

User Manual for

QTL.gCIMapping.GUI

QTL genome-wide Composite Interval Mapping GUI

(version 2.0)

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

Last updated on November 10, 2019

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.GUI/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 F₂. *Briefings in Bioinformatics*, 2018, online, doi: 10.1093/bib/bby058
3. Zhang Ya-Wen, 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*, 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.GUI v2.0 (**QTL** Genome-wide **C**omposite **I**nterval **M**apping **G**raphical **U**ser **I**nterface) 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.GUI v2.0 works well on the R environment on Windows, Linux (desktop) and MacOS.

1.2 Getting started

The software package QTL.gCIMapping.GUI v2.0 can be freely downloaded from <https://cran.r-project.org/web/packages/QTL.gCIMapping.GUI/index.html>, or request from the maintainer, Dr Yuan-Ming Zhang at Crop Information Center, College of Plant Science & Technology, Huazhong Agri Univ (soy Zhang@mail.hzau.edu.cn).

1.2.1 One-Click online installation

On R environment and network connection, the command

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

is used to directly install the software package QTL.gCIMapping.GUI v2.0.

1.2.2 Step-by-step offline installation

1.2.2.1 Install the add-on packages

First, users download thirty-nine R packages, including

```
"cmprsk", "corpcor", "data.table", "digest", "doParallel", "Epi", "etm", "fdrtool", "foreach",  
"GeneNet", "glmnet", "htmltools", "httpuv", "iterators", "jsonlite", "later", "longitudinal",  
"magrittr", "MASS", "mime", "numDeriv", "openxlsx", "parcor", "plyr", "ppls", "promises",  
"QTL.gCIMapping", "qtl", "R6", "Rcpp", "shiny", "sourcetools", "stringi", "stringr", "testt  
hat", "utf8", "xtable", "zip", "zoo",
```

from CRAN, github (<https://github.com/>), or google search.

On the R environment, then, users select all the 39 packages and install them offline.

1.2.2.2 Install QTL.gCIMapping.GUI v2.0

On R GUI environment, users first select "Packages"—"Install package(s) from local files...", then find the software package [QTL.gCIMapping.GUI v2.0](#) on user's desktop computer or mobile device, and launch [QTL.gCIMapping.GUI v2.0](#).

1.2.3 Run QTL.gCIMapping.GUI v2.0

Once the software package QTL.gCIMapping.GUI v2.0 is installed, users may run it using two commands:

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

```
QTL.gCIMapping.GUI()
```

If users re-use the software QTL.gCIMapping.GUI v2.0, 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. 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

Marker genotype	Code	Meaning
AA	A	Homozygous genotype (P ₁)
Aa	H	Heterozygous genotype (F ₁)
aa	B	Homozygous genotype (P ₂)
Not AA (Aa + aa)	C	Dominance to P ₂
Not aa (AA + Aa)	D	Dominance to P ₁
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](#).

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 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](#).

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	B
DH6-170	B
DH7-124	C
DH7-125	C

4. Operation process

3.1 The graphical interface of QTL.gCIMapping.GUI v2.0

QTL.gCIMapping.GUI Start

QTL.gCIMapping.GUI (QTL genome-wide Composite Interval Mapping with Graphical User Interface)

Coding criteria			Dataset example					
Genotype	Code	Meaning	marker	chr	pos	DH6.10	DH6.101	DH6.102
AA	A	Homozygous genotype (P1)	RGAs(1)	1	0	B	-	B
Aa	H	Heterozygous genotype (F1)	wPL-6358	1	3.034	B	-	-
aa	B	Homozygous genotype (P2)	Hpk2	1	8.8291	A	A	B
AA+Aa(Not aa)	D	Dominance to P1
Aa+aa(Not AA)	C	Dominance to P2	gwm437	21	162.5218	A	B	-
Missing	-	Missing or unclear genotype	gwm121	21	180.2878	A	B	-
			wmc157	21	197.9196	A	B	A
			trait1	T19	75.33	105	96.33	
			trait2	T191	74	105.68	97.16	
			Covar	CovarName	A	B	B	

Reference

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 F2. Briefings in Bioinformatics doi:10.1093/bib/bby058.
3. Zhang Ya-Wen, 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, in revision.

Authors: Zhang Ya-Wen, Wen Yang-Jun, Wang Shi-Bo, Zhang Yuan-Ming
 Maintainer: Zhang Yuan-Ming (soy Zhang at mail.hzau.edu.cn)
 QTL.gCIMapping version 2.0. Released November 2019

Figure 1. The Graphical User Interface of QTL.gCIMapping.GUI v2.0

3.2 Input dataset

Users must upload the dataset files with three kinds of formats (Figs 2 to 4). If users select the QTLciMapping format and there are the covariates, users should upload the covariate matrix (Fig 5).

QTL.gCIMapping.GUI

Please select data format

GCIM

WinQTLCart

QTLciMapping

Input dataset

Browse... GCIM_Format_DH.csv

Upload complete

Show dataset:

Genotype

Parameter Settings

Figure

User manual

Dataset Parameter Settings Figure

Show 25 entries Search:

marker	DH6-10	DH6-101	DH6-102	DH6-104	DH6-105	DH6-108	DH6-11	DH6-111	DH6-112	DH6-114
RGA3(1)	B	-	B	A	B	B	A	A	A	-
wPt-6358	B	-	-	-	-	B	A	A	A	-
Hplc2	A	A	B	A	B	B	A	A	A	B
wPt-9752	A	-	-	-	-	-	A	A	A	-
abc156a	A	A	B	A	B	B	B	B	A	B
RGA36b(2)	A	-	B	A	-	B	B	B	A	-
bcd98	B	A	B	A	B	B	B	B	A	B
wmc24	B	A	B	A	B	B	B	B	A	B
ksuG9c	B	A	B	A	B	B	B	B	A	B
wPt-2436	B	-	-	-	-	B	B	B	B	-
wPt-4886	B	-	-	-	-	B	B	B	B	-
wmc120	B	A	B	A	B	B	B	B	B	A
cdo105	B	A	B	A	B	B	B	B	B	A

Fig 2. The GCIM dataset format

QTL.gCIMapping.GUI

Please select data format

GCIM

WinQTLCart

QTLciMapping

Input dataset

Browse... env1-jun3_WinQTLCart_Format.mcd

Upload complete

Show dataset:

Genotype

Parameter Settings

Figure

User manual

Dataset Parameter Settings Figure

```
#FileID 1148497108 #bychromosome -type interval -function l -units cM -chromosomes 3 -maximum 20
-named yes -start -Chromosome Chr-1 c1m1 10.0000 c1m3 10.0000 c1m4
10.0000 c1m5 10.0000 c1m6
10.0000 c1m7 10.0000 c1m8
10.0000 c1m9 10.0000 c1m10
10.0000 c1m11 10.0000 c1m12
10.0000 c1m13 10.0000 c1m14
10.0000 c1m15 10.0000 c1m16
10.0000 c1m17 10.0000 c1m18
10.0000 c1m19 10.0000 c1m20
0.0000 -Chromosome Chr-2 c2m1 10.0000 c2m2
10.0000 c2m3 10.0000 c2m4
10.0000 c2m5 10.0000 c2m6
10.0000 c2m7 10.0000 c2m8
10.0000 c2m9 10.0000 c2m10
10.0000 c2m11 10.0000 c2m12
10.0000 c2m13 10.0000 c2m14
10.0000 c2m15 10.0000 c2m16
10.0000 c2m17 10.0000 c2m18
10.0000 c2m19 10.0000 c2m20
0.0000 -Chromosome Chr-3 c3m1 10.0000 c3m2
10.0000 c3m3 10.0000 c3m4
10.0000 c3m5 10.0000 c3m6
10.0000 c3m7 10.0000 c3m8
10.0000 c3m9 10.0000 c3m10
10.0000 c3m11 10.0000 c3m12
10.0000 c3m13 10.0000 c3m14
10.0000 c3m15 10.0000 c3m16
10.0000 c3m17 10.0000 c3m18
10.0000 c3m19 10.0000 c3m20
```

Fig 3. The WinQTLCart dataset format

QTL.gCIMapping.GUI Start

QTL.gCIMapping.GUI

Please select data format

GCIM
 WinQTLCart
 QTLiCiMapping

Input dataset

Browse... WheatDH_QTLiCiMapping_Format.xls
Upload complete

Input covariate file

Browse... No file selected

Show dataset:

Genotype

Parameter Settings

Figure

User manual

Dataset Parameter Settings Figure

Show 25 entries Search:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
RGA3(1)	0	-1	0	2	0	0	2	2	2	-1	0	0	0	
wPt-6358	0	-1	-1	-1	-1	0	2	2	2	-1	0	-1	-1	
Hplc2	2	2	0	2	0	0	2	2	2	0	0	0	0	
wPt-9752	2	-1	-1	-1	-1	-1	2	2	2	-1	0	-1	-1	
abc156a	2	2	0	2	0	0	0	0	2	0	-1	0	0	
RGA38b(2)	2	-1	0	2	-1	0	0	0	2	-1	0	0	0	
bcd98	0	2	0	2	0	0	0	0	2	0	2	0	0	
wmc24	0	2	0	2	0	0	0	0	2	0	2	0	0	
ksuG9c	0	2	0	2	0	0	0	0	2	0	2	0	0	
wPt-2436	0	-1	-1	-1	-1	0	0	0	0	-1	2	0	-1	
wPt-4886	0	-1	-1	-1	-1	0	0	0	0	-1	2	0	-1	
wmc120	0	2	0	2	0	0	0	0	0	2	2	0	0	
cdo105	0	2	0	2	0	0	0	0	0	2	2	0	0	
wPt-6074	0	-1	-1	-1	-1	0	0	0	0	-1	2	0	-1	
GluA1	0	2	0	2	0	2	0	0	0	2	2	0	0	

Fig 4. The QTLiCiMapping dataset format

QTL.gCIMapping.GUI Start

QTL.gCIMapping.GUI

Please select data format

GCIM
 WinQTLCart
 QTLiCiMapping

Input dataset

Browse... WheatDH_QTLiCiMapping_Format
Upload complete

Input covariate file

Browse... ICIM_Cov.csv
Upload complete

Show dataset:

Covariate

Parameter Settings

Figure

User manual

Dataset Parameter Settings Figure

Show 25 entries Search:

V1	V2
Individual ID	Covariate
DH6-10	A
DH6-101	A
DH6-102	A
DH6-104	A
DH6-105	A
DH6-108	A
DH6-11	B
DH6-111	A
DH6-112	A
DH6-114	A
DH6-119	A
DH6-124	A
DH6-125	B
DH6-128	B

Fig 5. Covariate input in the QTLiCiMapping dataset format

3.3 Parameter settings (Fig 6)

Select population: BC1 (F1 × P1), BC2 (F1 × P2), DH, RIL, and F2.

Select model: Random or Fixed model for QTL effects.

Walk Speed for Genome-wide Scanning (cM): Set walk speed for genome-wide scanning (centi-Morgan, cM), for example, 1 cM.

Critical LOD score: Critical LOD scores for significant QTL, for example, 2.5 or 3.0.

Likelihood function: This parameter is only for F₂ population, including restricted maximum likelihood (REML) and maximum likelihood (ML).

Completing CIM in one neighborhood: This parameter is only for F₂ population. In the first running, please set "FALSE". If the other software detects only one QTL in a neighborhood but the current software finds two linked QTLs (one with additive effect and another with dominant effect) in the **neighborhood**, please set "TRUE" and run again.

Traits analyzed: "2:2" or "2" indicates the analyses from the second trait, "2:4" indicates the analyses from the second to fourth traits, and "2,4" indicates the analyses of the second trait and the fourth trait.

Save path: The result will be written to the path in your computer.

Draw plot or not: This parameter setup includes FALSE and TRUE. "FALSE" indicates no figure output, and "TRUE" indicates the output of QTL mapping curve, for example, the LOD score [or $-\log_{10}(P\text{-value})$] curve against genome position.

Resolution of plot: Low or High: the low or high resolution for the figure file.

Plot format: Users can download the picture for different file formats: *.jpeg, *.png, *.tiff and *.pdf.

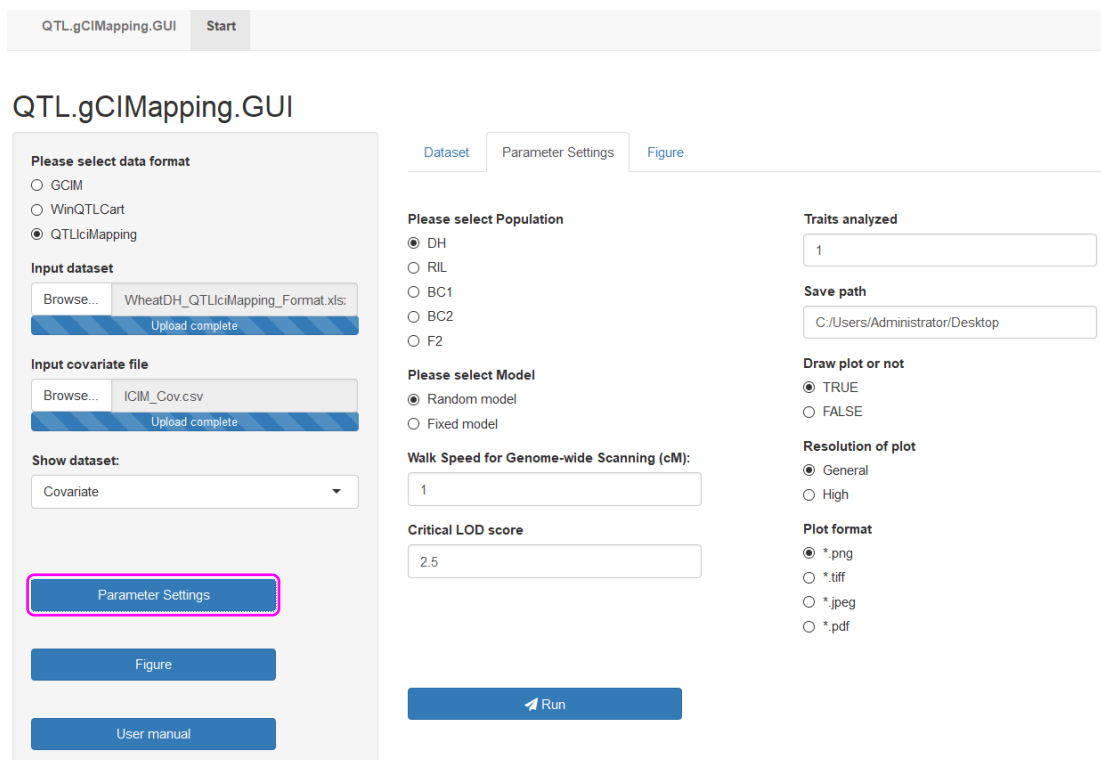


Fig 6. Parameter setting in the mapping of QTL for quantitative traits

3.4 Run the software

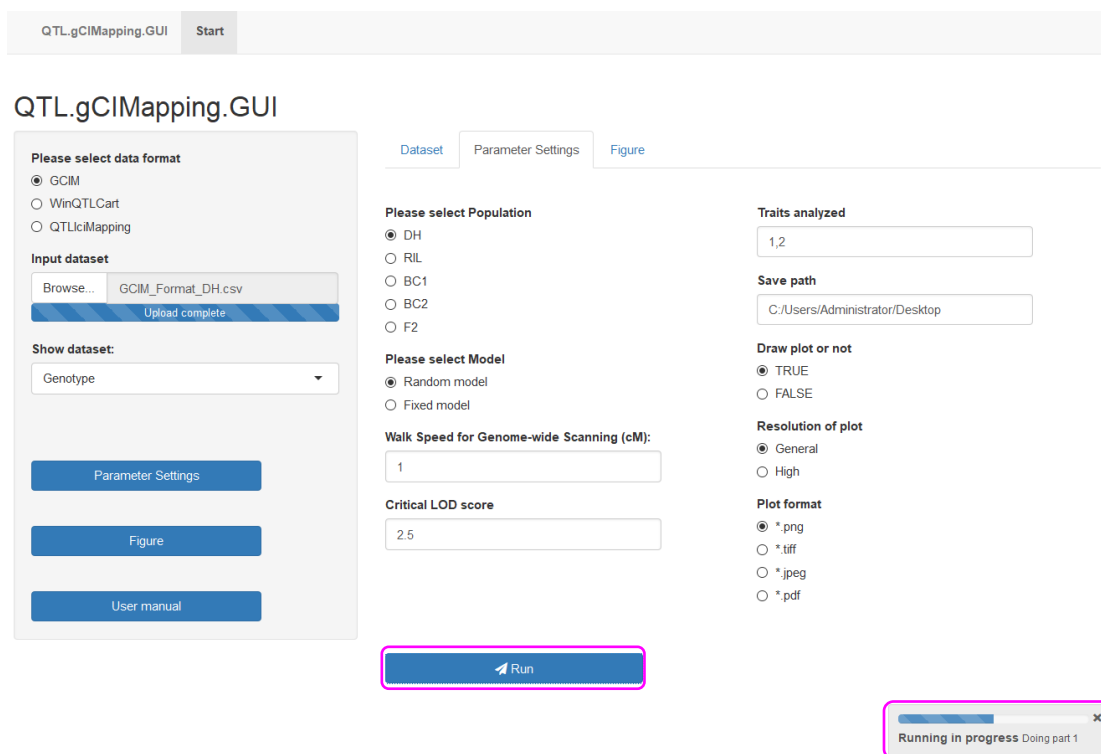


Fig 7. Run the software package QTL.gCIMapping.GUI v2.0

3.5 Re-draw the plot according to your own requirement

When users finish the running, users get the resultforplot.xlsx file. With this file information, users may redraw the curve figure {LOD score or $-\log_{10}(P\text{-value})$ }. With this Figure module, users may set all the figure parameters (Fig 8), including

Legend and tick marks: the size of the words in axis.

LOD line size: the size of the LOD line, the larger the coarse.

Size for $-\log_{10}(P\text{-value})$ curve: the size of $-\log_{10}(P\text{-value})$ curve, the larger the coarse.

Margin space: the space between the figure and the margin of the paper.

Critical LOD score: The critical LOD score for significant QTL.

Before saving this Figure, please set the related parameters: **width** and **height** [with the unit of pixel (px)], **word resolution** [with the unit of 1/72 inch, being pixels per inch (ppi)], and **figure resolution** [with the unit of pixels per inch (ppi)]. Users may set the colors for the LOD line color and $-\log_{10}(P\text{-value})$ curve, with a drop-down option. Use Download plot button to choose a path and to save the Figure, with four frequently used image formats: *.png, *.tiff, *.jpeg and *.pdf (Fig 9).

The screenshot shows the 'Parameter Settings' tab of the 'QTL.gCIMapping.GUI'. The interface is divided into two main sections: 'Please select data format' on the left and 'Genome-wide composite interval mapping (GCIM) figure' on the right. The 'Please select data format' section includes radio buttons for 'GCIM' (selected), 'WinQTLCart', and 'QTL IciMapping'. Below this is an 'Input dataset' section with a 'Browse...' button and a text field containing 'GCIM_Format_DH.csv', followed by an 'Upload complete' button. A 'Show dataset:' dropdown menu is set to 'Genotype'. At the bottom of this section are three buttons: 'Parameter Settings', 'Figure' (highlighted with a pink border), and 'User manual'. The 'Genome-wide composite interval mapping (GCIM) figure' section has two tabs: 'Parameter Settings' (selected and highlighted with a pink border) and 'Draw plot'. Under 'Parameter Settings', there are several configuration options: 'Select resolution of plot' with radio buttons for 'General resolution' (selected), 'High resolution', and 'Set by yourself'; 'Width (px):' with a text input field containing '1500'; 'Height (px):' with a text input field containing '600'; 'Word resolution (1/72 inch, ppi):' with a text input field containing '12'; 'Figure resolution (ppi):' with a text input field containing '72'; 'Critical LOD score:' with a text input field containing '2.5'; 'LOD line size:' with a text input field containing '1.0'; 'Size for $-\log_{10}(P)$ curve:' with a text input field containing '0.5'; 'Legend and tick marks:' with a text input field containing '1.0'; 'Margin space:' with a text input field containing '1.5'; 'Space between tick marks and axis:' with a text input field containing '1.0'; 'Times for max(-log10(P)):' with a text input field containing '1.5'; 'LOD line color:' with a dropdown menu set to 'red'; '-log10(P) curve color1:' with a dropdown menu set to 'gray50'; and '-log10(P) curve color2 (only for F2):' with a dropdown menu set to 'green'.

Figure 8. Parameter settings

QTL.gCIMapping.GUI

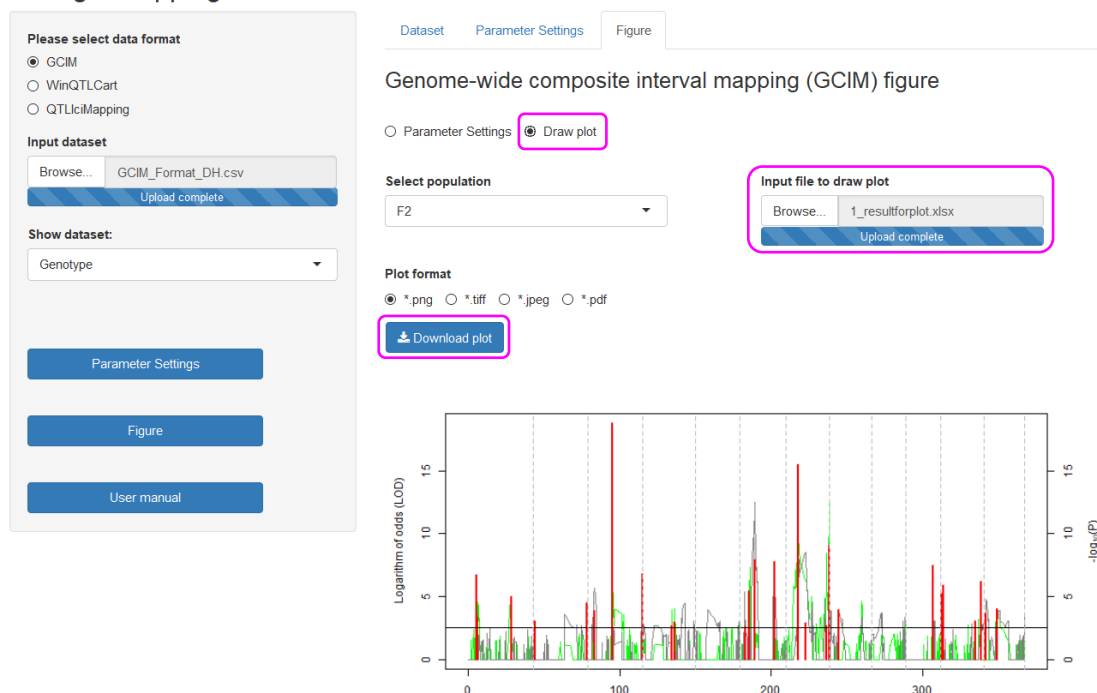


Fig 9. How to draw the LOD score figure in QTL mapping

5. 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.

r² (%): 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 F₂ population, the **Results** file has eleven columns. Trait, Chr, Position (cM), Left_Marker, Right_Marker, Var_Genet, LOD, r² (%), Var_Error and Var_phen are same as those in the above populations. In F₂ population, QTLs include additive (**Effect.a**) and dominant (**Effect.d**) effects.