

Efficient Use of Known Gene Information in Plant Breeding

Jiankang Wang

CIMMYT China and Institute of Crop Sciences, CAAS

E-mail: jkwang@cgiar.org; wangjk@caas.net.cn

*Related GCP project – SP3 Commissioned G4008.14: Breeding for drought tolerance with known gene information (**Co-PI: Scott Chapman**)*

Objectives of the project

- To develop simulation tools for direct application in the study of drought adaptation
- To improve the QTL detection power and to conduct intensive QTL mapping studies in order to acquire precise gene information and build increasingly realistic and robust genetic models
- To apply the simulation tools in ongoing breeding programs and to train breeders to be able to optimize their ongoing drought tolerance breeding programs

Inclusive Composite Interval Mapping (ICIM) of Quantitative Trait Genes or Loci

**(Genetics, 2007, 175: 361-374
TAG, 2008, 116: 243–260)**

Simple Interval mapping (IM) and the associated assumption

- IM proposed by Lander and Botstein (1989)
- 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) proposed by Zeng (1993, 1994)

GSA A Publication of The Genetics Society of America

GENETICS

Search

[Advanced Search](#)

Home [Journal Information](#) [Subscriptions & Services](#) [Collections](#) [Previous Issues](#) [Current Issue](#) [Future Issues](#)

The 50 Most-Frequently Cited Articles

in Genetics as of September 1, 2009 -- updated monthly

Most-cited rankings are recalculated at the beginning of the month. Rankings are based on citations to articles on this journal site from articles in HighWire-hosted journals.

- 11. Ziheng Yang, Rasmus Nielsen, Nick Goldman, Anne-Mette Krabbe Pedersen
Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites
Genetics May 01, 2000; 155: 431-449.
(In "INVESTIGATIONS") [\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)
- 12. Allan Force, Michael Lynch, F. Bryan Pickett, Angel Amores, Yilin Yan, John Postlethwait
Preservation of Duplicate Genes by Complementary, Degenerative Mutations
Genetics Apr 01, 1999; 151: 1531-1545.
(In "INVESTIGATIONS") [\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)
- 13. Y. X. Fu, W. H. Li
Statistical Tests of Neutrality of Mutations
Genetics Mar 01, 1993; 133: 693-709.
(In "INVESTIGATIONS") [\[Abstract\]](#) [\[PDF\]](#)
- 14. Z. B. Zeng
Precision Mapping of Quantitative Trait Loci
Genetics Apr 01, 1994; 136: 1457-1468.
(In "INVESTIGATIONS") [\[Abstract\]](#) [\[PDF\]](#)
- 15. Richard R. Hudson, Martin Kreitman, Montserrat Aguadé
A Test of Neutral Molecular Evolution Based on Nucleotide Data
Genetics May 01, 1987; 116: 153-159.
(In "INVESTIGATIONS") [\[Abstract\]](#) [\[PDF\]](#)

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.
- CIM cannot be extended to mapping epistasis.

Genetic and statistical basis of ICIM for additive and epistatic mapping

Genetic model

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

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

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

Statistical model

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

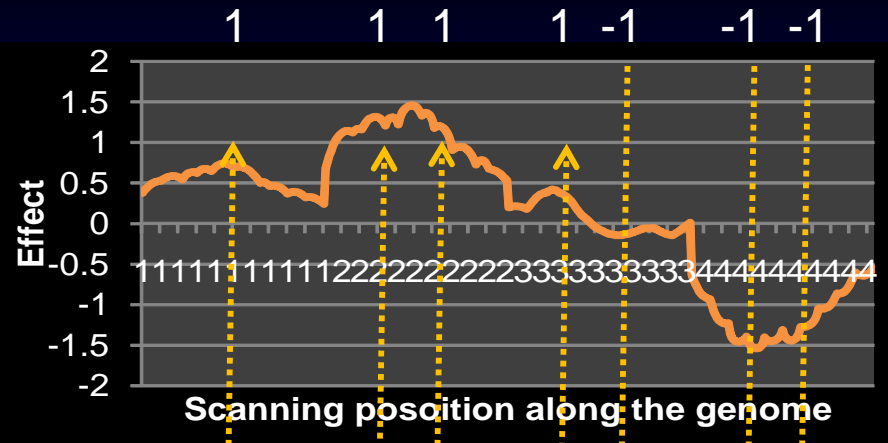
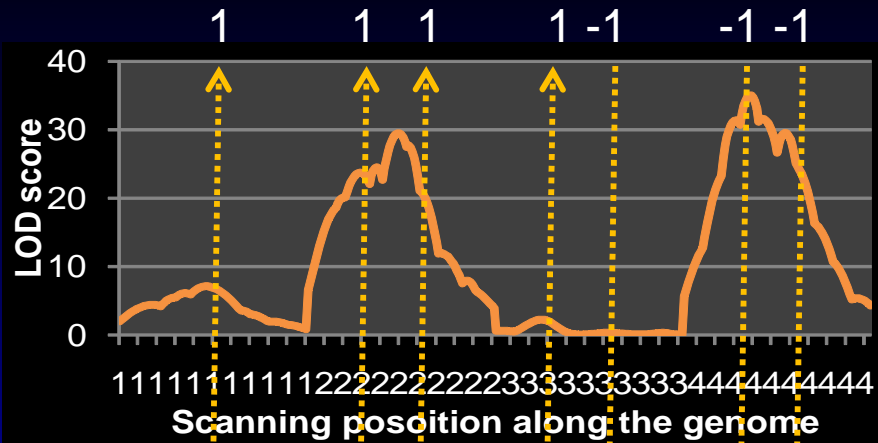
Genomic scanning for additive and interacting QTL

- One-dimensional scanning (interval mapping)

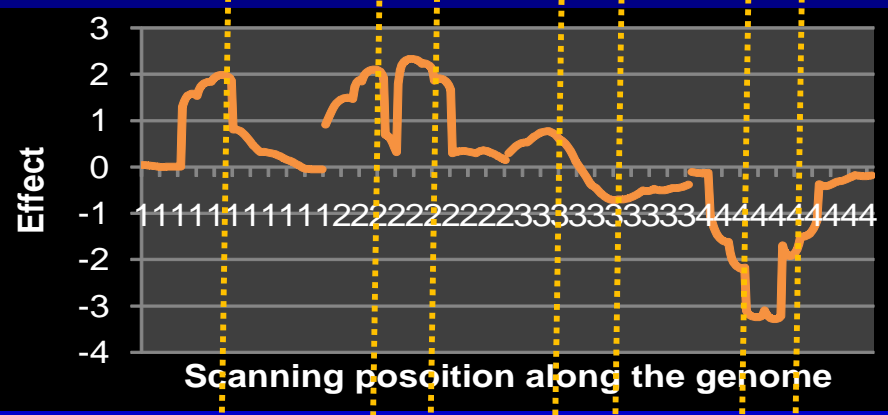
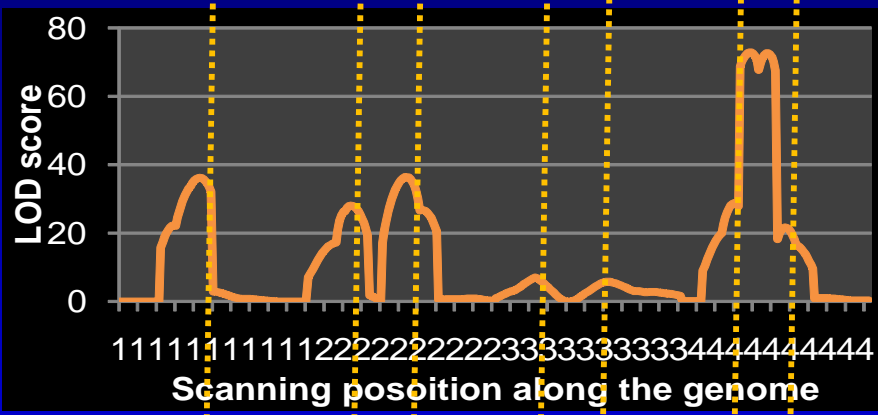
$$\Delta y_i = y_i - \sum_{j \neq k, k+1} \hat{b}_j x_{ij}$$

- Two-dimensional scanning (interval mapping)

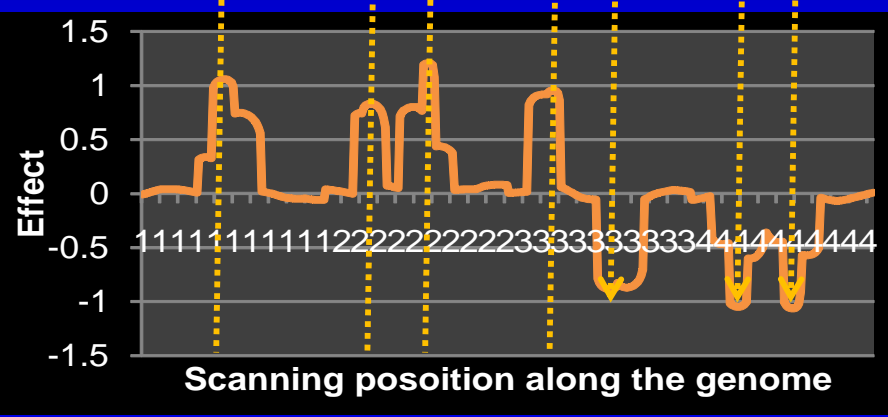
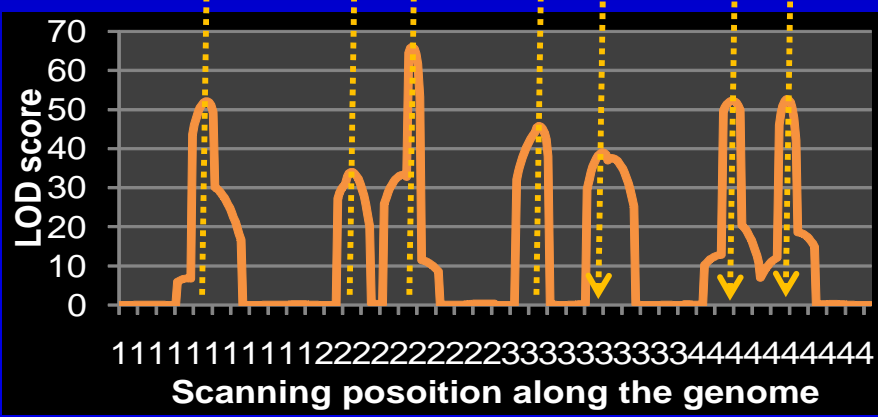
$$\Delta y_i = y_i - \sum_{r \neq j, j+1, k, k+1} \hat{b}_r x_{ir} - \sum_{\substack{r \neq j, j+1 \\ s \neq k, k+1}} \hat{b}_{rs} x_{ir} x_{is}$$



IM

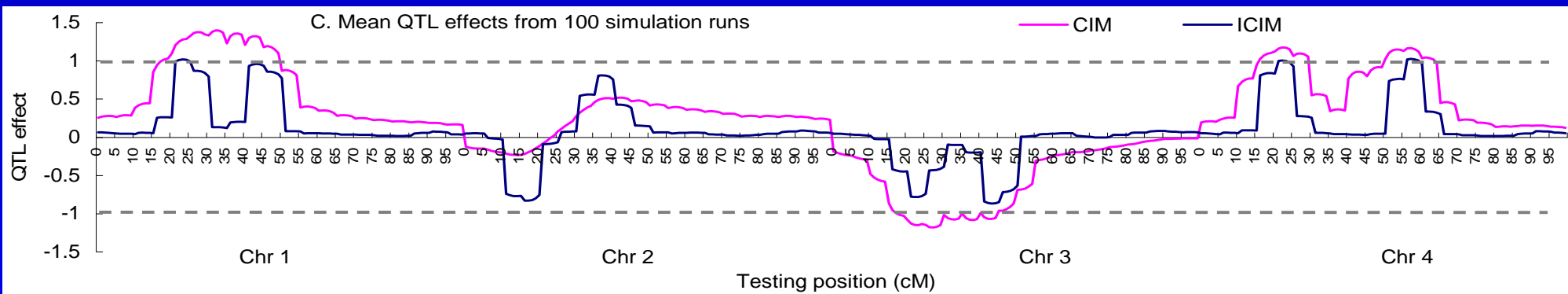
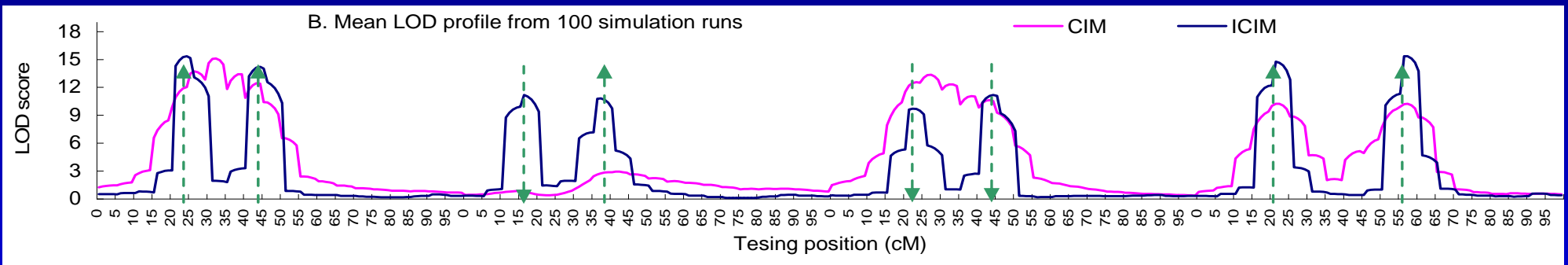
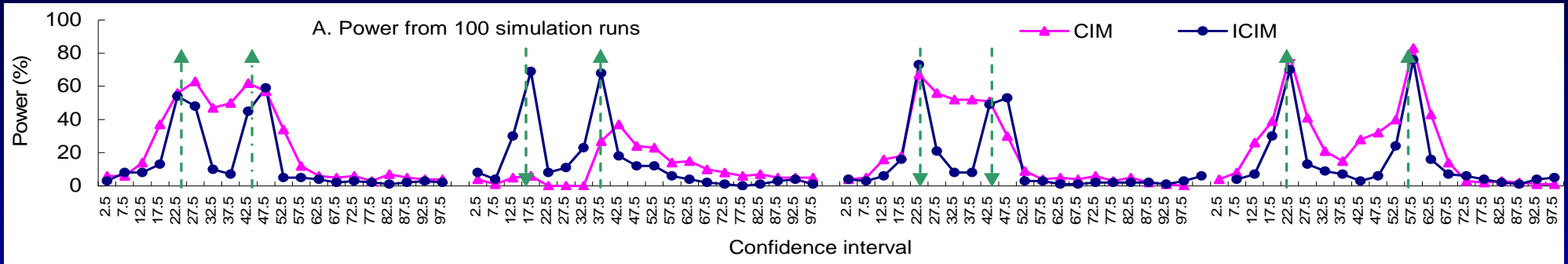


CIM



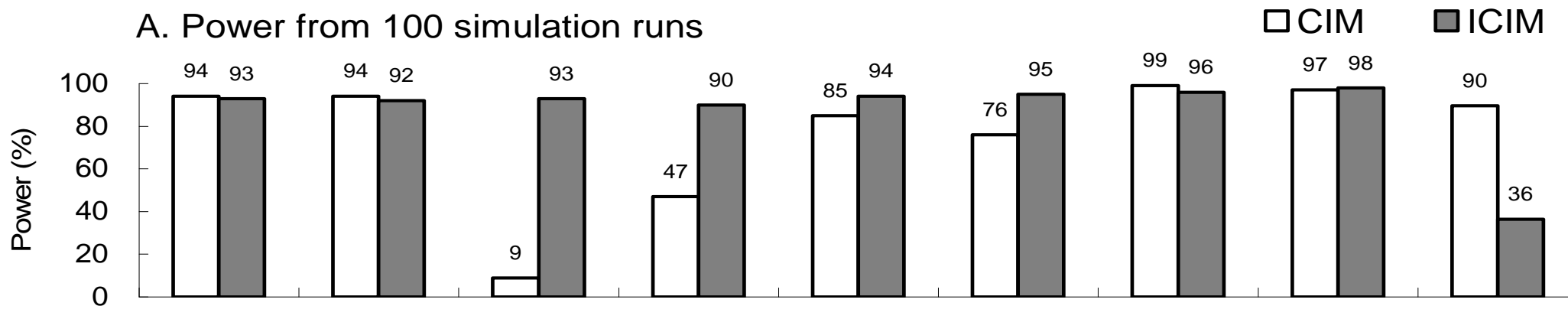
ICIM

Power comparison of ICIM and CIM

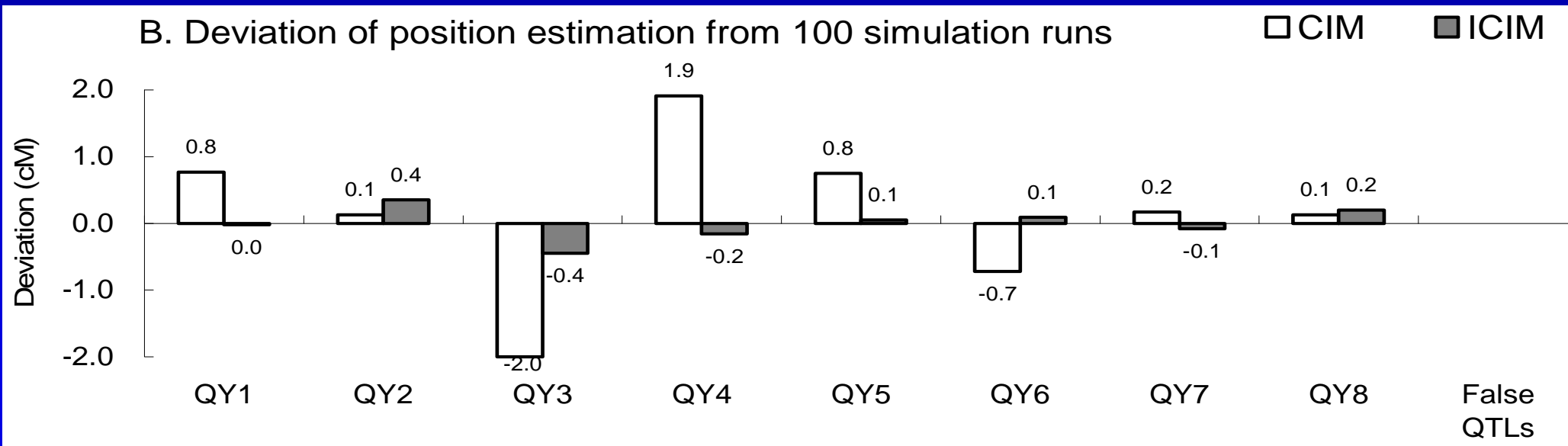


Power comparison of ICIM and CIM

A. Power from 100 simulation runs



B. Deviation of position estimation from 100 simulation runs

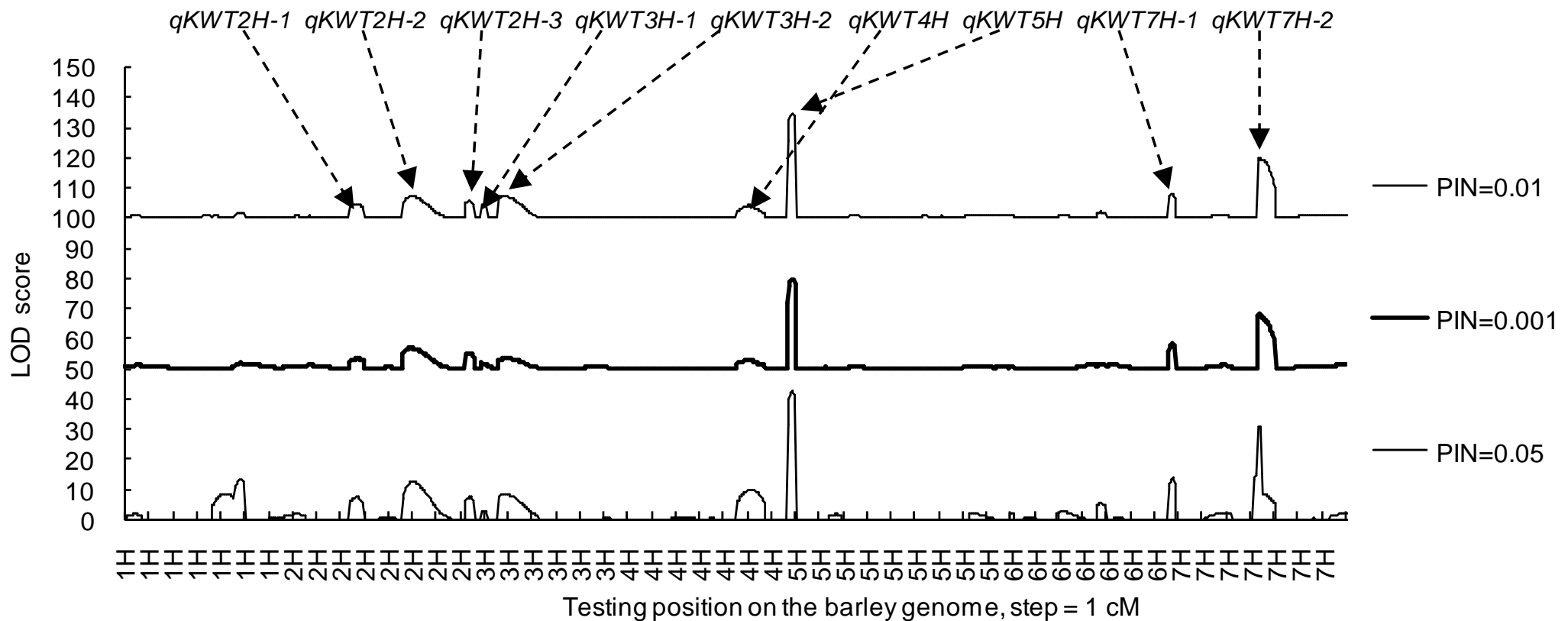


Advantages of ICIM over CIM

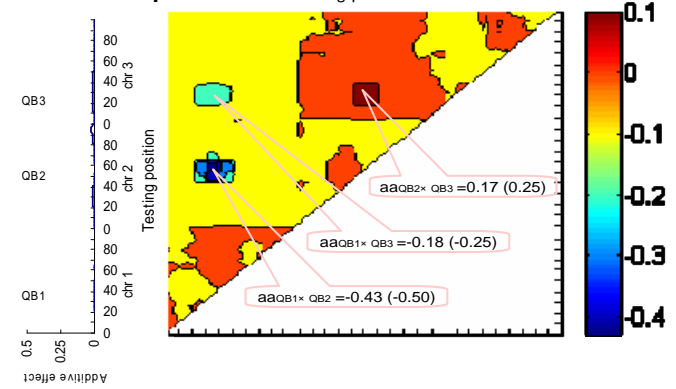
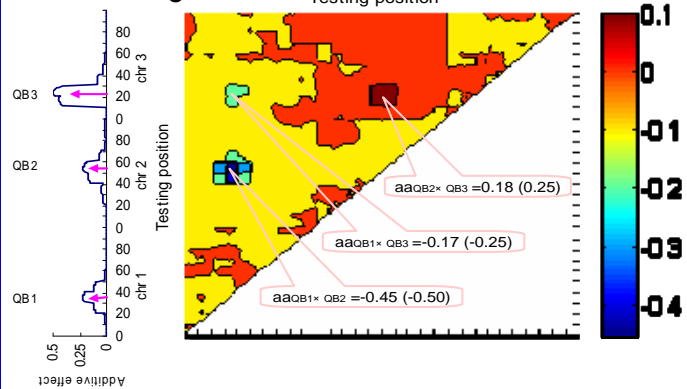
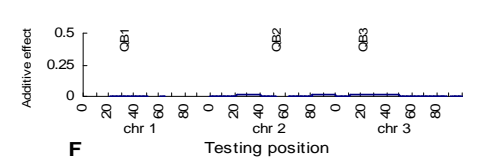
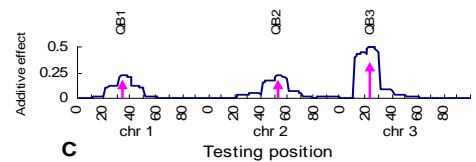
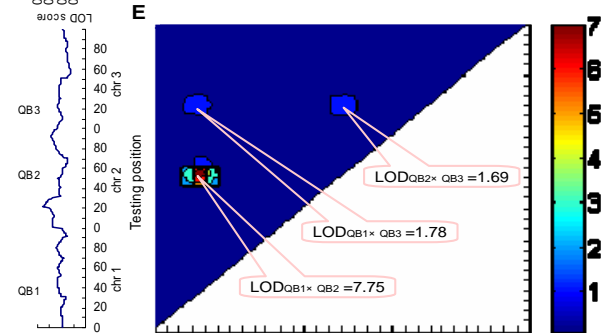
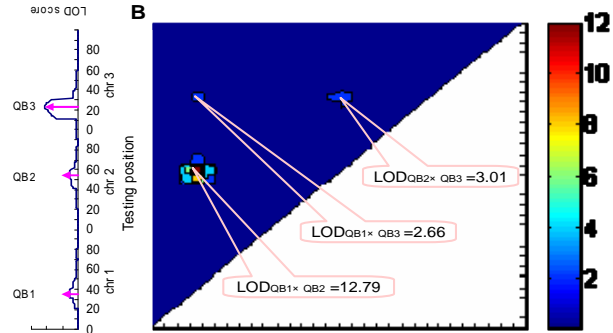
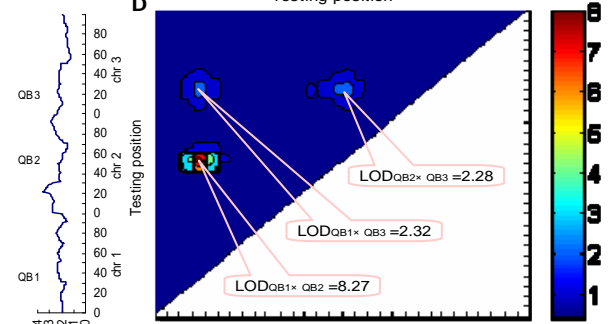
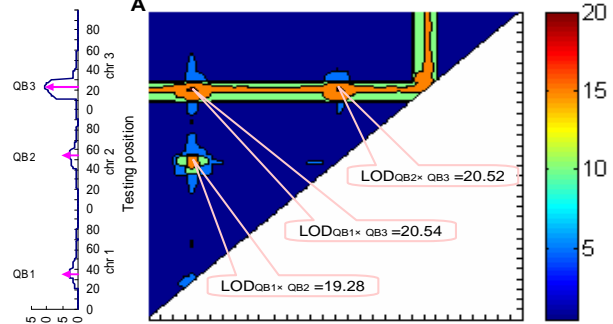
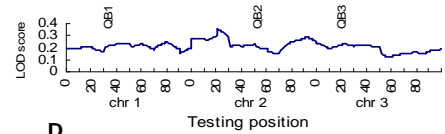
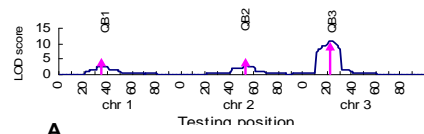
- ICIM has a simpler form and faster convergence speed (EM algorithm converges after 3 to 5 iterations), without losing the optimal properties associated with CIM.
- ICIM gives clearly high LOD scores at chromosomal regions with QTLs but rather low LOD scores where no QTL is located, and results in less biased estimates of QTL effects, thereby improving the mapping power and precision.

High stability of ICIM to input parameters

145 random DH lines in barley (Harrington×TR306), and 127 markers



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



ICIM with F2 populations

Interactions between marker variables can be caused simply by dominance effects of QTL!

$$\begin{bmatrix}
 \mu + (d)\mu_d \\
 (a)A_1 \\
 (a)A_2 \\
 (d)D_1 \\
 (d)D_2 \\
 (d)AA_{12} \\
 AD_{12} \\
 DA_{12} \\
 (d)DD_{12}
 \end{bmatrix}
 =
 \begin{bmatrix}
 \mu + \frac{1}{2}(g_1 + g_3)d \\
 f_2 a \\
 \frac{1}{2}(f_1 - f_3)a \\
 (-\frac{1}{2}g_1 - \frac{1}{2}g_3 + g_4)d \\
 (-\frac{1}{2}g_1 + g_2 - \frac{1}{2}g_3)d \\
 \frac{1}{2}(g_1 - g_3)d \\
 0 \\
 0 \\
 (\frac{1}{2}g_1 - g_2 + \frac{1}{2}g_3 - g_4 + g_5)d
 \end{bmatrix}$$

$$P = E(G) + \varepsilon = \beta + \sum_{j=1}^{m+1} \lambda_j x_j + \sum_{j=1}^{m+1} \rho_j y_j + \sum_{j=1}^m \lambda \lambda_{j,j+1} x_j x_{j+1} + \sum_{j=1}^m \rho \rho_{j,j+1} y_j y_{j+1} + \varepsilon$$

Interactions Between Markers Can Be Caused by the Dominance Effect of Quantitative Trait Loci

Luyan Zhang,^{*,†} Huihui Li,^{*,†} Zhonglai Li^{*} and Jiankang Wang^{†,1}

**School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China and [†]Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement and CIMMYT China Office, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

Manuscript received May 30, 2008
Accepted for publication July 24, 2008

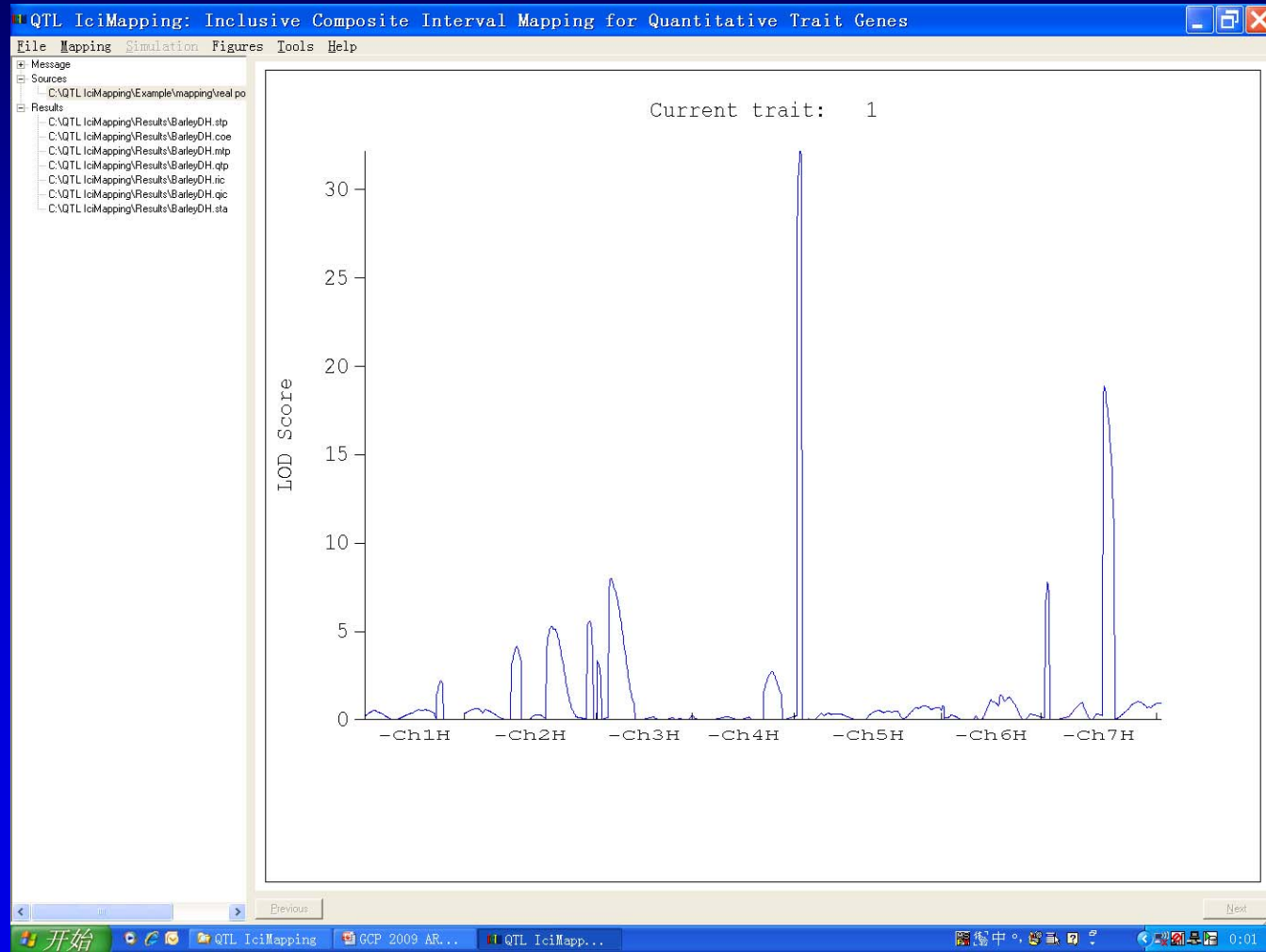
ABSTRACT

F₂ populations are commonly used in genetic studies of animals and plants. For simplicity, most quantitative trait locus or loci (QTL) mapping methods have been developed on the basis of populations having two distinct genotypes at each polymorphic marker or gene locus. In this study, we demonstrate that dominance can cause the interactions between markers and propose an inclusive linear model that includes marker variables and marker interactions so as to completely control both additive and dominance effects of QTL. The proposed linear model is the theoretical basis for inclusive composite-interval QTL mapping (ICIM) for F₂ populations, which consists of two steps: first, the best regression model is selected by stepwise regression, which approximately identifies markers and marker interactions explaining both additive and dominance variations; second, the interval mapping approach is applied to the phenotypic values adjusted by the regression model selected in the first step. Due to the limited mapping population size, the large number of variables, and multicollinearity between variables, coefficients in the inclusive linear model cannot be accurately determined in the first step. Interval

The software implementing ICIM additive and epistasis mapping called

QTL IciMapping

(available from <http://www.isbreeding.net>)



Purpose of modeling and simulation in plant breeding

- Optimize breeding schemes
 - Increase genetic gain and selection response
 - Improve breeding efficiency
- Investigate issues that cannot be tested through field experiments
- Better understand the genetics of quantitative traits

Available breeding simulation tools

- **QuLine**, a computer software that simulates breeding programs for developing inbred lines
- **QuHybrid**, a computer software that simulates breeding programs for developing hybrids
- **QuMARS**, a computer software that simulates marker-assisted recurrent selection and genome-wide selection

Nine major genes to be pyramided in wheat

Gene	<i>Rht-B1</i>	<i>Rht-D1</i>	<i>Rht8</i>	<i>Sr2</i>	<i>Cre1</i>	<i>VPM</i>	<i>Glu-B1</i>	<i>Glu-A3</i>	<i>tin</i>
Chr.	4BS	4DS	2DL	3BS	2BL	7DL	1BL	1AS	1AS
Marker	Codom	Codom	Codom	Codom	Dom	Dom	Codom	Codom	Codom
MK-gene distance	0	0	0.6	1.1	0	0	0	0	0.8
HM14BS	<i>Rht-B1a</i>	<i>Rht-D1a</i>	<i>Rht8</i>	<i>sr2</i>	<i>cre1</i>	<i>vpm</i>	<i>Glu-B1a</i>	<i>Glu-A3e</i>	<i>Tin</i>
Sunstate	<i>Rht-B1a</i>	<i>Rht-D1b</i>	<i>rht8</i>	<i>Sr2</i>	<i>cre1</i>	<i>VPM</i>	<i>Glu-B1i</i>	<i>Glu-A3b</i>	<i>Tin</i>
Silverstar + tin	<i>Rht-B1b</i>	<i>Rht-D1a</i>	<i>rht8</i>	<i>sr2</i>	<i>Cre1</i>	<i>vpm</i>	<i>Glu-B1i</i>	<i>Glu-A3c</i>	<i>tin</i>
Target	<i>Rht-B1a</i>	<i>Rht-D1a</i>	<i>Rht8</i>	<i>Sr2</i>	<i>Cre1</i>	<i>VPM</i>	<i>Glu-B1i</i>	<i>Glu-A3b</i>	<i>tin</i>

One strategy identified by QuLine to combine the nine genes from a topcross

- Selection of **Sunstate as the final parent** (having largest number of favorable alleles) in the topcross
- Stage I: Selection for *Rht-B1a* and *Glu-B1i* homozygotes, and enrichment of *rht8*, *Cre1*, and *tin* in TCF1
- Stage II: Selection of homozygotes for one target allele, e.g. *Rht8*, and enrich remaining target alleles in TCF2
- Stage III: Selection of the target genotype in DHs/RILs

Comparison with other strategies

- For this strategy, one target genotype can be selected by screening **< 600 individuals/lines**
- For one-stage selection in advanced generations, one target genotype can be selected by screening **> 3500 lines**
- For one-stage selection in early generations, say TCF2, one target genotype can be selected by screening **millions of individuals**
- Crop Science, 2007, 47: 580-588

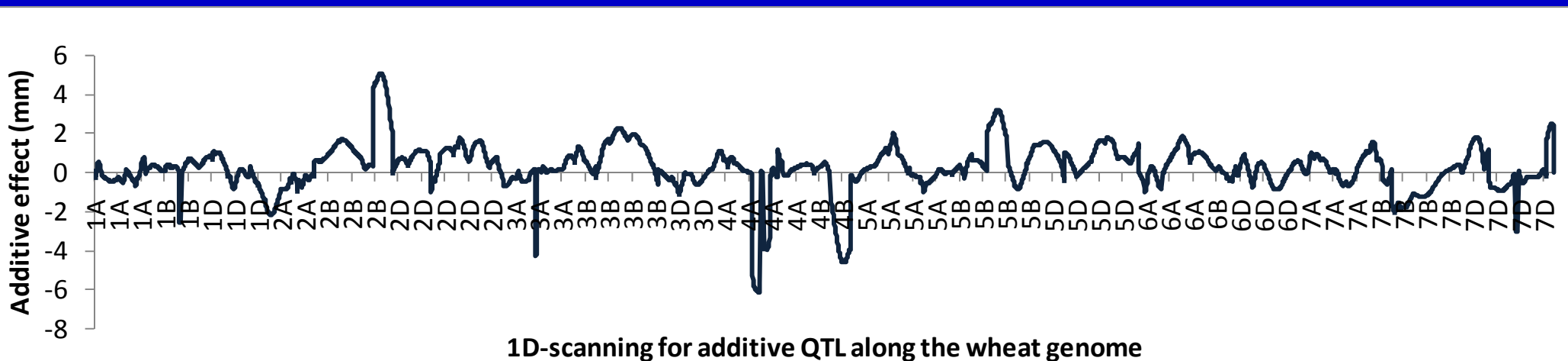
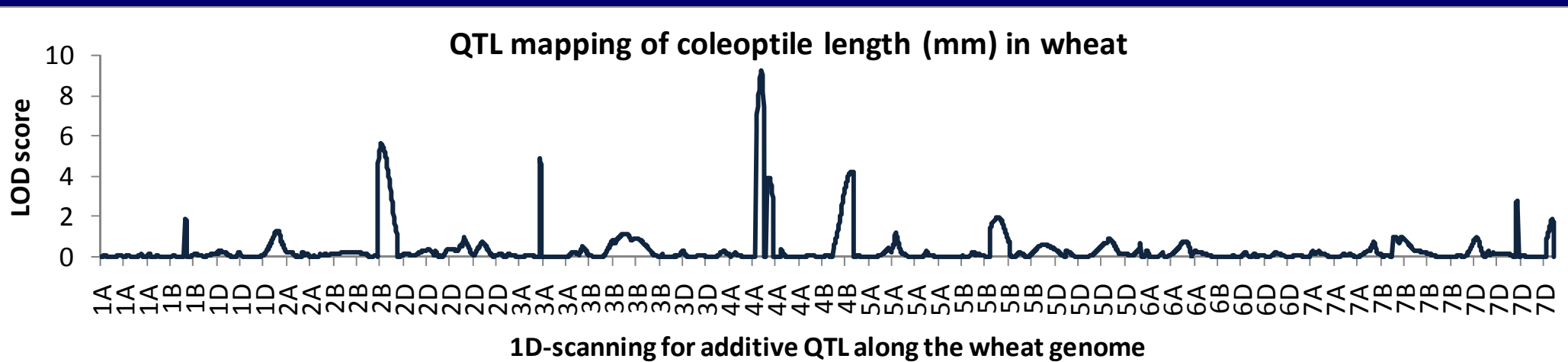
Coleoptile length (CL) of wheat ($h^2_{\text{entry-mean}} = 0.80$)



From Dr. David Bonnett, CIMMYT

CL is a typical quantitative trait

- 159 DH lines derived from two Australian wheat cultivars Cranbrook and Halbert (provided by Greg Rebetzke, CSIRO, Canberra, Australia) using Inclusive Composite Interval Mapping (ICIM)



Six major genes in two parental lines

Gene	<i>Rht-D1</i>	<i>Rht8</i>	<i>Sr2</i>	<i>VPM</i>	<i>Glu-B1</i>	<i>Glu-A3</i>
Chr.	4DS	2DL	3BS	7DL	1BL	1AS
Marker	Codom	Codom	Codom	Dom	Codom	Codom
MK-gene distance	0	0.6	1.1	0	0	0
HM14BS	<i>Rht-D1a</i>	<i>Rht8</i>	<i>sr2</i>	<i>vpm</i>	<i>Glu-B1a</i>	<i>Glu-A3e</i>
Sunstate	<i>Rht-D1b</i>	<i>rht8</i>	<i>Sr2</i>	<i>VPM</i>	<i>Glu-B1i</i>	<i>Glu-A3b</i>

Additive genetic effects of CL genes, genotypes of HM14BS, Sunstate and the genotype with the longest CL

Locus	Chromosome	Distance to the nearest marker (cM)	Additive effect (mm)	Additive variance explained (%)	HM14BS	Sunstate	Genotype with all increased CL alleles
<i>Rht-D1</i>	4DS	0.0	9.5	42.63	<i>Rht-D1a</i>	<i>Rht-D1b</i>	<i>Rht-D1a</i>
<i>qCL1</i>	1AS	8.1	2.9	3.97	+	-	+
<i>qCL2</i>	2BS	0.7	2.5	2.95	+	-	+
<i>qCL3</i>	2DS	1.1	4.1	7.94	+	-	+
<i>qCL4</i>	3BS	0.9	2.0	1.89	+	-	+
<i>qCL5</i>	5AL	6.2	4.9	11.34	+	-	+
<i>qCL6</i>	5DS	13.0	3.6	6.12	+	-	+
<i>qCL7</i>	Unidentified		4.0	7.56	-	+	+
<i>qCL8</i>	Unidentified		3.0	4.25	+	-	+
<i>qCL9</i>	Unidentified		3.0	4.25	-	+	+
<i>qCL10</i>	Unidentified		2.0	1.89	+	-	+
<i>qCL11</i>	Unidentified		2.0	1.89	+	-	+
<i>qCL12</i>	Unidentified		2.0	1.89	-	+	+
<i>qCL13</i>	Unidentified		1.0	0.47	+	-	+
<i>qCL14</i>	Unidentified		1.0	0.47	+	-	+
<i>qCL15</i>	Unidentified		1.0	0.47	+	-	+
<i>qCL16</i>	Unidentified		1.0	0.47	+	-	+
<i>qCL17</i>	Unidentified		1.0	0.47	-	+	+
Coleoptile length (mm)					158	97	178

Breeding target ...

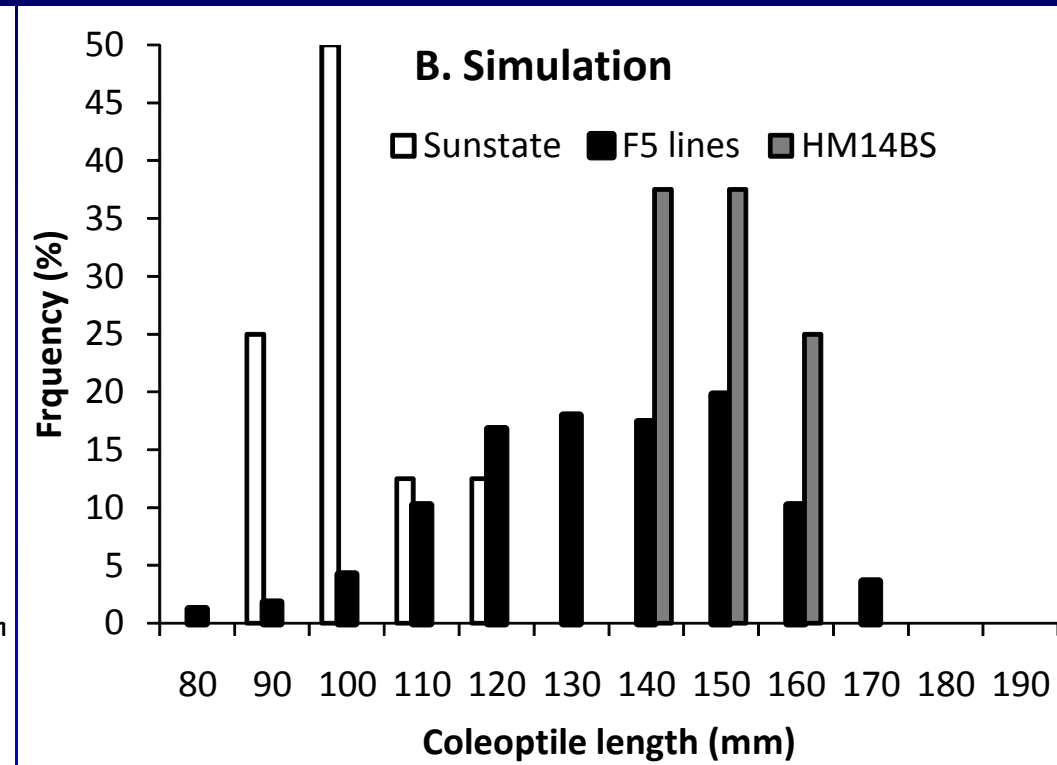
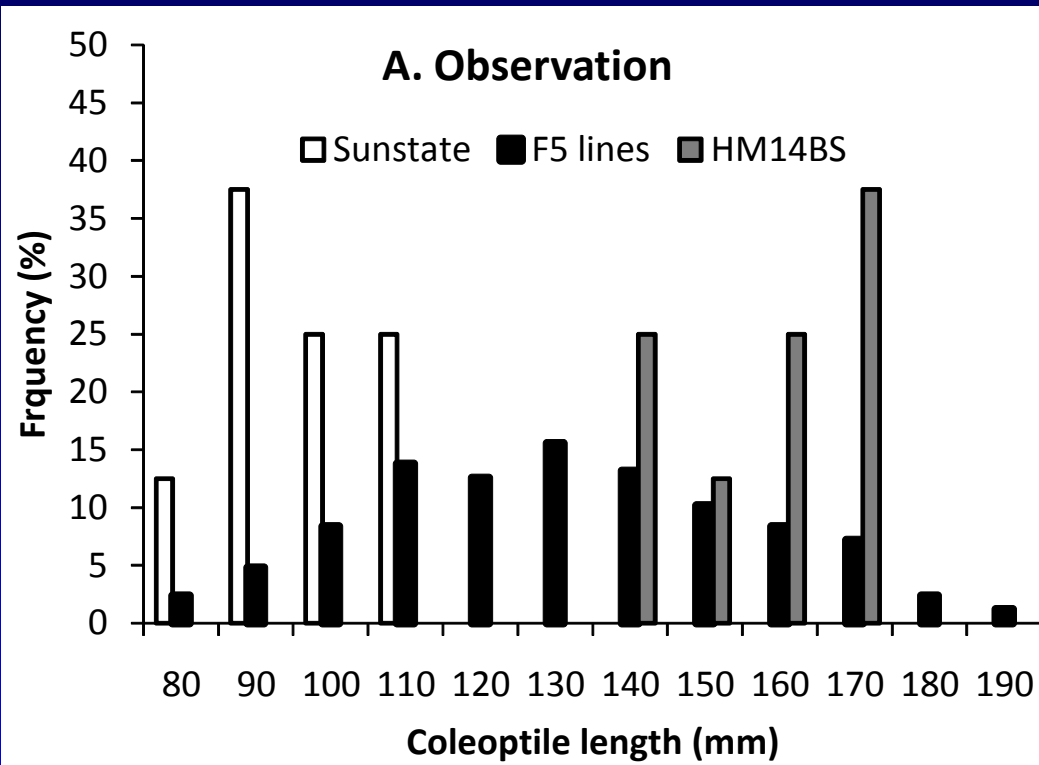
For the six major genes

Gene	<i>Rht-D1</i>	<i>Rht8</i>	<i>Sr2</i>	<i>VPM</i>	<i>Glu-B1</i>	<i>Glu-A3</i>
Target	<i>Rht-D1a</i>	<i>Rht8</i>	<i>Sr2</i>	<i>VPM</i>	<i>Glu-B1i</i>	<i>Glu-A3b</i>

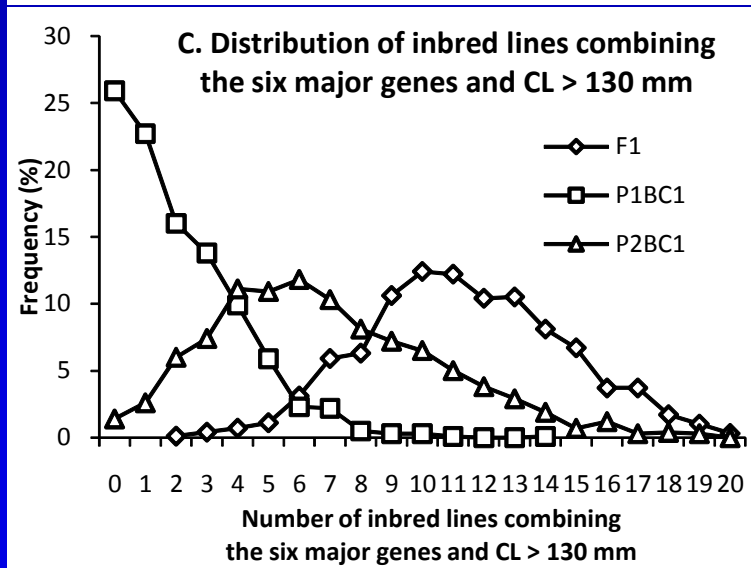
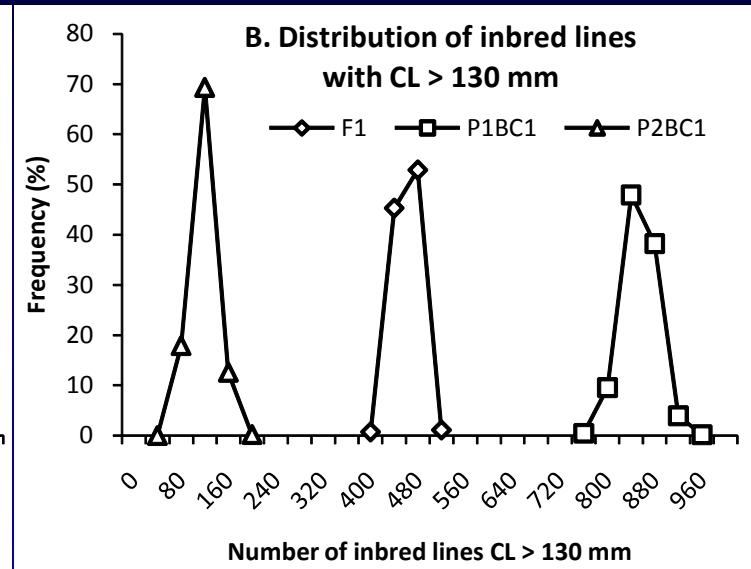
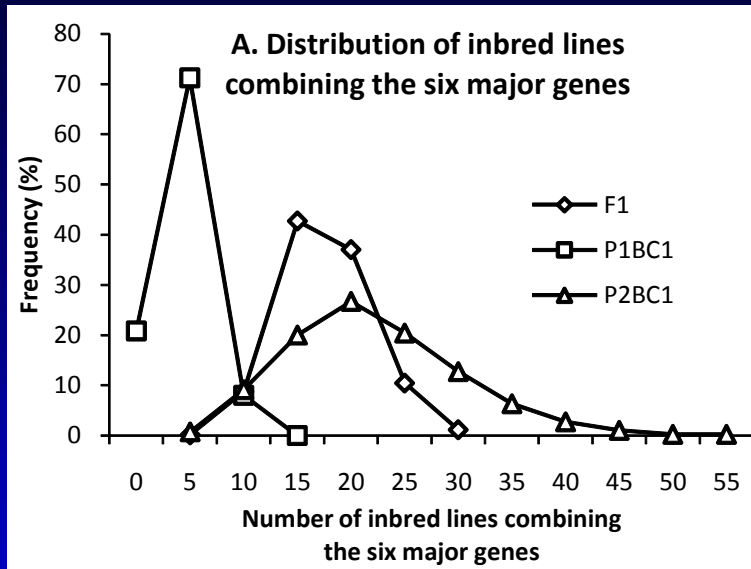
For coleoptile length (CL), $CL > 130$ mm, which is 30% longer than Sunstate's

Target: to select as many as possible inbred lines combining the six desired alleles and having $CL > 130$ mm, so as to have chances to select for other traits, i.e., yield and adaptation.

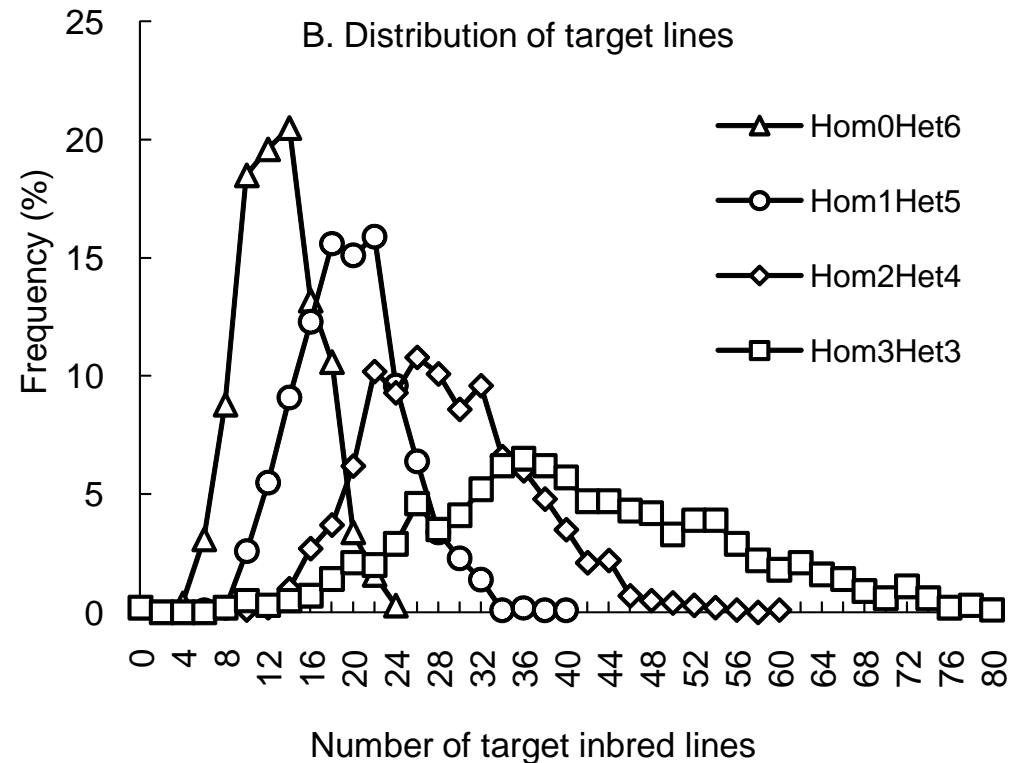
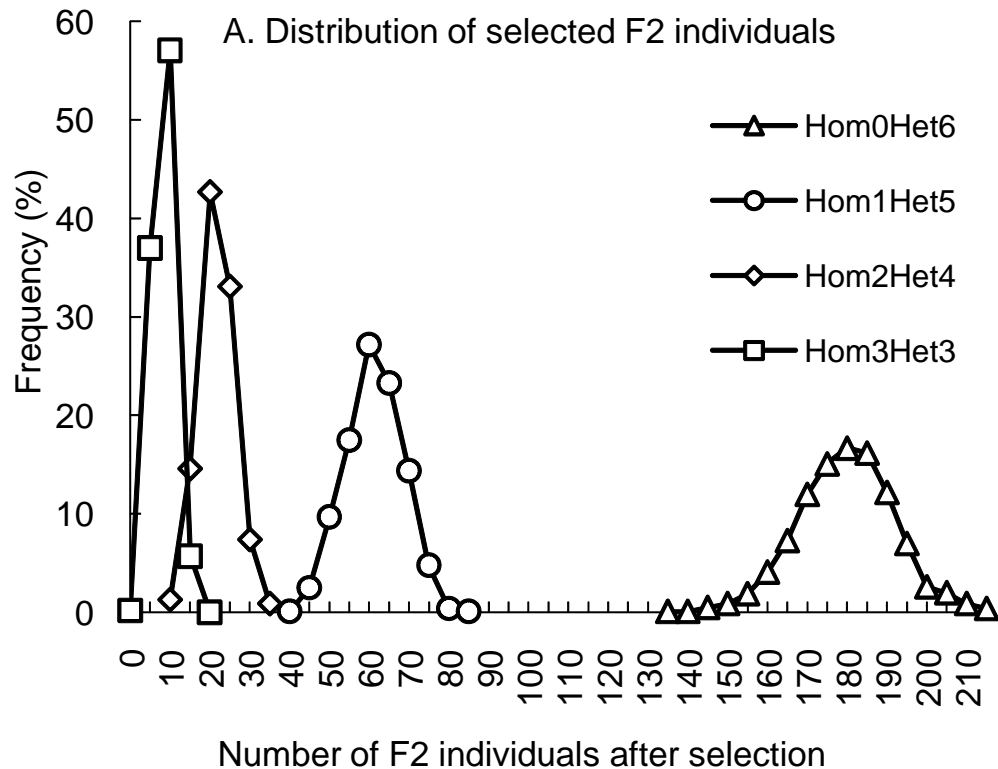
Validating the genetic model



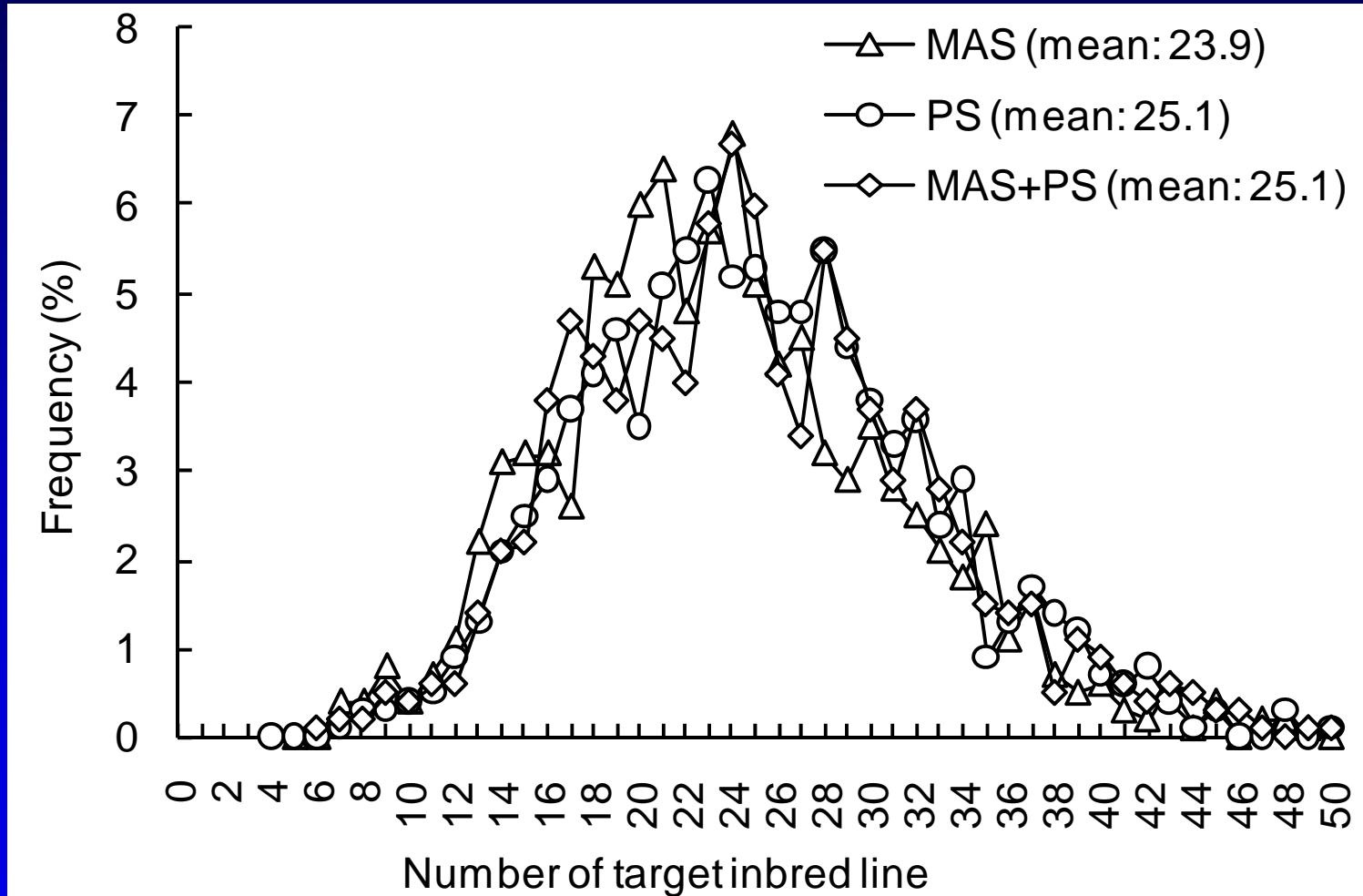
Do we need backcrossing?



Comparison of various MAS strategies in selecting the six major genes (1000 F2 individuals and 200 RILs)



Comparison of MAS and phenotypic selection (PS) for coleoptile length



Conclusions

- An average of **2.4** individuals with the target genotype were present in unselected F_1 -derived doubled haploid (DH) or recombinant inbred line (RIL) populations of size 200.
- A selection scheme for the six major genes in F_2 increased the number of target individuals to **19.1**, and additional marker-assisted selection (MAS) for CL increased the number to **23.9**.
- Phenotypic selection (PS) of CL outperformed MAS in this study due to the high heritability of CL, incompletely linked markers for known QTL, and the existence of unidentified QTL.
- However, a selection scheme combining MAS and PS was equally as efficient as PS (**25.1** target genotypes selected) and would result in net savings in production and time to delivery of long coleoptile wheats containing the six favorable alleles.

Tangible outputs delivered

- QTL IciMapping software package v2.2, available from <http://www.isbreeding.net>.
- Improved QU-GENE modules, available from <http://www.uq.edu.au/lcafs/qugene/>.
- Li, H, J.-M. Ribaut, Z. Li, and J. Wang. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. **Theor. Appl. Genet. 116: 243-260.**
- 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.**

Tangible outputs delivered (cont'd)

- Chapman, S. C., J. Wang, G. J. Rebetzke, D. G. Bonnett. 2008. Designing crossing and selection strategies to combine diagnostic markers and quantitative traits. **IWGS 2008**, Brisbane, Australia.
- Wang, J., S. C. Chapman, D. G. Bonnett, G. J. Rebetzke. 2009. Simultaneous selection of major genes and QTL: a case study using coleoptile length in wheat (*Triticum aestivum* L.). **Theor. Appl. Genet. 119: 65-74.**
- Wang, J., R. P. Singh, H.-J. Braun, W. H. Pfeiffer. 2009. Investigating the efficiency of the single backcrossing breeding strategy through computer simulation. **Theor. Appl. Genet. 118: 683-694.**
- H. Li, Wang, J., R. Messmer, J.-M. Ribaut. 2009. Interacting genes are less repeatable than additive genes in maize (*Zea mays* L.) (to be submitted soon)

Acknowledgements

- Grains Research and Development Corporation (GRDC), Australia (1999 - 2004)
- GCP Commissioned Project SP3-20: Optimizing marker-assisted breeding systems for drought tolerance in cereals through linkage of physiological and genetic models (July 2005 - December 2007; PI: Scott Chapman)
- GCP Commissioned Project G4008.14: Breeding for drought tolerance with known gene information (January 2008 - December 2009)
- Carry-on in Molecular Breeding Platform

For those who want to know more about how the breeding procedure can be modeled, simulated and optimized...

- A Half-day Training Course on using Simulation in Breeding Projects, 8:30-12:30, 24 Sep., 2009
 - Introduction of the principles of breeding simulation (0.5 hr, Jiankang Wang)
 - The genetics and breeding simulation tools and their applications in optimizing breeding strategies and improving breeding efficiency (1 hr, Jiankang Wang)
 - Use of the QU-GENE graphical user interfaces to design and complete simulation experiments, analyze and visualize the results (1.5 hrs, Scott Chapman)

第四届QTL作图和育种模拟研讨会合影留念
The 4th Training Course on QTL Mapping and
Breeding Simulation, 17-20 August, 2009,
Shandong Agricultural University, Tai'an



Thank you for your attention!