User Manual for

III V m r M L M

3 Variance-component multi-locus random-SNP-effect Mixed Linear Model C++ tool for genome-wide association study

(version 1.0)

Yu Xiang, Ming-Hui Liu, and Yuan-Ming Zhang

(soyzhang@mail.hzau.edu.cn)

Last updated on May, 2022

Disclaimer: While extensive testing has been performed by Yuan-Ming Zhang's Lab at College of Plant Science and Technology, Huazhong Agricultural University, the results are, in general, reliable, correct, and appropriate. However, results are not guaranteed for any specific datasets. We strongly recommend that users integrate the IIIVmrMLM results with those from other software packages, i.e., mrMLM, GEMMA, EMMAX, and PLINK.

Contacted information:

Yuan-Ming Zhang (soyzhang@mail.hzau.edu.cn)

College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

Citation:

1 Li Mei, Zhang Ya-Wen, Zhang Ze-Chang, Xiang Yu, Liu Ming-Hui, Zhou Ya-Hui, Zuo Jian-Fang, Zhang Han-Qing, Chen Ying, Zhang Yuan-Ming. A compressed variance component mixed model for detecting QTNs, and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. Molecular Plant 2022, 15(4): 630-650

2 Li Mei, Zhang Ya-Wen, Xiang Yu, Liu Ming-Hui, Yuan-Ming Zhang. IIIVmrMLM: An R tool for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. Submitted

The work was supported by the National Natural Science Foundation of China (32070557 and 31871242), and Huazhong Agricultural University Scientific & Technological Self-Innovation Foundation (2014RC020).

1. Introduction

1.1 Why IIIVmrMLM?

IIIVmrMLM (3 Variance-component multi-locus random-SNP-effect Mixed Linear Model) program is an C++ software package for multi-locus genome-wide association studies (GWAS), which identifies main-effect QTNs, QTN-by-environment interactions (QEIs), and QTN-by-QTN interactions (QQIs) for complex and multi-omics traits.

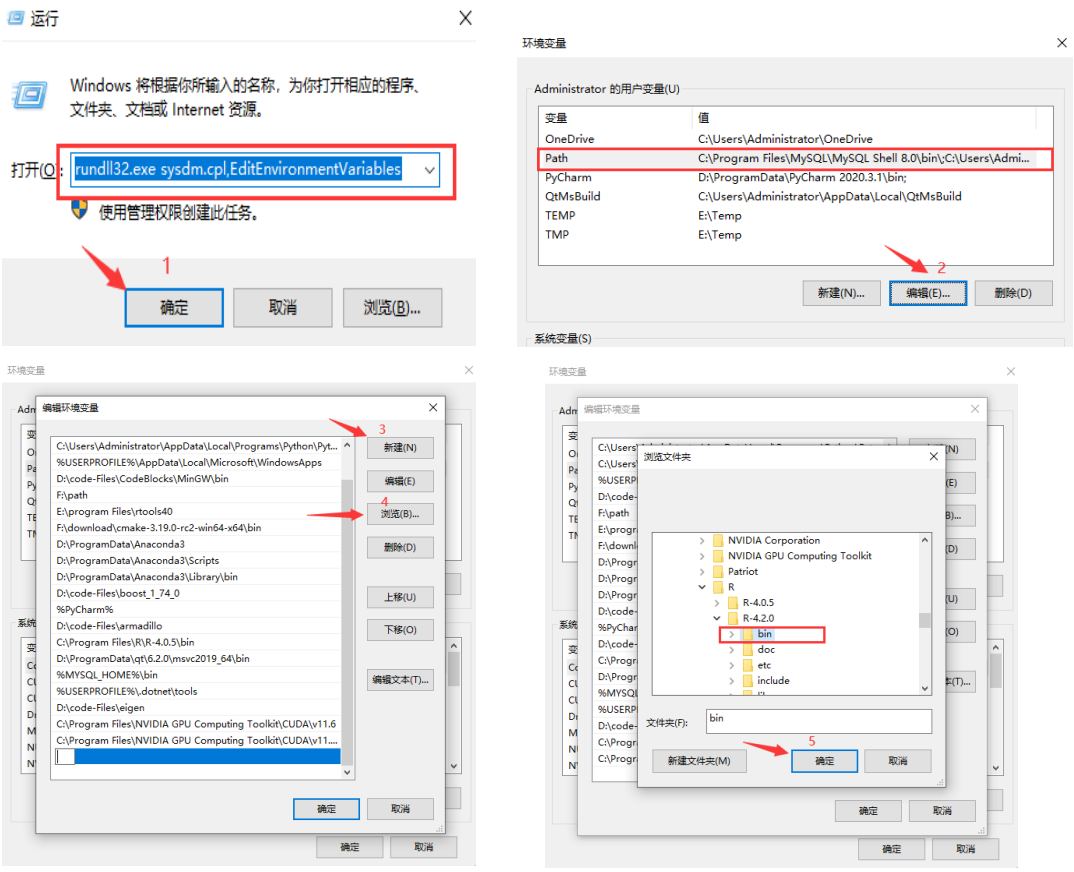
1.2 Getting started

Before users install our IIIVmrMLM software package, users should first install the newest versions of R, Rtools, and Rstudio software packages.

The software package IIIVmrMLM runs only in the Windows operating system and can be freely requested from the maintainer, Dr Yuan-Ming Zhang at College of Plant Science and Technology, Huazhong Agricultural University (soyzhang@mail.hzau.edu.cn).

1.2.1 One-Click installation

1) Open "Run" prompt (Win+R) ->"*rundll32.exe sysdm.cpl,EditEnvironmentVariables*". -> Path -> Edit. Add the bin folder in the R installation directory to Path.



2) Open Rstudio or Rgui and enter *install.packages(c("bigmemory","biglasso","data.table","lars"))*. Ignore this step if these R packages are already installed.

3) Unzip the IIIVmrMLMcpp.zip file with the path you want to save.

1.2.2 Run IIIVmrMLM

In the command line interface version, after setting the parameters in the config.json file, click IIIVmrMLM.exe to run the software. In the GUI version, parameter setting is performed directly through the software interface.

2. Function IIIVmrMLM

Parameter settings

In the command line interface version, open the config.json file in the IIIVmrMLM folder. All parameters are set in this file.

Parameter	Meaning	File format	Note
fileGen	Name & path of genotypic file in your computer, i.e., fileGen="D:/Users/Genotype.csv".	*.csv; *.txt (Genotypic values. **Row**: traits; **Column**: individual)	Table 1
filePhe	Name & path of trait phenotypic file in your computer, i.e., filePhe="D:/Users/Phenotype.csv".	*.csv; *.txt (Phenotypic values. **Row**: individual; **Column**: traits)	Table 2
fileKin	Name & path of individual kinship file in your computer, i.e., fileKin="D:/Users/Kinship.csv" or fileKin=NULL.	*.csv; *.txt (Kinship matrix. **Row** & **Column**: individuals)	Table 3
filePS	Name & path of population structure file in your computer, i.e., filePS="D:/Users/PopStr.csv" or filePS=NULL.	*.csv; *.txt [Population structure. **Row**: individual; **Column**: sub-populations 1, 2, ---, *k* (No. of sub-populations)]	Tables 4~6
GenSeparator	The separator types of genotype file include: tab('\t'), comma(',').		
PopStrType	The types of population structure include *Q* (*Q* matrix), PCA (principal components), and EvolPopStr (evolutionary population structure).		
fileCov	Name & path of covariate file in your computer, i.e., fileCov="D:/Users/Covariate.csv" or fileCov=NULL.	*.csv; *.txt (**Row**: individual; **Column**: covariates 1, 2, ---, *k* (no. of covariates)) **Cate**: categorical variable; **Con**: continuous variable	Table 7
method	Three GWAS methods: single- and multi-environment analysis, and epistasis detection, i.e., method="Single_env", method="Multi_env", and method="Epistasis".		
trait	Traits analyzed from all numbers, i.e., **trait=[1,2,3]** indicates that the first to third traits are analyzed.		
	If method="Multi_env", users need to add a parameter **n_en** to indicate the number of environments for each trait in the **filePhe**, i.e., **trait=[1,2]** (Analyzing the first and second traits)**; n_en=[2,2,3]** (The filePhe file contains the phenotypic values of three traits, and the environmental numbers of each trait are 2, 2, and 3, respectively)		

SearchRadius	In the setup of decollinearity parameter **SearchRadius**, only one parameter should be set in QTN and QEI detection, i.e., **SearchRadius=20**, indicating **the radius of 20 kb** in the determination of only one potentially associated QTN or QEI, while two parameters should be set in QQI detection, i.e., **SearchRadius=[10,20]**, the first parameter "10" (kb) is for main-effect QTNs and the second one "20" (kb) is for QQIs.
svpal	In the setup of critical P-value parameter **svpal**, the critical P-value 0.01 (**svpal=0.01**) is set to select all the potentially associated QTNs and QEIs in genome-wide single-marker scanning in QTN and QEI detection, while two critical P-values 0.01 and 0.01 are set to select main-effect QTNs and QQIs, respectively, in QQI detection, i.e., **svpal=[0.01,0.01]**, the first parameter is for main-effect QTNs and the second one is for QQIs.
sblgwas_t	Only needed **in QQI detection**, a number between [-3, 0] to control sparseness of the "sblgwas" function, default value is -1.
savePath	Save path of the results in your computer, i.e, "D:/Users".

Single_env: The detection of main-effect QTNs for complex traits

Users must set "fileGen", "filePhe", "method", "trait", and "savePath", while the other parameters may be default in *IIIVmrMLM*, including *fileKin=NULL; filePS=NULL; GenSeparator="comma"; PopStrType="Q"; fileCov=NULL; SearchRadius=20; svpal=0.01.*

2.2.1 Data input format

Format for genotypic dataset "fileGen" (Table 1)

The first eleven columns describe specific information of markers and individuals, and column names are "rs#", "alleles", "chrom", "pos", "strand", "assembly#", "center", "protLSID", "assayLSID", "panelLSID", and "QCcode". "rs#" (1st column): marker name, such as "PZB00859.1"; "alleles" (2nd): marker alleles; "chrom" (3rd): chromosome; "pos" (4th): marker position on genome (bp). The values of marker genotypes should be character, such as AA, TT, CC, GG, NN, AC and AG, where "NN" indicates the missing or unknown genotypes. In the 2nd and 5th to eleventh columns, "NA" indicates no information available. All the individual genotypic information will be showed from the 12th to last columns. In each column, individual name is listed in the first row, i.e., "33-16" is accession name, while its genotypes are listed in the other rows, such as "CC".

Table 1. Hapmap format of genotypic dataset

rs#	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panelLSID	QCcode	33-16	…
PZB00859.1	A/C	1	157104	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	…
PZA01271.1	C/G	1	1947984	+	AGPv1	Panzea	NA	NA	maize282	NA	CC	…
PZA03613.2	G/T	1	2914066	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	…
PZA03613.1	A/T	1	2914171	+	AGPv1	Panzea	NA	NA	maize282	NA	TT	…
PZA03614.2	A/G	1	2915078	+	AGPv1	Panzea	NA	NA	maize282	NA	GG	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…

Format for phenotypic dataset "filePhe" (Table 2)

The file type of phenotypes for complex trait is *.csv or *.txt, as shown below. The first row in the first column: "<Phenotype>"; the second to *n*th rows in the first column: individual IDs or names, such as B46. The first row in other columns: trait names, such as "trait1", and the second to *n*th rows in other columns: phenotypic values for complex traits. The missed phenotypes: "NA".

Table 2. The format of phenotypic dataset

<Phenotype>	trait1	trait2	trait3	…
B46	42	43.02	44.32	…
B52	72.5	71.88	72.8	…
B57	41	41.7	41.42	…
B64	74.5	74.43	NA	…
B68	65	66.4	65.33	…
⋮	⋮	⋮	⋮	…

The format for kinship dataset "fileKin" (Table 3)

The "fileKin" should be a file with *.csv or *.txt format. All the kinship coefficients are listed as an $n \times n$ matrix.

Table 3. The format of the Kinship dataset

1	0.700361011	0.599277978	0.675090253	0.620938628	…
0.700361011	1	0.620938628	0.666064982	0.653429603	…
0.599277978	0.620938628	1	0.561371841	0.5433213	…
0.675090253	0.666064982	0.561371841	1	0.615523466	…
0.620938628	0.653429603	0.5433213	0.615523466	1	…
⋮	⋮	⋮	⋮	⋮	…

fileKin=NULL indicates that the Kinship matrix is calculated by the "IIIVmrMLM" software. Here only the marker information of all the *n* individuals is used to calculate Kinship matrix.

***Q* matrix format for dataset "filePS" (Table 4)**

The Q matrix dataset in Table 4 consists of a $(n+2) \times (k+1)$ matrix, where *n* is sample size (the number of the common individuals across the marker genotype, phenotype, K matrix, and co-variable datasets), and *k* is the number of sub-populations. In the first column, "<PopStr>" and "<ID>" must present in the first and second rows, respectively; "33-16", "Nov-38" and "A4226" are individual IDs or names. In the 2nd to $(k+1)$-th columns, "Q1" to "Q*k*" indicate sub-populations. In the third row, "0.014", "0.972" and "0.014" are posterior probabilities that the individual "33-16" is belong to the 1st, 2nd, and 3rd subpopulations, respectively. When the Q matrix was uploaded to the software, the software would automatically delete the column in which the column sum is the smallest if their sums in each row are all equal to one.

Table 4. The Q matrix format of dataset filePS

<PopStr>			
<ID>	**Q1**	**Q2**	**Q3**
33-16	0.014	0.972	0.014
Nov-38	0.003	0.993	0.004
A4226	0.071	0.917	0.012
A4722	0.035	0.854	0.111
⋮	⋮	⋮	⋮

Principal component format for dataset "filePS" (Table 5)

The principal component dataset in Table 5 consists of a (n+2) × (k+1) matrix, where n is sample size (the number of the common individuals, as described above), and k is the number of principal components. In the first column, "<PCA>" and "<ID>" must present in the first and second rows, respectively; "33-16", "Nov-38", and "A4226" are individual IDs or names. In the 2nd to (k+1)-th columns, "PC1" to "PCk" indicate the first to kth principal components. In the second column, "0.306" is the score of the first principal component for the 1st individuals. Note that any principle components are not deleted in the software.

Table 5. The principal components format of the filePS dataset

<PCA>			
<ID>	**PC1**	**PC2**	**PC3**
33-16	0.306	0.029	0.226
Nov-38	-0.708	-0.271	1.413
A4226	-2.330	0.116	-0.824
A4722	1.059	0.470	-0.135
⋮	⋮	⋮	⋮

Evolutionary population structure format for dataset "filePS" (Table 6)

The evolutionary population structure dataset in Table 6 consists of a (n+2) × 2 matrix, where n is sample size (the number of common individuals, as described above). In the first column, "<EvolPopStr>" and "<ID>" must present in the first and second rows, respectively; "33-16", "Nov-38" and "A4226" are individual IDs or names. In the second column, "EvolType": evolutionary type, i.e., the evolutionary types for individuals "33-16" and "A4722" are "A" and "B", respectively, such as wild (A), landrace (B), and bred (C) soybeans.

Table 6. The evolutionary population format of the filePS dataset

<EvolPopStr>	
<ID>	EvolType
33-16	A
Nov-38	B
A4226	A
A4722	B
⋮	⋮

filePS=NULL indicates no population structure available in model. filePS="D:/Users/PopStr.csv" means that population structure dataset with name PopStr.csv is uploaded from the folder "D:/Users". If the number and order of individuals in PopStr.csv are not consistent with those of the above common individuals, our software may match the population structure matrix in order that the number and order of new matrix are consistent with those in the above common individuals.

The format for covariate dataset "fileCov" (Table 7)

The "**Covariate**" dataset consists of the $(n+2) \times (k+1)$ matrix, where n is sample size (the number of common individuals, as described above), and k is the number of covariates. In the first column, "<**Covariate**>" and "<**ID**>" must present in the first and second rows, respectively. If covariate is categorical, the names are Cate_covariate*. If covariate is continuous, the names are Con_covariate* (Table 6).

Table 7. The format of fileCov dataset

<Covariate>				
<ID>	Cate_covariate1	Cate_covariate2	Con_covariate1	Con_covariate2
33-16	A	C	349.5	374
Nov-38	B	C	205	452
A4226	A	D	300	374
A4722	A	D	190	452
⋮	⋮	⋮	⋮	

fileCov=NULL indicates no covariates available in model. fileCov="D:/Users/covariate.csv" means that the covariates with file name covariate.csv are uploaded from the folder "D:/Users". If the number and order of individuals in the uploaded file are not consistent with those in the above common individuals, our software may be used to change the number and order of individuals for matching the above genotypic and phenotypic datasets.

2.2.2 Result

The result files (result-main_QTN_detection) include three files: *_K.csv (Kinship matrix calculated by IIIVmrMLM), *_midresult.csv (intermediate results), and *_result.csv (final results).

*_midresult.csv: This is the results of single marker scanning on the genome in the first step. In this file, all the columns are named as Marker (marker name), Chromosome, Position (marker position (bp) on the genome), and pvalue.Q (the P-value for main-effect QTNs).

Marker	Chromosome	Position (bp)	pvalue.Q
PZB00859.1	1	157104	0.292043111
PZA01271.1	1	1947984	0.185246808
PZA03613.2	1	2914066	0.99208603
PZA03613.1	1	2914171	0.999987108
PZA03614.2	1	2915078	0.976018023
⋮	⋮	⋮	⋮

*_result.csv: The final results for significant and suggested QTNs. In this file, all the columns are named as Trait ID, Trait name, Marker (marker name), Chromosome, Position (marker position (bp) on the genome), LOD (LOD score), add (additive effect), dom (dominant effect), variance (variance of each QTN), r^2 (%) (proportion of total phenotypic variance explained by each QTN), P-value (calculated from LOD score using χ^2 distribution), and significance (significant (SIG) QTNs are based on Bonferroni correction, that is, critical P-value is $0.05/m$, where m is the number of tests or markers, whereas suggested (SUG) QTNs are based on LOD $\geq$ 3.0, default). If there are **no any dominant effects** to be listed in the below table, and it indicates that there are only two types of genotypes for this marker.

Trait ID	Trait name	Marker	Chromosome	Position (bp)	LOD	add	dom	variance	r2(%)	P-value	significance
1	trait1	PZA03214.3	1	245136244	11.7017	7.0691		26.4678	6.3725	2.12416E-13	SIG
1	trait1	PZA03188.4	1	280719882	10.1462	-7.1265	0.896	34.9545	8.4157	7.14861E-11	SIG
1	trait1	PZA03559.1	2	15810363	7.872	5.5235	17.383	31.2681	7.5282	1.34367E-08	SIG
1	trait1	PZB01892.1	3	161573186	20.8753	-9.9062		16.0224	3.8576	1.07536E-22	SIG
1	trait1	PZA03647.3	3	185318086	4.4864	4.3915		14.9657	3.6032	5.48594E-06	SIG
1	trait1	PZA00112.5	5	13664679	13.5465	-7.7296		24.641	5.9327	2.8295E-15	SIG
1	trait1	PZA03042.5	5	64413280	16.8506	-8.392	-38.9679	18.2141	4.3853	1.4125E-17	SIG
1	trait1	PZB00379.1	9	26661626	5.1546	-4.2882	-27.6411	21.0427	5.0663	7.00752E-06	SIG

Multi_env: Detection of QTN-by-environment interactions for complex traits

Users must set "fileGen", "filePhe", "method", "trait", "n.en", and "savePath", while the other

parameters may be default in function *IIIVmrMLM*, including *fileKin=NULL; filePS=NULL; GenSeparator="comma"; PopStrType="Q"; fileCov=NULL; SearchRadius=20; svpal=0.01*.

Compared to the detection of main-effect QTNs for complex traits in single environment (Single_env), there are three main changes in QEI detection (Multi_env).

1) The phenotype file should be arranged according to traits, the number of environments for each trait is greater than or equal to 2;

2) method="Multi_env";

3) Add a vector n_en to represent the number of environments for each trait in the *filePhe*. For example, n_en=[2,2,3] (The *filePhe* file contains the phenotypic values of three traits, and the environmental numbers of each trait are 2, 2, and 3, respectively).

2.3.1 Data input format

Format for the dataset "filePhe"

The phenotypic file for complex trait should be a *.csv or *.txt file, as shown below.

<Phenotype>	trait1Env1	trait1Env2	trait2Env1	trait2Env2	…
B46	38.04	34.45	38.71	35.72	…
B52	38.64	40.85	43.04	34.97	…
B57	41.54	33.82	45.10	33.23	…
B64	40.82	NA	39.14	30.93	…
B68	33.40	33.99	38.04	39.24	…
⋮	⋮	⋮	⋮	⋮	…

The first row in the first column: "<Phenotype>", while the second to *n*th rows in the first column: individual names (or IDs), such as B46. The first rows from the second column: trait ad environment names, such as "**trait1Env1**" (the phenotypes for the first trait in the first environment), while the other rows from the second column: phenotypic values of complex traits. The phenotypic file is arranged by traits, each trait has at least two columns, and each column is the phenotypes measured in an environment. The missed phenotypes are represented by "NA".

Format for datasets "fileGen", "fileKin", "filePS", "fileCov" are same as those in main-effect QTN detection (Single_env).

2.3.2 Result

The result file (result-QEI_detection) includes four files: *_K.csv (Kinship matrix calculated by IIIVmrMLM), *_midresult.csv (intermediate results), *_resultQ.csv (final results for main-effect QTNs) and *_resultQEI.csv (final result for QEIs). Note that there are two sheets: one is for significant and suggested QTNs (resultQ), and another is for significant and suggested QEIs (resultQEI).

*_midresult.csv: This is the results of single marker scanning on genome in the first step. In this file, all the columns are named as Marker (marker name), Chromosome, Position (marker position (bp) on genome), pvalue.Q (the P-value for QTNs), and pvalue.QE (the P-value for QEIs).

Marker	Chromosome	Position (bp)	pvalue.Q	pvalue.QE
PZB00859.1	1	157104	0.812972071	0.928177513
PZA01271.1	1	1947984	0.993594668	0.988087592
PZA03613.2	1	2914066	0.99306619	0.993440946
PZA03613.1	1	2914171	0.99997328	0.999970187
PZA03614.2	1	2915078	0.971495059	0.983226371
⋮	⋮	⋮	⋮	⋮

resultQ: The results are for significant and suggested QTNs. In this sheet, all the columns are named as Trait ID, Trait name, Marker (marker name), Chromosome, Position (marker position (bp) on genome), LOD (Q) (LOD score for QTNs), add (additive effect), dom (dominant effect), variance (variance of each QTN), r^2 (%) (proportion of total phenotypic variance explained by each QTN), P-value (transferred from LOD score in QTN detection using χ^2 distribution), and significance (significant (SIG) QTNs are based on Bonferroni correction, that is, critical P-value is 0.05/*m*, where *m* is the number of tests or markers, while suggested (SUG) QTNs are based on LOD ≥ 3.0, default).

Trait ID	Trait name	Marker	Chromosome	Position (bp)	LOD (Q)	add	dom	variance	r2(%)	P-value	significance
1	trait1	PZB01647.1	1	231039372	11.0446	1.1628	5.8057	1.2459	2.8372	9.03249E-12	SIG
1	trait1	PZA02812.34	1	267615649	12.721	-1.3274	4.7147	1.4823	3.3754	1.9032E-13	SIG
1	trait1	PZA02957.4	1	281818425	18.9611	1.6327		1.3099	2.9828	9.25024E-21	SIG
1	trait1	PZA03305.5	1	286642725	4.7749	-0.311	5.8961	0.658	1.4984	1.67971E-05	SIG
1	trait1	PZA00176.8	2	10533421	10.299	1.187	0.065	1.1969	2.7257	5.02727E-11	SIG
1	trait1	PZA03073.28	3	168443662	14.6457	-1.1235	4.7367	1.8954	4.3162	2.26368E-15	SIG
1	trait1	PZA01122.1	4	12618115	19.6559	-1.7468	4.5194	1.6788	3.8229	2.21205E-20	SIG
1	trait1	PZB01642.1	5	12337501	13.8005	1.3689		0.5059	1.1521	1.56227E-15	SIG

resultQEI: This is the results for significant and suggested QEIs. In this sheet, all the columns are named as Trait ID, Trait name, Marker (marker name), Chromosome, Position (marker position (bp) on genome), LOD (QE) (LOD score for QEIs). add*env*k* (additive effect in environment *k*), dom*env*k* (dominant effect in environment *k*), variance (variance of each QEI), r2 (%) (r2 (%) is proportion of total phenotypic variance explained by each QEI), P-value (transferred from LOD score in QEI detection using χ^2 distribution), and significance (significant (SIG) QEIs are based on Bonferroni correction, that is, critical P-value is 0.05/*m*, where *m* is the number of tests or markers, while suggested (SUG) QEIs are based on LOD ≥ 3.0, default).

Trait ID	Trait name	Marker	Chromosome	Position (bp)	LOD (QE)	add*env1	dom*env1	add*env2	dom*env2	variance	r2(%)	P-value	significance
1	trait1	tb1.15	1	264847721	9.7984	-1.152		1.152		1.3272	3.0223	1.85152E-11	SIG
1	trait1	PZA03191.3	3	185290309	10.0227	-1.1749		1.1749		1.3804	3.1435	1.09283E-11	SIG
1	trait1	PZA00281.1	5	9965510	12.9872	-1.3908	-0.356	1.3908	0.356	1.9268	4.3877	1.0311E-13	SIG
1	trait1	PZB00869.2	5	32366232	4.7824	-0.7892	-0.5069	0.7892	0.5069	0.6214	1.4151	1.65109E-05	SIG
1	trait1	PZA03042.1	5	64413079	5.7313	-0.814	-3.6012	0.814	3.6012	0.8164	1.8591	1.85763E-06	SIG

Epistasis: Detection of QTN-by-QTN interactions (QQIs) for complex traits

Users must set "fileGen", "filePhe", "method", "trait", and "savePath", while the other parameters may be default in function *IIIVmrMLM*, including *fileKin=NULL; filePS=NULL; GenSeparator="comma"; PopStrType="Q"; fileCov=NULL; SearchRadius=[10,20]; svpal=[0.01,0.01]; sblgwas_t=-1*.

In QQI detection *SearchRadius* and *svpal* are two-dimensional vectors, the first parameter is for QTNs and the second one is for QQIs. *sblgwas_t* is a number between [-3,0] to control sparseness of the “sblgwas” function, and the default value is -1.

Format for datasets “fileGen”, “fileKin”, “filePS”, “fileCov” are same as those in QTN detection (Single_env).

The result file (result-Epi_detection) includes three files: *_K.csv (Kinship matrix calculated by the IIIVmrMLM software), *_resultQ.csv (significant and suggested QTNs (resultQ)) and *_resultQQI.csv (significant and suggested QQIs (resultQQI)).

resultQ: The results are for significant and suggested QTNs. In this sheet, all the columns are named as Trait ID, Trait name, Marker (marker name), Chromosome, Position (marker position (bp) on genome), LOD (Q) (LOD score for QTN), add (additive effect), dom (dominant effect), variance (variance of each QTN), r^2 (%) (proportion of total phenotypic variance explained by each QTN), P-value (transferred from LOD score of QTNs in QQI detection using χ^2 distribution), and significance (significant (SIG) QTNs are based on Bonferroni correction, that is, critical P-value is 0.05/*m*, where *m* is the number of tests or markers, while suggested (SUG) QTNs are based on LOD ≥ 3.0, default).

Trait ID	Trait name	Marker	Chromosome	Position (bp)	LOD (Q)	add	dom	variance	r2(%)	P-value	significance
1	trait1	PZA03188.4	1	280719882	3.3317	-5.8445	0.5376	23.8219	5.7354	0.00046606	SUG

resultQQI: This is the results of significant and suggested QQIs. In this sheet, all the columns are named as Trait ID, Trait name, Marker_*i* (name of marker *i* in an interaction pair), Chr_*i*

(chromosome of marker *i* in an interaction pair), and Posi_*i* (position of marker *i* in an interaction pair, bp) (*i*=1,2), LOD (LOD score), aa.effect (additive-additive effect), ad.effect (additive-dominant effect), da.effect (dominant-additive effect), dd.effect (dominant-dominant effect), variance (variance of each QQI), r2 (%) (proportion of total phenotypic variance explained by each QQI), P-value (transferred from LOD score of QQIs in QQI detection using $\chi 2$ distribution), and significance (significant (SIG) QQIs are based on Bonferroni correction, that is, critical P-value is 0.05/*m*, where *m* is the number of tests or markers, while suggested (SUG) QQIs are based on LOD ≥ 3.0, default).

Trait ID	Trait name	Marker_1	Chr_1	Posi_1	Marker_2	Chr_2	Posi_2	LOD	aa.effect	ad.effect	da.effect	dd.effect	variance	r2(%)	P-value	significance
1	trait1	PZA03301.2	1	240574247	PZA03665.2	1	241430615	5.2653	-3.6773				12.9649	3.1215	8.47516E-07	SUG
1	trait1	PZA03214.1	1	245136387	PZA03336.3	2	11193471	3.5036	3.7019				12.986	3.1265	5.90067E-05	SUG
1	trait1	PZB01427.5	1	270364633	PZA02204.1	1	278690966	3.2263	-2.1569	0.0736			4.4934	1.0819	0.000594087	SUG
1	trait1	PZB00119.1	1	286269951	PZA00449.2	5	37511797	4.1247	-5.19	-1.0276			20.532	4.9434	7.50614E-05	SUG
1	trait1	PZA00108.4	2	13779970	PZB02017.3	2	20958616	3.8997	-3.9178		-0.7314		14.5959	3.5141	0.000126029	SUG
1	trait1	PZA03559.1	2	15810363	PZB01482.1	7	3671683	5.8034	5.3545		0.7908		27.1068	6.5263	1.57329E-06	SUG
1	trait1	PZD00027.1	3	169757113	PZA00281.13	5	9965534	3.5998	3.5533	0.3447			8.3321	2.0061	0.000251375	SUG
1	trait1	PZB01919.1	3	178235128	PZA03107.1	4	2851075	3.3939	-3.7499				13.9399	3.3562	7.70701E-05	SUG
1	trait1	PZB02080.1	4	4980568	PZA00188.1	7	2998280	3.7489	-4.9426				18.2097	4.3842	3.25358E-05	SUG
1	trait1	PZA00112.5	5	13664679	PZB01983.2	5	23437363	8.8132	-8.2499				35.1406	8.4606	1.88252E-10	SIG
1	trait1	PZA02792.16	5	21771297	PZA00710.1	5	61492543	3.7439	-4.5708		-0.1197		20.4298	4.9188	0.000180398	SUG

The IIIVmrMLM includes three steps. In the first step, single-marker genome-wide scanning is conducted, and its purpose is to obtain potentially associated markers. In the second step, all the selected markers are placed into a multi-locus genetic model, all the effects are estimated by empirical Bayes, and all the non-zero effects are further identified by likelihood ratio test for significant and suggested QTNs. In the third step, around all the significant and suggested loci, previously reported and candidate genes may be mined using multi-omics data and bioinformatics analyses. If there are known or candidate genes around a suggested locus, the suggested locus is reliable.

3. Reference

Li M, Zhang YW, Zhang ZC, Xiang Y, Liu MH, Zhou YH, Zuo JF, Zhang HQ, Chen Y, Zhang YM. A compressed variance component mixed model for detecting QTNs, and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. **Molecular Plant** 2022, 15(4): 630-650. doi: 10.1016/j.molp.2022.02.012