

**User Manual for**

# QTL.gCIMapping

**QTL genome-wide Composite Interval Mapping**

**(version 3.1)**

**Zhang Ya-Wen, Wen Yang-Jun, Wang Shi-Bo,  
Zhang Yuan-Ming (soyzzhang@mail.hzau.edu.cn)**

**Last updated on October 16, 2018**

**Disclaimer:** While extensive testing has been performed by Yuan-Ming Zhang's Lab (Statistical Genomics Lab) at Huazhong Agricultural University, the results are, in general, reliable, correct or appropriate. However, results are not guaranteed for any specific datasets. We strongly recommend that users validate the GCIM results with other software packages, such as [Windows QTL Cartographer V2.5\\_011](https://brcwebportal.cos.ncsu.edu/qtldcart/WQTLCart.htm) (<https://brcwebportal.cos.ncsu.edu/qtldcart/WQTLCart.htm>) and [QTL IciMapping V4.1](http://www.isbreeding.net/software/?type=detail&id=18) (<http://www.isbreeding.net/software/?type=detail&id=18>).

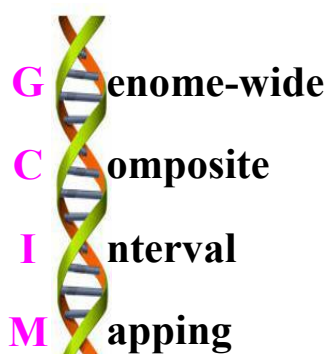
**Download website:**

<https://cran.r-project.org/web/packages/QTL.gCIMapping/index.html>

**References**

1. Wang Shi-Bo, Wen Yang-Jun, Ren Wen-Long, Ni Yuan-Li, Zhang Jin, Feng Jian-Ying, Zhang Yuan-Ming\*. Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. *Scientific Reports* 2016, 6: 29951.
2. Wen Yang-Jun, Zhang Ya-Wen, Zhang Jin, Feng Jian-Ying, Jim M. Dunwell, Zhang Yuan-Ming\*. An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F<sub>2</sub>. *Briefings in Bioinformatics*, 2018, doi: [10.1093/bib/bby058](https://doi.org/10.1093/bib/bby058)

## Quantitative Trait Loci



**Funding**

The works were supported by the National Natural Science Foundation of China (31571268 and U1602261), Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2014RC020), and State Key Laboratory of Cotton Biology Open Fund (CB2017B01).

## INTRODUCTION

### 1.1 Why GCIM?

**QTL.gCIMapping** (**QTL** Genome-wide **C**omposite **I**nterval **M**apping) is an R package for multi-QTL mapping.

QTL.gCIMapping v3.1 is able to work on the popular platforms, like Windows, Linux (desktop) and MacOS.

### 1.2 Getting started

QTL.gCIMapping is a package that runs in the R software environment, which can be freely downloaded from <https://cran.r-project.org/web/packages/QTL.gCIMapping/index.html>, or request from the maintainer, Dr Yuan-Ming Zhang at Huazhong Agricultural University ([soyzzhang@mail.hzau.edu.cn](mailto:soyzzhang@mail.hzau.edu.cn) or [soyzzhang@hotmail.com](mailto:soyzzhang@hotmail.com)).

#### 1.2.1 One-Click installation

Within R environment, the QTL.gCIMapping software can be installed directly using the below command:

```
install.packages(pkgs="QTL.gCIMapping")
```

#### 1.2.2 Step-by-step installation

##### 1.2.2.1 Install the add-on packages

**Online installation** Within R environment on the internet, the QTL.gCIMapping package can be installed online, using the below command:

```
install.packages(pkgs=c("qtl","doParallel","foreach","iterators","openxlsx","MASS",  
"stringr","parcor","data.table"))
```

**Offline installation** Users should download the below 27 packages from CRAN, github (<https://github.com/>), or google search:

```
"cmprsk","corpcor","data.table","doParallel","Epi","etm","fdrtool","foreach","GeneN  
et","glmnet","iterators","longitudinal","magrittr","MASS","numDeriv","openxlsx","p  
arcor","plyr","pppls","qtl","Rcpp","stringi","stringr","testthat","utf8","zip","zoo"
```

Then, install them offline (under the R environment, select all the 27 packages and install them offline).

### 1.2.2.2 Install QTL.gCIMapping

Open R GUI, select "Packages"—"Install package(s) from local files..." and then find the QTL.gCIMapping package which you have downloaded on your desktop.

Within R environment, launch the QTL.gCIMapping by command:

```
library(QTL.gCIMapping)
```

**User Manual file** Users can decompress the QTL.gCIMapping package and find the User Manual file (name: **Instruction.pdf**) in the folder of ".../QTL.gCIMapping/inst/doc".

## 2. Parameter settings

Parameter	Meaning	File format	Note
file	File path & name in your computer, i.e., file="D:/Users/GCIM_Format_DH.csv"	*.csv; *.txt	Table 1
fileFormat	Format for input file: GCIM (QTL.gCIMapping), ICIM (QTL IciMapping) and Cart (WinQTLCart)		
fileICIMcov	File path & name in your computer, i.e., fileICIMcov="D:/Users/GCIM_Cov.csv" or fileICIMcov=NULL	*.csv; *.txt (Covariate values: Row: individual; Column: covariate name)	Table 3
Population	BC1 (F1×P1), BC2 (F1×P2), DH, RIL, F2, i.e., Population="BC1"		
Model	Random (random model) or Fixed (fixed model) for QTL effects, i.e., Model="Random"		
WalkSpeed	Walk speed for Genome-wide Scanning (centi-Morgan, cM), i.e., WalkSpeed=1		
CriLOD	Critical LOD scores for significant QTL. CriLOD=2.5: the critical LOD score for significant QTL is set at 2.5		
Likelihood	This parameter is only for F2 population, including restricted maximum likelihood (REML) and maximum likelihood (ML). Likelihood="REML" or Likelihood="ML"		
flagrqt1	This parameter is only for F2 population, flagrqt1="FALSE" in the first running. If the other software detects only one QTL in a neighborhood but this software finds two linked QTLs (one with additive effect and another with dominant effect) in the region, let flagrqt1="TRUE"		
DrawPlot	This parameter is for all the populations, including FALSE and TRUE. DrawPlot=FALSE indicates no figure output; DrawPlot=TRUE: the LOD score [or $-\log_{10}(P\text{-value})$ ] figure against genome position.		
Plotformat	*.jpeg, *.png, *.tiff and *.pdf. For example, Plotformat="jpeg" indicates the *.jpeg format of the figure file.		
Resolution	Low or High. Resolution="Low" indicates the low resolution of the figure file.		
Trait	Trait=1:3 indicates the analyses from the first trait to the third trait.		
dir	Save path in your computer, i.e., "D:/Users"		

## Example

```
QTL.gCIMapping(file="D:/Users/GCIM_Format_DH.csv",fileFormat="GCIM",fileICIMcov=NULL,Population="DH",Model="Random",WalkSpeed=1,CriLOD=2.5,Likelihood="REML",flagrqt1="FALSE",DrawPlot="TRUE",PlotFormat="png",Resolution="Low",Trait=1:1,dir="D:/Users")
```

## Dataset format

**GCIM format for Dataset** The first column, named "**marker**", presents marker name. The second column, named "**chr**", stands for chromosome. The third column, named "**pos**", stands for marker position (cM) on the chromosome. Among the remaining columns, each column lists all the genotypes for one individual while the first row shows the individual names. For each marker, the coding criteria are shown as [Table 2](#). The phenotype and covariate information are followed the marker genotypes. The phenotypes start with the second column (named "**trait1**"), and the third column is trait name. The remaining columns are phenotypic values for all the individuals. If one more traits exist, more rows will be added (the phenotypic values for each trait are listed in one row). If covariates exist, all the information for the covariates will list after the trait information. The format is seen in Table 1. If there is no covariate, users should delete the last row in [Table 1](#).

**Table 1. The GCIM format of the dataset**

marker	chr	pos	DH6-10	DH6-101	DH6-102
RGA3(1)	1	0	B	-	B
wPt-6358	1	3.034	B	-	-
Hplc2	1	8.8291	A	A	B
wPt-9752	1	10.1452	A	-	-
abc156a	1	41.3408	A	A	B
⋮	⋮	⋮	⋮	⋮	⋮
gwm437	21	162.5218	A	B	-
gwm121	21	180.2878	A	B	-
wmc157	21	197.9196	A	B	A
*stm1actc	21	200.4216	-	-	-
	trait1	T19	75.33	105	96.33
	trait2	T191	74	105.68	97.16
	trait3	T192	75.37	104.67	95.55
	Covar1	CovarName	A	B	B

**Table 2. Coding criteria for GCIM format**

Genotype	Code	Meaning
AA	A	Homozygous genotype (P <sub>1</sub> )
Aa	H	Heterozygous genotype (F <sub>1</sub> )
aa	B	Homozygous genotype (P <sub>2</sub> )
AA + Aa (Not aa)	D	Dominance to P <sub>1</sub>
Aa + aa (Not AA)	C	Dominance to P <sub>2</sub>
Missing	-	Missing or unclear genotype

**ICIM format for Dataset** If users have the dataset files for QTL IciMapping format, these files are also available in our software. Details can be seen in the folder of “.../QTL.gCIMapping/inst/extdata”, i.e., [WheatDH\\_QTLIciMapping\\_Format.xlsx](#).

**WinQTLCart format for Dataset** If users have the dataset file for WinQTLCart format, this file is also available in our software. Details can be seen in the folder of “.../ QTL.gCIMapping /inst/extdata”, i.e., [env1-jun3\\_WinQTLCart\\_Format.mcd](#).

**The format for fileICIMcov dataset** If users select ICIM format and the covariate exists in the dataset, it needs to input a covariate file. In the file, the first column indicates individual name and the second column is the covariate information (Table 3). In Table 3, the values for covariate are A, B and C.

**Table 3. The covariate file format**

Individual ID	Covariate
DH6-10	A
DH6-101	A
DH6-102	A
DH6-104	A
DH6-164	B
DH6-165	B
DH6-166	B
DH6-170	B
DH7-124	C
DH7-125	C

### 3. Result

For BC1, BC2, DH and RIL populations, the **Results** file has ten columns, as shown below.

**Trait:** The trait name analyzed.

**Chr:** Chromosome, represented by an integer number.

**Position (cM):** The QTL position (cM) on the chromosome.

**Additive Effect:** Additive effect for significant QTL.

**LOD:** LOD score for significant QTL.

**Left\_Marker:** Left flanking marker name for significant QTL.

**Right\_Marker:** Right flanking marker name for significant QTL.

**Var\_Genet:** Genetic variance for each significant QTL.

**r<sup>2</sup> (%):** Proportion of phenotypic variance explained by single QTL.

**Var\_Error:** residual variance for the full model.

**Var\_Phen (total):** Phenotypic variance in the analyzed population.

For F<sub>2</sub> population, the **Results** file has eleven columns. Trait, Chr, Position (cM), Left\_Marker, Right\_Marker, Var\_Genet, LOD, r<sup>2</sup> (%), Var\_Error and Var\_phen are same as those in the above populations. The different columns are as follows.

**Effect.a:** Additive effect for significant QTL.

**Effect.d:** Dominant effect for significant QTL.