Comparison of Association Mapping Methods in Cattle Population

G. Sahana, B. Guldbrandtsen, L. Janss and M.S. Lund

Department of Genetics and Biotechnology
Faculty of Agricultural Sciences
University of Aarhus, Denmark

Approaches to Mapping

1. Candidate gene studies
    - Association
    - Resequencing approaches

2. Genome-wide studies
    - Linkage analysis
    - Genome-wide association studies [Linkage disequilibrium (LD) mapping]

Association mapping

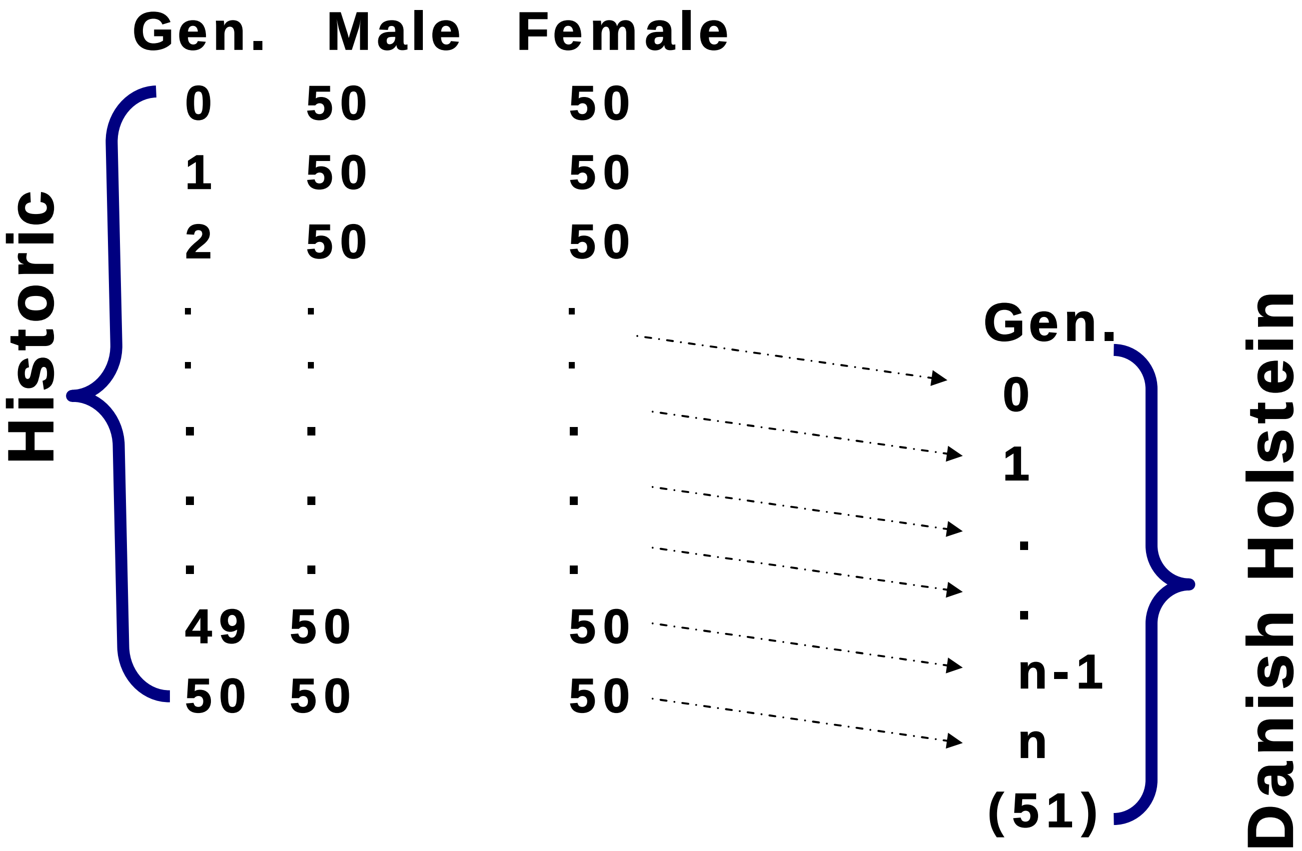
- Marker alleles are correlated with a trait on a population level
- Does not necessarily imply that markers are linked to (are close to) genes influencing the trait.
- Approaches
  - Case-control study
  - Family based methods

Limitation in cattle

- Complex pedigree
- Population structure
- Complex traits with very many genes individually of small effect
- Environmental influence
- Genetic heterogeneity underlying phenotype

Data Simulation

Generations and individuals



Historic

Gen.	Male	Female
0	50	50
1	50	50
2	50	50
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
49	50	50
50	50	50

Danish Holstein

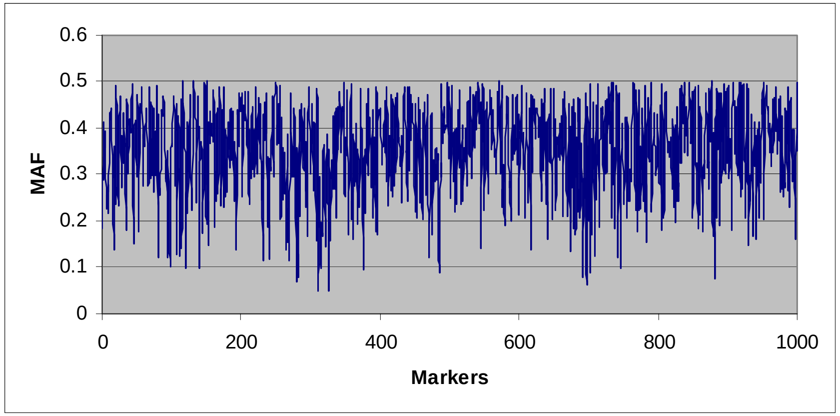
Gen.
0
1
.
.
n-1
n
(51)

Marker & QTL

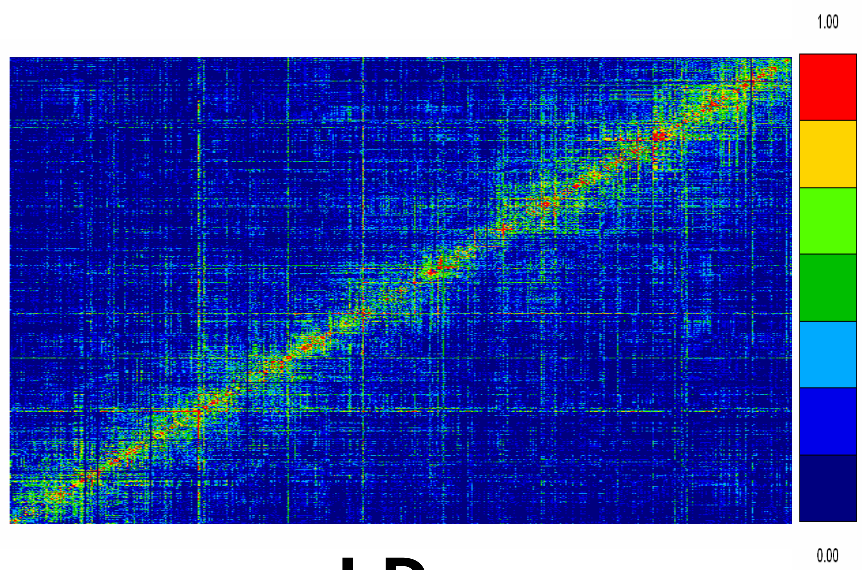
- 5 chromosomes 100 cM each
- 5000 biallelic markers
- 15 QTL with known position and effects
  - 1 QTL with 10%, 4 QTL with 5% and 10 QTL with 2% genetic variance
- Chromosome-5 had no QTL

Population evolves over 50 generations

- LD structure (drift, recombination)
- Allele frequencies change



MAF



LD

Phenotypes

- Heritability of the trait : 30%
- 50% genetic variance comes from 15 QTL, rest 50% was polygenic
- Breeding values were simulated assuming 100 daughters per sire
- 2000 individuals were selected randomly from 136 families
- 25 replications

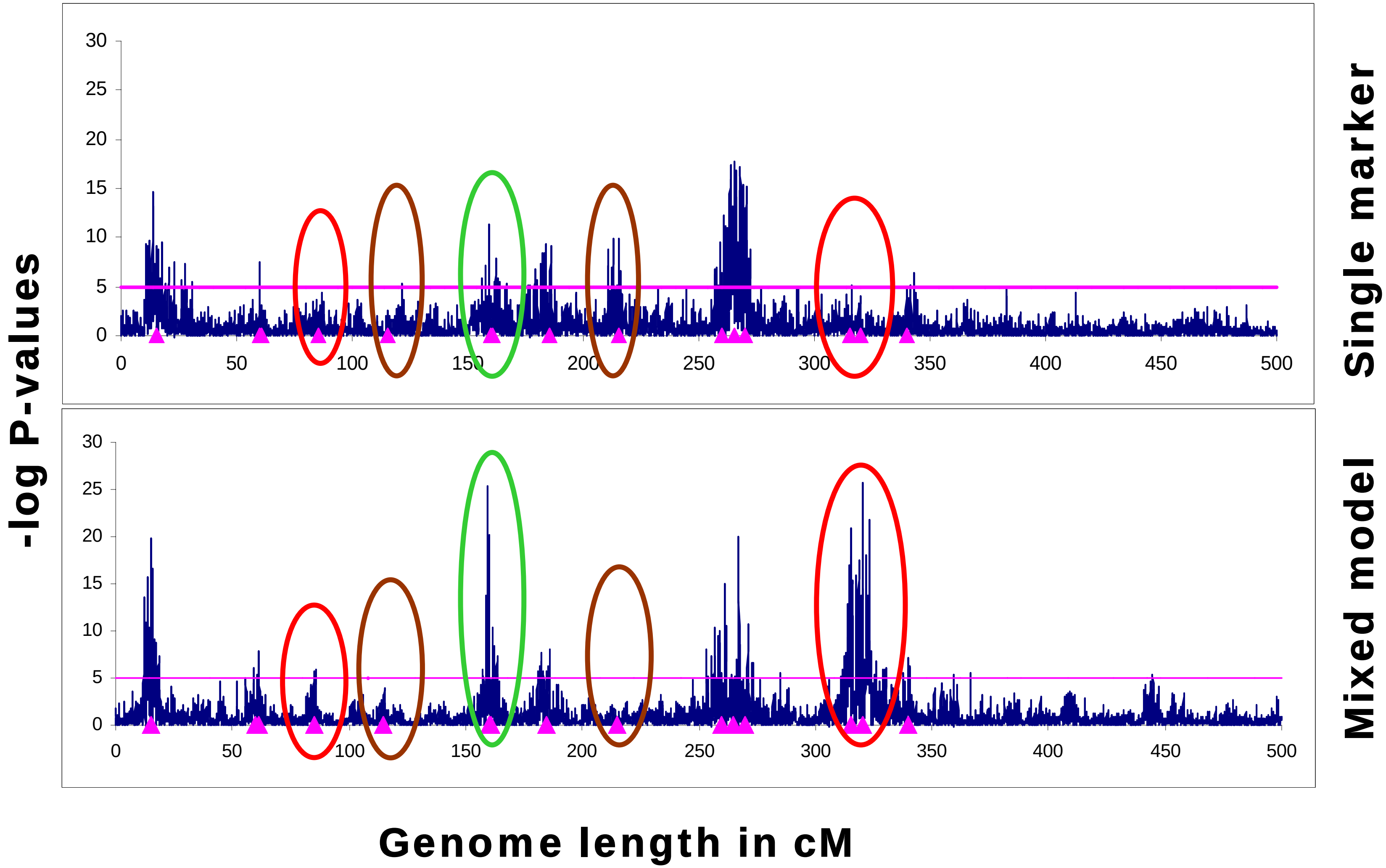
Association Analysis

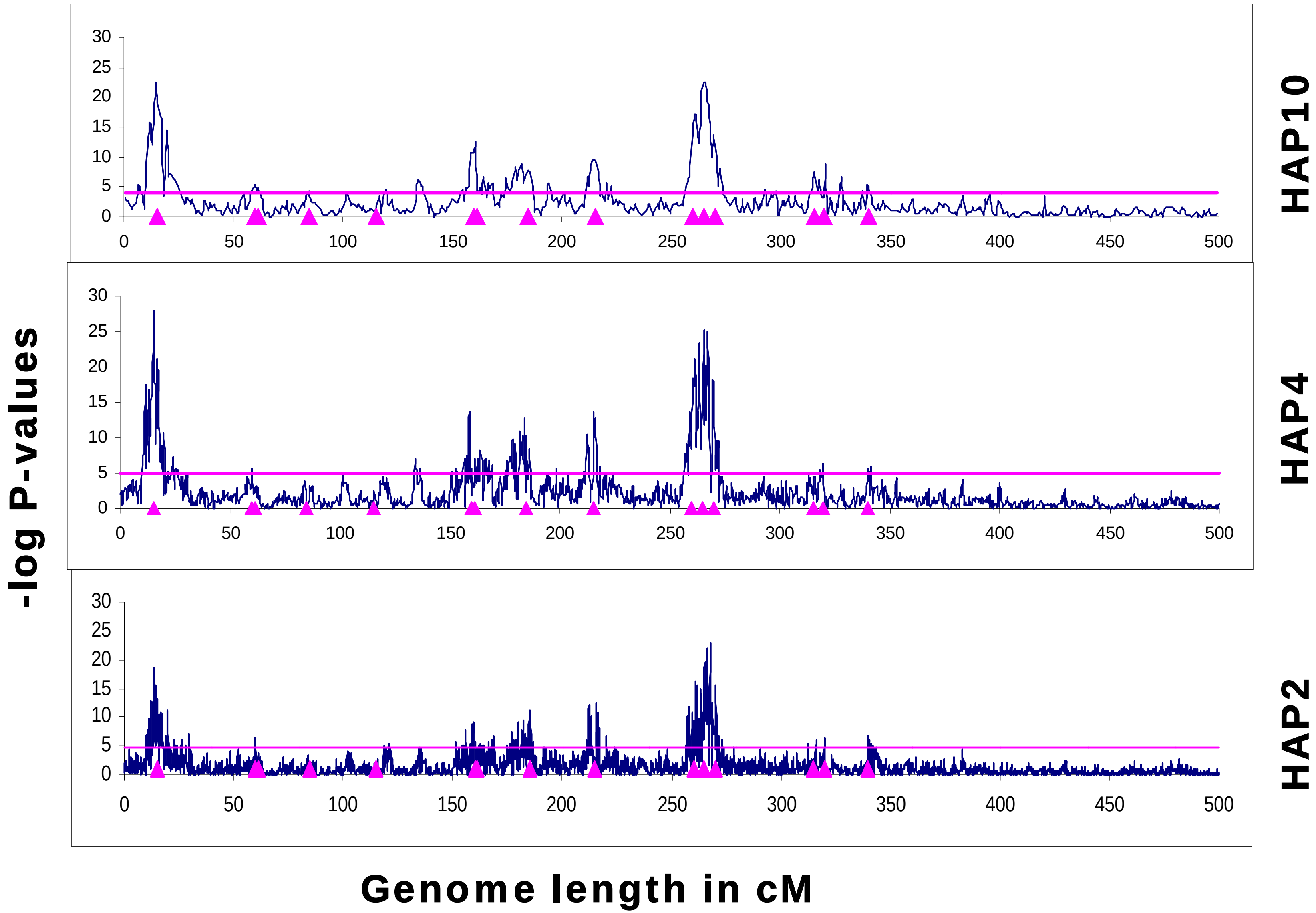
Methods

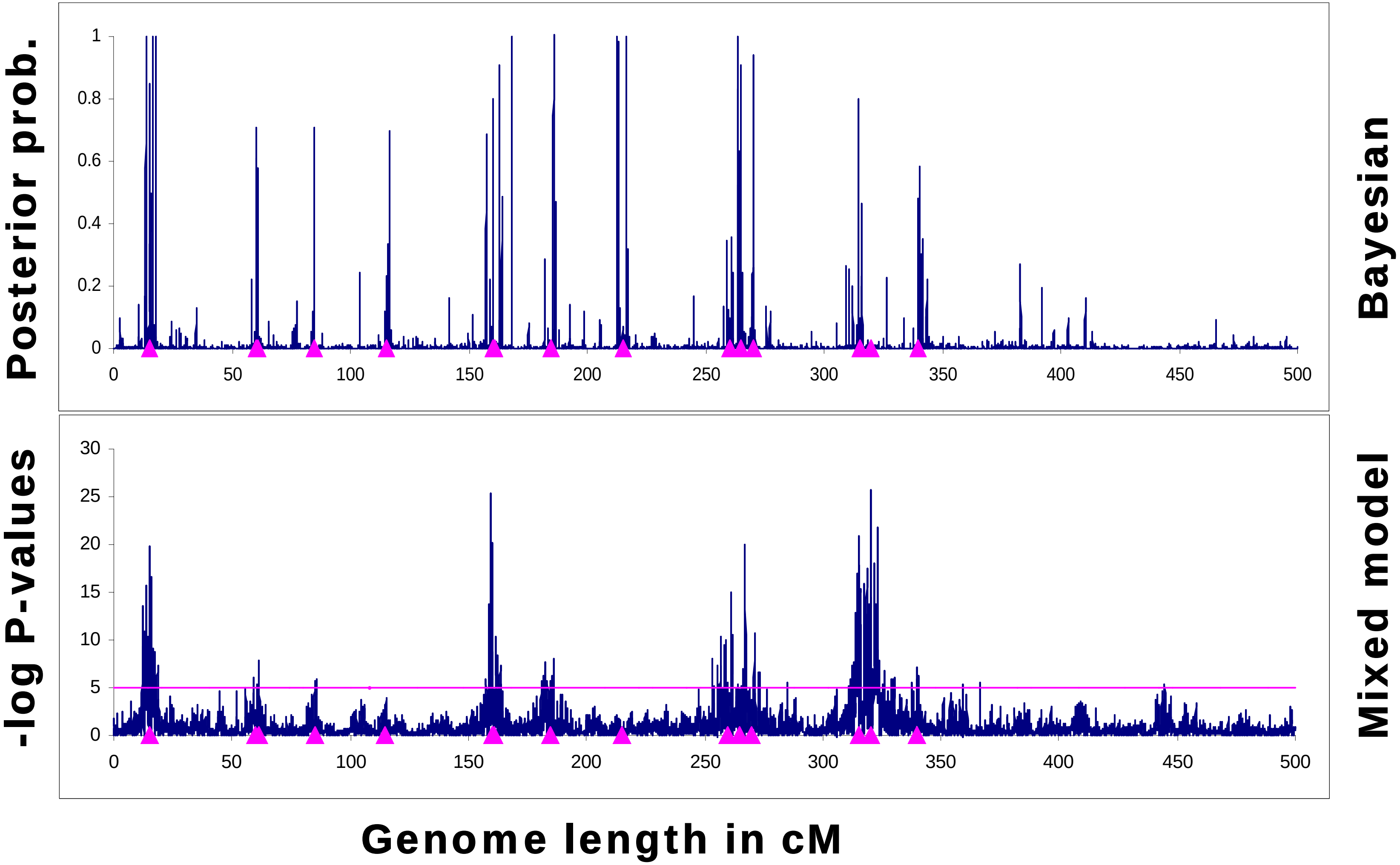
- Linear models (*TASSEL/R*)
  - Phenotype = Sire + genotype/haplotype

- Mixed model (*DMU*)
  - Phenotype= Fixed factors + polygene + SNP

- Bayesian approach (*iBays*)
  - Phenotype= Fixed factors + Polygene + ΣSNP

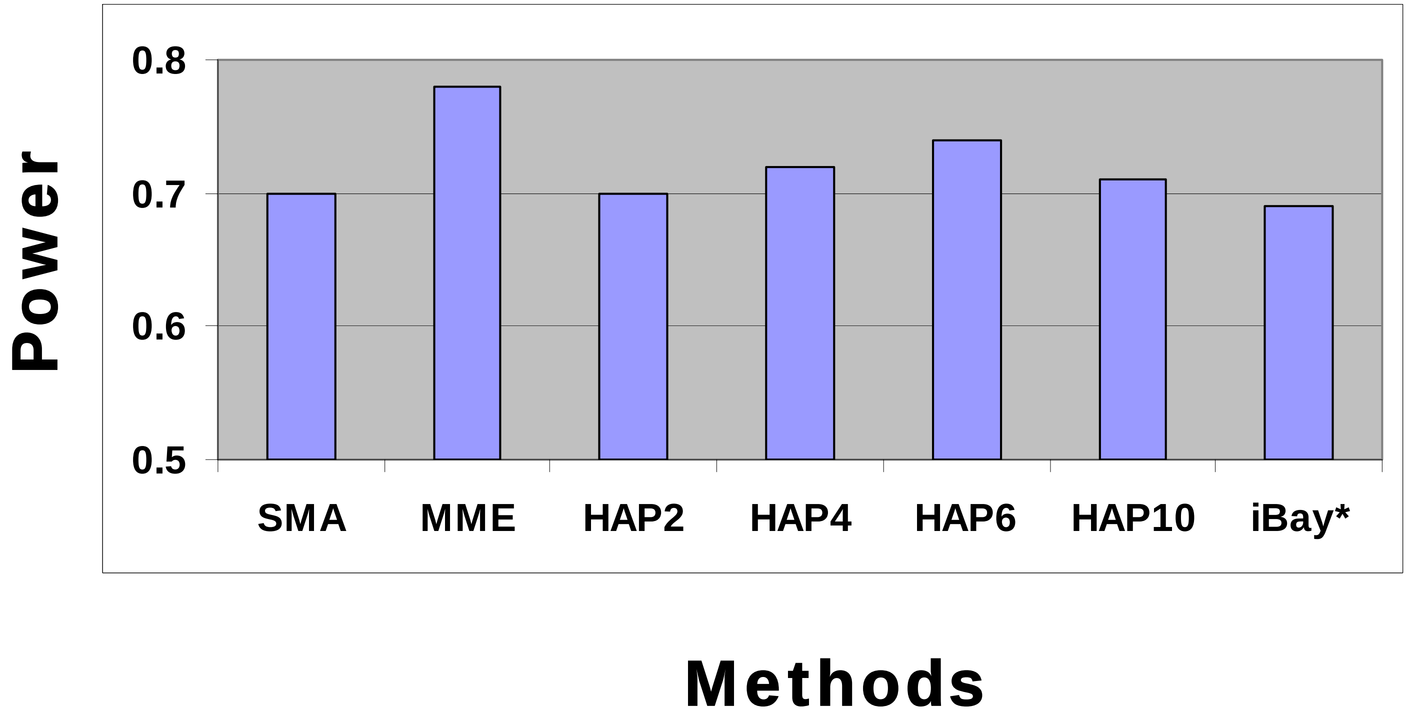

-log P-values
Single marker
Mixed model
Genome length in cM
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500


-log P-values
HAP10
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500
HAP4
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500
HAP2
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500
Genome length in cM


Posterior prob.
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
350
400
450
500
Bayesian
-log P-values
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500
Mixed model
Genome length in cM

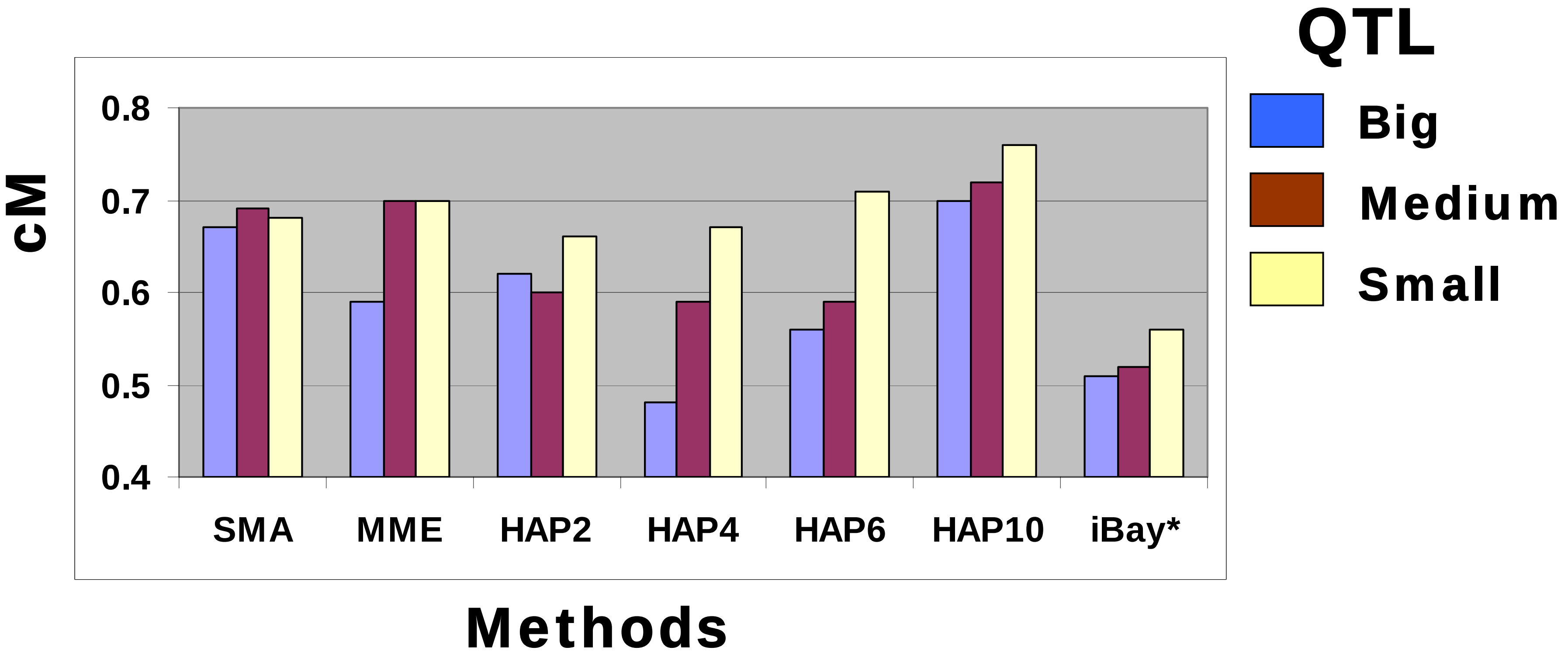
Power (QTL with 2% genetic variance)

A QTL was identified if a markers within 2cM region of the QTL location was significant



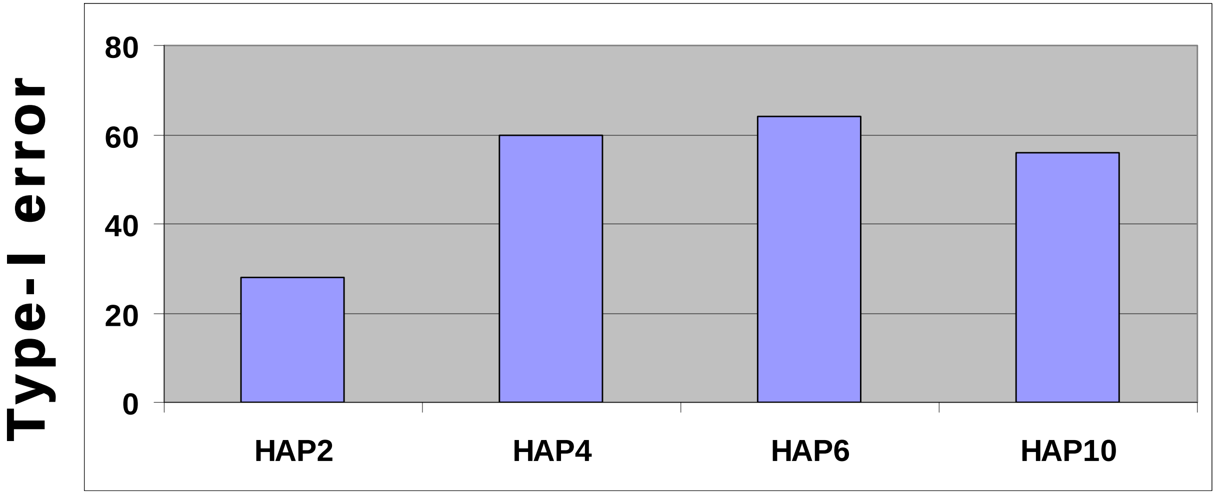
Mean error of QTL position

Precision was quantified as mean absolute error of position estimate



False Positives

- Single marker analysis and mixed model approach had type-I rate within expected limit
- Haplotype methods had very high Type-I error



Conclusions

- Mixed model approach had highest power in association mapping for QTL
- In general, Bayesian approach gave more precise location estimate
- The magnitude of QTL effect had little impact on precision
- Haplotype based methods had very high false positive rates