**User Manual for**

# QTL.gCIMapping

**QTL genome-wide Composite Interval Mapping**

**(version 3.3)**

**Zhang Ya-Wen, Wen Yang-Jun, Wang Shi-Bo,**

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

**Last updated on April 30, 2020**

**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/qtlcart/WQTLCart.htm) and QTL IciMapping V4.1 (http://www.isbreeding.net/software/?type=detail&id=18).

**Download website**:
https://cran.r-project.org/web/packages/QTL.gCIMapping/index.html

## References

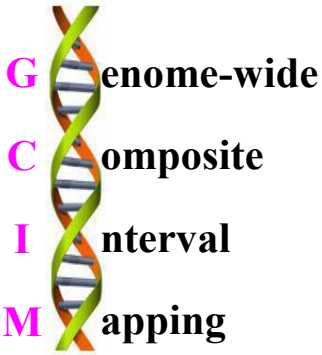

1 Wang Shi-Bo, Wen Yang-Jun, Ren Wen-Long, Ni Yuan-Li, Zhang Jin, Feng Jian-Ying, Zhang Yuan-Ming*. Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. ***Scientific Reports*** 2016, 6: 29951.

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 $F_2$. ***Briefings in Bioinformatics*** 2019, 20(5): 1913-1924.

3 Zhang Ya-Wen, Wen Yang-Jun, Jim M. Dunwell, Zhang Yuan-Ming*. QTL.gCIMapping.GUI v2.0: An R software for detecting small-effect and linked QTLs for quantitative traits in bi-parental segregation populations. ***Computational and Structural Biotechnology Journal*** 2020, 18: 59-65.

4 Wen Yang-Jun, Zhang Ya-Wen, Zhang Jin, Feng Jian-Ying, Zhang Yuan-Ming*. The improved FASTmrEMMA and GCIM algorithms for genome-wide association and linkage studies in large mapping populations. ***The Crop Journal***, in revision.


**Quantitative Trait Loci**



**Genome-wide Composite Interval Mapping**


## Funding

The works were supported by the National Natural Science Foundation of China (31571268, 31871242, 31701071 and U1602261), Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2014RC020), and State Key Laboratory of Cotton Biology Open Fund (CB2017B01 and CB2019B01).

# INTRODUCTION

## 1.1 Why GCIM?

**QTL.gCIMapping** v3.3 (**QTL G**enome-wide **C**omposite **I**nterval **M**apping) is an R package for multi-QTL mapping of quantitative traits in bi-parental segregation populations, including backcross (BC), doubled haploid (DH) lines, recombinant inbred lines (RIL), $F_2$, and immortalized $F_2$ ($IF_2$). QTL.gCIMapping v3.3 works well on the R environment (v3.6.3) on Windows, Linux (desktop) and MacOS.

## 1.2 Getting started

The software package QTL.gCIMapping v3.3 can be freely downloaded from https://cran.r-project.org/web/packages/QTL.gCIMapping/index.html, or BioCode (https://bigd.big.ac.cn/biocode/tools/BT007078), or request from the maintainer, Dr Yuan-Ming Zhang at Crop Information Center, College of Plant Science & Technology, Huazhong Agri Univ (soyzhang@mail.hzau.edu.cn).

### 1.2.1 One-Click online installation

On R environment and network connection, the command,
install.packages(pkgs="QTL.gCIMapping")
is used to directly install the software package QTL.gCIMapping v3.3.

### 1.2.2 Step-by-step offline installation

#### 1.2.2.1 Install the add-on packages

First, users download twenty-seven R packages, including
"cmprsk","corpcor","data.table","doParallel","Epi","etm","fdrtool","foreach","GeneNet","glmnet","iterators","longitudinal","magrittr","MASS","numDeriv","openxlsx","parcor","plyr","ppls","qtl","Rcpp","stringi","stringr","testthat","utf8","zip","zoo"
from CRAN, github (https://github.com/), or google search.

On the R environment, then, users select all the 27 packages and install them offline.

#### 1.2.2.2 Install QTL.gCIMapping v3.3

On R GUI environment, users first select "Packages"—"Install package(s) from local files…", then find the software package QTL.gCIMapping v3.3 on user's desktop

computer or mobile device, and launch QTL.gCIMapping v3.3.

### 1.2.3 Run QTL.gCIMapping v3.3

Once the software package QTL.gCIMapping v3.3 is installed, users may run it using two commands:

library(QTL.gCIMapping)

QTL.gCIMapping(***) (*** are the parameter list: please see § 2 Example)

If users re-use the software QTL.gCIMapping v3.3, users use the above two commands as well.

**User Manual file** Users can decompress the QTL.gCIMapping.GUI package and find the User Manual file (name: **Instruction.pdf**) in the folder of "…/QTL.gCIMapping.GUI/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, $F_2$, 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 $F_2$ population, including restricted maximum likelihood (REML) and maximum likelihood (ML). Likelihood="REML" or Likelihood="ML" | | |
| **SetSeed** | This parameter is only for $F_2$ population, in which the cross-validation experiment is needed | | |
| **flagrqtl** | This parameter is only for $F_2$ population, **flagrqtl="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 **flagrqtl="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

### The full codes

QTL.gCIMapping(file="D:/Users/GCIM_Format_DH.csv",fileFormat="GCIM",fileICIMcov=NULL,Population="DH",Model="Random",WalkSpeed=1,CriLOD=2.5,Likelihood="REML",SetSeed=11001,flagrqtl="FALSE",DrawPlot="TRUE",PlotFormat="png",Resolution="Low",Trait=1:1,dir="D:/Users")

### The reduced codes

QTL.gCIMapping(file="D:/Users/GCIM_Format_F2.csv",Population="F2",WalkSpeed=1,CriLOD=2.5,Trait=1,dir="D:/Users")

It should be noted that users must set "file", "Population", "WalkSpeed", "CriLOD", "Trait" and "dir", and the other eight parameters can be default in function, including fileFormat="GCIM"; fileICIMcov=NULL; Model="Random"; Likelihood="REML"; SetSeed=11001 and flagrqtl="FALSE" only for $F_2$ population; DrawPlot=TRUE; Plotformat= "jpeg"; Resolution= "Low". Generally speaking, the random seed in the cross-validation experiment was set as 11001. If some known genes aren't identified by the seed, users may try to use some new random seeds. At this case, one better result may be obtained.

## Dataset format

**GCIM format for Dataset** The first three columns, named "**marker**", "**chr**" and "**pos**", stand for marker name, chromosome and marker position (cM) on the chromosome, respectively. Among the remaining columns, each column lists all the genotypes of one individual or line, while the first row shows the name of the individual or line. For the genotypes of each marker, the coding criteria are shown as Table 1.

**Table 1. Coding criteria for GCIM format**

| Genotype | Code | Meaning |
|---|---|---|
| AA | A | Homozygous genotype ($P_1$) |
| Aa | H | Heterozygous genotype ($F_1$) |
| aa | B | Homozygous genotype ($P_2$) |
| AA + Aa (Not aa) | D | Dominance to $P_1$ |
| Aa + aa (Not AA) | C | Dominance to $P_2$ |
| Missing | - | Missing or unclear genotype |

The genotypic, phenotypic and covariate datasets are located on the upper, middle,

lower sections, and each covariate or trait is presented on one row. On each row, the first column is empty followed by "**trait1**", "real trait name", and "phenotypic values for all the individuals or lines". If there are multiple traits, these traits occupy multiple lines. If there are covariates, the content lies below the trait dataset. The format is seen in Table 2. If there is no covariate, users should delete the last row in Table 2.

**The format of ICIM dataset** If users have the QTL IciMapping dataset, these files are also available in our software. Details can be found in the folder of "…/QTL.gCIMapping.GUI/inst/extdata", i.e., WheatDH_QTLIciMapping_Format.xlsx.

**The format of WinQTLCart dataset** If users have the WinQTLCart dataset, its file is also available in our software. Details can be found in the folder of "…/QTL.gCIMapping.GUI/inst/extdata", i.e., env1-jun3_WinQTLCart_Format.mcd.

**Table 2. 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 |

**The format of ICIM covariate dataset** If users use the ICIM dataset and there are covariates, users need 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 covariate values are indicated by such as 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 | C |
| DH7-124 | 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.

**$r^2$ (%):** 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 $F_2$ population, the **Results** file has eleven columns. Trait, Chr, Position (cM), Left_Marker, Right_Marker, Var_Genet, LOD, $r^2$ (%), Var_Error and Var_phen are same as those in the above populations. In $F_2$ population, QTL effects include additive (**Effect.a**) and dominant (**Effect.d**) effects.