User Manual for

S E A

a **R** software package of **SE**gregation **A**nalysis

for quantitative traits in plants

(**version 2.0**)

Wang Jing-Tian, Zhang Ya-Wen and

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

Last updated on February 24, 2021

Disclaimer: While extensive testing and Monte Carlo simulation studies have been performed by Yuan-Ming Zhang's Lab (Statistical Genomics Lab) at Crop Information Center, College of Plant Science and Technology, Huazhong Agricultural University. The results are, in general, reliable, correct or appropriate. However, results are not guaranteed for any specific datasets. You could consult us if you have any questions.

Download website:
https://cran.r-project.org/web/packages/SEA/index.html

SE gregation
A nalysis

This work was supported by Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2014RC020).

1 Introduction

1.1 Why SEA?

Quantitative traits are controlled by a few major genes and a series of polygenes. Major genes can be individually identified and polygenes are collectively detected. This is mixed major-gene plus polygenes inheritance model. This best model for complex trait can be obtained from the comparison between frequent distribution in real data and theoretical distribution of the above mixed inheritance model. This method, named "segregation analysis (SEA)", has been widely used in China since 1990s. To popularize this approach, we open R software SEA with interactive graphic user interface (GUI) under the framework of the RStudio-1.4.1103 platform. In the SEA, the packages **kolmim** and **KScorrect** were used to conduct the Lilliefors-corrected Kolmogorov-Smirnoff test; **doParallel** is used to carry out parallel computation; **data.table** is used to read and write the file quickly; **MASS** is used to perform parameter estimation from distribution parameters to the first-order genetic parameters.

The current software SEA v2.0 includes fourteen types of populations: 1) SEA-F_2 (F_2); 2) SEA-F_3 ($F_{2:3}$); 3) SEA-DH (DH or RIL); 4) SEA-BIL (BIL); 5) SEA-BC (B_1&B_2); 6) SEA-BCF ($B_{1:2}$ & $B_{2:2}$); 7) SEA-G4F_2 (P_1, P_2, F_1 and F_2); 8) SEA-G4F_3 (P_1, P_2, F_1 and $F_{2:3}$); 9) SEA-G3DH (P_1, P_2 and DH); 10) SEA-G5BC (P_1, P_2, F_1, B_1 and B_2); 11) SEA-G5BCF (P_1, P_2, F_1, $B_{1:2}$ and $B_{2:2}$); 12) SEA-G5 (P_1, P_2, F_1, F_2 and $F_{2:3}$); 13) SEA-G6 (P_1, P_2, F_1, F_2, B_1 and B_2); and 14) SEA-G6F (P_1, P_2, F_1, $F_{2:3}$, $B_{1:2}$ and $B_{2:2}$).

SEA is able to work on the Windows, Linux (desktop) and MacOS platforms.

1.2 Getting started

SEA is a package that runs in the R software environment, which can be freely downloaded from https://cran.r-project.org/web/packages/SEA/index.html, or request from the maintainer, Dr Yuan-Ming Zhang at Crop Information Center, College of

Plane Science and Technology, Huazhong Agricultural University (soyzhang@hotmail.com; soyzhang@mail.hzau.edu.cn).

1.2.1 One-Click installation

Within R environment, the SEA software can be installed directly using the below command:

```
install.packages(pkgs="SEA")
```

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 SEA package can be installed online, using the below command:

```
install.packages(pkgs=c("shiny","MASS","KScorrect","kolmim","data.table","doParallel"))
```

Offline install The following R packages are needed: bit64, curl, data.table, digest, doParallel, foreach, htmltools, httpuv, iterators, jsonlite, knitr, kolmim, KScorrect, MASS, mclust, mime, nanotime, R6, Rcpp, reshape2, shiny, sourcetools, testthat, xtable, xts, zoo, which can be downloaded from CRAN (https://cran.r-project.org/) on your desktop. Open R GUI, select "Packages"—"Install package(s) from local files…" (**Figure 1.1**), then find R packages (only *.zip and *.tar.gz files available) above and install them in order, as some depend on others.

1.2.2.2 Install SEA

Download the **SEA** package (only *.zip and *.tar.gz files available) on your desktop ahead, method of installing SEA is the same as §1.2.2.1 Within R environment, launch the SEA by command: library(SEA), To restart the GUI, the command SEA() can be issued (**Figure 1.2**).

1.2.2.3 Run and re-run SEA

Users may run the below two commands in R environment

```
library(SEA)
SEA()
```

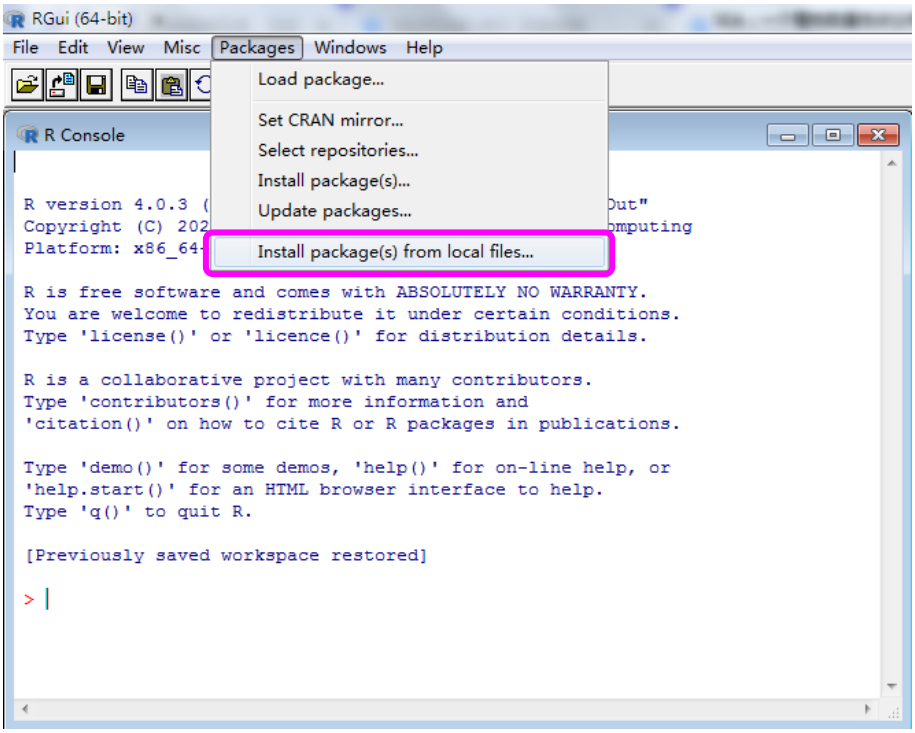

Figure 1.1 Offline installation of R software package.

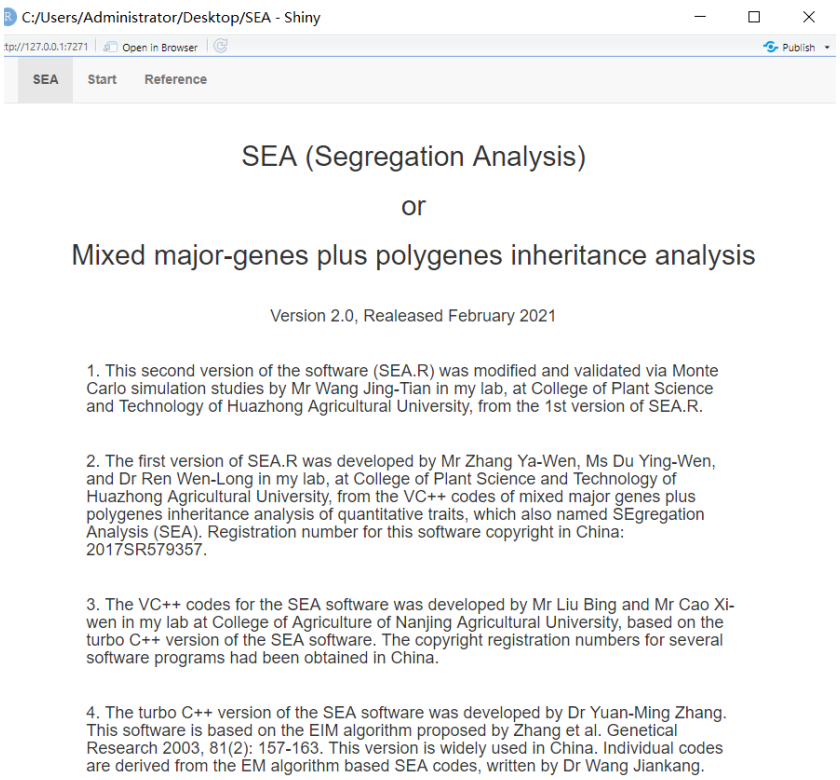

Figure 1.2 Screenshot of SEA.GUI v2.0 package

2 Phenotypic datasets and genetic models

2.1 Population types

Bi-parental segregation populations in the software package SEA v2.0

Population	Abbreviation	Population	Abbreviation
F_2	SEA-F2	P_1, F_1, P_2 and $F_{2:3}$	SEA-G4F3
$F_{2:3}$	SEA-F3	P_1, P_2 and DH	SEA-G3DH
DH or RIL	SEA-DH	P_1, F_1, P_2, B_1 and B_2	SEA-G5BC
BIL	SEA-BIL	P_1, F_1, P_2, $B_{1:2}$ and $B_{2:2}$	SEA-G5BCF
B_1 and B_2	SEA-BC	P_1, F_1, P_2, F_2 and $F_{2:3}$	SEA-G5
$B_{1:2}$ and $B_{2:2}$	SEA-BCF	P_1, F_1, P_2, F_2, B_1 and B_2	SEA-G6
P_1, F_1, $P_2$and F_2	SEA-G4F2	P_1, F_1, P_2, $F_{2:3}$, $B_{1:2}$ and $B_{2:2}$	SEA-G6F

Note: DH: doubled haploid; RIL: recombinant inbred line; BIL: backcross inbred line; $B_1 = F_1 \times P_1$; $B_2 = F_1 \times P_2$; $B_{1:2}$ and $B_{2:2}$: families derived from B_1 and B_2, respectively.

Citation:

Populations	References
SEA-F_2	Wang et al. ***Acta Genetica Sinica*** 1997, 24(5):432-440[4] Zhang et al. ***Journal of Biomathematics*** 2000, 15(3):358-366[7]
SEA-F_3	Zhang et al. ***Hereditas (Beijing)*** 2001, 23(4):329-776[10] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]
SEA-DH	Zhang et al. ***Hereditas (Beijing)*** 2001, 23(5):467-470[12]
SEA-BIL	Wang et al. ***Acta Agron Sin*** 2013, 39(2):198-206[5]
SEA-BC	Zhang et al. ***Journal of Biomathematics*** 2000, 15(3):358-366[7]
SEA-BCF	Zhang et al. ***Hereditas (Beijing)*** 2001, 23(4):329-776[5] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]
SEA-G4F2	Zhang et al. ***Journal of Southwest Agricultural University*** 2000, 42(1):6-9[11]
SEA-G4F3	Zhang et al. ***Journal of Southwest Agricultural University*** 2000, 42(1):6-9[11] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]
SEA-G3DH	Zhang et al. ***Hereditas (Beijing)*** 2001, 23(5):467-470[12]

SEA-G5BC	Zhang et al. ***Acta Agron Sin*** 2000, 26(6):699-706[13]
SEA-G5BCF	Zhang et al. ***Acta Agron Sin*** 2000, 26(6):699-706[13] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]
SEA-G5	Wang et al. ***Acta Agron Sin*** 1998, 24(6):651-659[6] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]
SEA-G6	Gai et al. ***Theor Appl Genet*** 1998, 97(7): 1162-1168[2] Gai et al. ***Acta Agron Sin*** 2000, 26(4):385-391[3]
SEA-G6F	Zhang et al. ***Acta Agron Sin*** 2001, 27(6):787-793[14] Zhang et al. ***Genetical Research*** 2003, 81(2): 157-163[8]

Note: The above references are listed in **References**.

2.2 Phenotypic dataset format

The Phenotypic file should be a *.csv format file, the phenotypic observations for all the populations are included in one *.csv file, and all the observations for each population are listed in one same column (**Table 2.1**). In each column, the first element must be population type, such as “F2”, “P1” (**Table 2.2**). This dataset can be uploaded into the software **SEA** by clicking the button “Browse” (**Figure 2.1**).

Table 2.1 The phenotypic file (*.csv)

P1	P2	F1	F2	B1	B2
56.33	74.21	62.2	55.19	60.83	58.79
58.96	85.3	54.02	57.64	57.84	64.58
62.93	79.95	52.72	53.24	56.45	62.86
55.97	82.56	55.47	50.58	55.06	60.96
59.28	81.12	51.4	47.03	53.64	70.42
59.45	75.86	55.69	58.71	56.89	56.8
65.56	82.47	53.08	54.78	67.26	51.23
57.67	77.58	58.81	48.04	57.72	74.88
62.68	76.81	51.85	53.71	54.9	67.86

Table 2.2 The column name in phenotypic file (*.csv)

Population	Column name	Population	Column name
F_2	F2	G4F3	P1, F1, P2, F23
$F_{2:3}$	F23	G3DH	P1, P2, DH
DH or RIL	DH	G5BC	P1, F1, P2, B1, B2
BIL	BIL	G5BCF	P1, F1, P2, B12, B22
BC	B1, B2	G5	P1, F1, P2, F2, F23
BCF	B12, B22	G6	P1, F1, P2, F2, B1, B2
G4F2	P1, F1, P2, F2	G6F	P1, F1, P2, F23, B12, B22

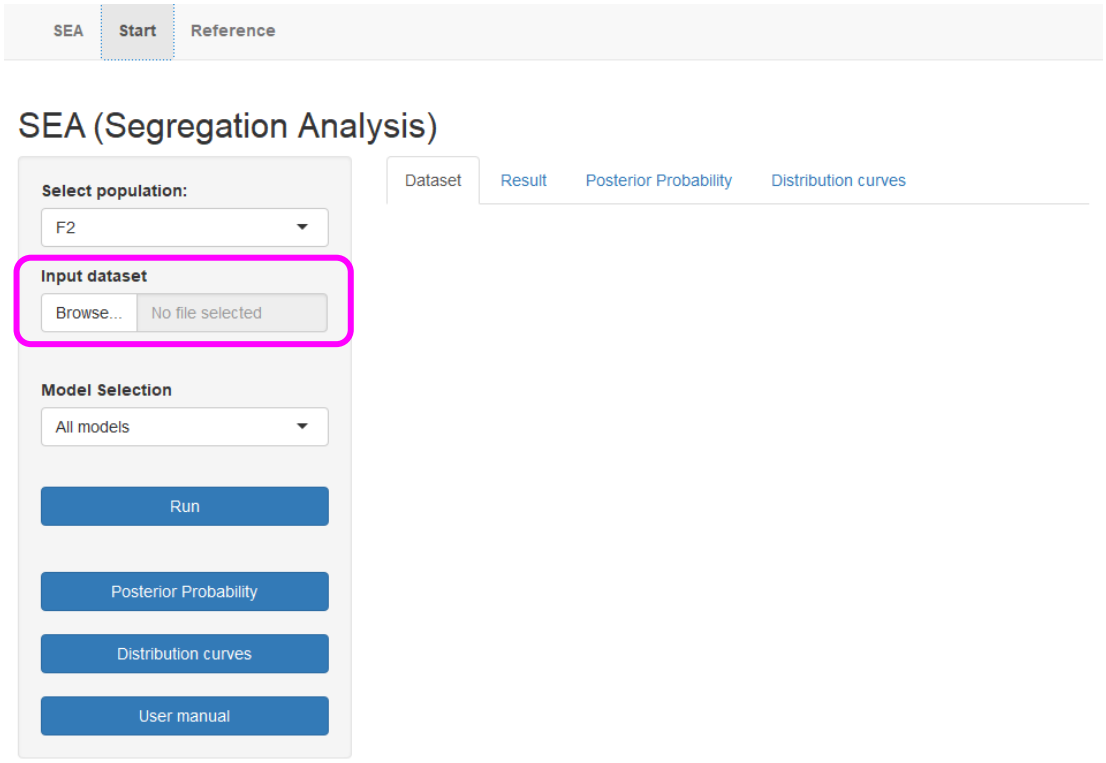

Figure 2.1 The interface of data input

2.3 Genetic models and meanings of model codes

Table 2.3 Genetic models in the joint segregation analysis of the five generations of P_1, F_1, P_2, F_2 and $F_{2:3}$

Class	Major gene	Polygenes	Model code	
			Only major gene	Mixed major gene & polygenes
Polygenes	-	Additive-dominant-epistasis, *[d], [h], [i], [j], [l]*	-	PG-ADI
	-	Additive-dominant, *[d], [h]*	-	PG-AD
A major gene	Additive-dominant, *d, h*	Additive-dominant-epistasis, *[d], [h], [i], [j], [l]*	1MG-AD	MX1-AD-ADI
	Additive-dominant, *d, h*	Additive-dominant, *[d], [h]*	1MG-AD	MX1-AD-AD
	Additive, *d (h=0)*	Additive-dominant, *[d], [h]*	1MG-A	MX1-A-AD
	Completely dominant, *d (h=d)*	Additive-dominant, *[d], [h]*	1MG-CD	MX1-CD-AD
	Completely negative dominant, *d (h=-d)*	Additive-dominant, *[d], [h]*	1MG-NCD	MX1-NCD-AD
Two major genes	Additive-dominant-epistasis, d_a, d_b, h_a, h_b, i, j_{ab}, j_{ba}, l	Additive-dominant-epistasis, *[d], [h], [i], [j], [l]*	2MG-ADI	MX2-ADI-ADI
	Additive-dominant-epistasis, d_a, d_b, h_a, h_b, i, j_{ab}, j_{ba}, l	Additive-dominant, *[d], [h]*	2MG-ADI	MX2-ADI-AD
	Additive-dominant, d_a, d_b, h_a, h_b, $i=j_{ab}=j_{ba}$, l	Additive-dominant, *[d], [h]*	2MG-AD	MX2-AD-AD
	Additive, d_a, d_b, $h_a=h_b=0$	Additive-dominant, *[d], [h]*	2MG-A	MX2-A-AD
	Equally additive, $d(=d_a=d_b, h_a=h_b=0)$	Additive-dominant, *[d], [h]*	2MG-EA	MX2-EA-AD
	Completely dominant, $d_a=h_a$, $d_b=h_b$	Additive-dominant, *[d], [h]*	2MG-CD	MX2-CD-AD
	Equally dominant, $d=d_a=h_a=d_b=h_b$	Additive-dominant, *[d], [h]*	2MG-EAD	MX2-EAD-AD

Meanings of model codes:

1) **1MG ~ 4MG:** One to four major genes
2) **MX1 ~ MX3:** One to three major genes plus polygenes
3) **A:** additive
4) **AD:** additive-dominance
5) **ADI:** additive-dominance-epistasis
6) **CD:** $d = h$, completely dominance
7) **NCD:** $d = -h$, negatively completely dominance
8) **EA:** $d_a = d_b$, equally additive
9) **EAD:** $d_a = d_b = h_a = h_b = d$, equally additive-dominance

10) **AI:** Additive-epistasis

11) **CEA:** $d_a = d_b = d_c = d$, completely equally additive

12) **PEA:** $d_a = d_b = d_1$, $d_c = d_2$, partially equally additive

13) Six types of interactions between two major genes in F_2 are as follows:

CE: 9 *A_B_*:7 (3*A_bb*+3*aaB_* + 1*aabb*), complementary effect.

AE: 9 *A_B_*:6 (3*A_bb*+3*aaB_*) **:** 1 *aabb*, additive effect.

DE: 15 (9*A_B_*+ 3*A_bb*+3*aaB_*) **:** 1 *aabb*, duplicate effect (The effect of *A* or *B* allele on the trait phenotype).

ED: 12 (9*A_B_* + 3*A_bb*)**:** 3 *aaB_***:** 1 *aabb*, epistatic dominance between two pairs of major genes (The effect of *A* allele on *Bb* locus).

RE: 9 *A_B_***:** 3 *A_bb***:** 4 (3*aaB_*+ 1*aabb*), epistatic recessiveness of *aa* on *Bb* locus.

IE: 13 (9*A_B_***:** 3*aaB*+ 1*aabb*)**:** 3 *A_bb,* inhibiting effect of *B* allele on *Aa* locus.

In DH population, the above DE and IE models are indistinguishable, and same results are obtained from the two models, so we delete the IE model.

3 Running and Results

3.1 Start the software

Select the option "Start" (for example), the following interface will appear.

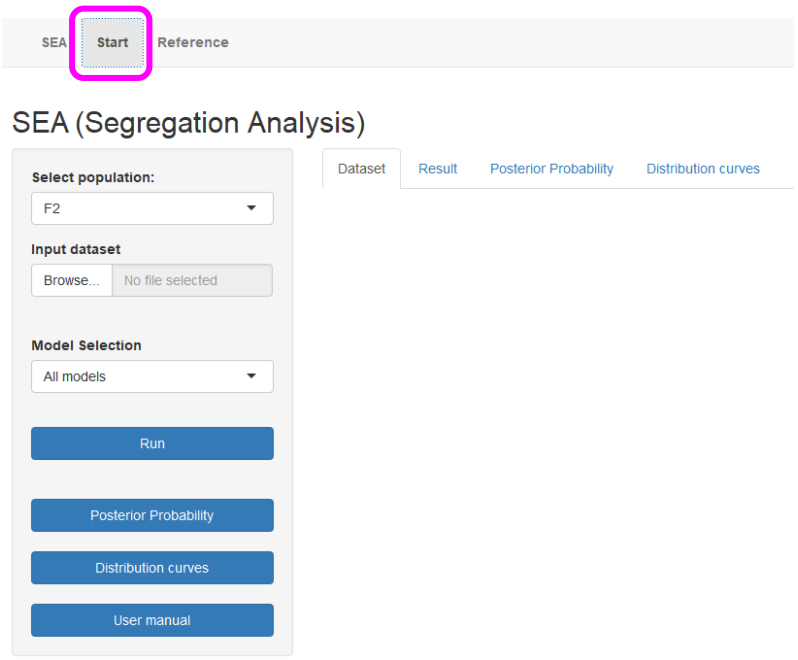

Figure 3.1 Screenshot of SEA.GUI v2.0

3.2 Select population

Use the drop down menu to select which population will be analyzed (Figure 3.2).

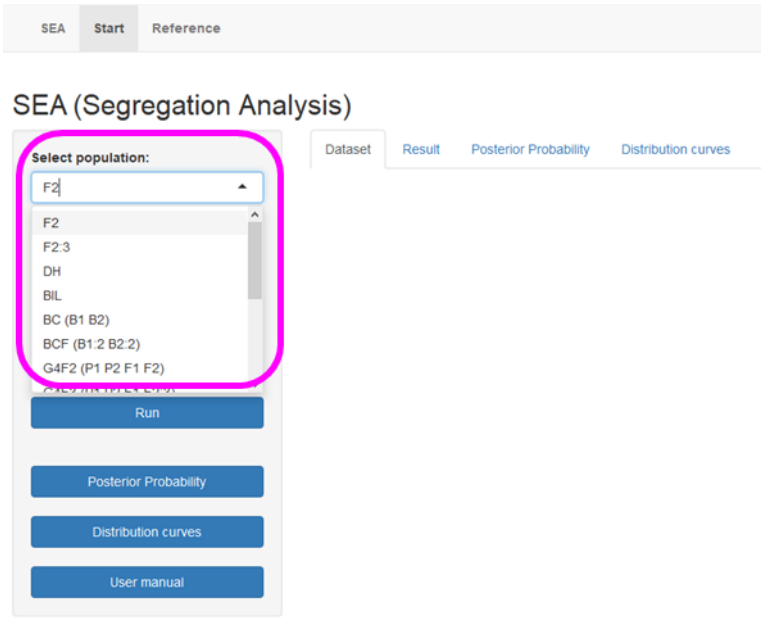

Figure 3.2 Select population

3.3 Input dataset

Use the **Browse** button to input dataset files. Once the file is successfully uploaded, the result will be shown as **Figure 3.3.**

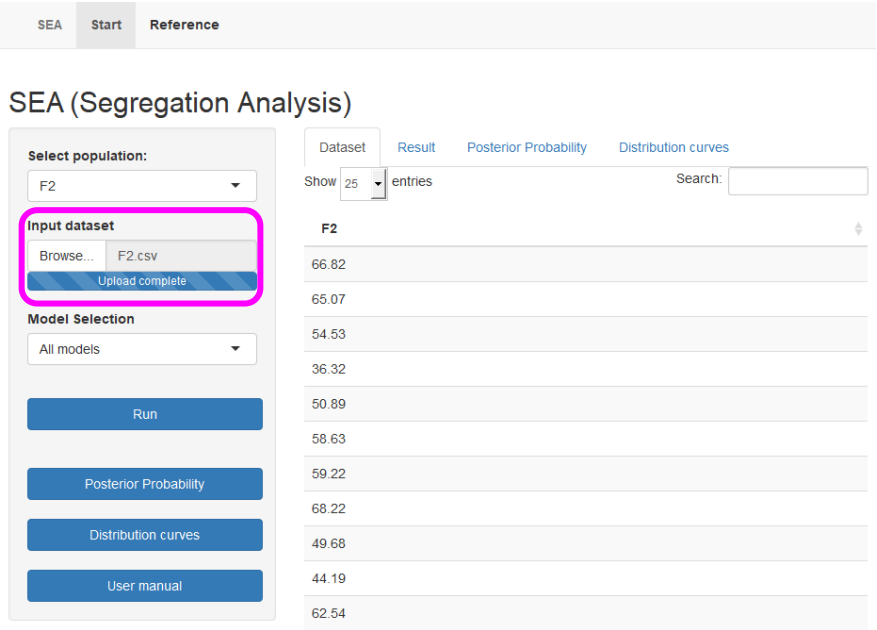

Figure 3.3 The Input Dataset module for SEA v2.0

3.4 Parameter settings

Use the Model Selection option menu to choose model before running the program, **"All models",** which means to run all the models under this population (**Figure 3.4**). In such population like: F3 (F2:3), BCF (B1:2, B2:2), G4F3 (P1, P2, F1, F2:3), G3DH (P1, P2, DH), G5BCF (P1, P2, F1, B1:2, B2:2), G5 (P1, P2, F1, F2, F2:3), G6F (P1, P2, F1, F2:3, B1:2, B2:2), you need to set the parameter "**No. of plants measured in each family**" (1 is default value) as well (**Figure 3.5**). In BIL population, you also need to select The **BIL type**: BIL1 (F1 × P1) and BIL2 (F1 × P2) (**Figure 3.6**).

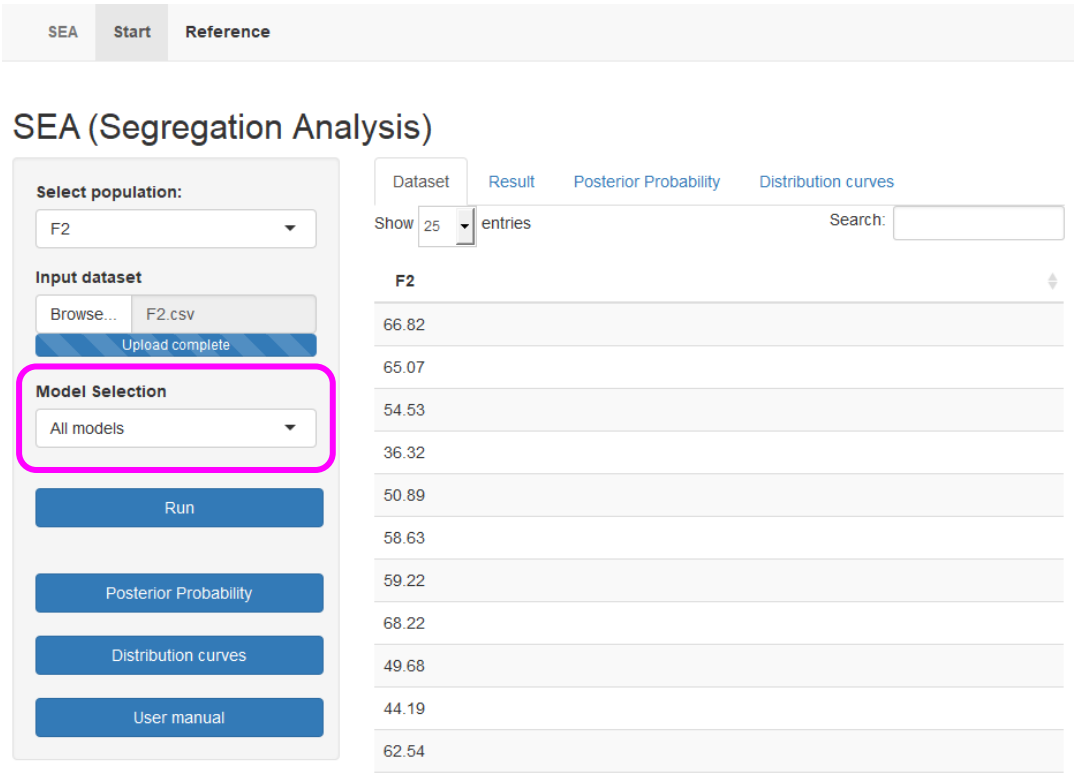

Figure 3.4 Model selection

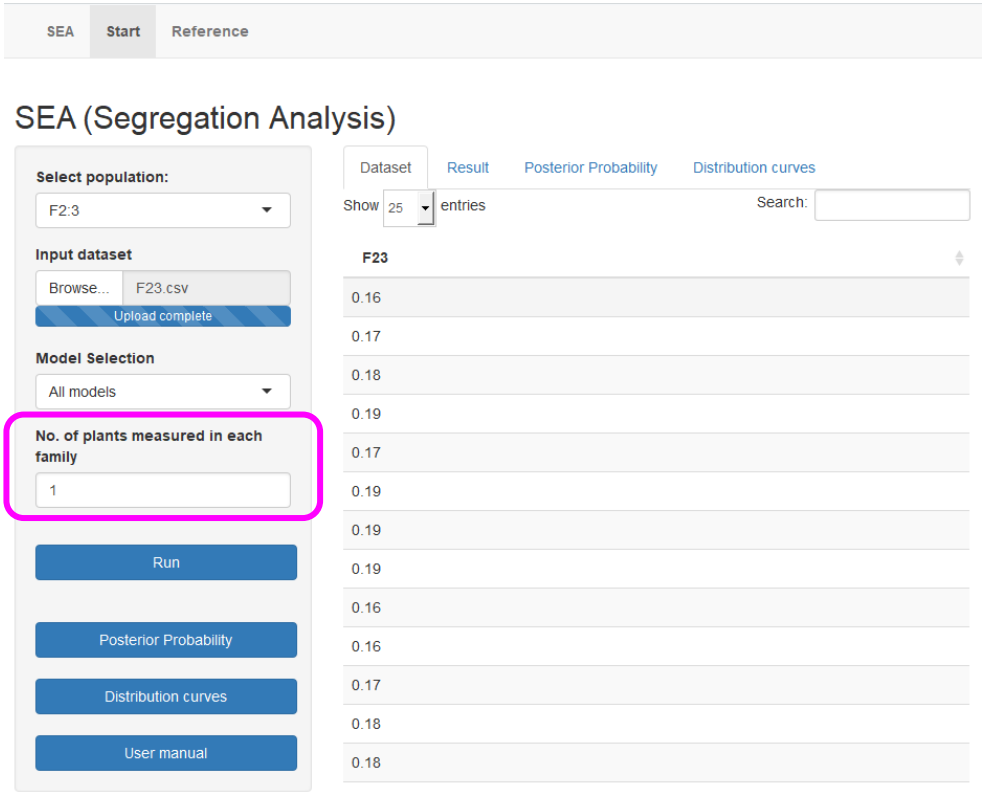

Figure 3.5 Set "No. of plants measured in each family"

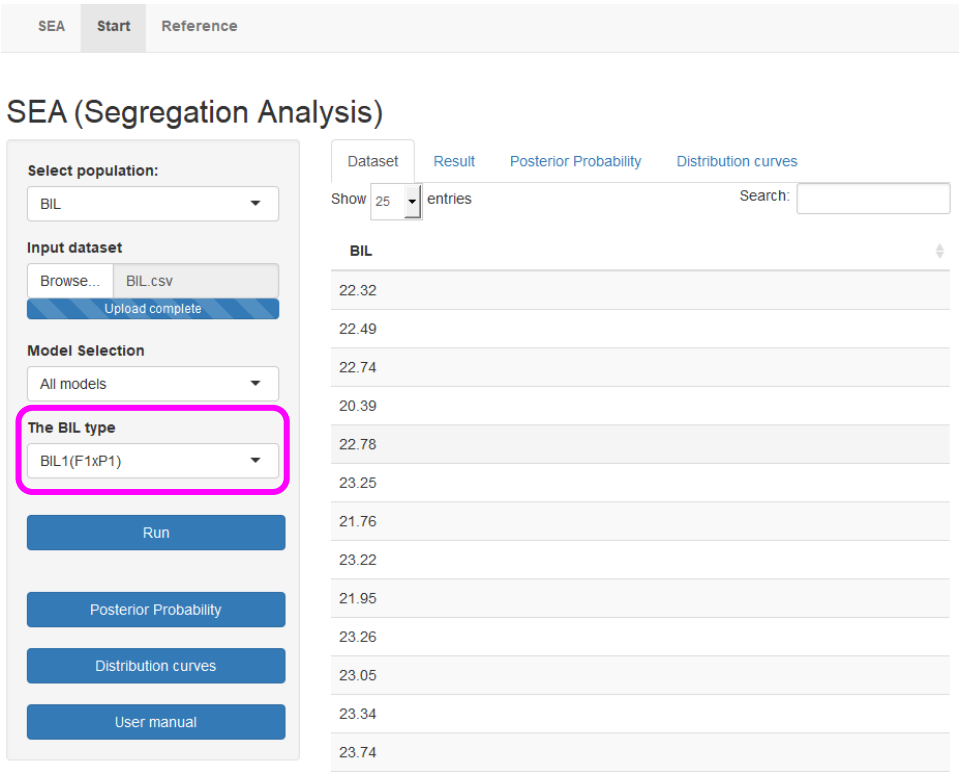

Figure 3.6 Select population types for BIL

3.5 Run the program

Use the **Run** Button to run the program, the result will be shown in the result module, and then it can be download use the **Download result** Button (**Figure 3.7**).

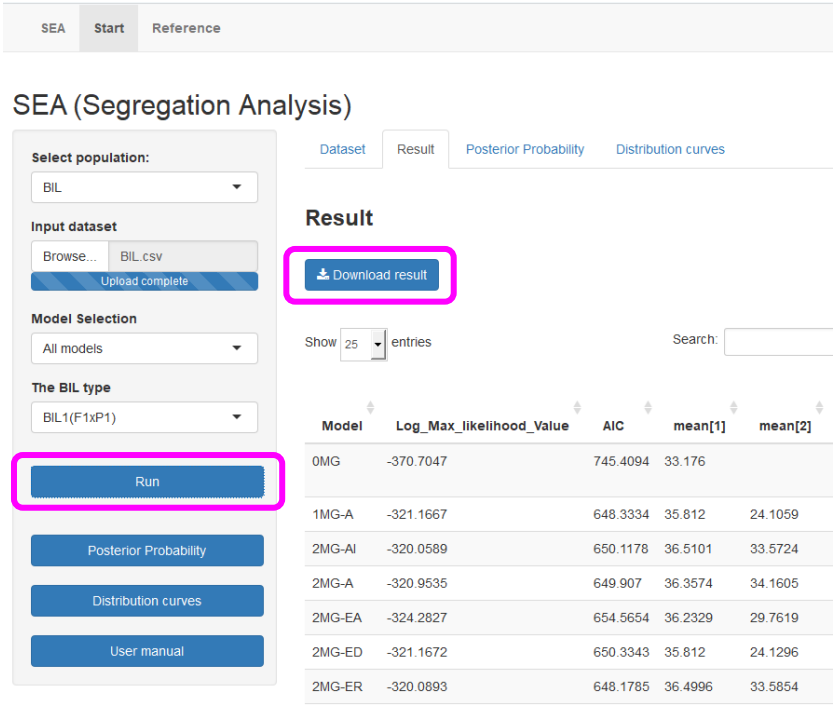

Figure 3.7 Run the SEA.GUI v2.0 programme

Descriptions in Results file

Descriptions in Results interface in §**3.6** are as follows:

1) **Model:** genetic model.
2) **Log_Max_likelihood_Value:** $\log_{10}$(maximum likelihood function value).
3) **AIC:** $AIC = -2L(Y|\Theta) + 2k$; $L(Y|\Theta)$: logarithm likelihood function, Θ : parameter in the logarithm likelihood function, *k*: the number of independent parameters in the model.
4) **mean:** mean for each component distribution in one segregation population.
5) **Proportion:** genotypic proportions or proportions of component distributions in one segregation population, which are calculated from posterior probability.
6) ***m*, *d*, *h*:**total average, additive effect, dominant effect for major gene.
7) ***i*, j_{ab}, j_{ba}, *l*:** additive × additive, additive × dominance, dominance × additive, and dominance × dominance interaction effects between two major genes.
8) **[*d*], [*h*]:** additive effect, dominance effect for polygenes; **[*i*], [*j*], [*l*]:** additive × additive, additive × dominance (or dominance × additive) and dominance × dominance interaction effects for polygenes.
9) **Major-Gene Var, Polygenes Var: genetic variances for** major genes and polygenes (second-order genetic parameter), respectively.
10) **Heritability (%):** the proportion of genetic variance in total phenotypic variance
11) **U* square-** (**U_1^2, U_2^2** and **U_3^2**; or **$U_1^2\_P_1$**), **nW square-** (**${}_nW^2$** or **${}_nW^2\_F_1$**), **Dn-** (**D_n** or **$D_n\_F_2$**): uniform, Smirnov and Kolmogorov statistics. *: numbers 1 to 3; **: population notion, i.e., P_1, F_1, P_2, F_2 and F_3.
12) **P(U1 square-*), P(U2 square-*), P(U3 square-*), P(nW square-*), P(Dn-*):** *P*-values of the above corresponding statistics.

3.6 Posterior probability

Posterior probability module shows the posterior probability of the *i*th individual (line) in *j*th major-gene genotype for the optimal model, implemented by the **Calculate Posterior Probability** button, and downloaded the results by the **Save Posterior Probability** button (**Figure 3.8**). If only one major-gene genotype is involved in this model, **Warning** notices "No posterior probability!" will be appeared (**Figure 3.9**).

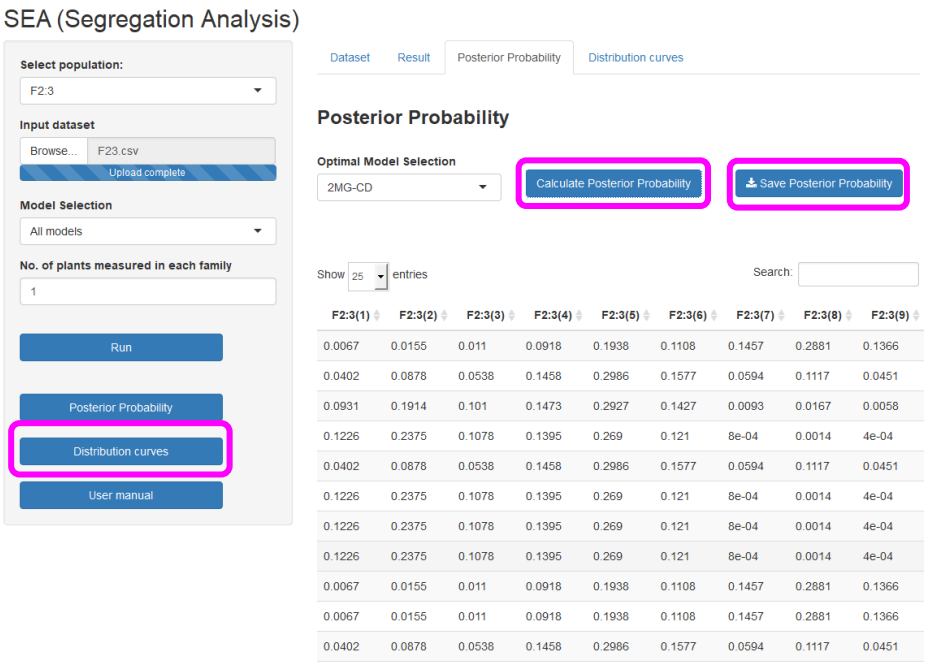

Figure 3.8 Posterior Probability module

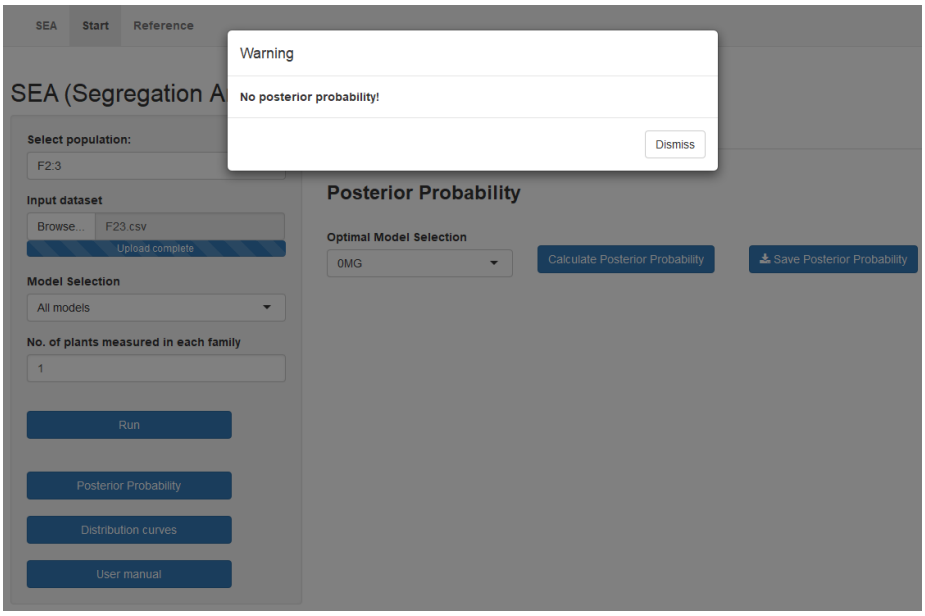

Figure 3.9 The model with only one major-gene genotype

3.7 Distribution curve

In Parameter Settings module, users can use **Draw distribution curves** button to preview the plot (**Figure3.10**). In download plot module, users can select general or high resolution plot to download, and use **Save Distribution Curves** button to save the results as ***.png, *.tiff, *.jpeg,** and ***.pdf** files in your selected pathway (**Figure3.11**).

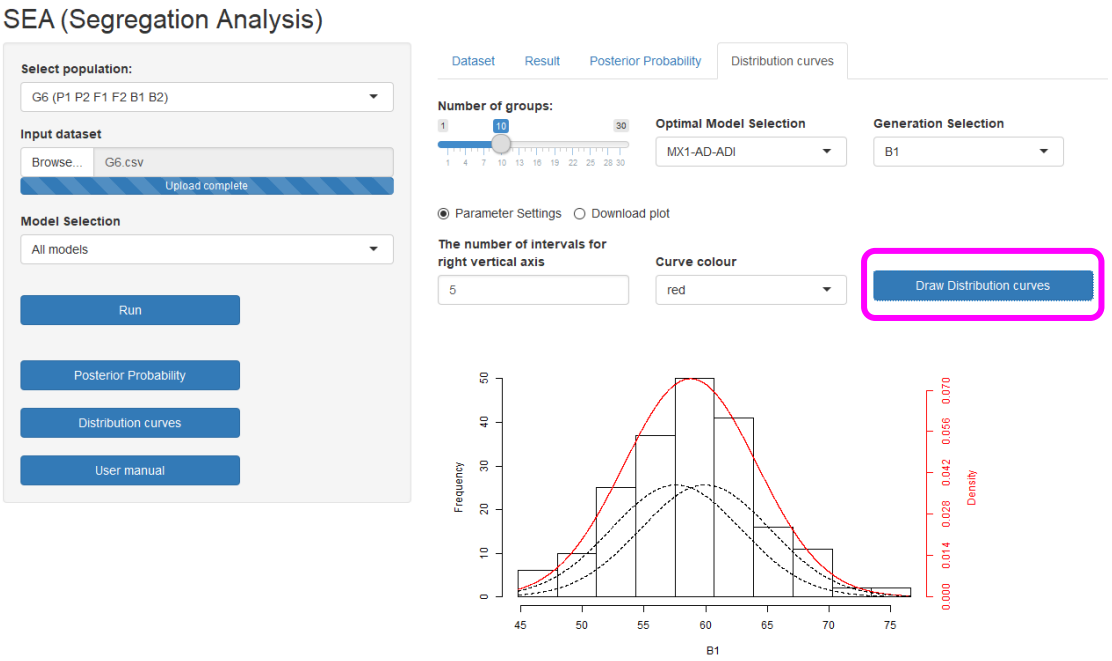

Figure 3.10 Draw distribution curves

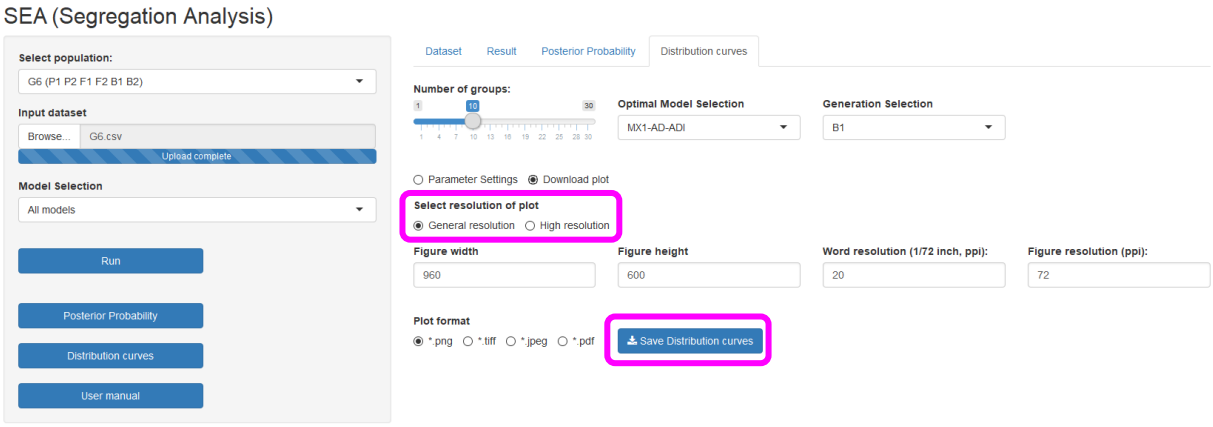

Figure 3.11 Download distribution curves

Parameter settings

1) Number of groups: Frequent distribution for quantitative traits can be indicated as the ideal figure that users want, if users change the number of groups.

2) Optimal Model Selection: the optimal model will be list in the drop down menu after calculation. Users can choose one model to draw the distribution curve.

3) Generation selection: For the single segregating population, users could draw the distribution curves directly. For the multi-generation populations, users need to select the populations, and then users can obtain the corresponding plots.

4) Curve color: The colors of the density curve can be changed via the combo box, with a drop-down option.

5) The **width** and **height** of the Figure, with the unit of pixel (px).

6) The **word resolution** in the Figure, with the unit of 1/72 inch, being pixels per inch (ppi).

7) The **figure resolution** in the Figure, with the unit of pixels per inch (ppi).

4 References

1 Gai Junyi, Zhang Yuan-Ming, Wang Jiankang. Genetic system of quantitative traits in plants. Beijing: Science Press, 2003.

2 Gai Junyi, Wang Jiankang. Identification and estimation of a QTL model and its effects. ***Theor Appl Genet*** 1998, 97(7): 1162-1168.

3 Gai Junyi, Zhang Yuan-Ming, Wang Jiankang. A joint analysis of multiple generations for QTL models extended to mixed two major genes plus polygene. ***Acta Agron Sin*** 2000, 26(4): 385-391.

4 Wang Jiankang, Gai Junyi. Identification of major gene and polygene mixed inheritance model and estimation of genetic parameters of a quantitative trait from F_2 Progeny. ***Acta Genetica Sinica*** 1997, 24(5):432-440.

5 Wang Jinshe, Zhao Tuanjie, Gai Junyi. Establishment of segregation analysis of mixed inheritance model with four major genes plus polygenes in backcross inbred lines (BIL) populations. ***Acta Agron Sin*** 2013, 39(2): 198-206.

6 Wang Jiankang, Gai Junyi. Identification of major gene and polygene mixed inheritance model of quantitative traits by using joint analysis of P_1, F_1, P_2, F_2 and $F_{2:3}$ generations. ***Acta Agron Sin*** 1998, 24(6): 651-659.

7 Zhang Yuan-Ming, Gai Junyi, Wang Jiankang. Identification of two major genes plus polygenes mixed inheritance model of quantitative traits in B_1 and B_2, and F_2. ***Journal of Biomathematics*** 2000, 15(3): 358-366.

8 Zhang Yuan-Ming, Gai Junyi, Yang Yonghua. The EIM algorithm in the joint segregation analysis of quantitative traits. ***Genetical Research*** 2003, 81(2): 157-163.

9 Zhang Yuan-Ming, Gai Junyi. Identification of mixed major genes and polygenes inheritance model of quantitative traits by using DH or RIL population. ***Acta Genetica Sinica*** 2000, 27(7):634-640.

10 Zhang Yuan-Ming, Gai Junyi, Qi Cunkou. Detection of genetic system of quantitative traits using backcross and selfing families. ***Hereditas (Beijing)*** 2001, 23(4): 329-776.

11 Zhang Yuan-Ming, Gai Junyi, Zhang Mengchen. Jointly segregating analysis of P_1 P_2 F_1 and F_2 or $F_{2:3}$ families. ***Journal of Southwest Agricultural University*** 2000, 42(1): 6-9.

12 Zhang Yuan-Ming, Gai Junyi, Wang Yongjun. An expansion of joint segregation analysis of quantitative trait for using P_1, P_2 and DH or RIL populations. ***Hereditas (Beijing)*** 2001, 23(5): 467-470.

13 Zhang Yuan-Ming, Gai Junyi. The IECM algorithm for estimation of component distribution parameters in segregating analysis of quantitative traits. ***Acta Agron Sin*** 2000, 26(6): 699-706.

14 Zhang Yuan-Ming, Gai Junyi, Qi Cunkou. The precision of segregating analysis of quantitative trait and its improving methods. ***Acta Agron Sin*** 2001, 27(6): 787-793.

15 Zhang Yuan-Ming. Segregation analysis of quantitative traits and its R software. Golden Light Academic Publishing, 2017, Beau Bassin 71504, Mauritius (ISBN: 978-620-2-41066-3).