User Manual for

QTL.gCIMapping

QTL genome-wide Composite Interval Mapping

(version 2.0)

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

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

Last updated on April 20, 2018

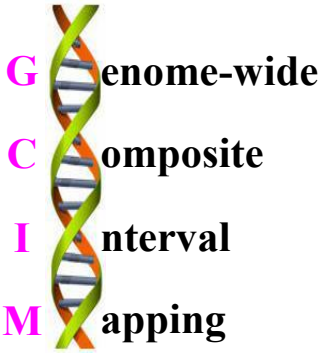
Disclaimer: While extensive testing has been performed by Yuan-Ming Zhang's Lab (Statistical Genomics Lab) at Huazhong Agricultural University, the results are, in general, reliable, correct or appropriate. However, results are not guaranteed for any specific datasets. We strongly recommend that users validate the GCIM results with other software packages, such as Windows QTL Cartographer V2.5_011 (https://brcwebportal.cos.ncsu.edu/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. Submitted

Quantitative Trait Loci



Genome-wide Composite Interval Mapping

Funding

The works were supported by the National Natural Science Foundation of China (31571268, 31701071 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** **G**enome-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.

QTL.gCIMapping v1.0 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 (soyzhang@mail.hzau.edu.cn or soyzhang@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("shiny","qtl","doParallel","foreach","iterators","openxlsx","
MASS","stringr","parcor","data.table"))
```

Offline installation Users should download the below 35 packages from CRAN, github (https://github.com/), or google search:

"cmprsk","corpcor","data.table","digest","doParallel","Epi","etm","fdrtool","foreach", "GeneNet","glmnet","htmltools","httpuv","iterators","jsonlite","longitudinal","magritt r","MASS","mime","numDeriv","openxlsx","parcor","plyr","ppls","qtl","R6","Rcpp", "shiny","sourcetools","stringi","stringr","testthat","utf8","xtable","zoo"

Then, install them offline (under the R environment, select all the 35 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. 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 while the first row shows the individual name. For the genotypes of each marker, the coding criteria are shown as Table 1. The phenotype and covariate information are followed the marker genotypes, and each covariate or trait are listed on one row. On each row, the first column is empty followed by "**trait1**", "real trait name", and "phenotypic values for all the individuals". If multiple traits exist, more rows will be added. If covariates exist, all the information for the covariates 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 2.

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

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

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, its 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 dataset fileICIMcov format 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 covariate values 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. Operation process

3.1 The graphical interface of QTL.gCIMapping

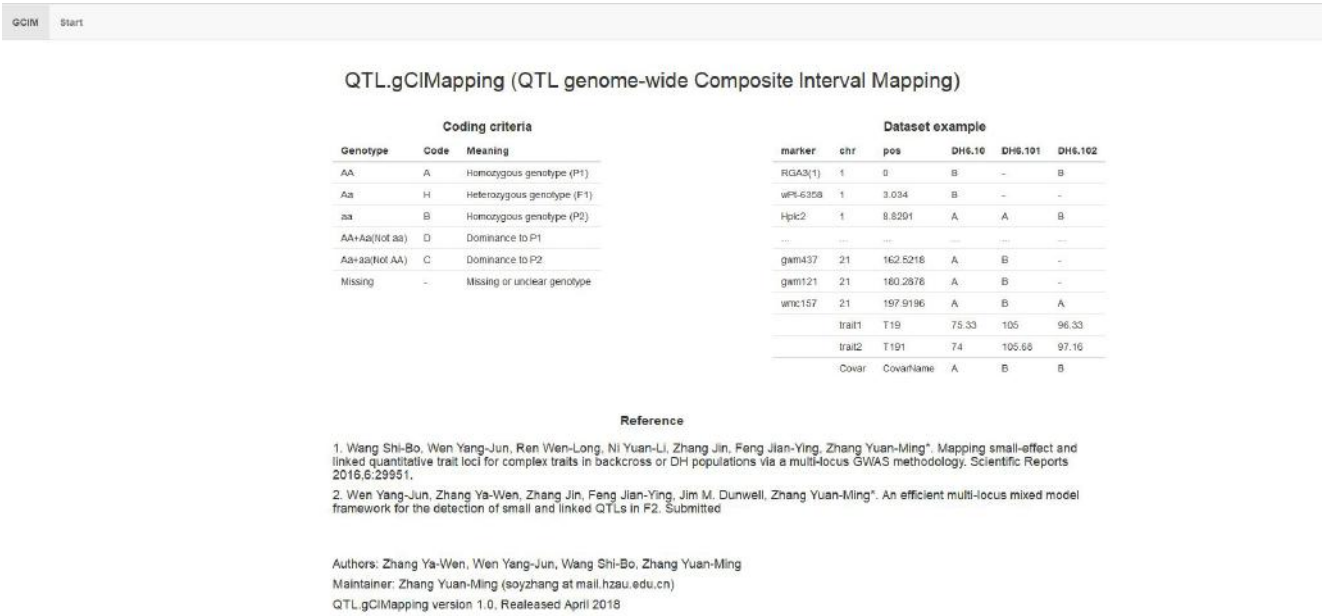


Figure 1. The Graphical User Interface of QTL.gCIMapping

3.2 Input dataset

Users must upload the dataset files with three formats (**Figs 2 to 4**). If users select the QTLIciMapping format and the covariate exists in the dataset, users should upload the covariate matrix (**Fig 5**).

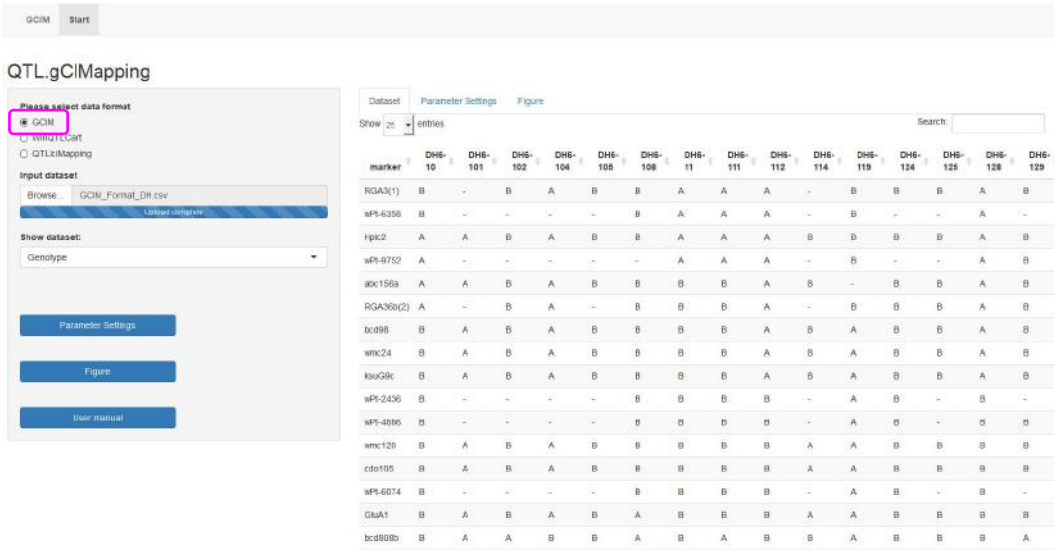


Fig 2. Dataset GCIM format

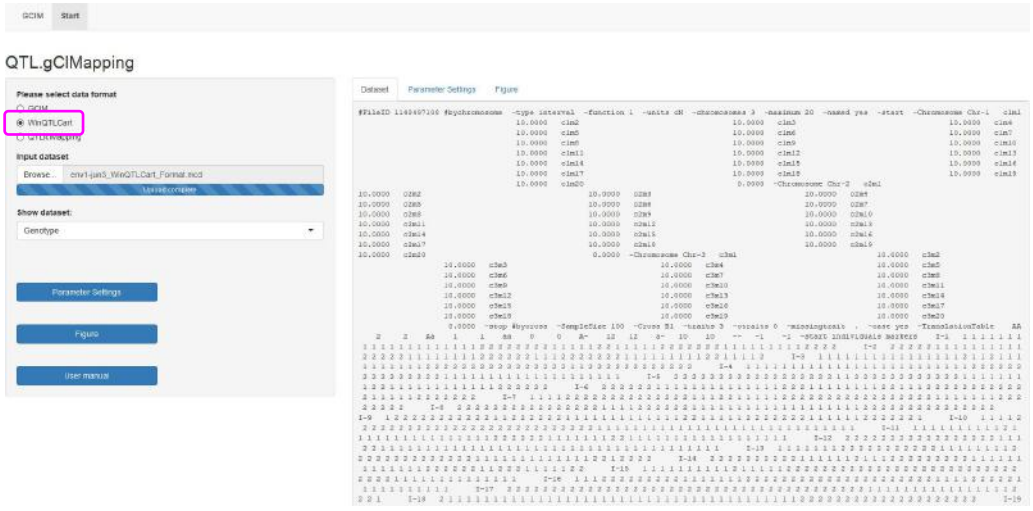


Fig 3. Dataset WinQTLCart format

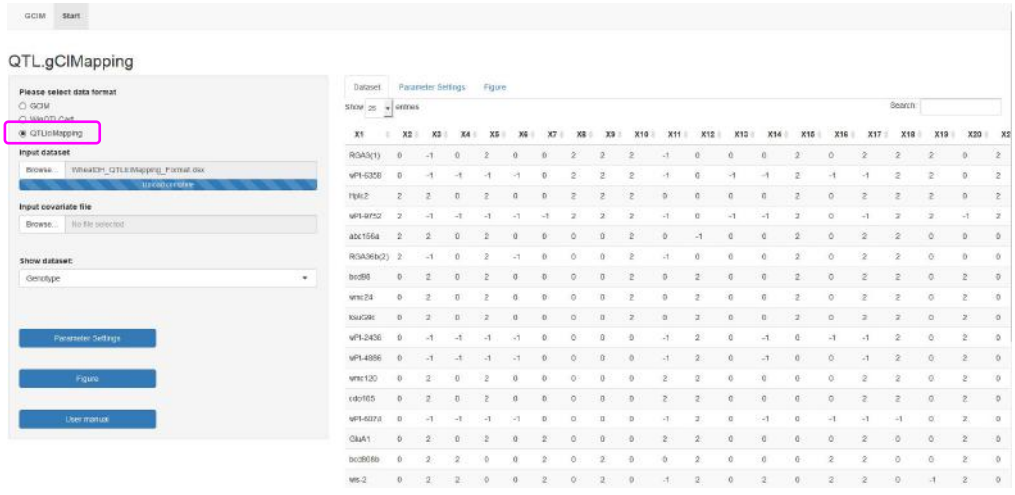


Fig 4. Dataset QTLIciMapping format

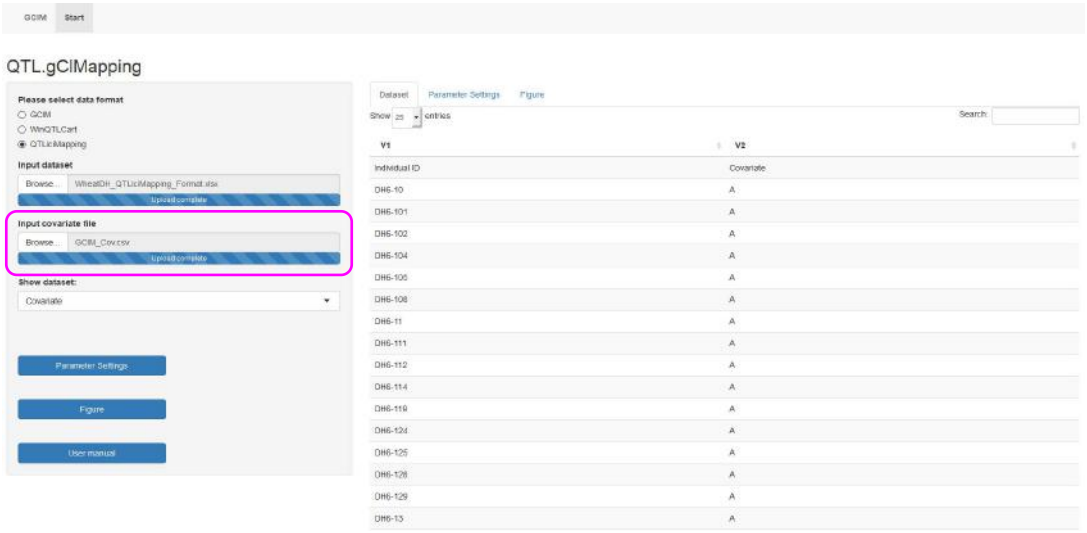


Fig 5. Covariate input in the QTLIciMapping dataset format

3.3 Parameter settings (Fig 6)

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

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 F_2 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".

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.

Select trait ID: "**2:2**" indicates the analyses from the second trait, and "**2:4**" indicates the analyses from the second to fourth traits.

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

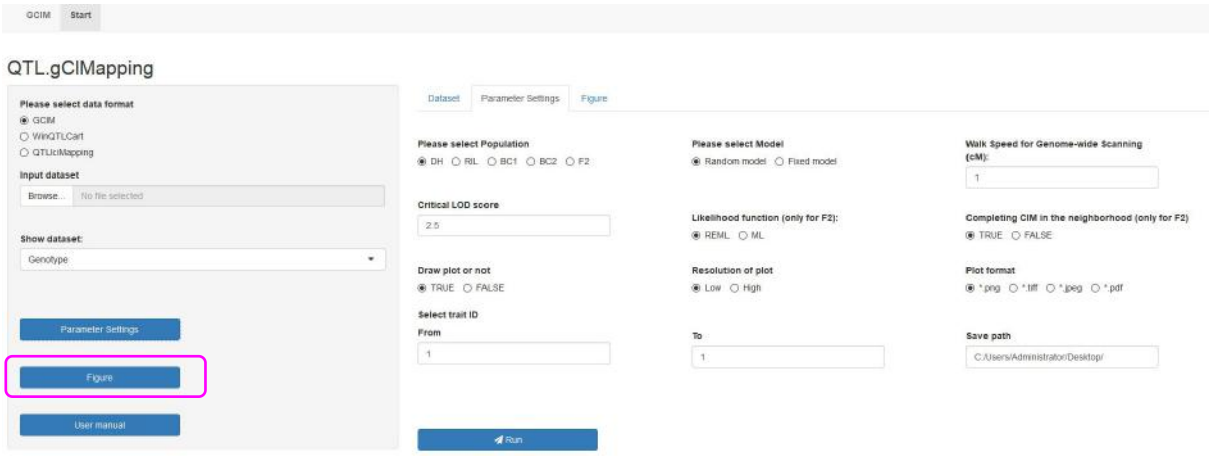


Fig 6. Parameter setting in the mapping of QTL for quantitative traits

3.4 Run the software

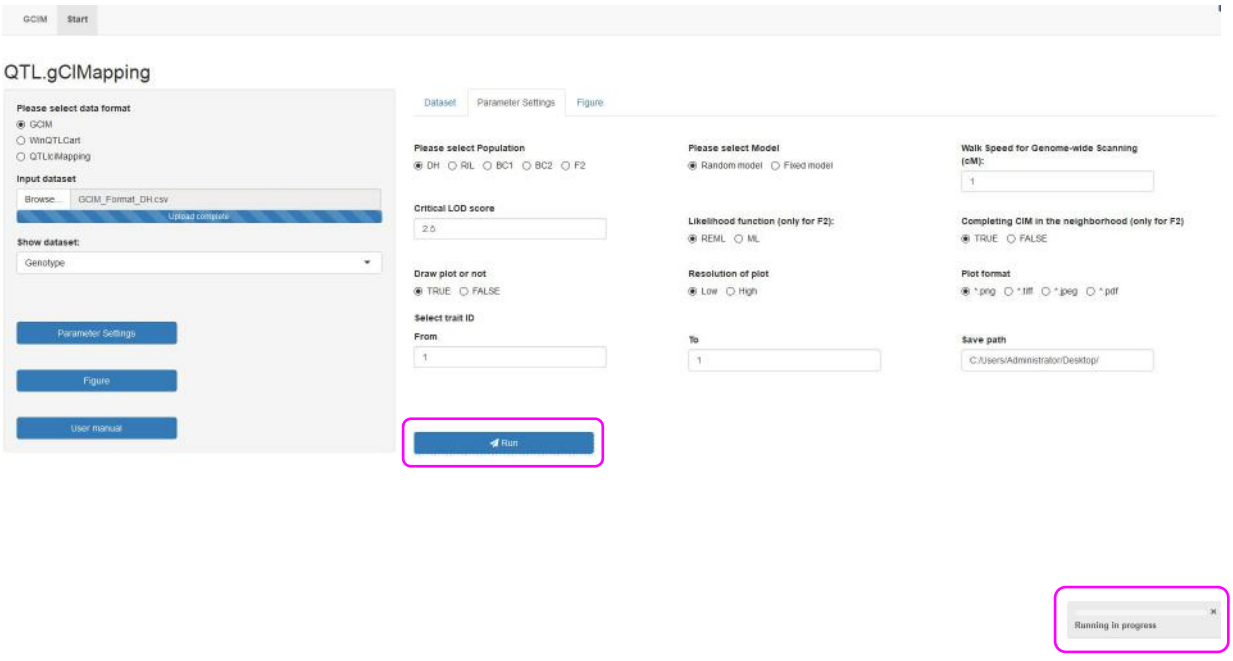


Fig 7. Run the software QTL.gCIMapping

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-value) curve: the size of $-\log_{10}$(P-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-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**).

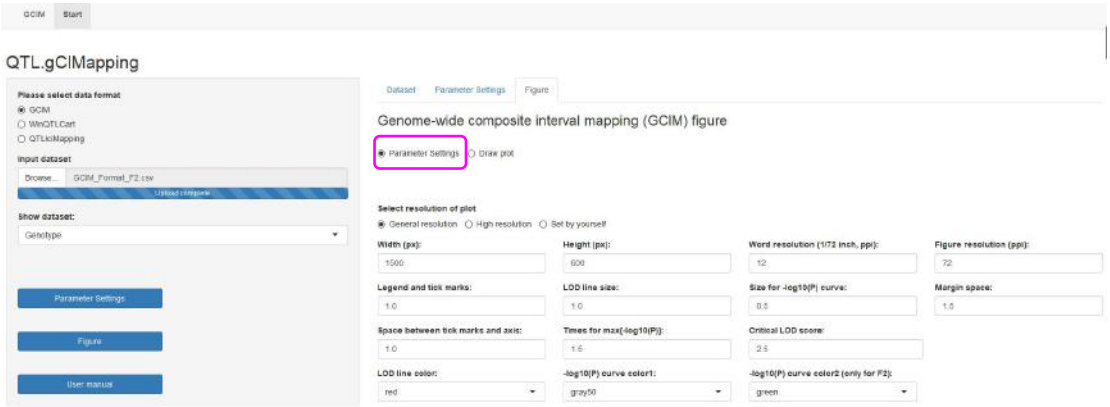


Figure 8. Parameter settings

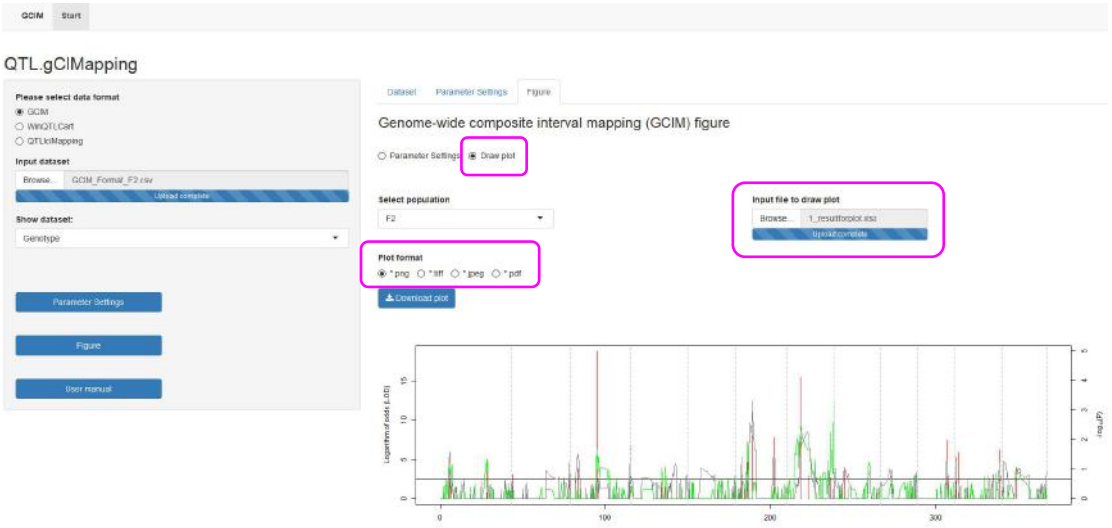


Fig 9. How to draw the QTL mapping figure

4. 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. The different columns are as follows.

Effect.a and **Effect.d:** Additive and dominant effects for significant QTL.