


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



# Principle of QTL mapping and Inclusive Composite Interval Mapping (ICIM)

Jiankang Wang  
CIMMYT China and CAAS

E-mail: [wangjk@caas.net.cn](mailto:wangjk@caas.net.cn) or [jkwang@cgiar.org](mailto:jkwang@cgiar.org)



# Outlines

---

- **Quantitative traits and QTL mapping**
- **Inclusive composite interval mapping (ICIM) for additive and interacting QTL**
- **Selected publications using ICIM**
- **The BIP functionality in QTL IciMapping**



---

# 1. Quantitative traits and QTL mapping



# Quantitative traits

---

- **Continuous phenotypic variation**
- **Affected by many genes**
- **Affected by environment**
- **Epistasis**
- **Polygene (or multi-factorial ) hypothesis**
- **Classical quantitative genetics**



# What is QTL Mapping?

---

- **The procedure to map individual genetic factors with small effects on the quantitative traits, to specific chromosomal segments in the genome**
- **The key questions in QTL mapping studies are:**
  - **How many QTL are there?**
  - **Where are they in the marker map?**
  - **How large an influence does each of them have on the trait of interest?**



# Dataset of QTL mapping

---

- Mapping population
- Marker data of each individual in the mapping population
- Linkage map
- Phenotypic data





# Classification of mapping populations

---

- Bi-parental mapping populations (**linkage mapping**)
  - Temporary population: F2 and BC
  - Permanent population: RIL, DH, CSSL
  - Secondary population
- **Association mapping**
  - Natural populations: human and animals





# Overview on QTL mapping methods

---

## ➤ **Single marker analysis (Sax 1923; Soller et al. 1976)**

The single marker analysis identifies QTLs based on the difference between the mean phenotypes for different marker groups, but cannot separate the estimates of recombination fraction and QTL effect.

## ➤ **Interval mapping (IM) (Lander and Botstein 1989)**

IM is based on maximum likelihood parameter estimation and provides a likelihood ratio test for QTL position and effect. The major disadvantage of IM is that the estimates of locations and effects of QTLs may be biased when QTLs are linked.

## ➤ **Regression interval mapping (RIM)**

**(Haley and Knott 1992; Martinez and Curnow 1992 )**

RIM was proposed to approximate maximum likelihood interval mapping to save computation time at one or multiple genomic positions.



➤ **Composite interval mapping (CIM) (Zeng 1994)**

CIM combines IM with multiple marker regression analysis, which controls the effects of QTLs on other intervals or chromosomes onto the QTL that is being tested, and thus increases the precision of QTL detection.

➤ **Multiple interval mapping (MIM) (Kao et al. 1999)**

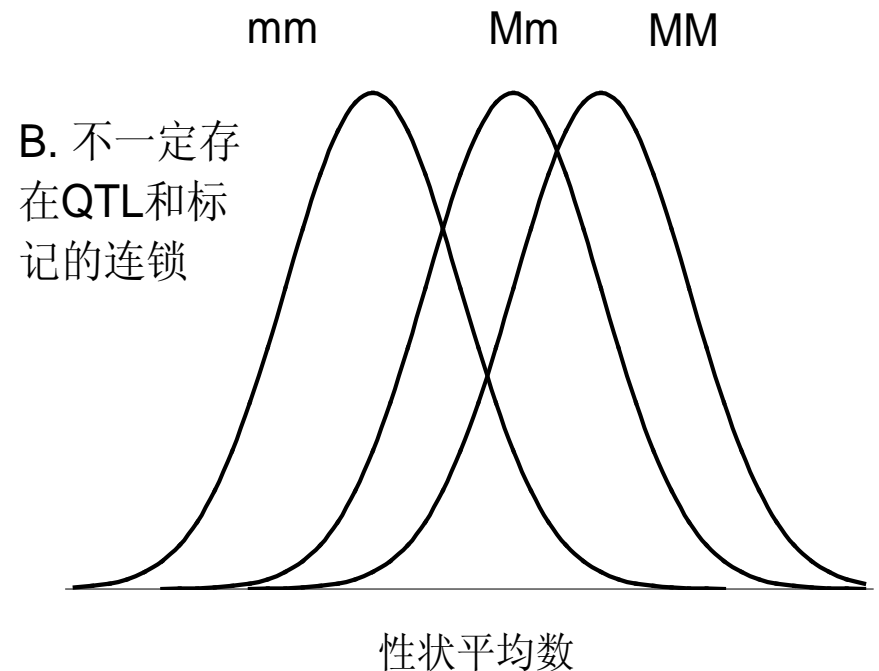
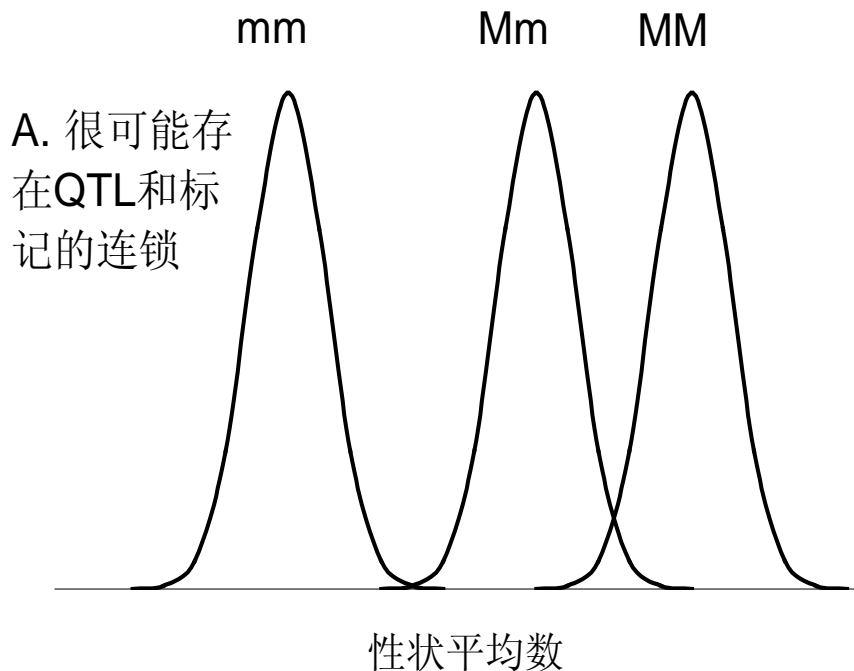
MIM is a state-of-the-art gene mapping procedure. But implementation of the multiple-QTL model is difficult, since the number of QTL defines the dimension of the model which is also an unknown parameter of interest.

➤ **Bayesian model (Sillanpää and Corander 2002)**

In any Bayesian model, a prior distribution has to be considered. Based on the prior, Bayesian statistics derives the posterior, and then conduct inference based on the posterior distribution. However, Bayesian models have not been widely used in practice, partially due to the complexity of computation and the lack of user-friendly software.

# Principle of QTL mapping

- Three marker types at one marker locus



# Backcrosses (P1BC1 and P2BC1) of P1: MMQQ and P2: mmqq

BC <sub>1</sub>			BC <sub>2</sub>		
Genotype	Frequency	Genotypic value	Genotype	Frequency	Genotypic value
MMQQ	$\frac{1}{2}(1-r)$	$m+a$	MmQq	$\frac{1}{2}(1-r)$	$m+d$
MMQq	$\frac{1}{2}r$	$m+d$	Mmqq	$\frac{1}{2}r$	$m-a$
MmQQ	$\frac{1}{2}r$	$m+a$	mmQq	$\frac{1}{2}r$	$m+d$
MmQq	$\frac{1}{2}(1-r)$	$m+d$	mmqq	$\frac{1}{2}(1-r)$	$m-a$



# Principle of single marker analysis (P1BC1 as example)

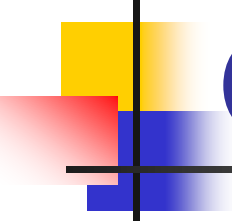
- Two marker types:

$$\begin{aligned}\mu_{MM} &= (1-r)\mu_{MMQQ} + r\mu_{MMQq} \\ &= (1-r)(m+a) + r(m+d) = m + (1-r)a + rd\end{aligned}$$

$$\begin{aligned}\mu_{Mm} &= r\mu_{MmQQ} + (1-r)\mu_{MmQq} \\ &= r(m+a) + (1-r)(m+d) = m + ra + (1-r)d\end{aligned}$$

- Difference in phenotype between the two types

$$\mu_{MM} - \mu_{Mm} = (1-2r)(a-d)$$



# Interval mapping (IM) (Lander and Botstein 1989)

---

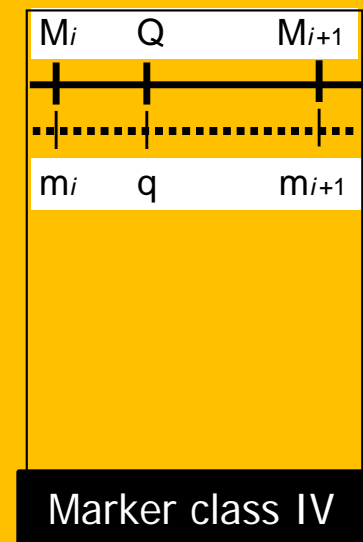
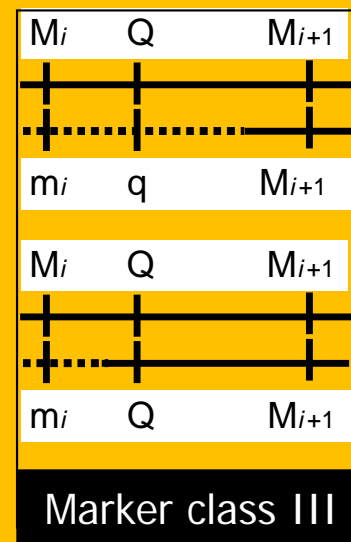
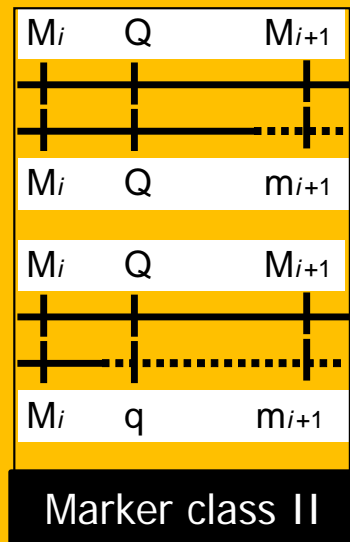
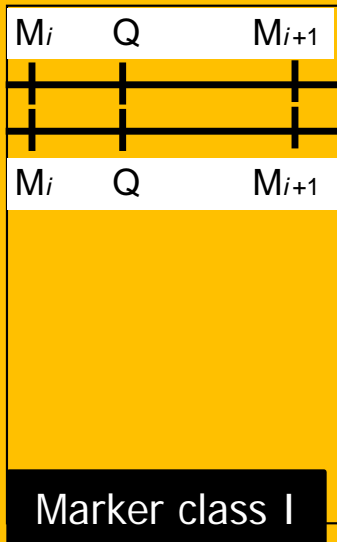
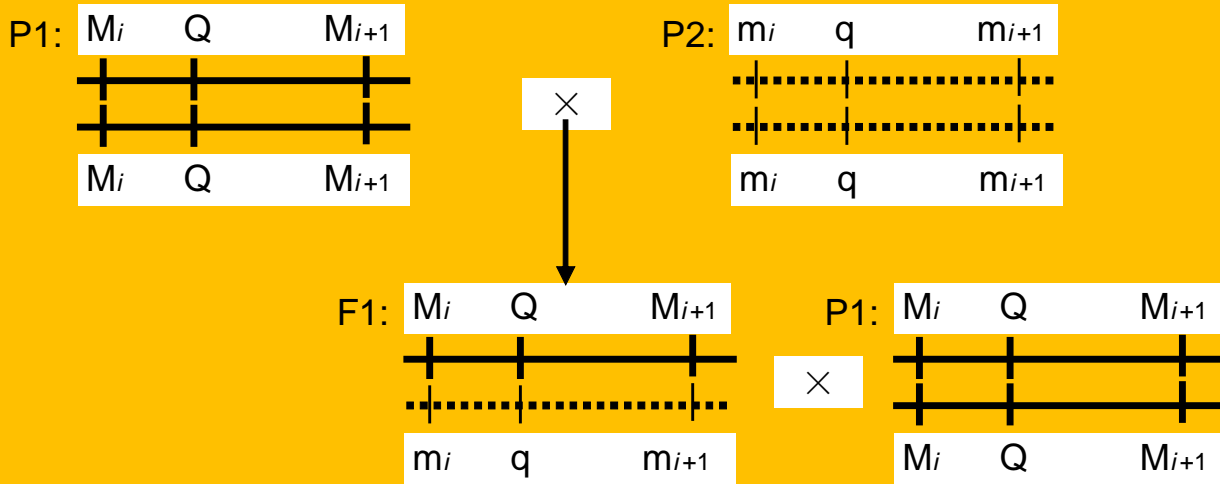
- Linear model ( $j=1, 2, \dots, n$ )

$$y_i = b_0 + b^* x_j^* + e_j$$

$b^*$  represent QTL effect,  $x_j^*$  is the indicator variable (0 or 1) for QTL genotype

- Likelihood profile
- Support interval: One-LOD interval

# QTL genotypes under each marker type in P1BC1 (double crossover not considered)





---

## **2. Inclusive Composite Interval Mapping (ICIM)**





# Problems with IM

---

- Assumption: No more than one QTL per chromosome or linkage group
- “Ghost QTL” for linked QTL
- Large confidence interval
- Biased effect estimation
  
- Composite interval mapping (CIM) (Zeng 1994)



# Problems with CIM

---

- In the algorithm of CIM, both QTL effect at the current testing position and regression coefficients of the marker variables used to control genetic background were estimated simultaneously in an expectation and maximization (EM) algorithm.
- Thus, this algorithm could not completely ensure that the effect of QTL at current testing interval was not absorbed by the background marker variables and therefore may result in biased estimation of the QTL effect.



# Theoretical basis of ICIM

---

$$G = \sum_{j=1}^m a_j g_j + \sum_{j < k} a a_{jk} g_j g_k$$

$$E(g_j | \mathbf{X}) = \lambda_j x_j + \rho_j x_{j+1}$$

$$E(g_j g_k | \mathbf{X}) = \lambda_j \lambda_k x_j x_k + \lambda_j \rho_k x_j x_{k+1} + \rho_j \lambda_k x_{j+1} x_k + \rho_j \rho_k x_{j+1} x_{k+1}$$

$$y_i = b_0 + \sum_{j=1}^{m+1} b_j x_{ij} + \sum_{j < k} b_{jk} x_{ij} x_{ik} + e_i$$

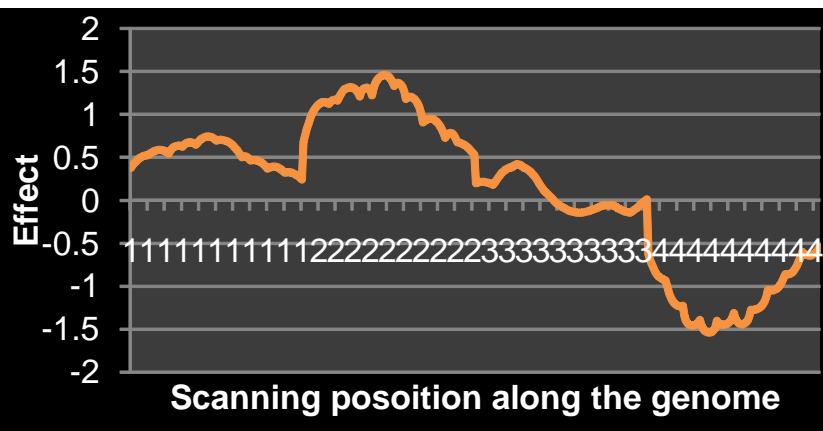
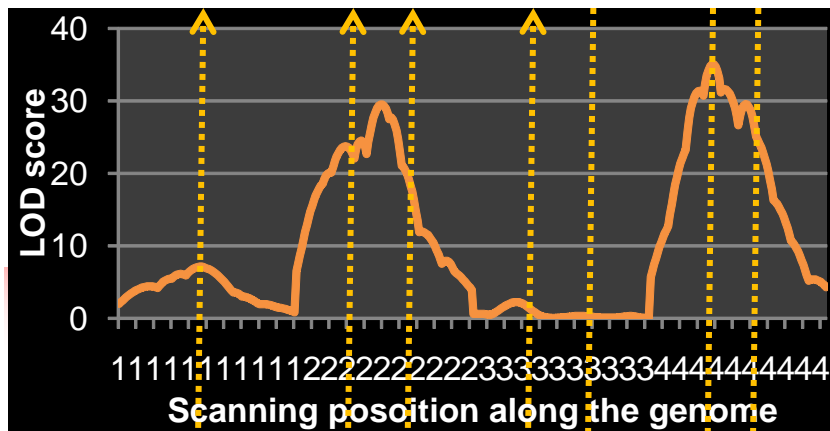
# Genomic scanning for additive and interacting QTL

- One-dimensional scanning (interval mapping)

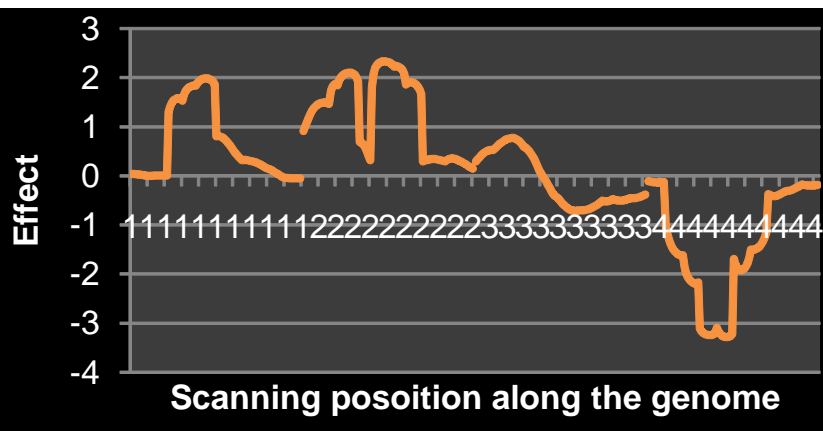
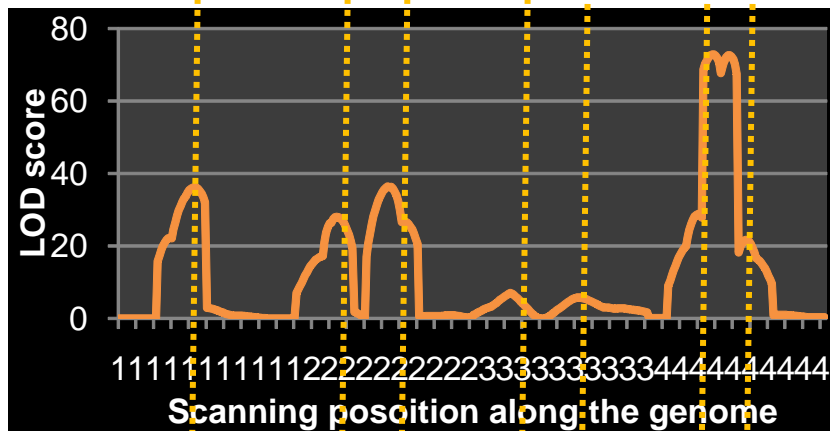
$$\Delta y_i = y_i - \sum_{j \neq k, k+1} \hat{b}_j x_{ij}$$

- Two-dimensional scanning (interval mapping)

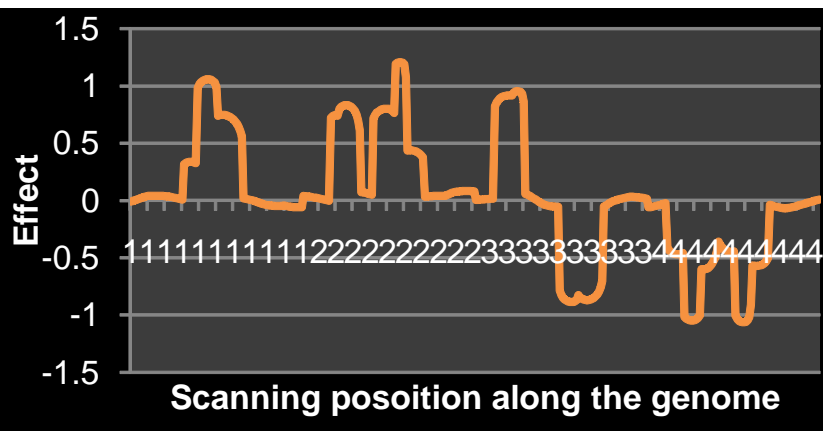
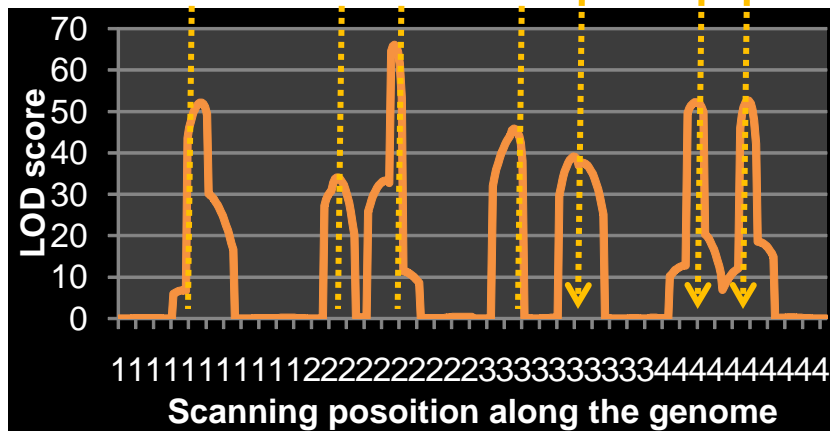
$$\Delta y_i = y_i - \sum_{r \neq j, j+1, k, k+1} \hat{b}_r x_{ir} - \sum_{\substack{r \neq j, j+1 \\ s \neq k, k+1}} \hat{b}_{rs} x_{ir} x_{is}$$



IM

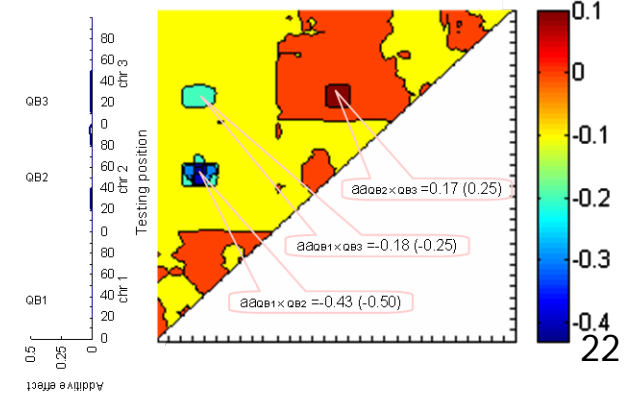
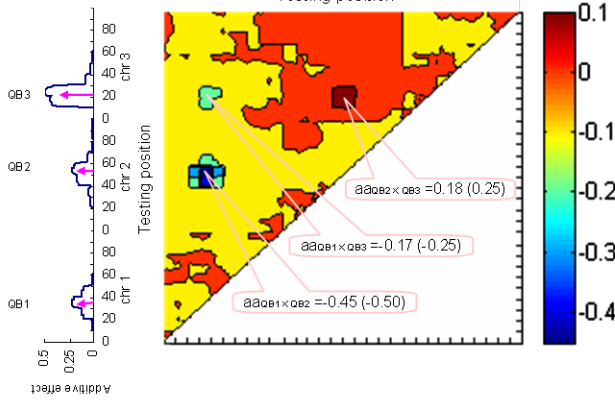
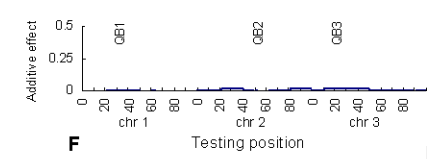
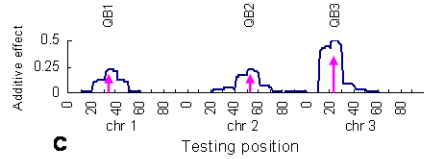
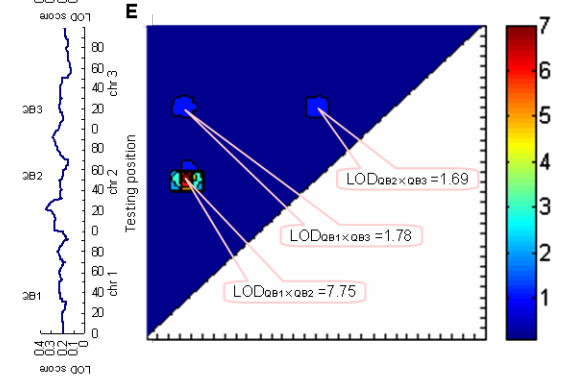
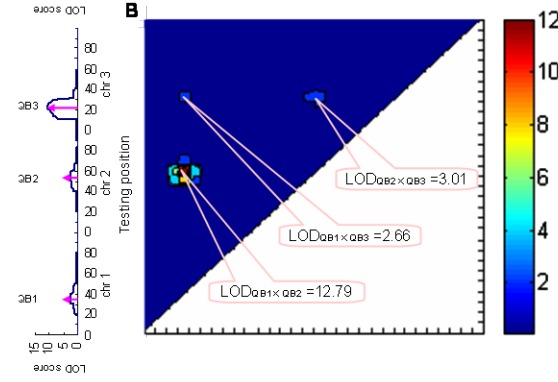
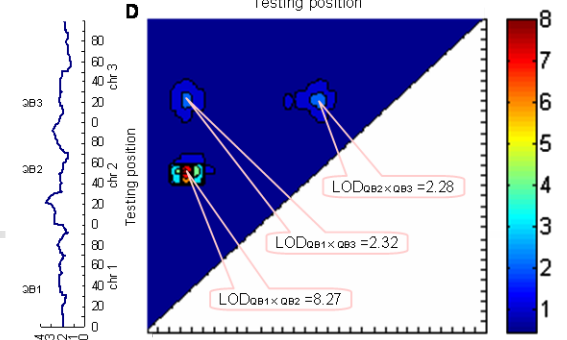
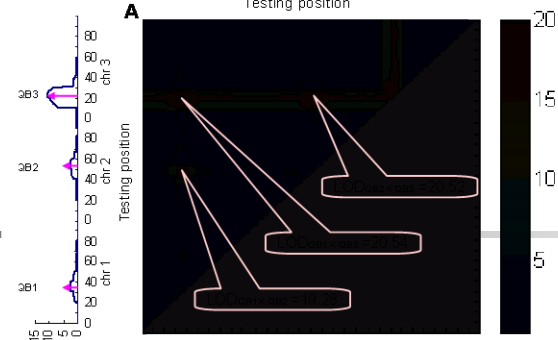
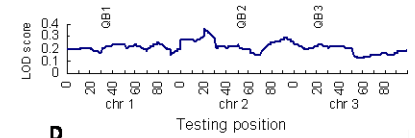
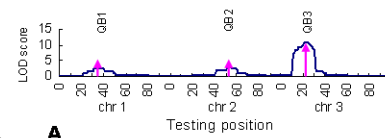


CIM



ICIM

➤ Detecting epistasis where the interacting QTL don't have significant additive effects







---

## **3. Selected publications using ICIM**





# In rice

---

- Crop Science (2008) 48: 1799-1806; **Tiller angle**
- Hereditas (2009) 146: 67-73; **Brown planthopper resistance**
- Mol. Breeding (2010) 25: 287-298; **Heading date**
- Scientia Agricultura Sinica 2010,43(21): 4331-4340; **Nitrogen efficiency**

# In wheat

## ➤ Euphytica (2009) 165: 435-444; flour and noodle color components and yellow pigment content

**Table 3** QTL for flour colour components, CWSN colour and yellow pigment content detected in the RIL population from PH82-2/Neixiang 188 based on phenotypic data averaged from six environments

Trait <sup>a</sup>	QTL <sup>b</sup>	Marker interval	A <sup>c</sup>	LOD	PVE (%)
Fa*	<i>QFa-1A</i>	<i>Xwmc120-Xbarc269</i>	0.03	4.2	2.4
	<i>QFa-1B</i>	<i>Sec1-HVM23</i>	0.11	33.7	26.1
	<i>QFa-3B</i>	<i>Xbarc84-Xbarc77</i>	0.02	2.5	1.4
	<i>QFa-4A</i>	<i>Xwmc468-Xbarc170</i>	0.03	3.5	2.5
	<i>QFa-7A</i>	<i>Xwmc809-YP7A</i>	0.13	41.8	35.9
Fb*	<i>QFb-1B</i>	<i>Sec1-HVM23</i>	-0.37	5.0	7.8
	<i>QFb-7A</i>	<i>Xwmc809-YP7A</i>	-0.44	7.4	12.6
YPC	<i>QYpc-1A</i>	<i>Xwmc278-Xcfd59</i>	-0.07	2.7	1.5
	<i>QYpc-1B</i>	<i>Sec1-HVM23</i>	-0.32	40.3	31.9
	<i>QYpc-4A</i>	<i>Xcfa2026-Xgwm160</i>	-0.11	3.2	4.1
	<i>QYpc-7A</i>	<i>Xwmc809-YP7A</i>	-0.31	40.3	33.9
FCI	<i>QFci-7A</i>	<i>Xwmc809-YP7A</i>	0.43	3.1	5.5
	<i>QFci-7B</i>	<i>Xbarc50-DuPw398</i>	0.38	2.7	4.3
KJ	<i>QKj-1A</i>	<i>Xwmc120-Xbarc269</i>	0.06	3.1	3.3
	<i>QKj-1B</i>	<i>Sec1-HVM23</i>	0.13	14.3	17.2
	<i>QKj-7A</i>	<i>Xwmc809-YP7A</i>	0.14	15.7	19.3
NL*	<i>QNI-1B</i>	<i>Sec1-HVM23</i>	0.25	2.6	4.5
	<i>QNI-7A</i>	<i>Xwmc809-YP7A</i>	0.23	2.5	4.3
Na*	<i>QNa-7A</i>	<i>Xwmc809-YP7A</i>	0.13	4.7	8.6
Nb*	<i>QNb-1B</i>	<i>Sec1-HVM23</i>	-0.65	9.5	13.7
	<i>QNb-7A</i>	<i>Xwmc809-YP7A</i>	-0.77	14.1	22.0

<sup>a</sup> For abbreviations see footnote Table 1

<sup>b</sup> Nomenclature for QTL in wheat: the *Q* for QTL followed by a trait designator, a hyphen (-) and the chromosome on which the QTL is located

<sup>c</sup> Positive additive effects are associated with an increased effect from PH82-2 alleles and negative additive effects are associated with an increased effect from Neixiang 188 alleles



# More in wheat

---

- Acta Agronomica Sinica (2011) 37 (2): 294-301; **Coleoptile Length and Radicle Length**
- Crop & Pasture Science (2009) 60: 587-597; **White salted noodle quality**
- Crop & Pasture Science (2011) 62: 625-638; **Kernel morphology traits**
- Mol. Breeding (2010) 25: 615-622; **Adult-plant resistance to powdery mildew**
- Theor. Appl. Genet. (2009) 119: 1349-1359; **Adult-plant resistance to stripe rust**
- Mol. Breeding (2011) on line published; **Grain protein content and grain yield component**
- Scientia Agricultura Sinica 2011,44(14):2857-2867; **Grain yield per plant and plant height**

# In soybean

- Breeding Science (2008) 58: 355-359 ; **Salt tolerance**
- ACTA AGRONOMICA SINICA/ 2009, 35(12): 2139-2149; **Protein Related Traits**

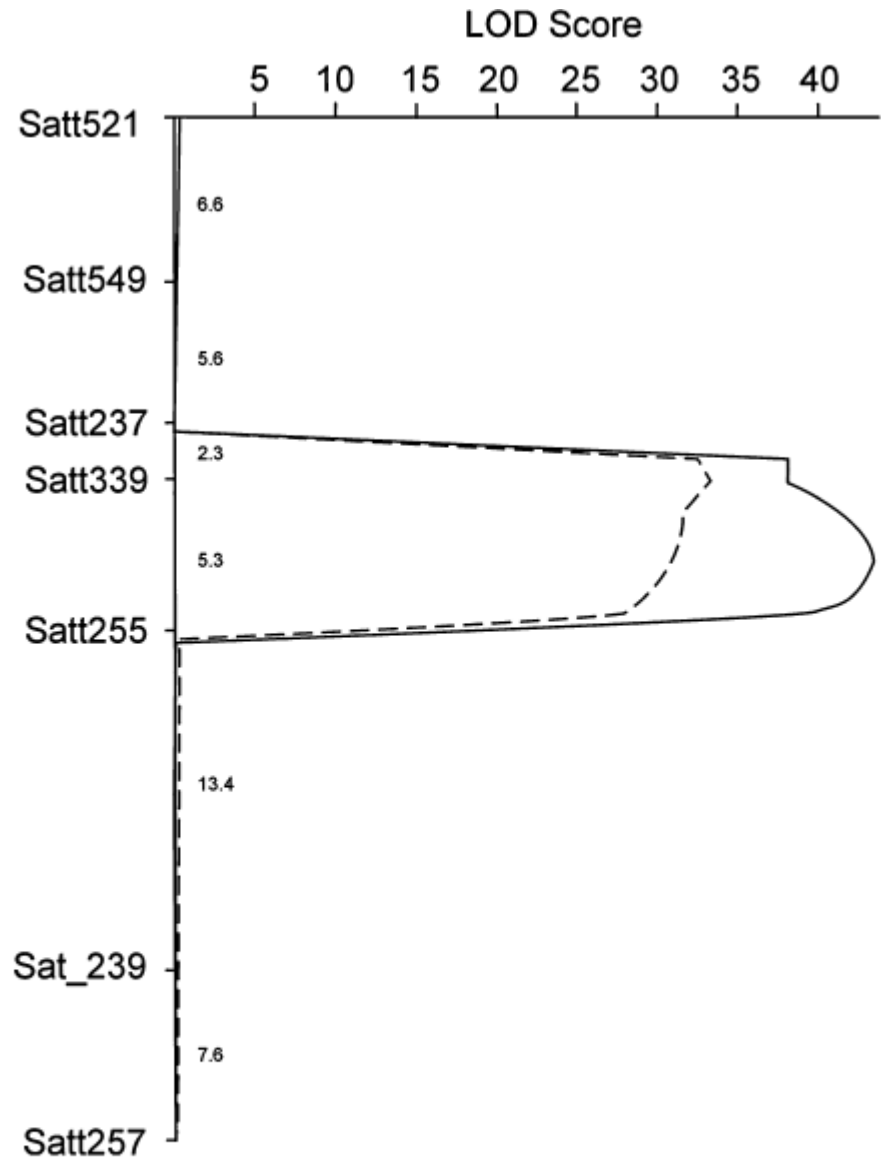


Fig. 3. QTL LOD score plots for salt-tolerance rating (STR) (solid line) and plant leaf SPAD values (dotted line) of the 225 F<sub>2</sub> plants derived from a cross between the soybean cultivar Jackson and wild soybean accession JWS156-1.



# In Maize

---

- Theor. Appl. Genet. (2011) 123: 327-338;  
**Partial restoration of male fertility of C-type cytoplasmic male sterility**
- Plant Mol. Biol. Rep. (2011) on line published;

## **Nitrogen Use Efficiency**

- HEREDITAS 32(6): 625-631; **The area of leaves**

# In melon

➤ Mol Breeding (2011) 27: 181-192; **Powdery mildew**

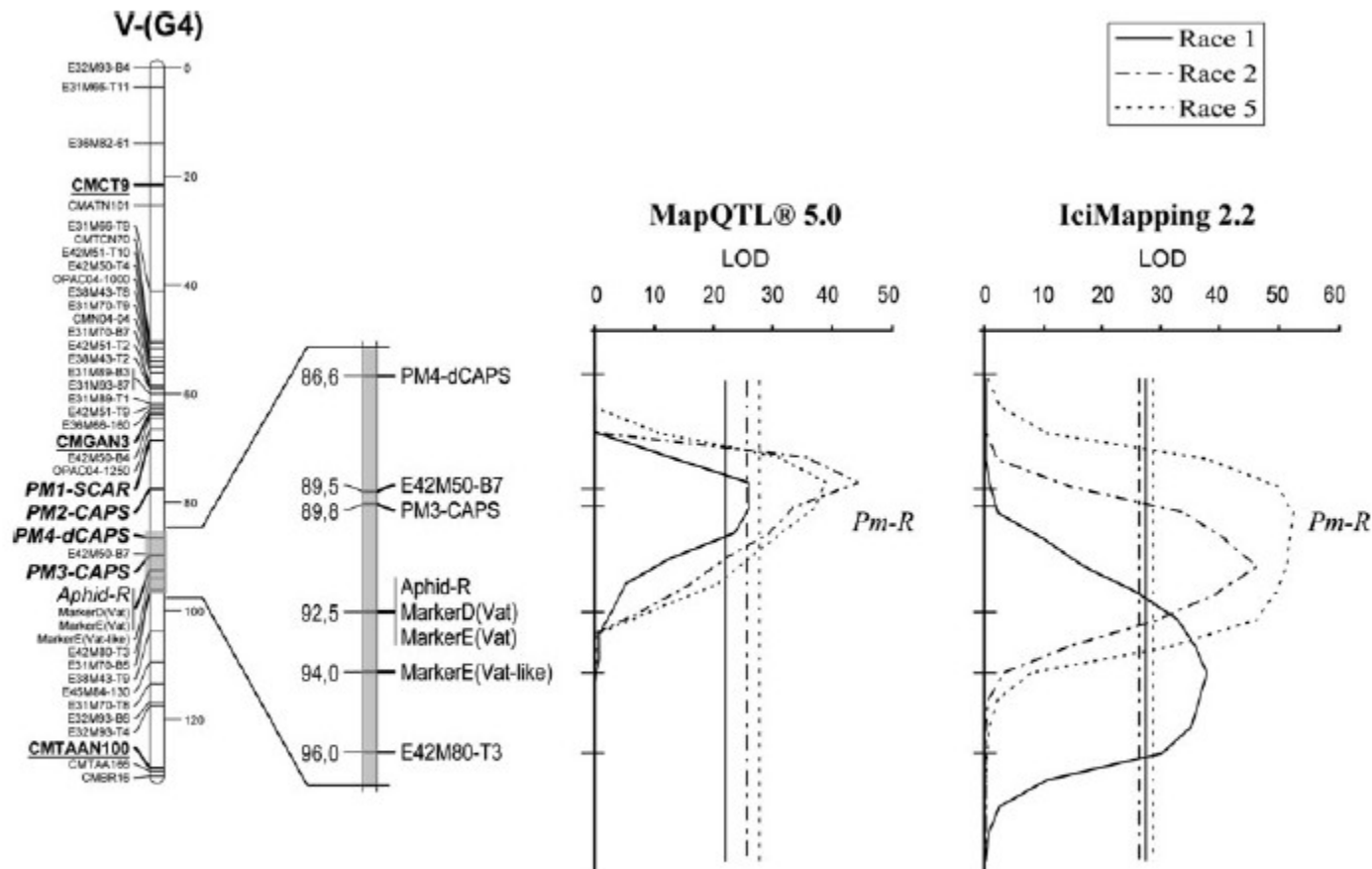


Fig. 2 *Pm-R* QTL (LG V) for powdery mildew resistance to races 1, 2, and 5 detected by using MapQTL 5.0 and IciMapping 2.2 software. The LOD values for the permutation test are also presented. Distances among the markers are indicated in cM



# Publications using RSTEP-LRT

---

- In Rice:

- Theor Appl Genet (2006) 112: 1258-1270; **Grain length**
- Plant Cell Report (2009) 28: 247-256; **Mature seed culturability**
- Mol. Breeding (2010) 25: 287-298; **Heading date**
- Plant Cell Rep. (2009) 28: 247-256; **Mature seed culturability**

- In Maize:

- Scientia Agricultura Sinica (2011),44(17):3508-3519; **Yield**

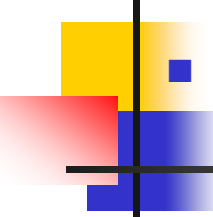


---

## 4. The BIP functionality in QTL IciMapping



# Six methods in BIP

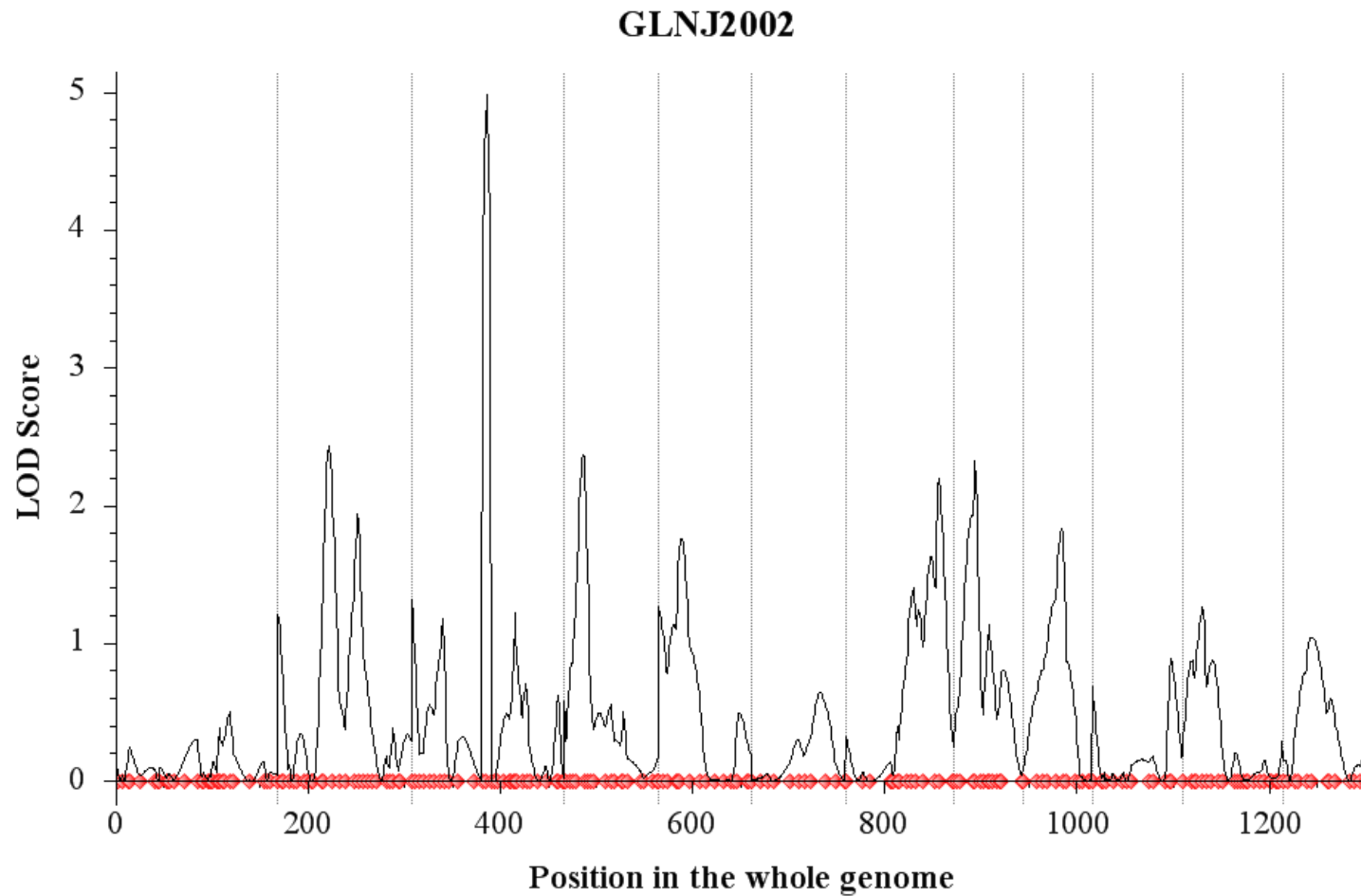
- 
- SMA: single marker analysis (Soller et al., 1976. Theor. Appl. Genet. 47: 35-39)
  - IM-ADD: the conventional simple interval mapping (Lander and Botstein, 1989. Genetics 121: 185-199)
  - ICIM-ADD: inclusive composite interval mapping of additive (and dominant) QTL (Li et al., 2007. Genetics 175: 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

The screenshot displays the QTL IciMapping Version 3.2 software interface. The window title is "QTL IciMapping Version 3.2" and the current file is "D:\test32". The interface is divided into several key areas:

- Menu Bar:** Located at the top, containing "File", "Task", "Figures", "View", and "Help".
- Tool Bar:** Below the menu bar, featuring icons for "Open", "Save", "Task", "Start", "Clear", "MAP", "ADD", "EPI", "EPI(Q)", and "Manual".
- Project Window:** A tree view on the left side showing a project structure for "BarleyDH.bip" and "MaizeRIL.bip".
- Display Window:** The central area showing a text-based configuration file for "BarleyDH.bip". It includes a list of 20 population types and their descriptions, such as "1, P1BC1F1 = P1 x F1, the first backcrossing where P1 is used as the recurrent parent".
- Parameter Setting Window:** A panel at the bottom right for configuring mapping parameters. It includes sections for "Missing Phenotype" (with radio buttons for "Deletion" and "Mean Replacement"), "Mapping Method" (set to "ICIM-ADD"), "Mapping Parameters" (with sliders for "Step (cM)" at 1.0000 and "Probability in Stepwise Regression" at 0.0010), "LOD Threshold" (with radio buttons for "Manual Input" at 2.5000 and "Permutation" at 1,000), and "Selected Methods" (containing "ICIM-ADD").
- Message Button:** A button at the bottom left labeled "Message".
- Task List Button:** A button at the bottom left labeled "Task List".

# LOD profile of ICIM additive mapping (ICIM-ADD)



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

