One-week Course on Genetic Analysis and Plant Breeding
21 - 25 January 2013, CIMMYT, Mexico

Inclusive Composite Interval Mapping (ICIM)

Jiankang Wang, CIMMYT China and CAAS
E-mail: jkwang@cgiar.org; wangjiankang@caas.cn
Web: http://www.isbreeding.net

Outlines

- **Importance of the background control**
- **ICIM for additive QTL**
- **ICIM for additive and dominance QTL**
- **The BIP functionality in QTL IciMapping**

Importance of the background control

Variance of the sample mean

- Random samples $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$
- Sample mean:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Distribution of the sample mean

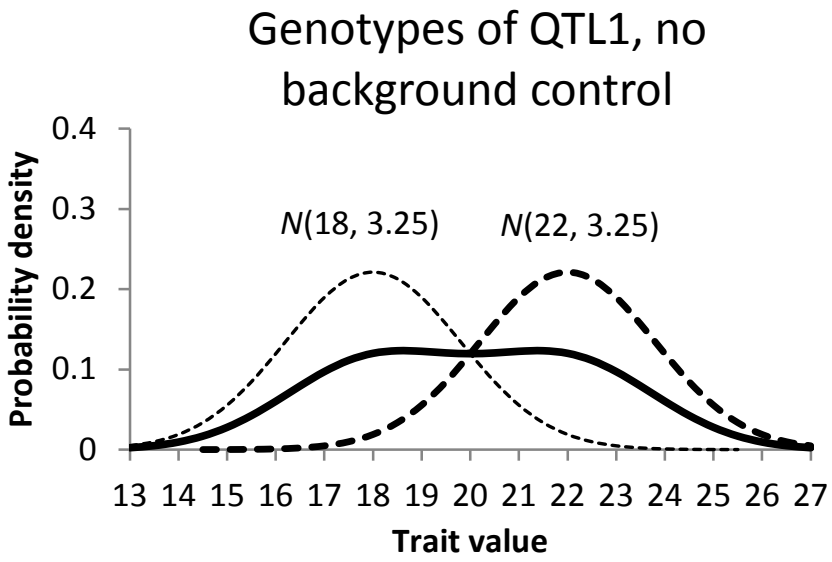
$$\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$$

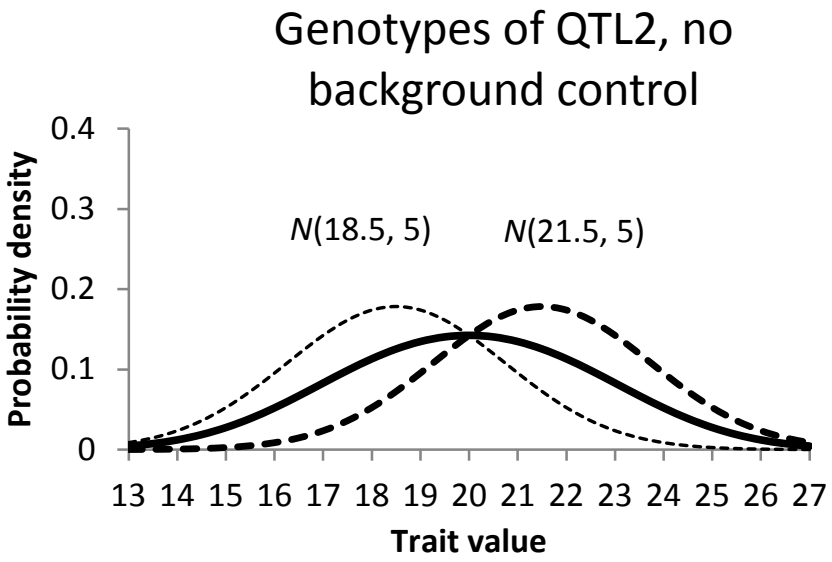
- The smaller the variance, the higher significance in hypothesis test!

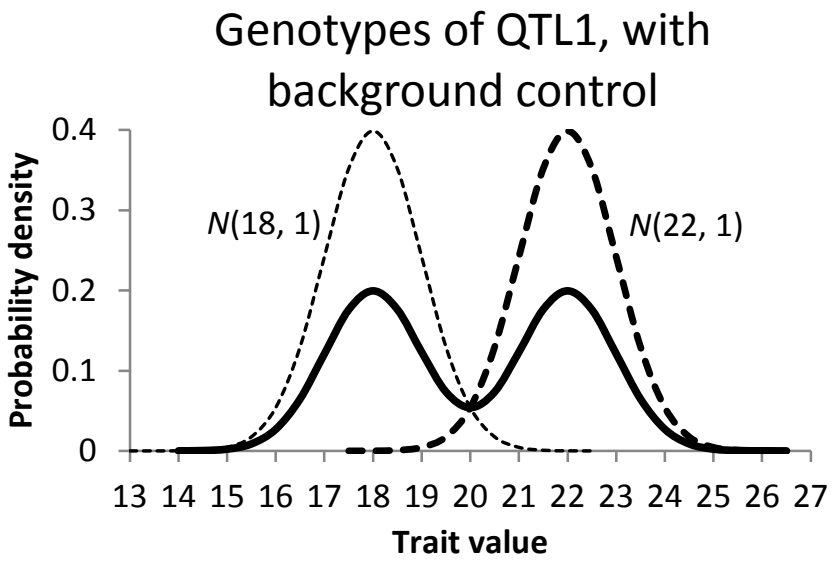
Two putative unlinked QTL

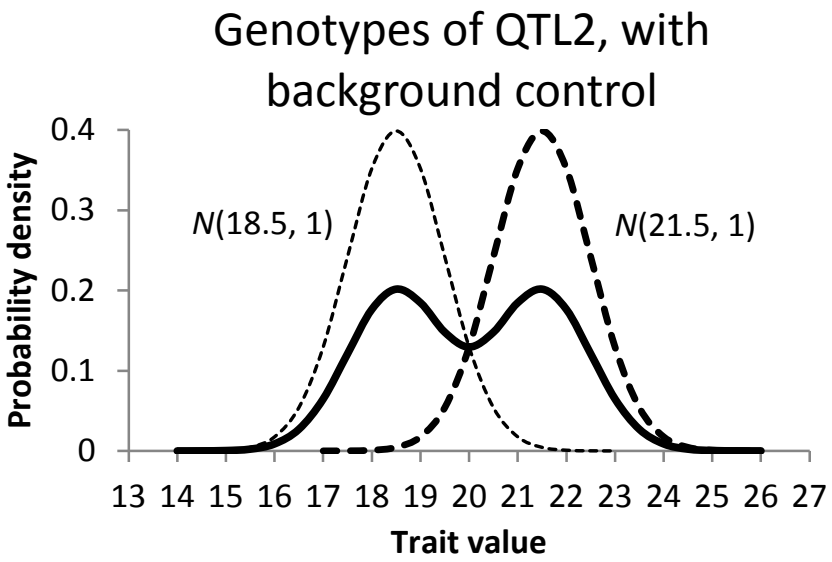
Overall mean	Additive effect		Additive variance		Genetic variance	Error variance	H^2
	QTL1	QTL2	QTL1	QTL2			
20	2	1.5	4	2.25	6.25	1	0.86

QTL distribution under no background control (upper) and background control (lower)









Inclusive Composite Interval Mapping (ICIM) for additive QTL

Li, H., G. Ye, and **J. Wang***. 2007. A modified algorithm for the improvement of composite interval mapping. **Genetics** 175: 361-374.

The single additive QTL model

$$G = \mu + aw$$

- Take DH population as an example
- G: mean genotypic value of QTL genotypes QQ or qq
- μ: mean of the two homozygous QTL genotypes
- *a*: additive effect of QTL
- w: indicator of QTL genotype. 1 for QQ; -1 for qq

Derive QTL genotype from its flanking markers

Marker interval		Sample size	Freq.	QTL genotype	
Left	Right			QQ (w=1)	Qq (w=-1)
AA (x_L=1)	BB (x_R=1)	n_1	½(1-r)	½(1-r_L-r_R+$r_L r_R$)	½$r_L r_R$
AA (x_L=1)	bb (x_R=-1)	n_2	½r	½(1-r_L)r_R	½r_L(1-r_R)
Aa (x_L=-1)	BB (x_R=1)	n_3	½r	½r_L(1-r_R)	½(1-r_L)r_R
Aa (x_L=-1)	bb (x_R=-1)	n_4	½(1-r)	½$r_L r_R$	½(1-r_L-r_R+$r_L r_R$)

$$E(w \mid x_L, x_R) = \lambda_L x_L + \lambda_R x_R$$

$$\lambda_L = \tfrac{1}{2}(f_1 + f_2) = \frac{r - r_L + r_R - 2r_L r_R}{2r(1-r)} \qquad \lambda_R = \tfrac{1}{2}(f_1 - f_2) = \frac{r + r_L - r_R - 2r_L r_R}{2r(1-r)}$$

Mean of flanking markers

$$E(G \mid x_{\mathrm{L}}, x_{\mathrm{R}}) = \mu + aE(w \mid x_{\mathrm{L}}, x_{\mathrm{R}}) = \mu + a\lambda_{\mathrm{L}} x_{\mathrm{L}} + a\lambda_{\mathrm{R}} x_{\mathrm{R}}$$

- Properties
  - The QTL additive effect is completely absorbed by its two nearest flanking markers

The linear model for multiple QTLs

$$G = \mu + \sum_{j=1}^{m} a_j w_j$$

$$E(w_j \mid \mathbf{X}) = \lambda_{j(\mathrm{L})} x_j + \lambda_{j(\mathrm{R})} x_{j+1}$$

$$E(G \mid \mathbf{X}) = \mu + \sum_{j=1}^{m} a_j (\lambda_{j(\mathrm{L})} x_j + \lambda_{j(\mathrm{R})} x_{j+1}) \hat{=} \beta_0 + \sum_{j=1}^{m+1} \beta_j x_j$$

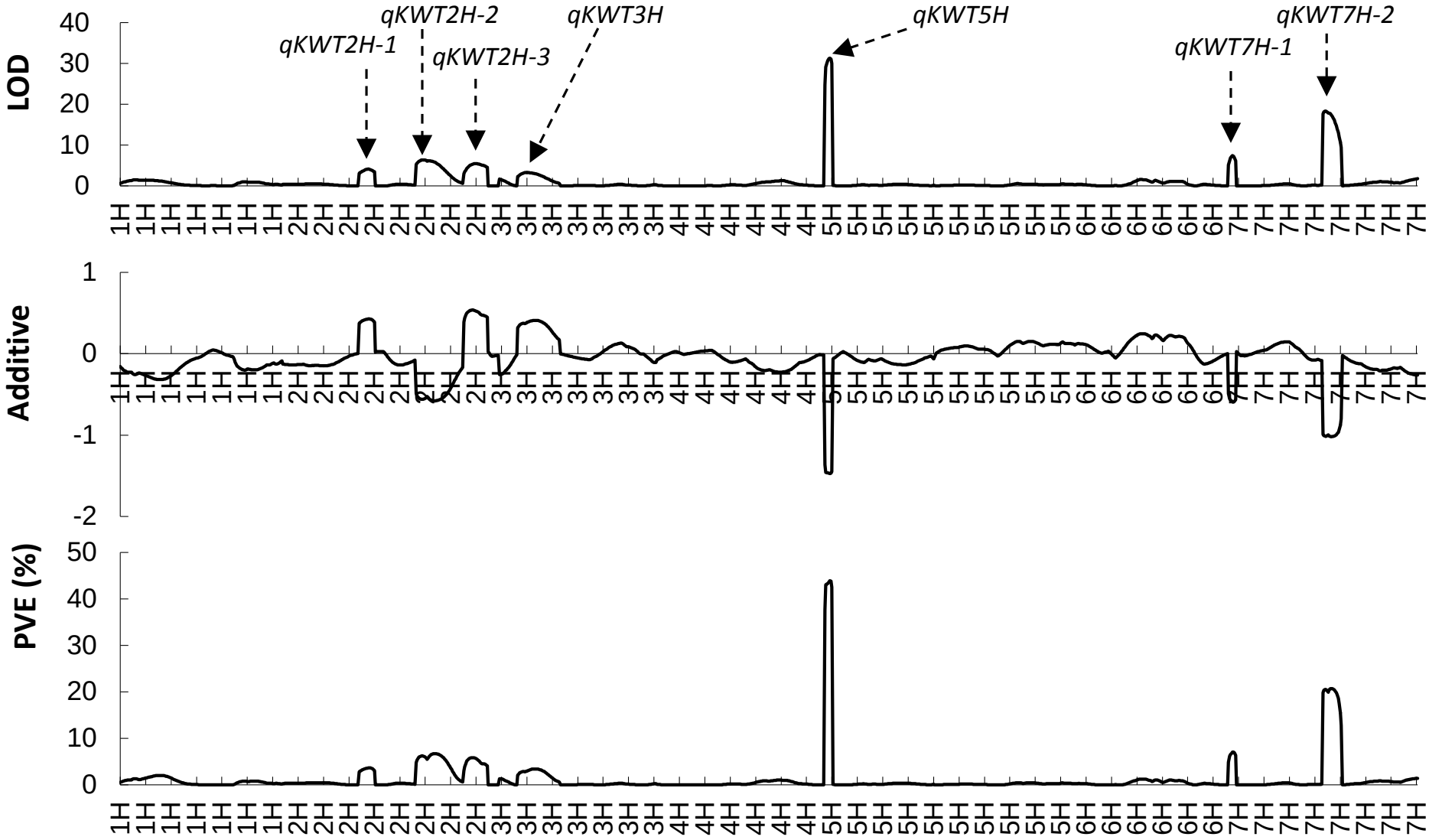
$$y_i = E(G \mid \mathbf{X}) + \varepsilon_i = \beta_0 + \sum_{j=1}^{m+1} \beta_j x_{ij} + \varepsilon_i$$

Genomic scanning for additive QTL

$$\Delta y_i = y_i - \sum_{j \neq k, k+1} \hat{b}_j x_{ij}$$

- Phenotype adjusted by the previous linear model
- Properties of the adjusted phenotypes
  - QTL on the current interval is retained
  - QTLs on other intervals are eliminated, thus the background genetic variation is controlled

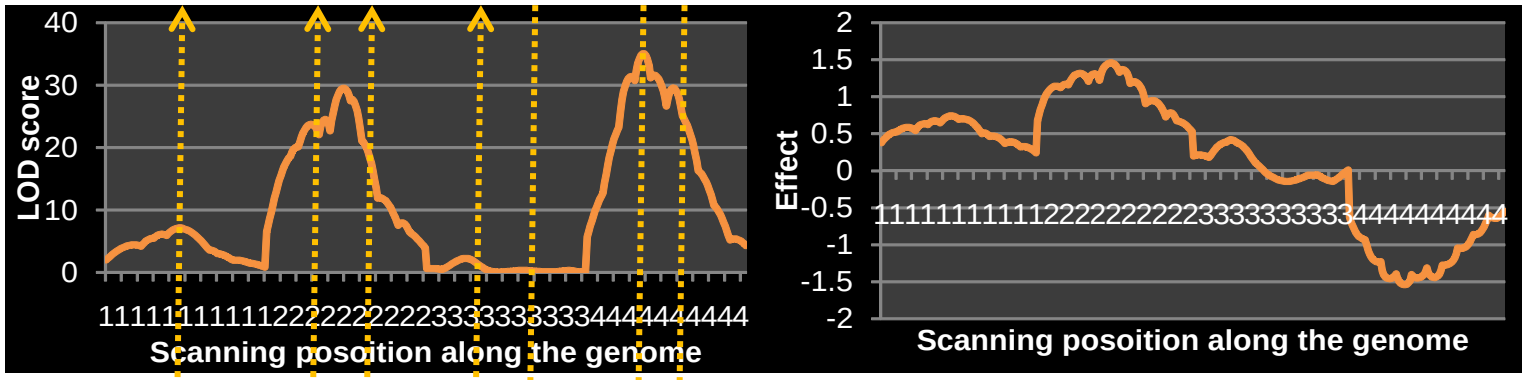
ICIM in a barley DH population



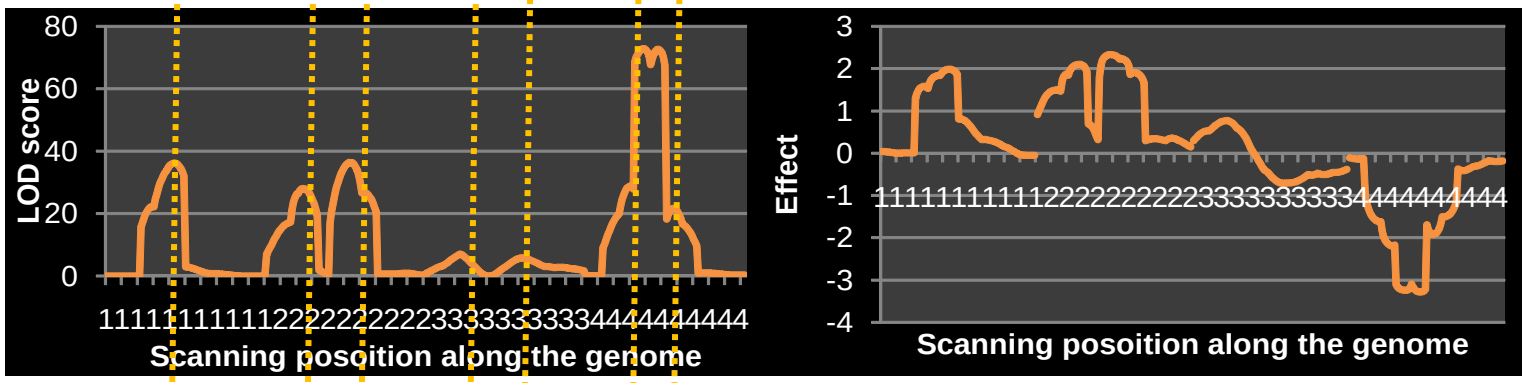
QTL identified by ICIM in the barley DH population

QTL name	Position (cM)	Left marker	Right marker	LOD	PVE (%)	Additive
qKWT2H-1	84	Pox	BCD351B	4.10	0.43	3.63
qKWT2H-2	138	ABC620	MWG882	6.36	-0.56	6.21
qKWT2H-3	190	BCD453B	ABG317	5.45	0.53	5.6
qKWT3H	27	Ugp2	Ugp1	3.23	0.38	2.99
qKWT5H	5	Act8B	MWG502	31.33	-1.47	43.92
qKWT7H-1	4	iPgd1A	BCD129	7.41	-0.59	7.07
qKWT7H-2	96	MWG626	VAtp57A	18.35	-1.01	20.53

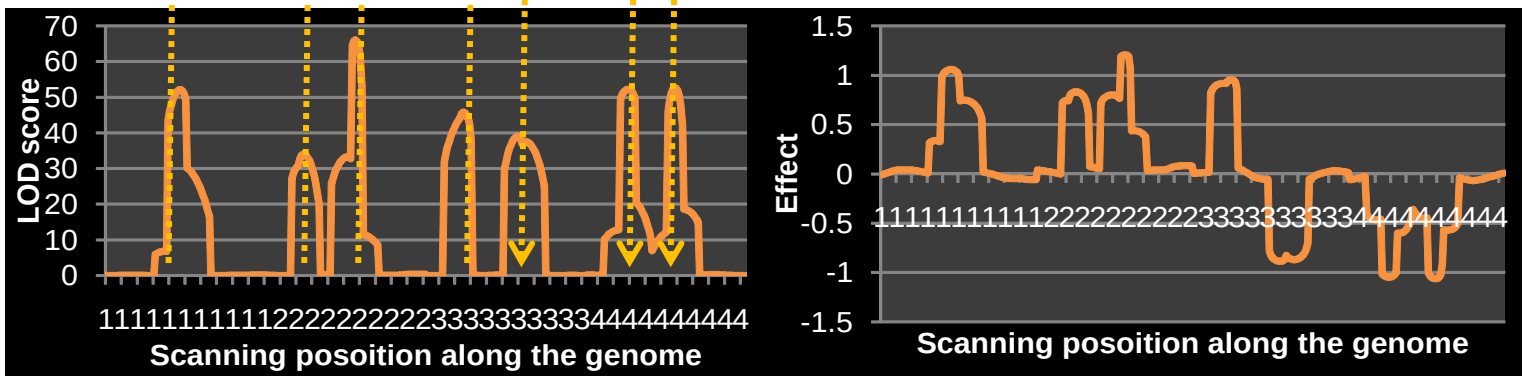
Results from a simulated population



IM



CIM



ICIM

Inclusive Composite Interval Mapping (ICIM) for additive and dominance QTL

Zhang, L., H. Li, Z. Li, and **J. Wang***. 2008. Interactions between markers can be caused by the dominance effect of QTL. **Genetics** 180: 1177-1190.

The single additive and dominance QTL model

$$G = \mu + aw + dv$$

- Take F2 population as an example
- G: mean genotypic value of QTL genotypes QQ or qq
- μ: mean of the two homozygous QTL genotypes
- *a*: additive effect of QTL
- *d*: dominance effect of QTL
- w: indicator of QTL genotypes
  - 1 for QQ; 0 for Qq; -1 for qq
- v: indicator of QTL genotypes
  - 0 for QQ; 1 for Qq; 0 for qq

Probability of the three QTL genotypes under each of the nine marker types

Left marker	Right marker	QQ (w=1, v=0) (*m*+*a*)	Qq (w=0, v=1) (*m*+*d*)	qq (w=-1, v=0) (*m*-*a*)
AA	BB	$\frac{1}{4}(1-r_1)^2(1-r_2)^2$	$\frac{1}{2}r_1(1-r_1)r_2(1-r_2)$	$\frac{1}{4}r_1^2r_2^2$
AA	Bb	$\frac{1}{2}(1-r_1)^2r_2(1-r_2)$	$\frac{1}{2}r_1(1-r_1)(1-r_2)^2+\frac{1}{2}r_1(1-r_1)r_2^2$	$\frac{1}{2}r_1^2r_2(1-r_2)$
AA	bb	$\frac{1}{4}(1-r_1)^2r_2^2$	$\frac{1}{2}r_1(1-r_1)r_2(1-r_2)$	$\frac{1}{4}r_1^2(1-r_2)^2$
…	…	…	…	…

The expected genotypic value of an individual with known marker types

$$E(G \mid x_1, x_2, y_1, y_2) = \mu + a \times E(w \mid x_1, x_2, y_1, y_2) + d \times E(v \mid x_1, x_2, y_1, y_2)$$

where x_1 and y_1 are the indicators for the left marker, x_2 and y_2 are the indicators for the right marker, similarly defined to the QTL genotypes

Relationship between marker class mean and marker effect

(including marker interactions)

$$\begin{bmatrix} \mu + f_1 a + g_1 d \\ \mu + f_2 a + g_2 d \\ \mu + f_3 a + g_3 d \\ \mu + f_4 a + g_4 d \\ \mu + g_5 d \\ \mu - f_4 a + g_4 d \\ \mu - f_3 a + g_3 d \\ \mu - f_2 a + g_2 d \\ \mu - f_1 a + g_1 d \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \mu + \Delta \\ A_{\mathrm{L}} \\ A_{\mathrm{R}} \\ D_{\mathrm{L}} \\ D_{\mathrm{R}} \\ AA \\ AD \\ DA \\ DD \end{bmatrix}$$

Relationship between marker effects and QTL effects

$$\begin{bmatrix} \mu+\Delta \\ A_{\mathrm{L}} \\ A_{\mathrm{R}} \\ D_{\mathrm{L}} \\ D_{\mathrm{R}} \\ AA \\ AD \\ DA \\ DD \end{bmatrix} = \begin{bmatrix} \mu+\frac{1}{2}(g_1+g_3)d \\ f_2 a \\ \frac{1}{2}(f_1-f_3)a \\ (-\frac{1}{2}g_1-\frac{1}{2}g_3+g_4)d \\ (-\frac{1}{2}g_1+g_2-\frac{1}{2}g_3)d \\ \frac{1}{2}(g_1-g_3)d \\ 0 \\ 0 \\ (\frac{1}{2}g_1-g_2+\frac{1}{2}g_3-g_4+g_5)d \end{bmatrix}$$

The additive QTL effect "*a*" only causes additive marker effects.

But the dominance QTL effect "*d*" causes additive by additive, and dominance by dominance marker interactions as well as dominance marker effects

The linear model of genotypic values on markers in F_2

- The expectations of *w* and *v* (indicators of QTL genotype) under each marker class can be proved as:

$$E(w \mid x_{\mathrm{L}}, x_{\mathrm{R}}, y_{\mathrm{L}}, y_{\mathrm{R}}) = \lambda_{\mathrm{L}} x_{\mathrm{L}} + \lambda_{\mathrm{R}} x_{\mathrm{R}}$$

$$E(v \mid x_{\mathrm{L}}, x_{\mathrm{R}}, y_{\mathrm{L}}, y_{\mathrm{R}}) = \delta + \rho_{\mathrm{L}} y_{\mathrm{L}} + \rho_{\mathrm{R}} y_{\mathrm{R}} + \lambda\lambda x_{\mathrm{L}} x_{\mathrm{R}} + \rho\rho y_{\mathrm{L}} y_{\mathrm{R}}$$

$$E(G \mid x_{\mathrm{L}}, x_{\mathrm{R}}, y_{\mathrm{L}}, y_{\mathrm{R}}) = \mu + \Delta + A_{\mathrm{L}} x_{\mathrm{L}} + A_{\mathrm{R}} x_{\mathrm{R}} + + D_{\mathrm{L}} y_{\mathrm{L}} + D_{\mathrm{R}} y_{\mathrm{R}} + AA x_{\mathrm{L}} x_{\mathrm{R}} + DD y_{\mathrm{L}} y_{\mathrm{R}}$$

Properties of the linear model in F_2

- The additive and dominance effects of the flanked QTL are completely absorbed by the six variables in the model above.
- Interactions between marker variables may be declared as interaction between QTL by mistake when using ANOVA.
- But from our analysis, interactions between marker variables can be caused simply by dominance effects of QTL .

Multiple QTL model in F_2

- For multiple QTL, assume there are *m* QTL located on *m* intervals defined by *m*+1 markers on one chromosome, then the genotypic value of an F_2 individual is defined as:

$$G = \mu + \sum_{j=1}^{m} [a_j w_j + d_j v_j]$$

where w_j and v_j are the indicators for genotypes at the j^{th} QTL.

The linear model in F_2 for multiple QTL

- The genotypic value of an F_2 individual with known marker types can be re-organized as:

$$E(G \mid \mathbf{X}, \mathbf{Y}) = \beta_0 + \sum_{j=1}^{m+1} \beta_j x_j + \sum_{j=1}^{m+1} \gamma_j y_j + \sum_{j=1}^{m} \beta_{j,j+1} x_j x_{j+1} + \sum_{j=1}^{m} \gamma_{j,j+1} y_j y_{j+1}$$

$$y = E(G \mid \mathbf{X}, \mathbf{Y}) + \varepsilon$$

$$= \beta_0 + \sum_{j=1}^{m+1} \beta_j x_j + \sum_{j=1}^{m+1} \gamma_j y_j + \sum_{j=1}^{m} \beta_{j,j+1} x_j x_{j+1} + \sum_{j=1}^{m} \gamma_{j,j+1} y_j y_{j+1} + \varepsilon$$

Property of the linear model for QTL mapping in F_2

- The QTL effects will be completely absorbed by the six variables of the two closest markers.
- The linear model is suitable for QTL mapping in F_2 populations, as it completely explains both additive and dominance variations.

Genomic scanning for additive and dominance QTL

$$\Delta y_i = y_i - \sum_{j \neq k, k+1} [\hat{\beta}_j x_{ij} + \hat{\gamma}_j y_{ij}] - \sum_{j \neq k} [\hat{\beta}_{j,j+1} x_{ij} x_{i,j+1} + \hat{\gamma}_{j,j+1} y_{ij} y_{i,j+1}]$$

- Phenotype adjusted by the previous linear model
- Propertied of the adjusted phenotypes
  - QTL on the current interval is retained
  - QTLs on other intervals are eliminated, thus the background genetic variation is controlled

Hypothesis test of QTLs in F_2

- **The two hypotheses used to test the existence of QTL at the scanning position are**

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_A : \text{at least two of } \mu_1, \mu_1 \text{ and } \mu_3 \text{ are not equal}$$

- **The logarithm likelihood under H_A is**

$$L_A = \sum_{j=1}^{9} \sum_{i \in S_j} \log[\sum_{k=1}^{3} \pi_{jk} f(\Delta P_i; \mu_k, \sigma^2)]$$

where S_j denotes individuals belonging to the j^{th} marker class (j=1, 2, …, 9), π_{jk} (k=1, 2, 3) is the proportion of the k^{th} QTL genotype in the j^{th} class, and $f(\cdot; \mu_k, \sigma^2)$ is the density function of the normal distribution $N(\mu_k, \sigma^2)$.

Estimation of genotypic means and genetic effects in F_2

- Use EM algorithm to have the estimation of μ_1, μ_2 and μ_3

- So the genetic effects can be estimated by

$$\hat{\mu} = \tfrac{1}{2}(\hat{\mu}_1 + \hat{\mu}_3)$$

$$\hat{a} = \tfrac{1}{2}(\hat{\mu}_1 - \hat{\mu}_3) \quad \hat{d} = \hat{\mu}_2 - \hat{\mu}$$

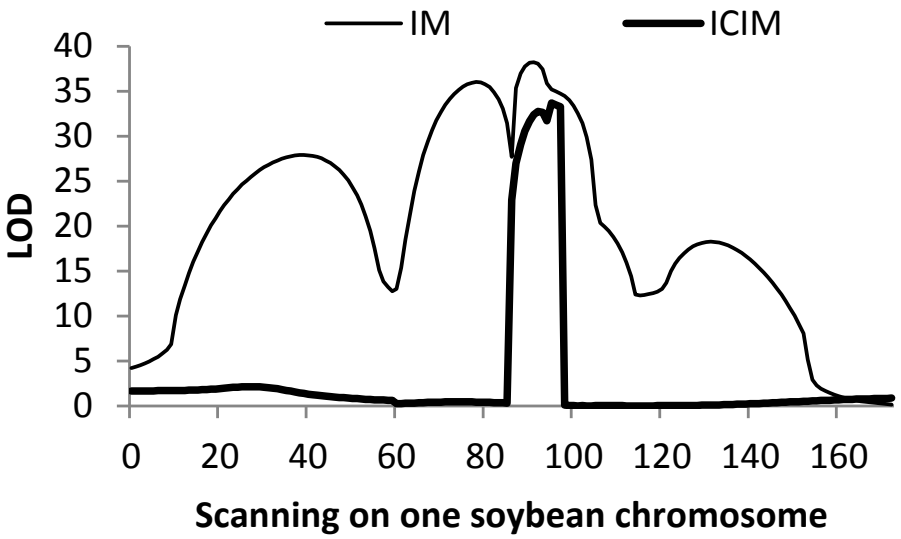
LOD score in F_2

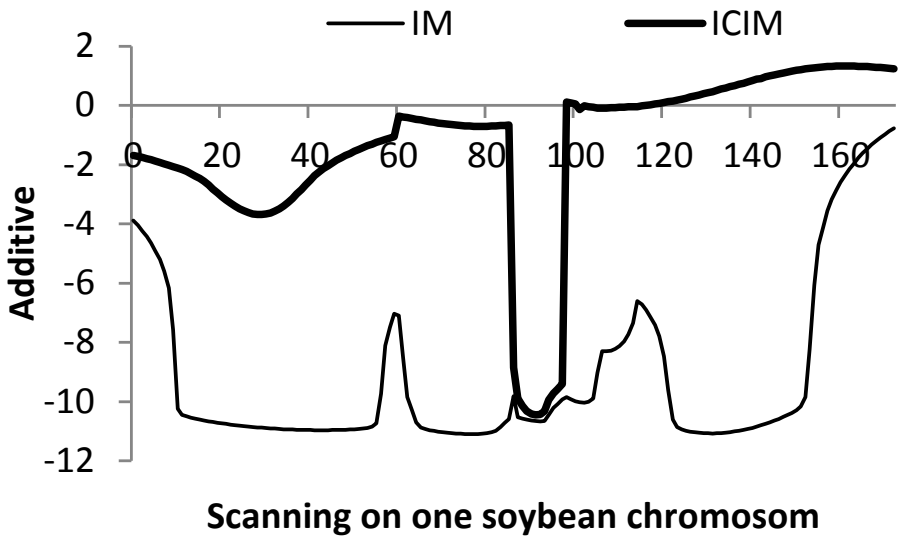
- Parameters under H_0 were calculated as:

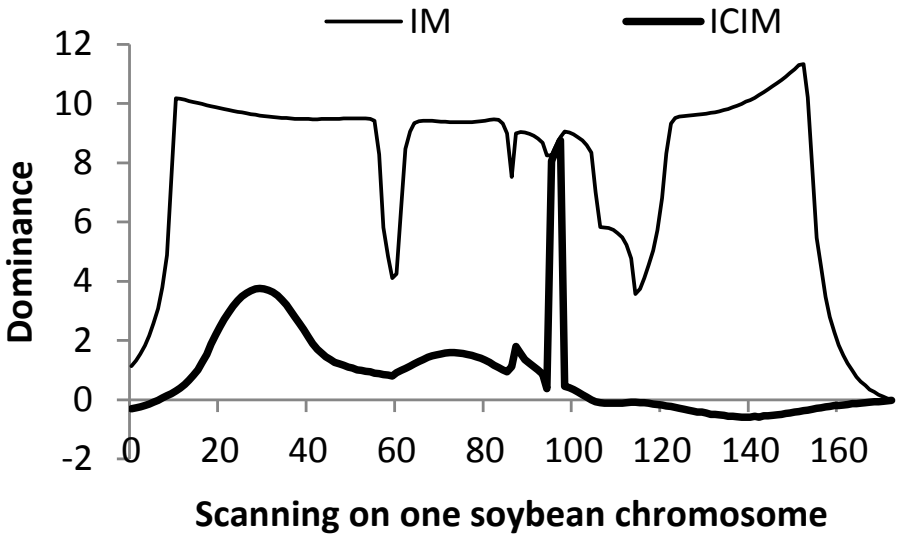
$$\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n}\Delta P_i \qquad \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(\Delta P_i - \hat{\mu}_0)^2$$

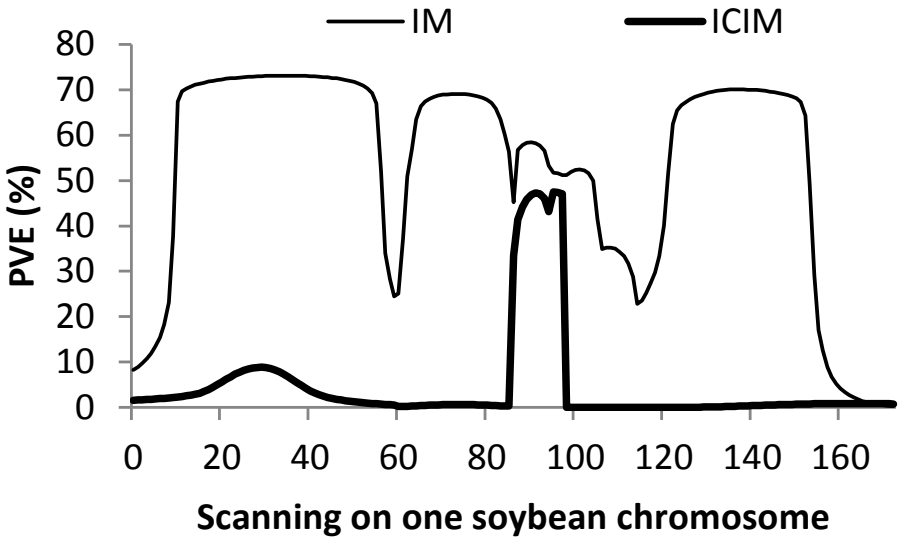
- From which the maximum likelihood under H_0, and the LOD score between H_A and H_0 can be calculated.

Interval mapping in previous soybean F2 population









QTL identified in the soybean F2 population

Position (cM)	Left marker	Right marker	LOD	PVE (%)	Additive	Dominance
92	*Satt255	*Satt339	32.73	47.07	-10.42	1.01
95	*Satt339	*Satt237	33.69	47.50	-9.72	8.06

QTL distribution models in simulation

- Six QTL with different levels of dominance and a genome consisting of 8 chromosomes.
- Each chromosome is 140 cM long, with 15 evenly distributed co-dominant markers.
- Three QTL distribution models.

QTL distribution models in simulation

QTL	*a*	*d*	Model I		Model II		Model III		Genotypic variation explained (%)	Phenotypic variation explained (%)
			Chr	cM	Chr	cM	Chr	cM		
QTL1	1	0	1	25	1	25	1	25	11.43	8.00
QTL2	0	1	2	55	1	55	2	55	5.71	4.00
QTL3	1	1	3	25	2	25	3	25	17.14	12.00
QTL4	1	-1	4	55	2	55	1	55	17.14	12.00
QTL5	1	1.5	5	25	3	25	2	25	24.29	17.00
QTL6	1	-1.5	6	55	3	55	3	55	24.29	17.00

Simulation of the mapping populations

- F_2 populations were simulated by the genetics and breeding simulation tool of QuLine (to be introduced on Day 5).
- QTL mapping using ICIM was implemented by the software QTL IciMapping.

Theoretical marker effects in the genetic model used in simulation

- The expected additive, dominance, additive by additive, and dominance by dominance effects of the two flanking markers associated with each QTL is shown in the following table.
- It indicated that the dominance of a QTL could complicate the coefficients of the two markers flanking a QTL, and cause the interactions between markers.

Expected marker effects in simulation

QTL	$(d)\mu_d$	$(a)A_1$	$(a)A_2$	$(d)D_1$	$(d)D_2$	$(d)AA_{12}$	$(d)DD_{12}$	Interaction variation (%)
QTL1	0.000	0.498	0.498	0.000	0.000	0.000	0.000	0.0
QTL2	0.253	0.000	0.000	0.248	0.248	-0.248	0.243	21.8
QTL3	0.253	0.498	0.498	0.248	0.248	-0.248	0.243	5.7
QTL4	-0.253	0.498	0.498	-0.248	-0.248	0.248	-0.243	5.7
QTL5	0.379	0.498	0.499	0.371	0.371	-0.371	0.364	9.6
QTL6	-0.379	0.498	0.498	-0.371	-0.371	0.371	-0.364	9.6

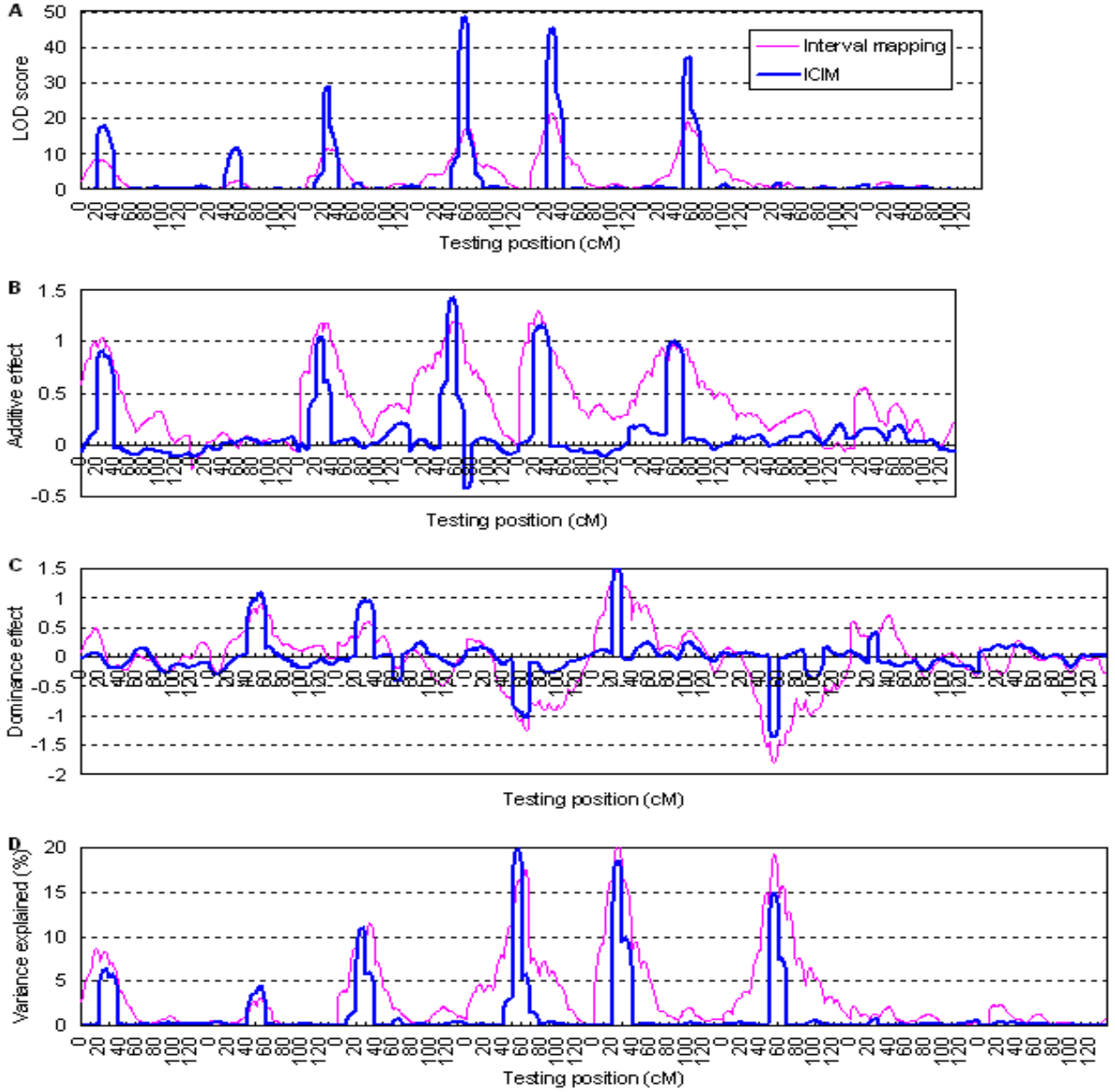
QTL mapping in simulated F_2 populations

- There is no clear classification on the phenotype, and it is impossible to deduce the number of QTL without the assistance of molecular markers.
- Set the LOD threshold at 3.0. The two probabilities for entering and removing variables were set at 0.01 and 0.02. The scanning step is 1 cM.

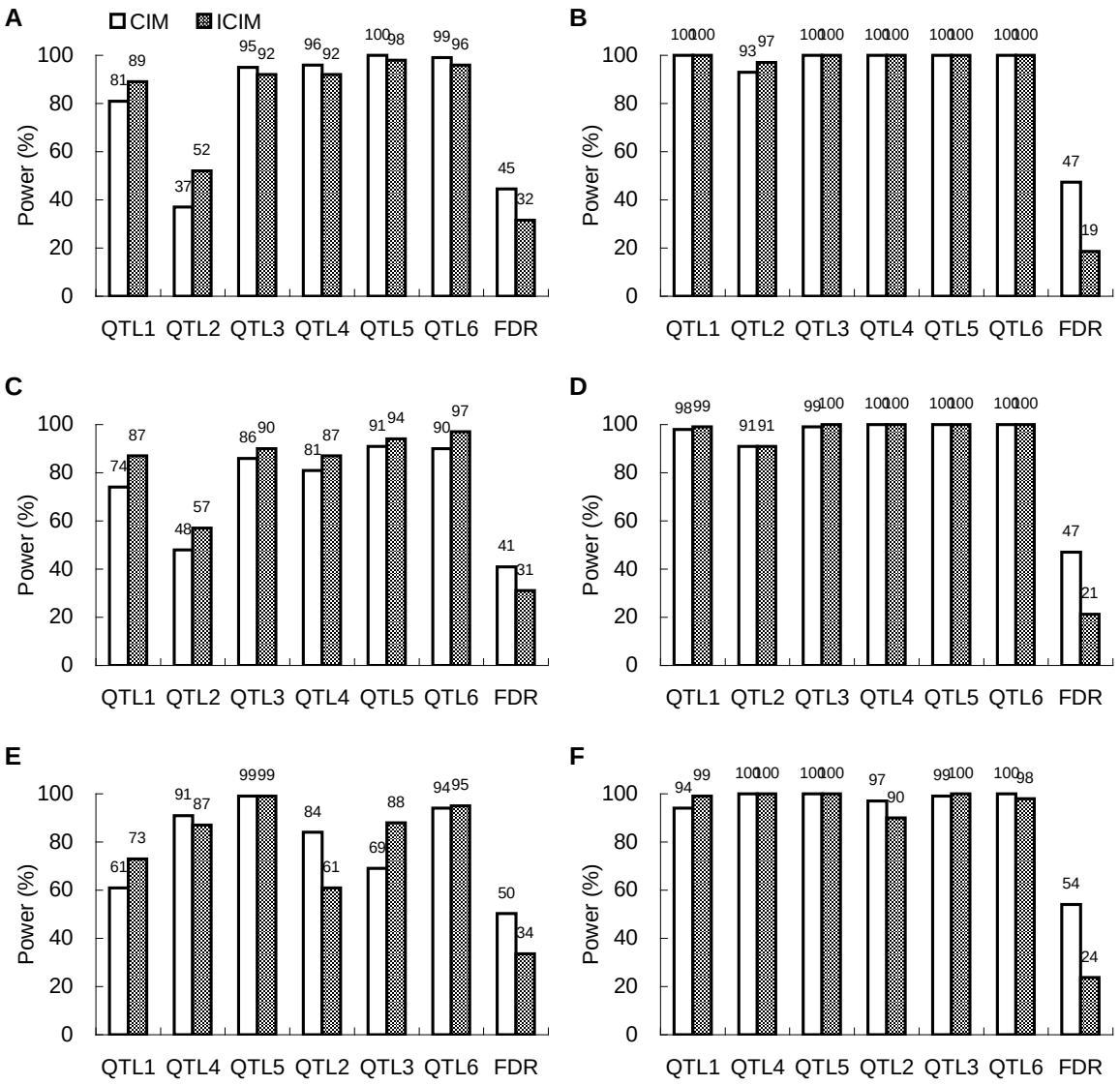
Mapping results of the simulated F2 populations

QTL	LOD score	PVE (%)	True Position (cM)	Est. Position (cM)	True add. effect	Est. add. effect	True dom. effect	Est. dom. effect
QTL distribution model I								
QTL1	16.52	6.67	25	28	1	0.88	0	-0.11
QTL2	7.67	3.27	55	53	0	0.03	1	0.85
QTL3	25.11	11.28	25	24	1	0.86	1	1.08
QTL4	35.46	16.43	55	57	1	0.74	-1	-1.58
QTL5	37.12	16.74	25	26	1	1.05	1.5	1.38
QTL6	28.44	13.16	55	55	1	0.84	-1.5	-1.22

Comparison of ICIM with IM



Power analysis of CIM and ICIM from 100 simulations



An actual rice F_2 population

- 180 individuals
- The cross was made in Chengdu, China, in July 2002 between the *indica* rice variety and Nipponbare.
- 137 SSR markers.
- The whole genome was of 2046.2 cM, and the average marker distance was 17.1 cM.
- A number of agronomic traits were investigated in the field.

QTL mapping in the actual F_2 population

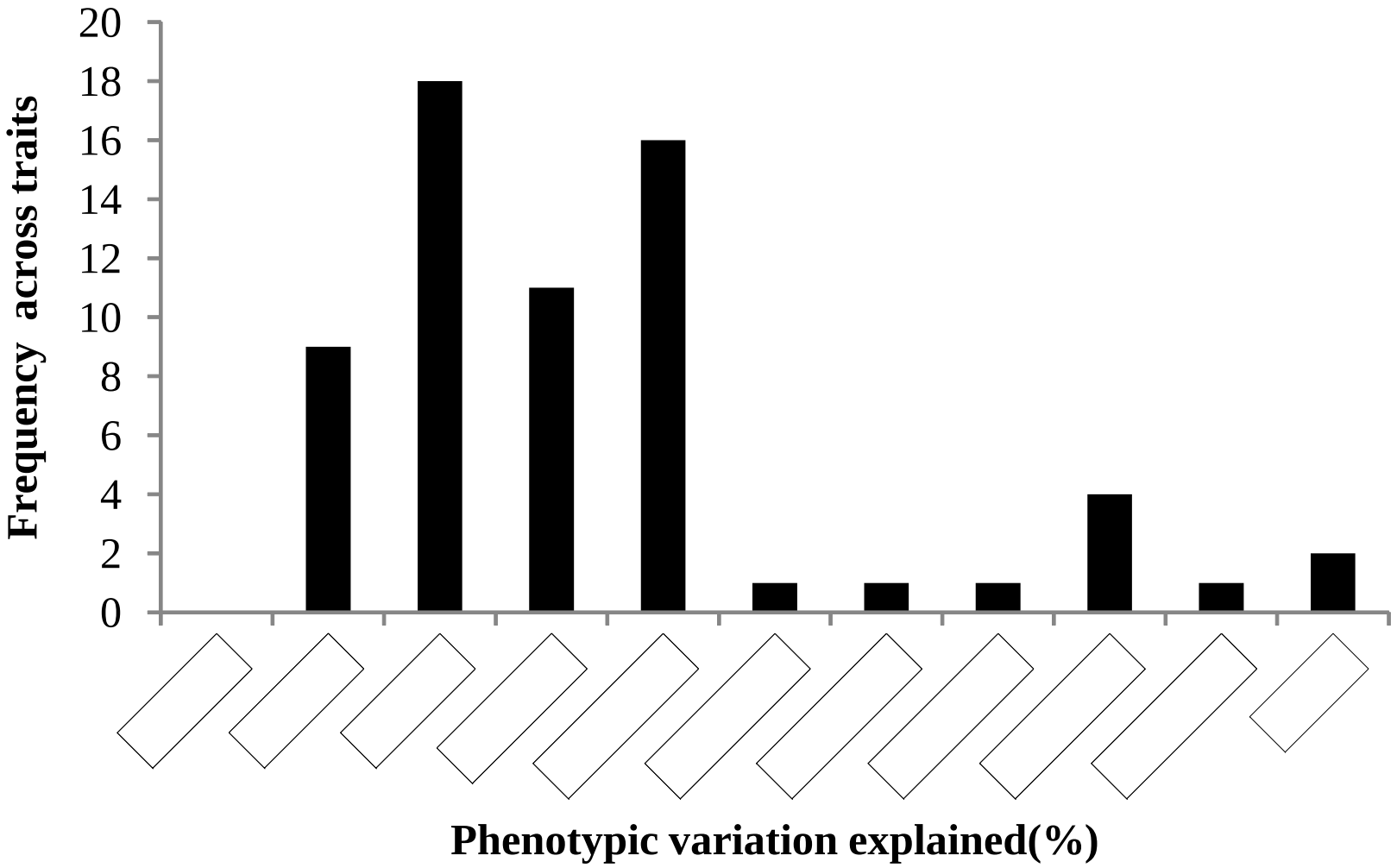
- In the real population, few F_2 individuals are shorter than PA64s, indicating most, if not all, reduced height alleles are harbored by PA64s.
- Set the LOD threshold at 3.0. The two probabilities for entering and removing variables were set at 0.01 and 0.02. Scanning step is 1 cM.

QTL distribution

Trait	R^2 of additive (%)	R^2 of additive and dominance (%)	Absolute degree of dominance (|d/a|)				Total
			<=0.25	(0.25, 0.75]	(0.75, 1.25]	>1.25	
PH	25.84	51.56	2	1	1	5	9
HD	16.12	41.37	1	1	1	3	6
PL	25.58	61.26	5	3	1	8	17
FL	20.86	40.00	0	2	0	3	5
SPK	25.64	27.09	1	1	1	1	4
TKW	20.11	20.11	2	0	2	1	5
DP	19.45	24.87	1	1	0	1	3
GL	30.69	41.96	1	1	0	0	2
GW	26.63	26.63	2	2	0	0	4
RLW	37.63	45.70	1	3	1	1	6
		Total	16	15	7	23	61

Notes: PH=plant height, HD=heading date, PL= panicle length, FL=flag leaf length, SPK=spikelets per panicle, TKW=thousand kernel weight, DP=density of panicle, GL=grain length, GW=grain width, RLW=ratio of grain length and width.

Distribution of PVE (%) of QTLs



QTL for height and maturity

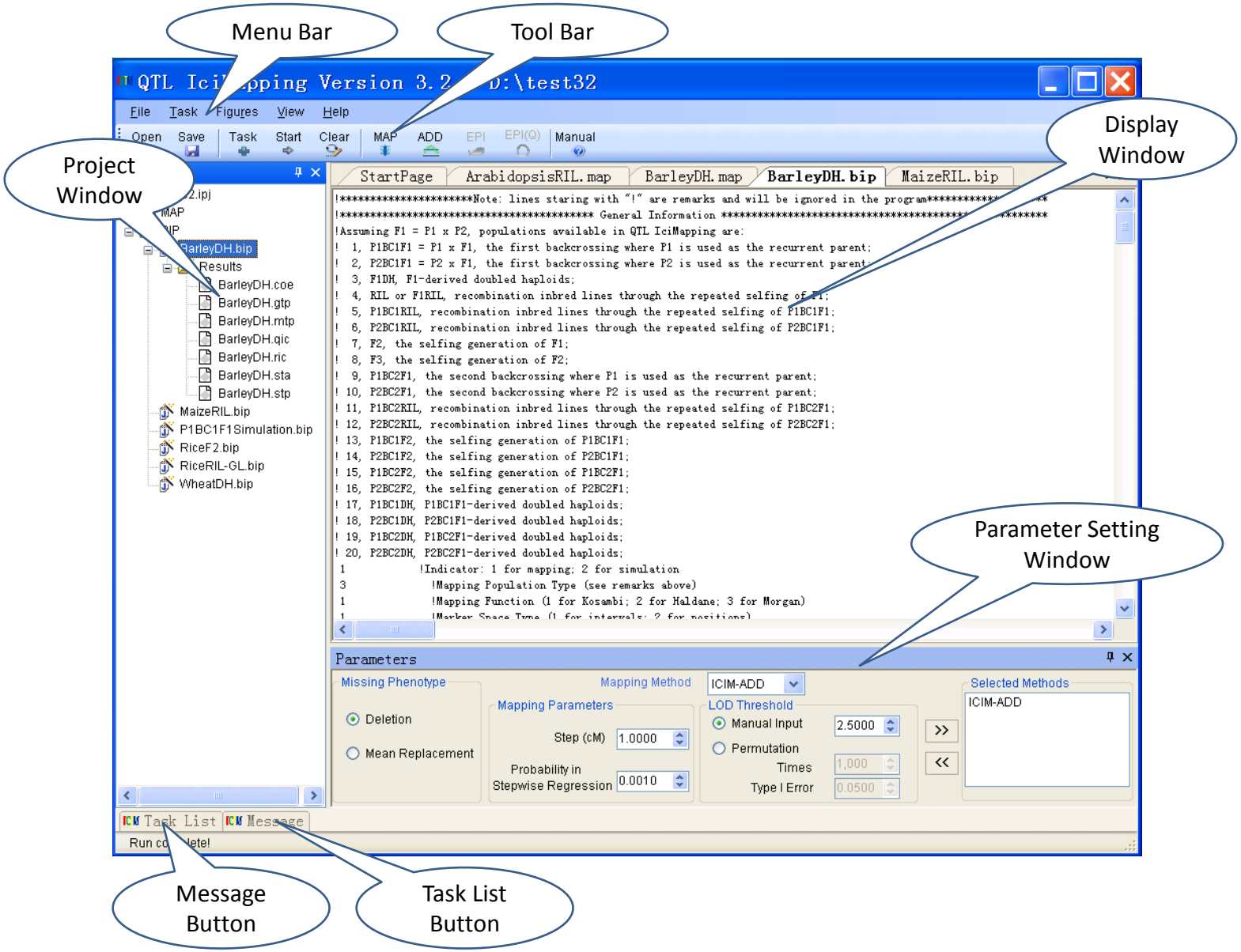
Trait	QTL	Chr	Distance to left marker	Add	Dom	LOD	PVE(%)
Plant height (Ph)	QPh1-1	1	12	-0.57	-7.98	8.04	12.03
	QPh1-2	1	19.5	-8.59	0.59	15.54	25.57
	QPh3-1	3	16.9	4.35	-4.86	6.51	13.30
	QPh3-2	3	11.4	-4.69	-1.00	5.04	6.84
	QPh4	4	13.7	-3.56	-2.09	4.61	5.53
	QPh5	5	13	-0.44	-4.48	3.13	3.86
	QPh6	6	6.2	-0.79	-5.05	3.17	4.96
	QPh7	7	7	0.26	6.48	5.27	7.56
	QPh12	12	2.4	-1.66	3.93	3.98	5.44
Heading date (Hd)	QHd1	1	22.1	1.74	-0.30	3.65	7.27
	QHd3	3	19.9	0.88	-3.70	6.04	21.09
	QHd4	4	0.2	-0.77	1.85	3.58	5.24
	QHd8	8	5.7	-1.41	-1.46	4.79	8.20
	QHd10	10	0.3	-1.78	-0.80	4.85	7.21
	QHd11	11	6.2	0.15	-3.03	5.71	11.70

The BIP functionality in QTL IciMapping

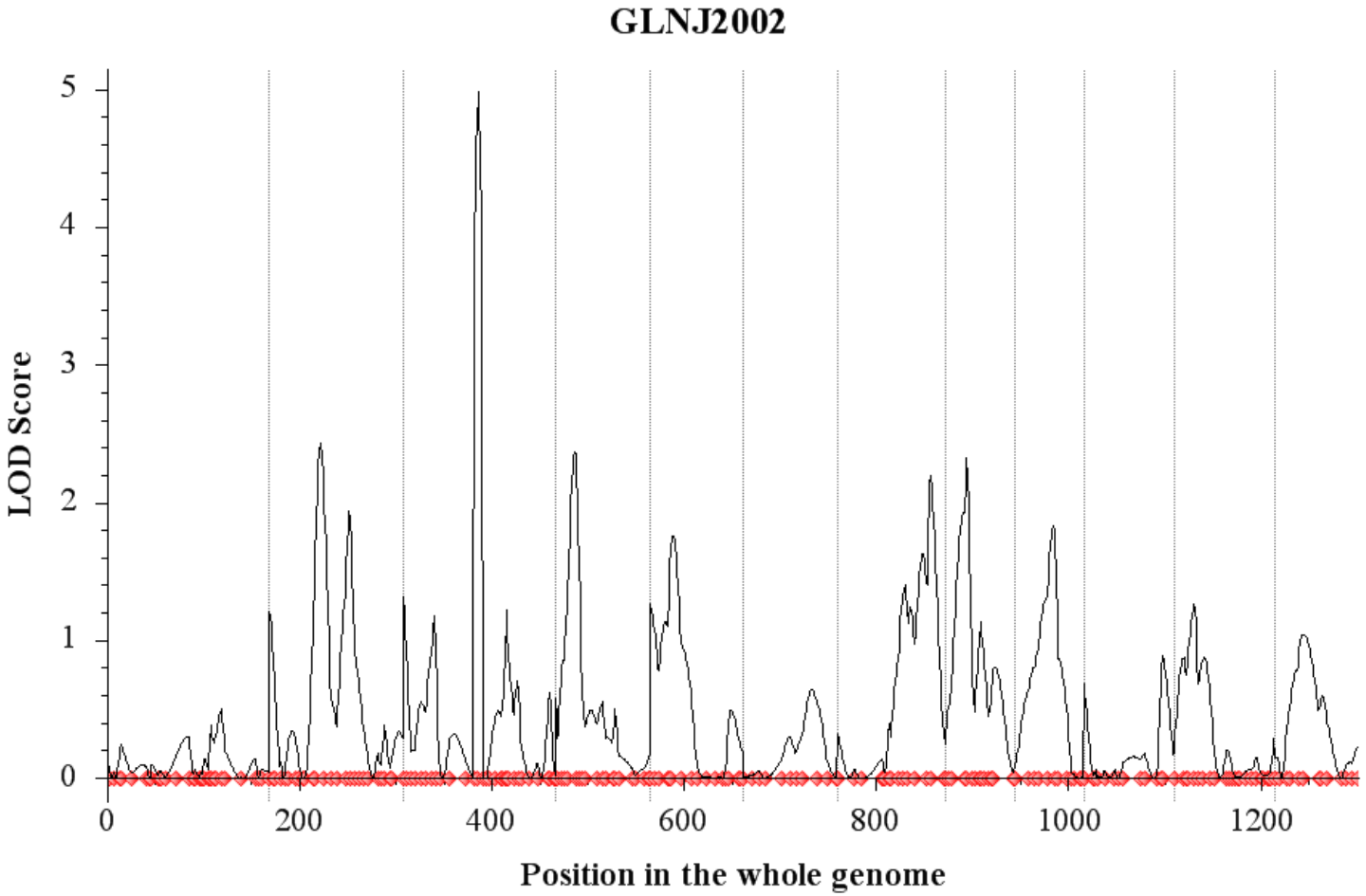
Six methods in BIP

- SMA: single marker analysis (Soller et al., 1976. Theor. Appl. Genet. 47: 35-39)
- IM-ADD: the conventional simple interval mapping (Lander and Botstein, 1989. Genetics 121: 185-199)
- 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)
- IM-EPI: interval mapping of digenic epistatic QTL
- ICIM-EPI: inclusive composite interval mapping of digenic epistatic QTL (Li et al., 2008. Theor. Appl. Genet. 116: 243-260)
- SGM: selective genotyping mapping (Lebowitz et al., 1987. Theor. Appl. Genet. 73: 556–562)

Interface of the BIP functionality



LOD profile of ICIM additive mapping (ICIM-ADD)



Figures of interacting QTL from ICIM epistatic mapping (ICIM-EPI)