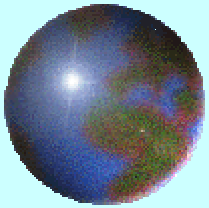


COMBINED LINKAGE DISEQUILIBRIUM AND CO-SEGREGATION BASED FINE MAPPING OF QUANTITATIVE TRAIT LOCI

M. Z. Firat[1]
Hélène Gilbert[2], L.R.Totir[3], J.C.M. Dekkers[3], R.L. Fernando[3]

[1]Akdeniz University, Faculty of Agriculture, Department of Animal Science, 07059 Antalya, Turkey

[2]Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas cedex, France

[3]Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, IA 50011, USA

Identity by Descent Method (IBD)

The QTL fine mapping method requires

- **an IBD matrix between QTL effects**
- **to calculate variance comps. and likelihood of the putative QTL.**

This IBD matrix is calculated conditional on

- **the marker information and**
- **the position of the QTL**

Identity by Descent Method (IBD)

In the case of linkage analysis mapping,

- **the first (base) generation of marker genotyped animals is assumed to be unrelated.**
- **Therefore, IBD regions occur within the marker genotyped data set**
- **This sets the base generation further back in time.**

Identity by Descent Method (IBD)

However,

- **there are IBD probabilities at the putative QTL locus between the first genotyped animals given the information from their marker haplotypes.**

Therefore the pedigree of genotyped animals split into two parts

- **The first part, referred to as the linkage disequilibrium (LD)**
- **The second part, referred to as the co-segregation (CS)**

The fine mapping method combining LD and CS

- **provide mapping resolution accurate enough to narrow down the QTL confidence interval to a few cM of the genomic region**

The earlier findings (Grapes et al. submitted)

IBD method for LD fine mapping (M&G 2000) using a haplotype consisting of 1, 2, 4, 6, or 10 markers

- **incorporating 10 markers made IBD insensitive to QTL position**
- **thus decreased the accuracy of the method.**
- **IBD method with 4 or 6 markers gave greatest mapping accuracy.**

This indicates :

- **an optimum number of markers to include in the haplotype.**

Objectives

- **to determine the optimum haplotype size for combined LDCS based fine mapping**

- **to evaluate the impact of different designs with varying numbers of pedigreed generations and QTL effects on optimal haplotype number**

Model under co-segregation

The genotypic value of candidate i, g_i:

$$g_i = v_i^m + v_i^p + u_i$$

and

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{W}\mathbf{v} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$Var(\mathbf{v}) = \Sigma_v \sigma_v^2 \qquad Var(\mathbf{u}) = \Sigma_u \sigma_u^2 \qquad Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$$

Where

v_i^p **and** v_i^m **= effects of maternal/ paternal alleles at MQTL**

u_i **= residual polygenic effect**

v **= vector of gametic effects at the MQTL,**

u **= vector of residual polygenic effects**

Model for combined LDCS analysis

The MQTL effect is divided between genotypic fixed effect and random effects

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{ZX}_g\boldsymbol{\alpha} + \mathbf{Wv} + \mathbf{Zu} + \mathbf{e}$$

Where

$E(\mathbf{g} \mid \mathbf{M}) = \mathbf{ZX}_g\boldsymbol{\alpha}$ = cond. Exp. of **g** given marker haplotypes **M**

$\boldsymbol{\alpha}$ = MQTL substitution effect

Comparison of methods

absolute differences between the estimated and true QTL positions for each design from each replicate of a simulation study

$$\left|\hat{\Theta}_i - \Theta\right|$$

Where

$\hat{\Theta}_i$ **: the estimated QTL position**

Θ **: the true QTL position**

Bias of position estimates for a given design

$$\frac{\sum_{i=1}^{n} \hat{\Theta}_i}{n} - \Theta$$

Identity by Descent Method (IBD)

Pedigree of genotyped animals split into two parts

1. **Linkage disequilibrium part - develops the population in a historical sense beyond the recorded pedigree.**

2. **Co-segregation part – describes the population in the last generations with a family structure and phenotypic data.**

Part I of the Simulation

Designed to generate t generations of random mating.

- **number of male parent = female parents**
- **alleles are inherited based on Mendelian segregation.**
- **Unique numbers are assigned to QTL alleles in generation 0**
- **In the last generation, QTL allele with frequency closest to 0.5 is chosen and treated as the favorable allele of a biallelic QTL.**

Part I of the Analysis

- Algorithm of M&G (2001) used to compute IBD probabilities between founders conditional on marker haplotypes
  - This requires
    - number of generations since the base generation, t
    - effective size of the population, Ne.
- Similarity of marker haplotypes used to see whether haplotypes are IBD.

Part II of the Simulation

- Designed for particular family structures with recorded data
- 5 sires and 25 dams randomly selected in generation t.
- Each male mated with 5 females to produce 8 offspring
- repeated 1, 3 or 5 generations until generation t+1, t+3 or t+5.
- descendents in (t+1) - (t+5) given genotypic/phenotypic records
- pedigree is only known for these animals
- animals from these generations used for fine mapping.

Part II of the Simulation

- **Phenotypic values for single trait simulated as**

$$y = \mu + v + u + e$$

$$N(0, \sigma_v^2) \qquad N(0, \sigma_u^2) \qquad N(0, \sigma_e^2)$$

- **Number of QTL loci = 1**
- **Number of unlinked polygenic loci = 50**

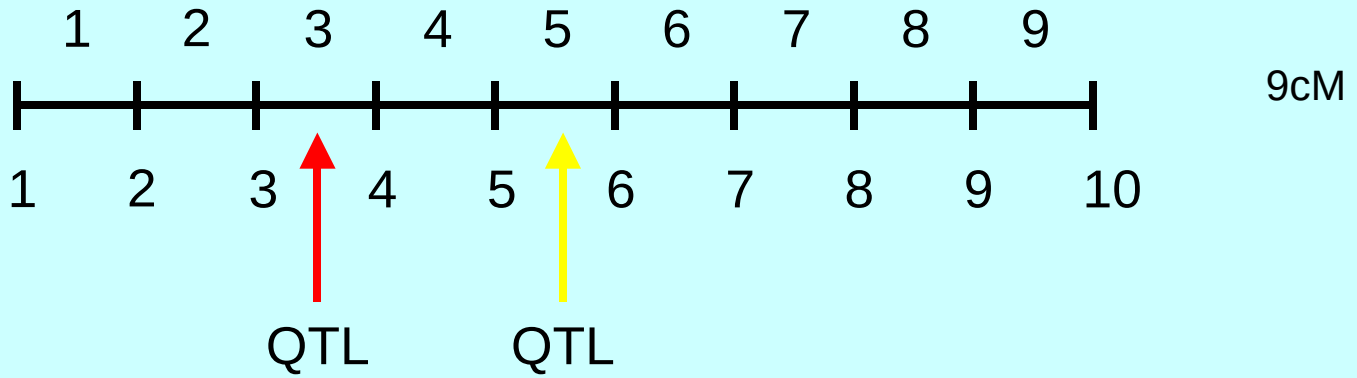
Case	σ_v^2	σ_u^2	σ_e^2
I	30	30	40
II	2.5	22.5	75

Part II of the Simulation

- Marker genotypes and phases available for animals in generations t to t+1, t+3 or t+5
- IBD probabilities for t+1 through t+5 calculated recursively based on conditional segregation probabilities given marker info.
  - estimated by MCMC
  - depend on the haplotype size.

Part I-II of the Simulation

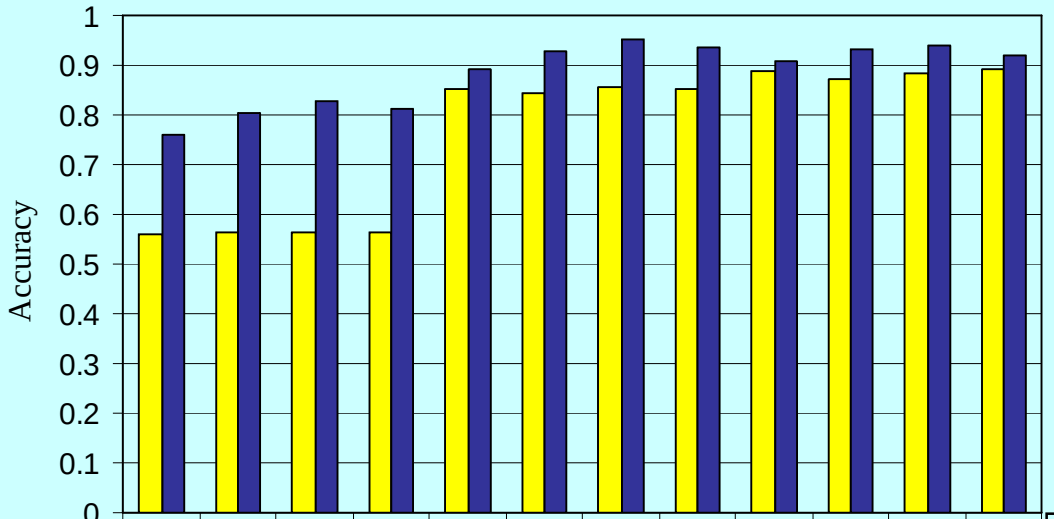
- Number of generations t=100 and
- Effective population size N_e=100
- Previous linkage analysis mapped QTL to 9 cM region
- 10 bi-allelic markers at 1 cM interval
- QTL centered between markers 3 and 4 or between 5 and 6,



- Haplotypes of 2, 4, 6 or 10 markers used as sliding window
- Number of replicates > 1,000

Results : Efficient designs for fine mapping

Proportion of replicates positioning QTL <3cM from true location

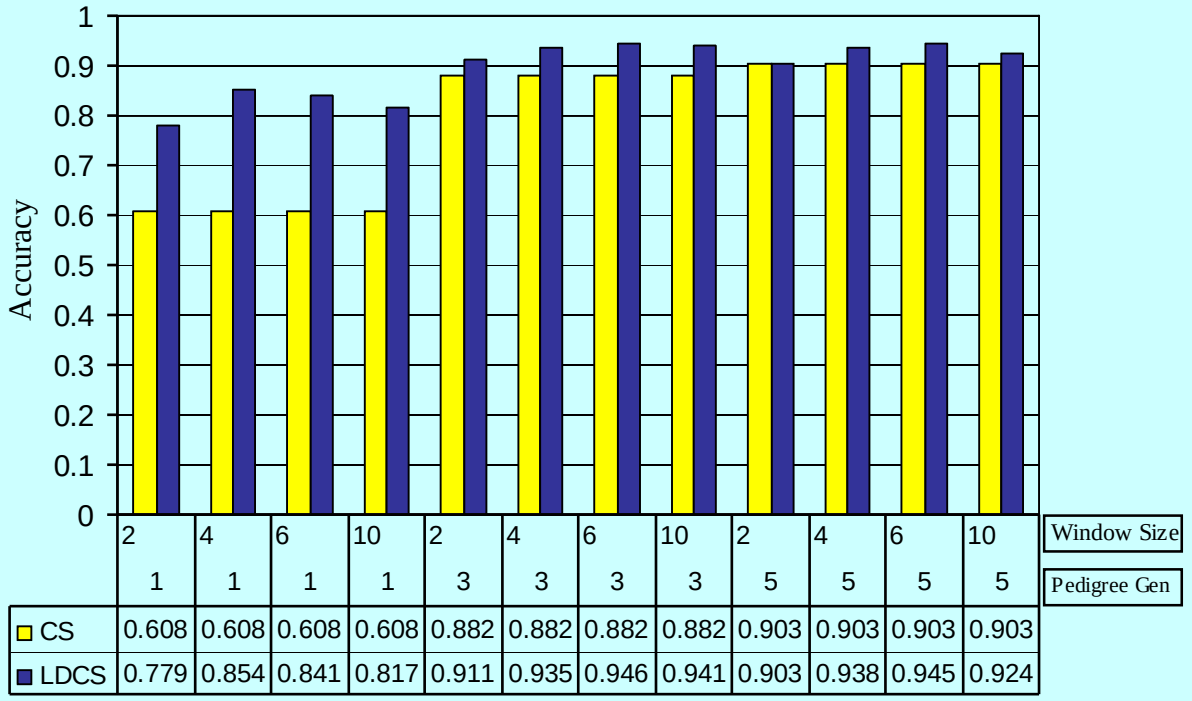


Window Size	2	4	6	10	2	4	6	10	2	4	6	10
Pedigree Gen	1	1	1	1	3	3	3	3	5	5	5	5
CS	0.562	0.564	0.565	0.563	0.85	0.846	0.855	0.852	0.89	0.87	0.885	0.89
LDCS	0.76	0.803	0.826	0.811	0.891	0.929	0.953	0.938	0.909	0.934	0.941	0.92

$\sigma_v^2=30$, $\sigma_u^2=30$, $\sigma_e^2=40$ **5M < QTL < 6M**

Results : Efficient designs for fine mapping

Proportion of replicates positioning QTL <3cM from true location

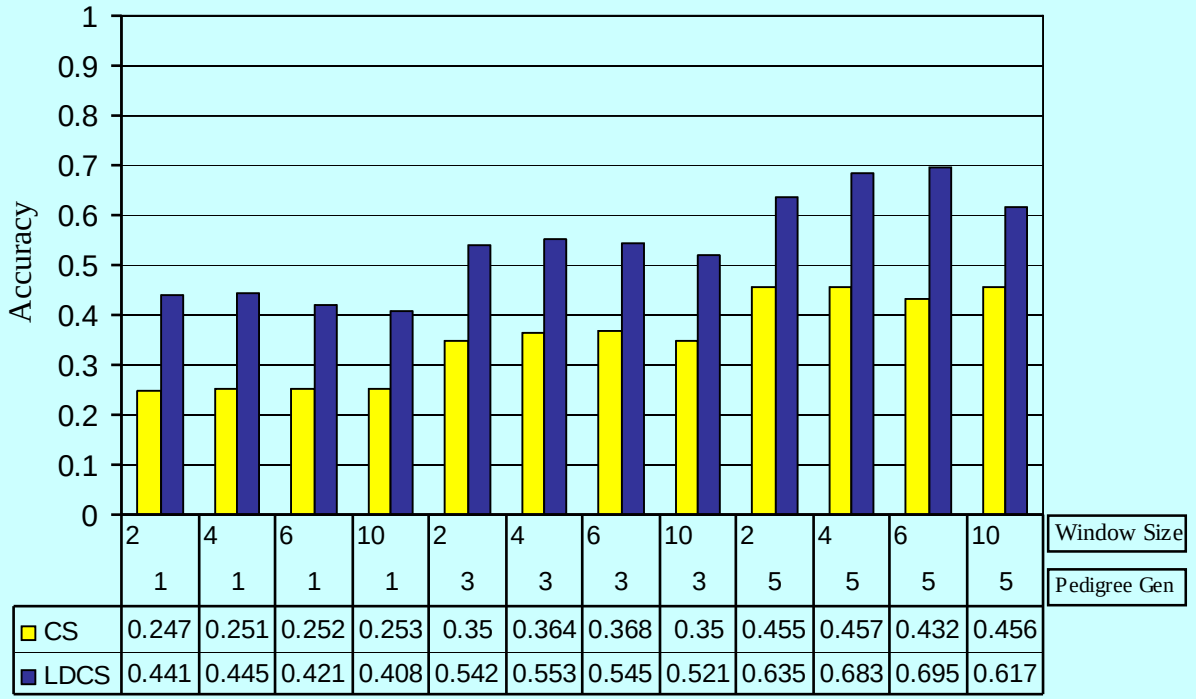


Window Size	2	4	6	10	2	4	6	10	2	4	6	10
Pedigree Gen	1	1	1	1	3	3	3	3	5	5	5	5
CS	0.608	0.608	0.608	0.608	0.882	0.882	0.882	0.882	0.903	0.903	0.903	0.903
LDCS	0.779	0.854	0.841	0.817	0.911	0.935	0.946	0.941	0.903	0.938	0.945	0.924

$\sigma_v^2=30$, $\sigma_u^2=30$, $\sigma_e^2=40$ 3M < QTL < 4M

Results : Efficient designs for fine mapping

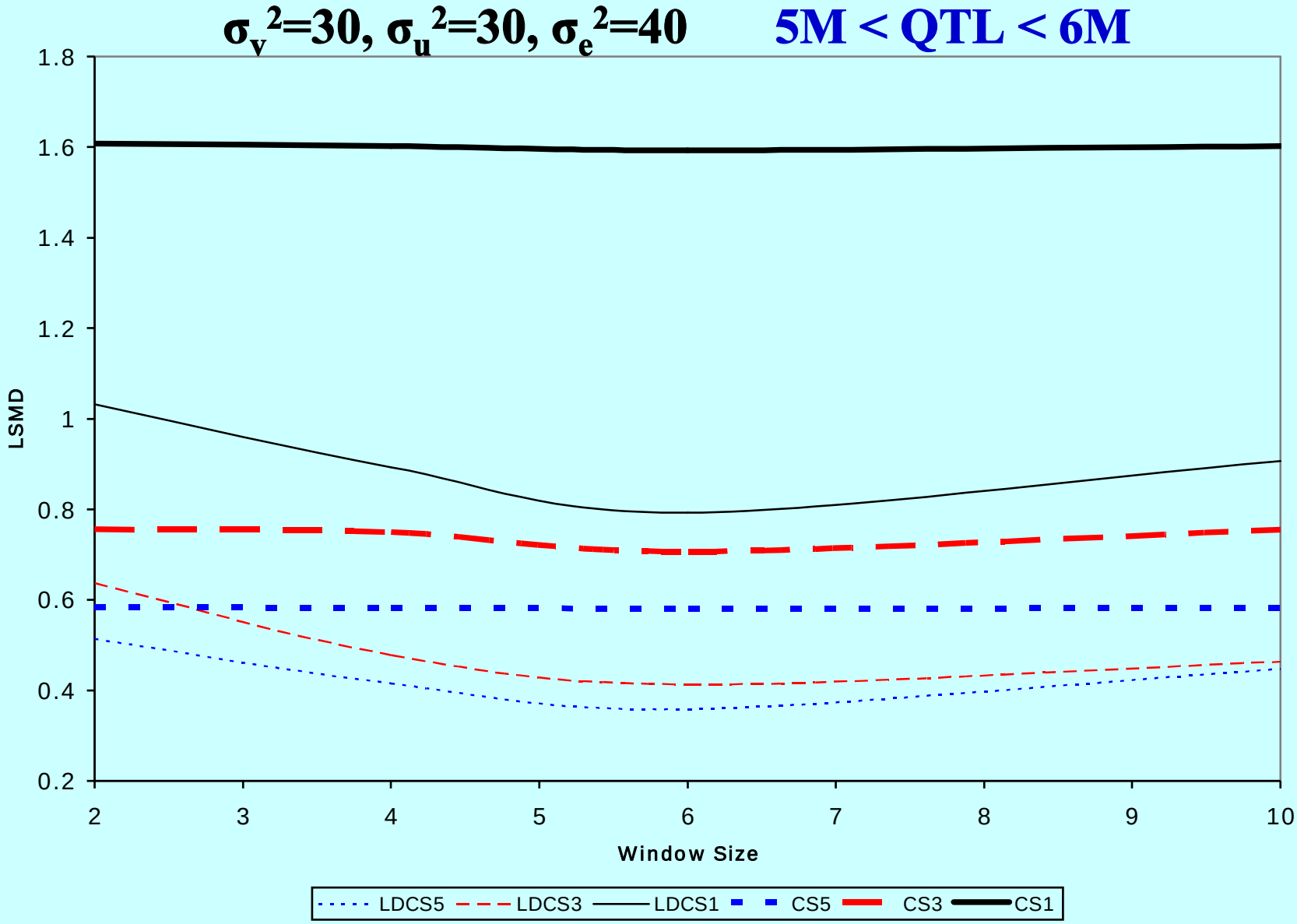
Proportion of replicates positioning QTL <3cM from true location



Window Size	2	4	6	10	2	4	6	10	2	4	6	10
Pedigree Gen	1	1	1	1	3	3	3	3	5	5	5	5
CS	0.247	0.251	0.252	0.253	0.35	0.364	0.368	0.35	0.455	0.457	0.432	0.456
LDCS	0.441	0.445	0.421	0.408	0.542	0.553	0.545	0.521	0.635	0.683	0.695	0.617

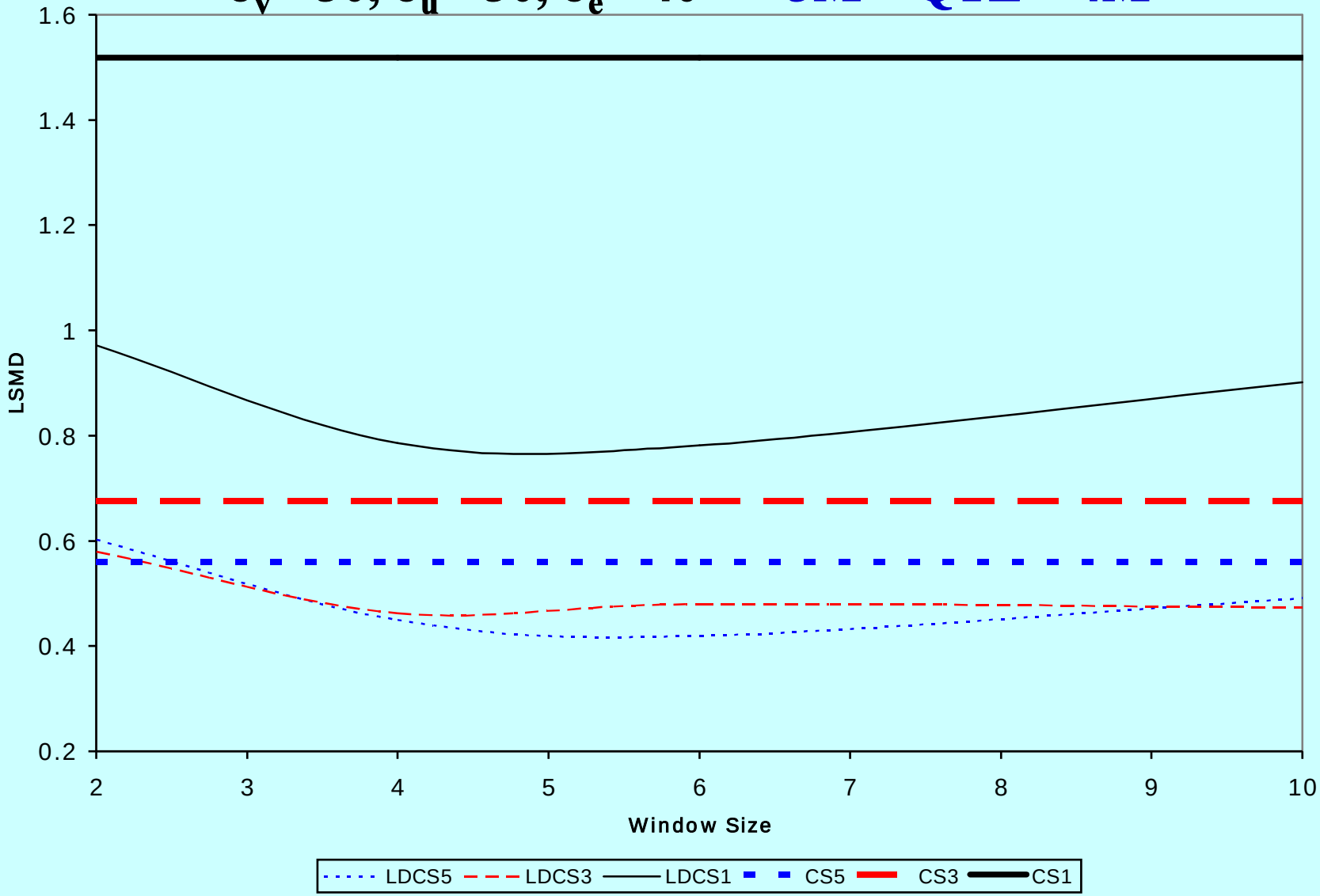
$\sigma_v^2=2.5$, $\sigma_u^2=22.5$, $\sigma_e^2=40$ **5M < QTL < 6M**

Results : Least squares mean absolute differences

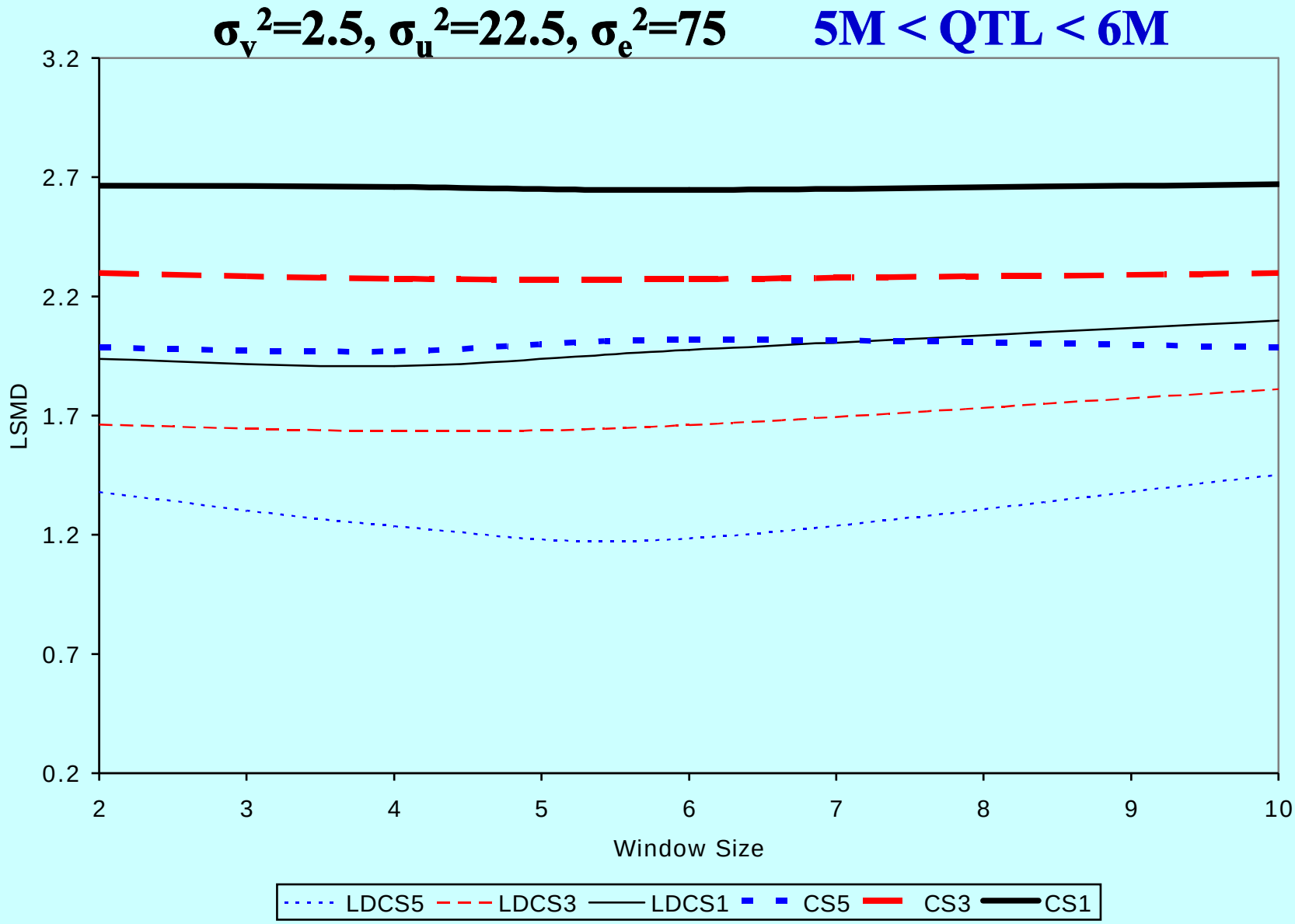


Results : Least squares mean absolute differences

$\sigma_v^2=30, \sigma_u^2=30, \sigma_e^2=40$ **3M < QTL < 4M**



Results : Least squares mean absolute differences



Results : Optimal haplotype size

		Type of	Pedigreed Generations		
Position of QTL		mapping	1	3	5
			$\sigma_v^2 = 30$, $\sigma_v^2 = 30$, $\sigma_v^2 = 40$		
halfway between markers 5 and 6	D1(a)	CS	6.68	6.21	5.16
		LDCS	6.82	7.18	6.58
halfway between markers 3 and 4	D1(b)	CS	-	-	-
		LDCS	6.32	7.25	6.72
			$\sigma_v^2 = 2.5$, $\sigma_v^2 = 22.5$, $\sigma_v^2 = 75$		
halfway between markers 5 and 6	D2	CS	5.87	5.95	6.92
		LDCS	3.14	4.11	5.69

When the IBD method is used for fine mapping using LDCS,

- **it appears preferable to fit a smaller haplotype instead of all available markers.**

- **A haplotype size of 4-6 markers was optimal for the experimental designs and parameter sets simulated here.**

Thanks…

PIC/Sygen **Monsanto Co.**

Department of Animal Science and
Center for Integrated Animal Genomics
Iowa State University

Akdeniz University,
department of animal science