	2973508	G	C	C	N	C	C
9	PZA02962.13	1	3205252	T	T	T	T	T	T
10	PZA02962.14	1	3205262	C	C	C	C	C	C
11	PZA00599.25	1	3206090	C	T	C	T	T	T
12	PZA02129.1	1	3706018	T	C	C	C	C	T
13	PZA00393.1	1	4175293	T	T	T	C	T	T
14	PZA02869.8	1	4429897	C	T	C	N	C	C
15	PZA02869.4	1	4429927	C	C	C	N	G	C
16	PZA02869.2	1	4430055	N	T	T	C	T	T
17	PZA02032.1	1	4490461	A	T	A	A	A	T
18	zagl1.5	1	4835434	A	N	A	A	A	A
19	zagl1.2	1	4835558	C	C	C	C	C	C
20	zagl1.6	1	4835658	T	T	T	T	T	T

Figure D1.1.2 The **genotypic** file with **mrMLM character** format

D1.1.3 Hapmap (TASSEL) format

Please see the TASSEL software in details. Here we introduce simply. The first eleven columns describe the specific information of markers and individuals, and their column names must be **"rs#"**, **"alleles"**, **"chrom"**, **"pos"**, **"strand"**, **"assembly#"**, **"center"**, **"protLSID"**, **"assayLSID"**, **"panel"** and **"QCcode"**. In the **"rs#"** (1)**, "chrom"** (3) **and "pos"** (4) columns, the information for each marker (row) must be listed, and their meanings are same as those in the above. The values for marker genotypes should be character, such as **AA, TT, CC, GG, NN, AC** and **AG**, where the "**NN**" indicates missing or unknown genotypes. In the 2 and 5 to 11 columns, the **no available** information must be marked by **"NA"**. All the individuals will be showed from the 12 to last columns, the first element in each column is individual ID (name) and the others are the genotypes (character).

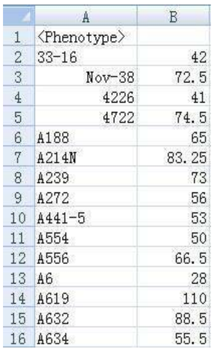


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	rs#	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panel	QCcode	33-16	Nov-38	4226
2	PZB00859.1	A/C	1	157104	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC
3	PZA01271.1	C/G	1	1947984	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	GG	CC
4	PZA03613.2	G/T	1	2914066	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	GG	GG
5	PZA03613.1	A/T	1	2914171	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT
6	PZA03614.2	A/G	1	2915078	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	GG	GG
7	PZA03614.1	A/T	1	2915242	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT
8	PZA00258.3	C/G	1	2973508	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	CC	CC
9	PZA02962.13	A/T	1	3205252	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT
10	PZA02962.14	C/G	1	3205262	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC
11	PZA00599.25	C/T	1	3206090	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	TT	CC
12	PZA02129.1	C/T	1	3706018	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	CC	CC
13	PZA00393.1	C/T	1	4175293	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT
14	PZA02869.8	C/T	1	4429897	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	TT	CC
15	PZA02869.4	C/G	1	4429927	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC
16	PZA02869.2	C/T	1	4430055	+	AGPv1	Panzea	NA	NA	maize282	NA	NN	TT	TT
17	PZA02032.1	A/T	1	4490461	+	AGPv1	Panzea	NA	NA	maize282	NA	AA	TT	AA
18	zagl1.5	A/T	1	4835434	+	AGPv1	Panzea	NA	NA	maize282	NA	AA	NN	AA
19	zagl1.2	A/C	1	4835558	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	CC	CC
20	zagl1.6	C/T	1	4835658	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	TT	TT

Figure D1.1.3 The **genotypic** file with **Hapmap (TASSEL)** format

D1.2 The Phenotypic file

The **Phenotypic** file should be a ***.csv** format file. The first column stands for individual ID, such as 33-16, Nov-38 and 4226. The second column is phenotypic values for the trait. Note that the phenotypic file includes only one trait. The first element in the first column must be **"<Phenotype>" or "<Trait>"**.

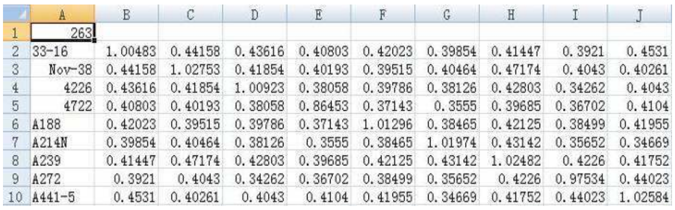


	A	B
1	<Phenotype>	
2	33-16	42
3	Nov-38	72.5
4	4226	41
5	4722	74.5
6	A188	65
7	A214N	83.25
8	A239	73
9	A272	56
10	A441-5	53
11	A554	50
12	A556	66.5
13	A6	28
14	A619	110
15	A632	88.5
16	A634	55.5

Figure D1.2. The **Phenotypic** file

D1.3 The Kinship file

The **Kinship** file should be a ***.csv** file. The number of rows (or columns) equals to the number of the common individuals between the phenotypic and genotypic datasets. If the Kinship matrix is calculated by this software, we calculate only the Kinship matrix between the common individuals. If the Kinship matrix has been obtained and uploaded from a known file, it is possible that the number and order of individuals in the known file are not consistent with those of the common (valid) individuals in the further analysis. At this situation, the software will change the known K matrix in order that the number and order of new K matrix match the number and order of common (valid) individuals in the Phenotypic and Genotypic files. In the known K matrix, the **first element** in the first column must be **the number of valid individuals**, such as 263; and the other elements in the first column are individual ID (names).



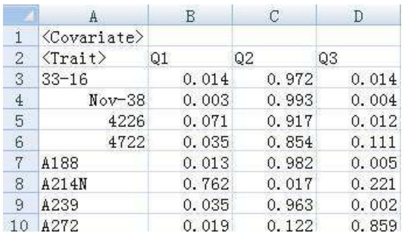
	A	B	C	D	E	F	G	H	I	J
1	263									
2	33-16	1.00483	0.44158	0.43616	0.40803	0.42023	0.39854	0.41447	0.3921	0.4531
3	Nov-38	0.44158	1.02753	0.41854	0.40193	0.39515	0.40464	0.47174	0.4043	0.40261
4	4226	0.43616	0.41854	1.00923	0.38058	0.39786	0.38126	0.42803	0.34262	0.4043
5	4722	0.40803	0.40193	0.38058	0.86453	0.37143	0.3555	0.39685	0.36702	0.4104
6	A188	0.42023	0.39515	0.39786	0.37143	1.01296	0.38465	0.42125	0.38499	0.41955
7	A214N	0.39854	0.40464	0.38126	0.3555	0.38465	1.01974	0.43142	0.35652	0.34669
8	A239	0.41447	0.47174	0.42803	0.39685	0.42125	0.43142	1.02482	0.4226	0.41752
9	A272	0.3921	0.4043	0.34262	0.36702	0.38499	0.35652	0.4226	0.97534	0.44023
10	A441-5	0.4531	0.40261	0.4043	0.4104	0.41955	0.34669	0.41752	0.44023	1.02584

Figure D1.3. The **Kinship** file

D1.4 The Population Structure file

The **Population Structure** file is a ***.csv** file. Using the **Structure** software, the population structure matrix may be calculated. The first column stands for the valid individual ID (names). The first and second elements in the first column must be **"<Covariate>"** and **"<Trait>"** respectively. If population structure matrix has **k** columns, please input all the **k** columns. In the second row, it must be **"Q1"**, **"Q2"**, ... , **"Qk"** following the **"<Trait>"**.

If the **Population Structure** file has been obtained and uploaded from a known file, it is possible that the number and order of individuals in the known file are not consistent with those of the common (valid) individuals in the further analysis. At this situation, the software will change the known matrix in order that the number and order of new **Population Structure** matrix match the number and order of common (valid) individuals in the Phenotypic and Genotypic files.



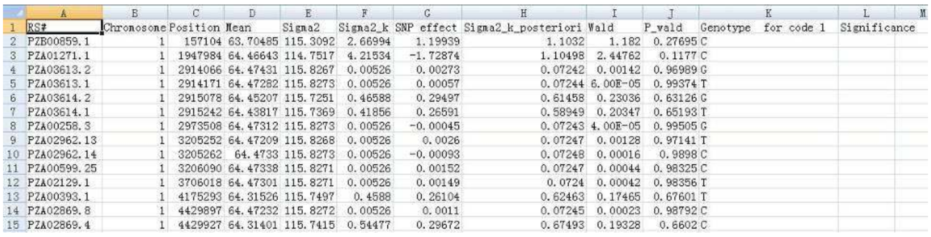
	A	B	C	D
1	<Covariate>			
2	<Trait>	Q1	Q2	Q3
3	33-16	0.014	0.972	0.014
4	Nov-38	0.003	0.993	0.004
5	4226	0.071	0.917	0.012
6	4722	0.035	0.854	0.111
7	A188	0.013	0.982	0.005
8	A214N	0.762	0.017	0.221
9	A239	0.035	0.963	0.002
10	A272	0.019	0.122	0.859

Figure D1.4. The **Population Structure** file

Direction 2: Explanation of Result1 and Result2 in details

D2.1 Explanation of Result1 file

The **Result1** table with twelve columns shows the results from the rMLM (random-SNP-effect mixed linear model) method. The corresponding column names are as follows: reference sequence number (rs#, marker name), chromosome, marker's position (bp) in the chromosome, population mean value (Mean), residual variance (σ^2, Sigma2), priori variance of the kth SNP effect (ϕ_k^2, Sigma2_k), SNP effect (γ_k, Effect), posteriori variance of SNP effect ($\mathrm{var}\left(\gamma_k\right)$, Sigma2_k_posteriori), Wald test statistic value, the P-value of Wald test, significance and genotype for code 1, respectively. In the significance column, only significant markers under the critical value 0.05/m$_e$ are marked.

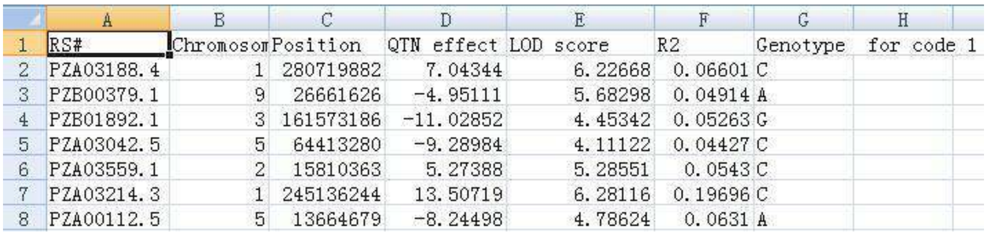


RS#	Chromosome	Position	Mean	Sigma2	Sigma2_k	SNP effect	Sigma2_k_posteriori	Wald	P_wald	Genotype for code 1	Significance
PZB00859.1	1	157104	63.70485	115.3092	2.66994	1.19939	1.1032	1.182	0.27695	C	
PZA01271.1	1	1947984	64.46643	114.7517	4.21534	-1.72874	1.10498	2.44762	0.1177	C	
PZA03613.2	1	2914066	64.47431	115.8267	0.00526	0.00273	0.07242	0.00142	0.96989	G	
PZA03613.1	1	2914171	64.47282	115.8273	0.00526	0.00057	0.07244	6.00E-05	0.99374	T	
PZA03614.2	1	2915078	64.45207	115.7251	0.46588	0.29497	0.61458	0.23036	0.63126	G	
PZA03614.1	1	2915242	64.43817	115.7369	0.41856	0.26591	0.58949	0.20347	0.65193	T	
PZA00258.3	1	2973508	64.47312	115.8273	0.00526	-0.00045	0.07243	4.00E-05	0.99505	G	
PZA02962.13	1	3205252	64.47209	115.8268	0.00526	0.0026	0.07247	0.00128	0.97141	T	
PZA02962.14	1	3205262	64.4733	115.8273	0.00526	-0.00093	0.07248	0.00016	0.9898	C	
PZA00599.25	1	3206090	64.47338	115.8271	0.00526	0.00152	0.07247	0.00044	0.98325	C	
PZA02129.1	1	3706018	64.47301	115.8271	0.00526	0.00149	0.0724	0.00042	0.98356	T	
PZA00393.1	1	4175293	64.31526	115.7497	0.4588	0.26104	0.62463	0.17465	0.67601	T	
PZA02869.8	1	4429897	64.47232	115.8272	0.00526	0.0011	0.07245	0.00023	0.98792	C	
PZA02869.4	1	4429927	64.31401	115.7415	0.54477	0.29672	0.67493	0.19328	0.6602	C	

Figure D2.1. Results in the rMLM (**Result1**)

D2.2 Explanation of Result2 file

The Result2 table with seven columns shows the final results of the mrMLM (multi-locus random-SNP-effect mixed linear model) method. The corresponding column names are as follows: reference sequence number (rs#, marker names), chromosome, marker's position (bp) in the chromosome, QTN effect, LOD score, the proportion of phenotypic variance explained by the putative QTN, and genotype for code 1, respectively.



RS#	Chromosom	Position	QTN effect	LOD score	R2	Genotype for code 1
PZA03188.4	1	280719882	7.04344	6.22668	0.06601	C
PZB00379.1	9	26661626	-4.95111	5.68298	0.04914	A
PZB01892.1	3	161573186	-11.02852	4.45342	0.05263	G
PZA03042.5	5	64413280	-9.28984	4.11122	0.04427	C
PZA03559.1	2	15810363	5.27388	5.28551	0.0543	C
PZA03214.3	1	245136244	13.50719	6.28116	0.19696	C
PZA00112.5	5	13664679	-8.24498	4.78624	0.0631	A

Figure D2.2. Results in the mrMLM (**Result2**)

Reference

Wang Shi-Bo, Feng Jian-Ying, Ren Wen-Long, Huang Bo, Zhou Ling, Wen Yang-Jun, Zhang Jin, Jim M. Dunwell, Xu Shizhong (*), Zhang Yuan-Ming (*). 2016. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Scientific Reports 6: 19444.