The 9th Workshop on QTL Mapping and Breeding Simulation
The University of Sydney, Cobbitty NSW, 7-9 March 2012

QTL by environment interaction analysis and Segregation distortion loci mapping

Outlines

- QTL by environment interaction analysis
- The MET functionality in QTL IciMapping
- Segregation distortion loci mapping
- The SDL functionality in QTL IciMapping

QTL by environment interaction analysis

Linear regression model

- Take RIL population as an example. Assume E environments, n individuals in the population and m markers. The linear regression model in each environment is

$$y_{ij} = b_{0j} + \sum_{s=1}^{m+1} b_{sj} x_{is} + e_{ij}$$

  - Note: y_{ij} is the phenotypic data of the *i*th individual in the *j*th environment, *i*=1, 2, …, *n*， *j*=1, 2, …, *E*; b_{0j} is the constant term in the *j*th environment; x_{is} is the indicator of the *i*th individual in the *s*th marker; b_{sj} is the regression coefficient of phenotypic data of the *j*th environment for the *s*th marker variable.

Adjusted phenotypic values and hypothesis test of QTL mapping

- Assume the scanning position is in the interval between the *k*th and (*k+1)*th marker. Using the results of stepwise regression for adjusting phenotypic values:

$$\Delta y_{ij} = y_{ij} - \sum_{s \neq k, k+1} \hat{b}_{sj} x_{is}$$

- Traditional interval mapping was conducted on adjusted phenotypic values. The hypotheses used to test the existence of QTL at the scanning position are:
- H_0: $\mu_1^{(1)} = \mu_2^{(1)}, \cdots, \mu_1^{(E)} = \mu_2^{(E)}$
- H_1: not H_0;
- H_2: not H_0, and $\bar{a} = 0$

The logarithm likelihood

- The logarithm likelihood under H_1

$$L_1 = \sum_{j=1}^{E} \sum_{l=1}^{4} \sum_{i \in S_l} \log[\sum_{k=1}^{2} f_{lk} f(\Delta y_i(j); \mu_k(j), \sigma^2)]$$

- The logarithm likelihood under H_2

$$L_2 = -\sum_{j=1}^{E} \sum_{i} \sum_{k=1}^{2} \left[\frac{(\Delta y_i(j) - \mu_k(j))^2}{2\sigma^2(j)} + \ln\left(\sqrt{2\pi}\sigma(j)\right) \right] \omega_{ik} - \lambda \left[\frac{1}{2n} \sum_{j} (\mu_1(j) - \mu_2(j)) \right]$$

- Use EM algorithm to find the estimation of μ_1, μ_2

LOD score

- LOD(A)=L_1-L_2. It reflects the significance of main effects of QTL.
- LOD(AbyE)=L_2-L_0 . It reflects the significance of QTL by environment interaction.
- Similarly, we can calculate LOD score in each environment.
- Estimation of parameters were calculated by the maximum likelihood method.

Estimation of effects

- Mean values under H_1 were calculated as (N is the population size):

$$\mu_1(j) = \frac{\sum_{i=1}^{N} w_{i1}\Delta P_i(j)}{\sum_{i=1}^{N} w_{i1}} \qquad \mu_2(j) = \frac{\sum_{i=1}^{N} w_{i2}\Delta P_i(j)}{\sum_{i=1}^{N} w_{i2}}$$

- Estimation of additive effect in the *j*th environment

$$a(j) = (\mu_1(j) - \mu_2(j))/2$$

- Average value of additive effects in all environments

$$\bar{a} = (\mu_1(j) - \mu_2(j))/(2E)$$

- Interaction of additive effects by environment

$$ae(j) = a(j) - \bar{a}$$

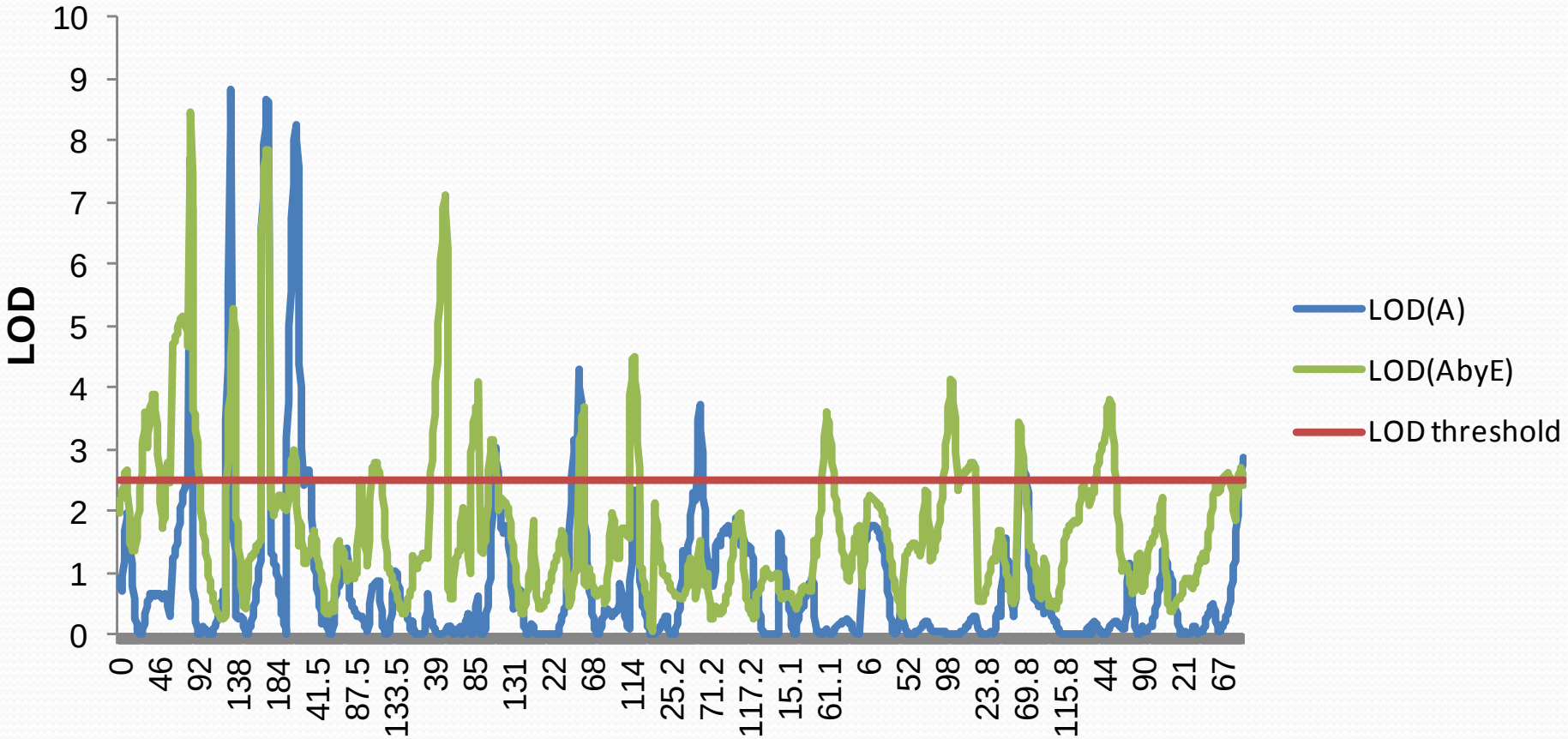
An actual maize RIL population

- Parents: B73 and B97
- Population size: 194
- 10chromosomes with 237 SSR markers. 8.0% of marker types are missing。
- 7 environments (E_1 to E_7). Trait: heading date. PIN is 0.001. LOD threshold is 2.5.
- At the same time, we consider the 7 environments as 7 traits and use the QTL mapping method of single environment.

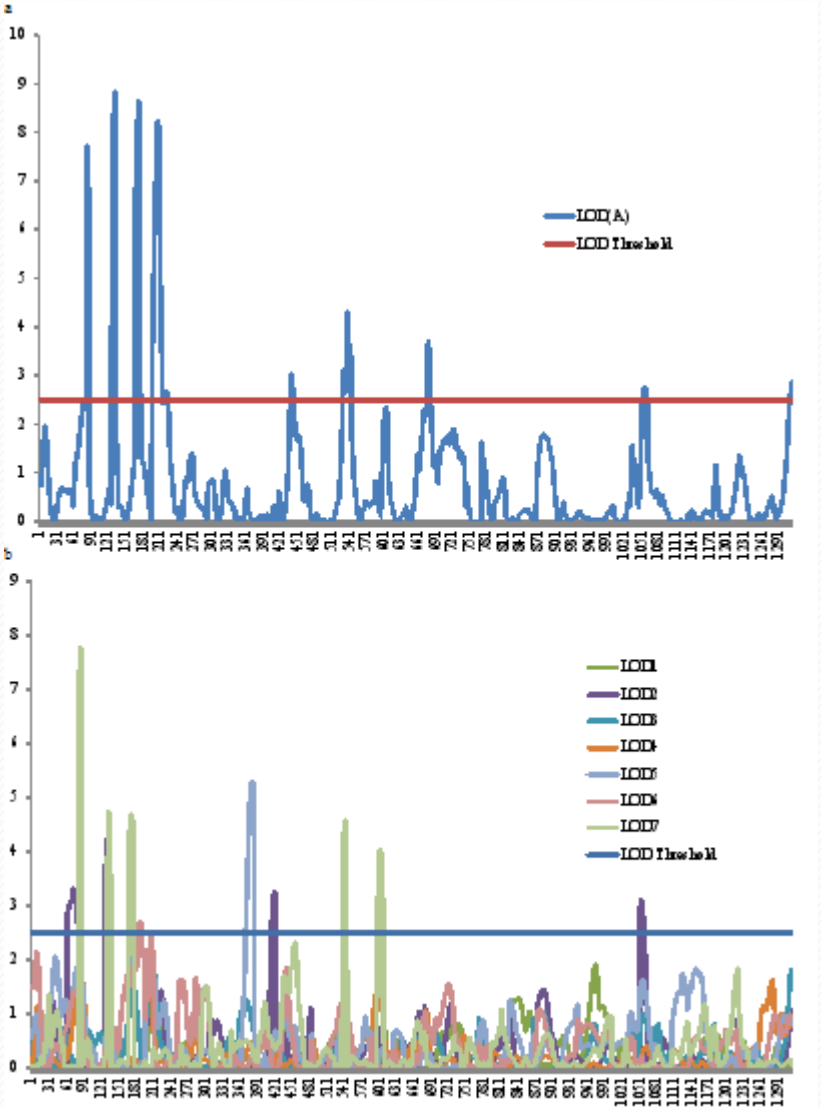
Mapping results

- 9 QTL are detected through QTL by environment interaction analysis which are on chromosome1~5, 8 and 10.
- Compared to mapping results in single environment, QTL by environment interaction analysis provides more all-around genetic information about positions and effects of QTL.

LOD profile of QTL by environment interaction analysis



QTL by environment interaction analysis (up) and QTL mapping in single environment (down)

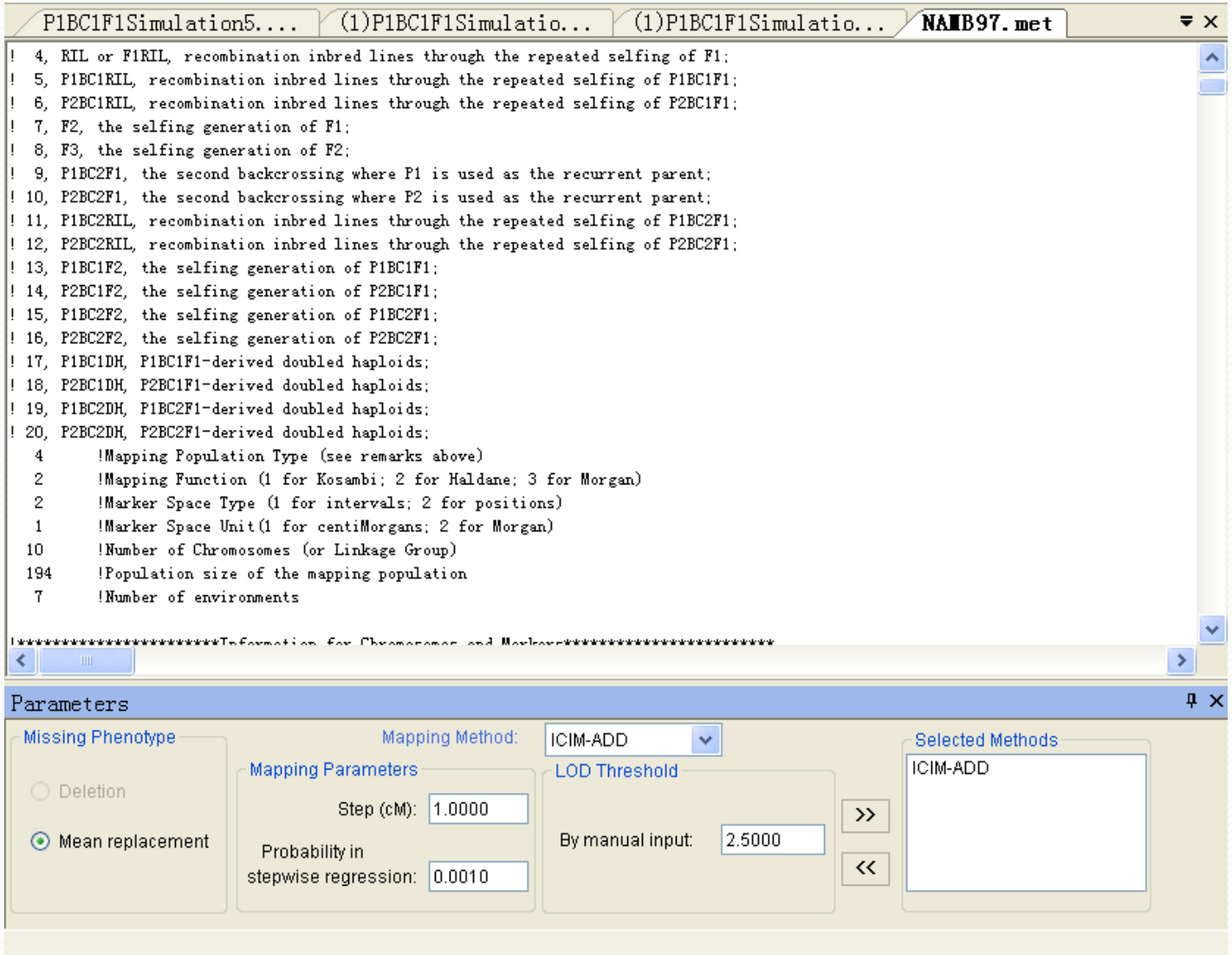


The MET functionality in QTL IciMapping

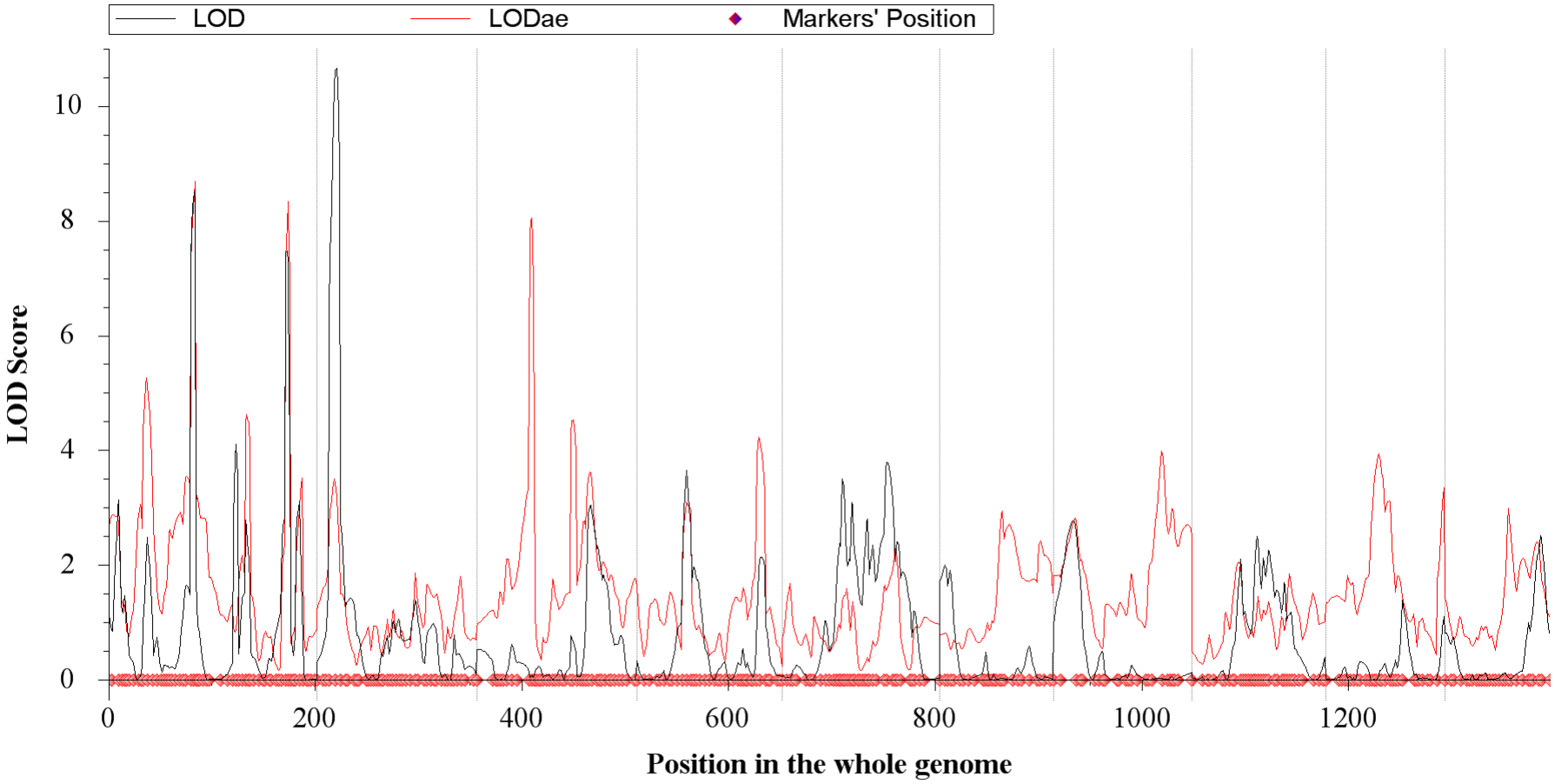
Two methods available in MET

- ICIM-ADD: inclusive composite interval mapping of additive (and dominant) QTL (Li et al., 2007. Genetics 175: 361-374. Zhang et al., 2008. Genetics 180: 1177-1190)
- ICIM-EPI: inclusive composite interval mapping of digenic epistatic QTL (Li et al., 2008. Theor. Appl. Genet. 116: 243-260)

Interface of the MET functionality



LOD profile of ICIM additive mapping (ICIM-ADD in MET)



Segregation distortion loci (SDL) mapping

Interval mapping

- Taking F_2 as an example， we calculate the probability of the *i*th individual with genotype MM, Mm and mm, i.e. $P_i(MM)$, $P_i(Mm)$, $P_i(mm)$.

Left marker	Right marker	$P_i(MM)$	$P_i(Mm)$	$P_i(mm)$
AA	BB	$\frac{(1-r_1)^2(1-r_2)^2}{(1-r_{12})^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{(1-r_{12})^2}$	$\frac{r_1^2 r_2^2}{(1-r_{12})^2}$
AA	Bb	$\frac{(1-r_1)^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$	$\frac{r_1(1-r_1)(1-r_2)^2+\frac{1}{2}r_1(1-r_1)r_2^2}{r_{12}(1-r_{12})}$	$\frac{r_1^2 r_2(1-r_2)}{r_{12}(1-r_{12})}$
AA	bb	$\frac{(1-r_1)^2 r_2^2}{r_{12}^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1^2(1-r_2)^2}{r_{12}^2}$

Interval mapping

- Calculating the number of individuals for each genotype in the population

$$N_1 = \sum_{i=1}^{N} P_i(MM) \qquad N_2 = \sum_{i=1}^{N} P_i(Mm) \qquad N_3 = \sum_{i=1}^{N} P_i(mm)$$

- N is the population size.
- Calculating the number of individuals for each genotype under Mendelian ratio

$$ExpN_1 = N * 0.25 \qquad ExpN_2 = N * 0.5 \qquad ExpN_3 = N * 0.25$$

Hypothesis test

- H_0: there is no segregation distortion in the scanning position
- H_1: not H_0
- LOD score for detecting SDL

$$LOD = \sum_{k=1}^{3} N_k \log_{10} \frac{N_k}{N} - \sum_{k=1}^{3} ExpN_k \log_{10} \frac{ExpN_k}{N}$$

Fitness of SDL

- It stands for the surviving ability of each genotype

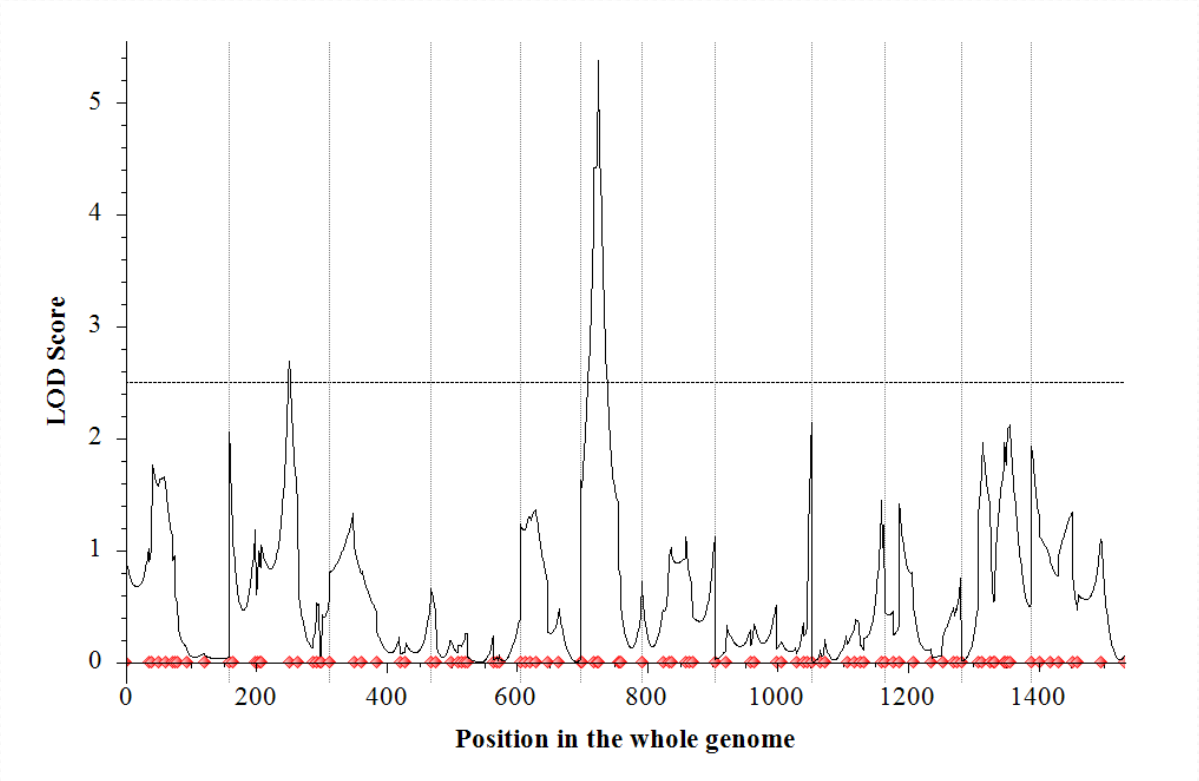
$$Fitness(MM) = \frac{N_1}{Max(N_1, N_2/2, N_3)}$$

$$Fitness(Mm) = \frac{N_2/2}{Max(N_1, N_2/2, N_3)}$$

$$Fitness(mm) = \frac{N_3}{Max(N_1, N_2/2, N_3)}$$

An actual rice F_2 population

- Population size: 180
- 12 chromosomes with 117 markers
- Scanning step:1cM. LOD threshold: 2.5



SDL in the actual population

- 2 SDL are detected in the population on chromosome 2 and 6.

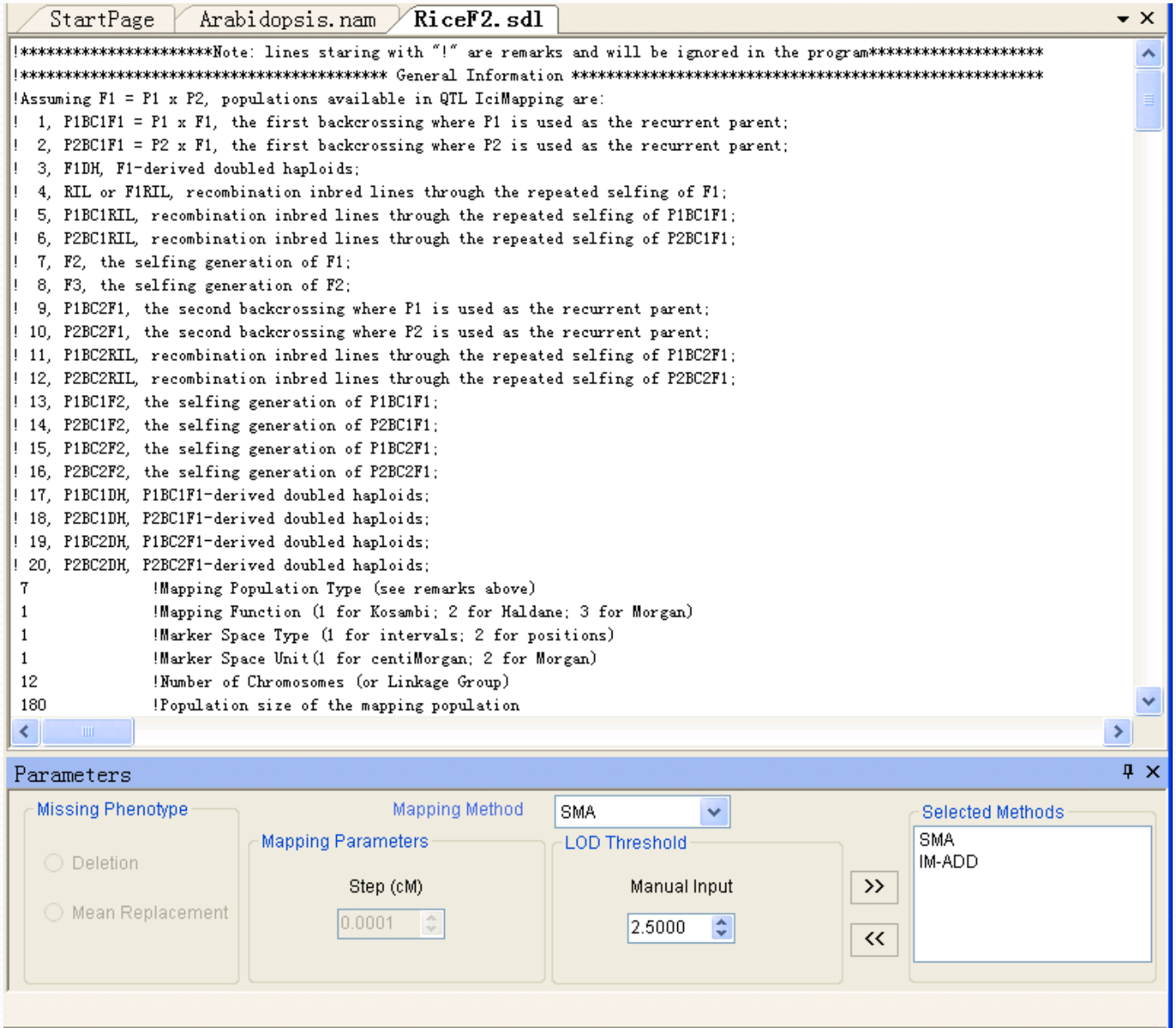
Chr.	Pos.	Left marker	Right marker	LOD	Fitness (MM)	Fitness (Mm)	Fitness (mm)
2	92	RM6318	RM525	2.68	0.68	1	0.52
6	26	RM6_19	RM6_30	5.37	0.32	0.69	1

The SDL functionality in QTL IciMapping

Two methods available in SDL

- SMA: single marker analysis (Soller et al., 1976. Theor. Appl. Genet. 47: 35-39)
- SIM: the conventional simple interval mapping (Lander and Botstein, 1989. Genetics 121: 185-199)

Interface of the SDL functionality



LOD profile from interval mapping (IM in SDL)