QTL Mapping, MAS, and Genomic Selection

Dr. Ben Hayes

Department of Primary Industries

Victoria, Australia

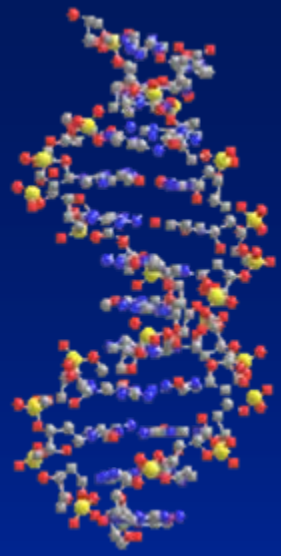
A short-course organized by

Animal Breeding & Genetics
Department of Animal Science
Iowa State University
June 4-8, 2007

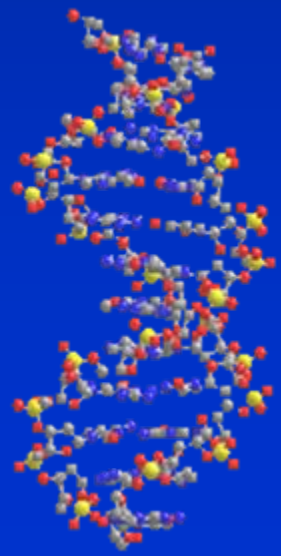
Animal Breeding & Genetics

With financial support from
Pioneer Hi-bred Int.

ANIMAL SCIENCE
IOWA STATE UNIVERSITY



Linkage Disequilbrium to Genomic Selection



Course overview

- Day 1
  - Linkage disequilibrium in animal and plant genomes
- Day 2
  - QTL mapping with LD
- Day 3
  - Marker assisted selection using LD
- Day 4
  - Genomic selection
- Day 5
  - Genomic selection continued

Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

Marker Assisted Selection using LD

- Marker assisted selection (MAS) can be based on DNA markers
  - in linkage equilibrium with a QTL (LE-MAS)
  - in linkage disequilibrium with a QTL (LD-MAS)
  - actual mutation causing QTL effect (Gene-MAS).
- All three types of MAS are currently used in the livestock industries (Dekkers 2004).

Table 1. Examples of gene tests used in commercial breeding for different species (D = dairy cattle, B = beef cattle, C = poultry, P = pigs, S = sheep) by trait category and type of marker

Trait category	Direct marker	Linkage disequilibrium marker	Linakge equilibrium marker
Congenital defects	BLAD (D[a])		
	Citrulinaemia (D,B[b])		
	DUMPS (D[c])		
	CVM (D[d])		
	Maple syrup urine (D,B[e])		
	Mannosidosis (D,B[f])		
	RYR (P[g])	RYR (P[h])	
Appearance	CKIT (P[i])		Polled (B[n])
	MC1R/MSHR (P[j],B[k],D[l])		
	MGF (B[m])		
Milk quality	-Casein (D[o])		
	ß-lactoglobulin (D[o])		
	FMO3 (D[p])		
Meat quality	RYR (P[g])	RYR (P[h])	
	RN/PRKAG3 (P[q])	RN/PRKAG3 (P[r])	
		A-FABP/FABP4 (P[s])	
		H-FABP/FABP3 (P[t])	
		CAST (P[u], B[v])	
>15 PICmarq (P[w])			
		THYR (B[x])	
		Leptin (B[y])	
Feed intake	MC4R (P[z])		
Disease	Prp (S[aa])	B blood group (C[bb])	
	F18 (P[cc])	K88 (P[dd])	
Reproduction	Booroola (S[ee])	Booroola (S[ff])	
	Inverdale(S[gg])	ESR (P[hh])	
	Hanna (S[ii])	PRLR (P[ij])	
		RBP4 (P[kk])	
Growth and composition	MC4R (P[z])	CAST (P[u])	QTL (P[ll])
	IGF-2 (P[mm])	IGF-2 (P[nn])	
	Myostatin (B[oo])		QTL (B[pp])
	Callipyge (S[qq])	Carwell (S[rr])	
Milk yield and composition	DGAT (D[ss])	PRL (D[tt])	QTL (D[uu])
	GRH (D[vv])		
	-Casein (D[o])		

Marker Assisted Selection using LD

- LE-MAS is most difficult to implement.
  - marker-QTL phase within each family must be established before an increase in selection response can be realised.
- LD-MAS now very attractive due to very large numbers of single nucleotide polymorphism (SNP) markers suitable for LD mapping now available.
- Gene-MAS requires enormous amount of work and resources!!

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
- In many cases, the selection candidates will have no phenotypic information of their own, eg young dairy bulls which are progeny test candidates.

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

LD-MAS with single markers

- Estimate effects of marker or markers in reference population

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}g + \mathbf{Zu}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X} & \mathbf{1_n'Z} \\ \mathbf{X'1_n} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{g}$$

LD-MAS with single markers

- Example

Animal	Sire	Dam	Phenotpe	SNP allele 1	SNP allele 2
1	0	0	3.53	1	1
2	0	0	3.54	1	2
3	0	0	3.83	1	2
4	0	0	4.87	2	2
5	0	0	1.91	1	2
6	0	0	2.34	1	1
7	0	0	2.65	1	1
8	0	0	3.76	1	2
9	0	0	3.69	1	2
10	0	0	3.69	1	2
11	1	2	-	1	2
12	1	4	-	2	1
13	5	6	-	1	1
14	5	7	-	2	1
15	5	8	-	2	2

LD-MAS with single markers

- The data was simulated as a SNP effect of 1 for 2 allele plus effect of sire 1 of 3 and sire 5 of -3 + random effect

LD-MAS with single markers

- Example

Animal	Sire	Dam	Phenotpe	SNP allele 1	SNP allele 2
1	0	0	3.53	1	1
2	0	0	3.54	1	2
3	0	0	3.83	1	2
4	0	0	4.87	2	2
5	0	0	1.91	1	2
6	0	0	2.34	1	1
7	0	0	2.65	1	1
8	0	0	3.76	1	2
9	0	0	3.69	1	2
10	0	0	3.69	1	2
11	1	2	-	1	2
12	1	4	-	2	1
13	5	6	-	1	1
14	5	7	-	2	1
15	5	8	-	2	2

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

LD-MAS with single markers

- Build:

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X} & \mathbf{1_n'Z} \\ \mathbf{X'1_n} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

LD-MAS with single markers

- Example
- $\mathbf{1_n}$ and **X**

Animal	Sire	Dam	Phenotpe	SNP allele 1	SNP allele 2
1	0	0	3.53	1	1
2	0	0	3.54	1	2
3	0	0	3.83	1	2
4	0	0	4.87	2	2
5	0	0	1.91	1	2
6	0	0	2.34	1	1
7	0	0	2.65	1	1
8	0	0	3.76	1	2
9	0	0	3.69	1	2
10	0	0	3.69	1	2
11	1	2	-	1	2
12	1	4	-	2	1
13	5	6	-	1	1
14	5	7	-	2	1
15	5	8	-	2	2

record	**1n**	**X**
1	1	0
2	1	1
3	1	1
4	1	2
5	1	1
6	1	0
7	1	0
8	1	1
9	1	1
10	1	1

LD-MAS with single markers

- Example
- **Z**

Animal	Sire	Dam	Phenotpe	SNP allele 1	SNP allele 2
1	0	0	3.53	1	1
2	0	0	3.54	1	2
3	0	0	3.83	1	2
4	0	0	4.87	2	2
5	0	0	1.91	1	2
6	0	0	2.34	1	1
7	0	0	2.65	1	1
8	0	0	3.76	1	2
9	0	0	3.69	1	2
10	0	0	3.69	1	2
11	1	2	-	1	2
12	1	4	-	2	1
13	5	6	-	1	1
14	5	7	-	2	1
15	5	8	-	2	2

		animal														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
record	1	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	2	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	3	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	4	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	5	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	6	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	7	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
	10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**

LD-MAS with single markers

- Example
- **A**
- $\lambda = 1 / 2$

Animal	Sire	Dam	Phenotpe	SNP allele 1	SNP allele 2
1	0	0	3.53	1	1
2	0	0	3.54	1	2
3	0	0	3.83	1	2
4	0	0	4.87	2	2
5	0	0	1.91	1	2
6	0	0	2.34	1	1
7	0	0	2.65	1	1
8	0	0	3.76	1	2
9	0	0	3.69	1	2
10	0	0	3.69	1	2
11	1	2	-	1	2
12	1	4	-	2	1
13	5	6	-	1	1
14	5	7	-	2	1
15	5	8	-	2	2

		Animal														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
animal	1	**1**														
	2	**0**	**1**													
	3	**0**	**0**	**1**												
	4	**0**	**0**	**0**	**1**											
	5	**0**	**0**	**0**	**0**	**1**										
	6	**0**	**0**	**0**	**0**	**0**	**1**									
	7	**0**	**0**	**0**	**0**	**0**	**0**	**1**								
	8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**							
	9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**						
	10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**					
	11	**0.5**	**0.5**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**				
	12	**0.5**	**0**	**0**	**0.5**	**0**	**0**	**0**	**0**	**0**	**0**	**0.25**	**1**			
	13	**0**	**0**	**0**	**0**	**0.5**	**0.5**	**0**	**0**	**0**	**0**	**0**	**0**	**1**		
	14	**0**	**0**	**0**	**0**	**0.5**	**0**	**0.5**	**0**	**0**	**0**	**0**	**0**	**0.25**	**1**	
	15	**0**	**0**	**0**	**0**	**0.5**	**0**	**0**	**0.5**	**0**	**0**	**0**	**0**	**0.25**	**0.25**	**1**

LD-MAS with single markers

- Example
- Solve equations..

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X} & \mathbf{1_n'Z} \\ \mathbf{X'1_n} & \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$\hat{\mu}$		2.69
$\hat{g}$		0.87
$\hat{u}$	1	0.56
	2	-0.01
	3	0.19
	4	0.3
	5	-1.1
	6	-0.23
	7	-0.03
	8	0.14
	9	0.09
	10	0.09
	11	0.28
	12	0.43
	13	-0.67
	14	-0.56
	15	-0.48

Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{g}$$

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{g}$$

$\hat{\mathbf{u}}$

0.28
0.43
-0.67
-0.56
-0.48

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{g}$$

$\hat{\mathbf{u}}$		$\mathbf{X}$	$\hat{g}$
0.28		1	0.87
0.43		1	
-0.67	+	0	
-0.56		1	
-0.48		2	

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{g}$$

$\hat{\mathbf{u}}$		$\mathbf{X}$	$\hat{g}$		**MEBV**
0.28		1	0.87		1.14
0.43		1			1.3
-0.67	+	0		=	-0.67
-0.56		1			0.3
-0.48		2			1.26

LD-MAS with single markers

- The data was simulated as a SNP effect of 1 for 2 allele plus effect of sire 1 of 3 and sire 5 of -3 + random effect

$\hat{\mathbf{u}}$		$\mathbf{X}$	$\hat{g}$		**MEBV**
0.28		1	0.87		1.14
0.43		1			1.3
-0.67	+	0		=	-0.67
-0.56		1			0.3
-0.48		2			1.26

LD-MAS with single markers

- Corr(MEBV,TBV) =0.93

$\hat{\mathbf{u}}$		$\mathbf{X}$	$\hat{g}$		**MEBV**	**TBV**
0.28		1	0.87		1.14	1.75
0.43		1			1.3	1.75
-0.67	+	0		=	-0.67	-0.75
-0.56		1			0.3	0.25
-0.48		2			1.26	1.25

LD-MAS with single markers

- Corr(MEBV,TBV) =0.93
- Corr(EBV,TBV)=?

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{Z'y} \end{bmatrix}$$

LD-MAS with single markers

- Corr(MEBV,TBV) =0.93
- Corr(EBV,TBV)=0.88

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{Z'y} \end{bmatrix}$$

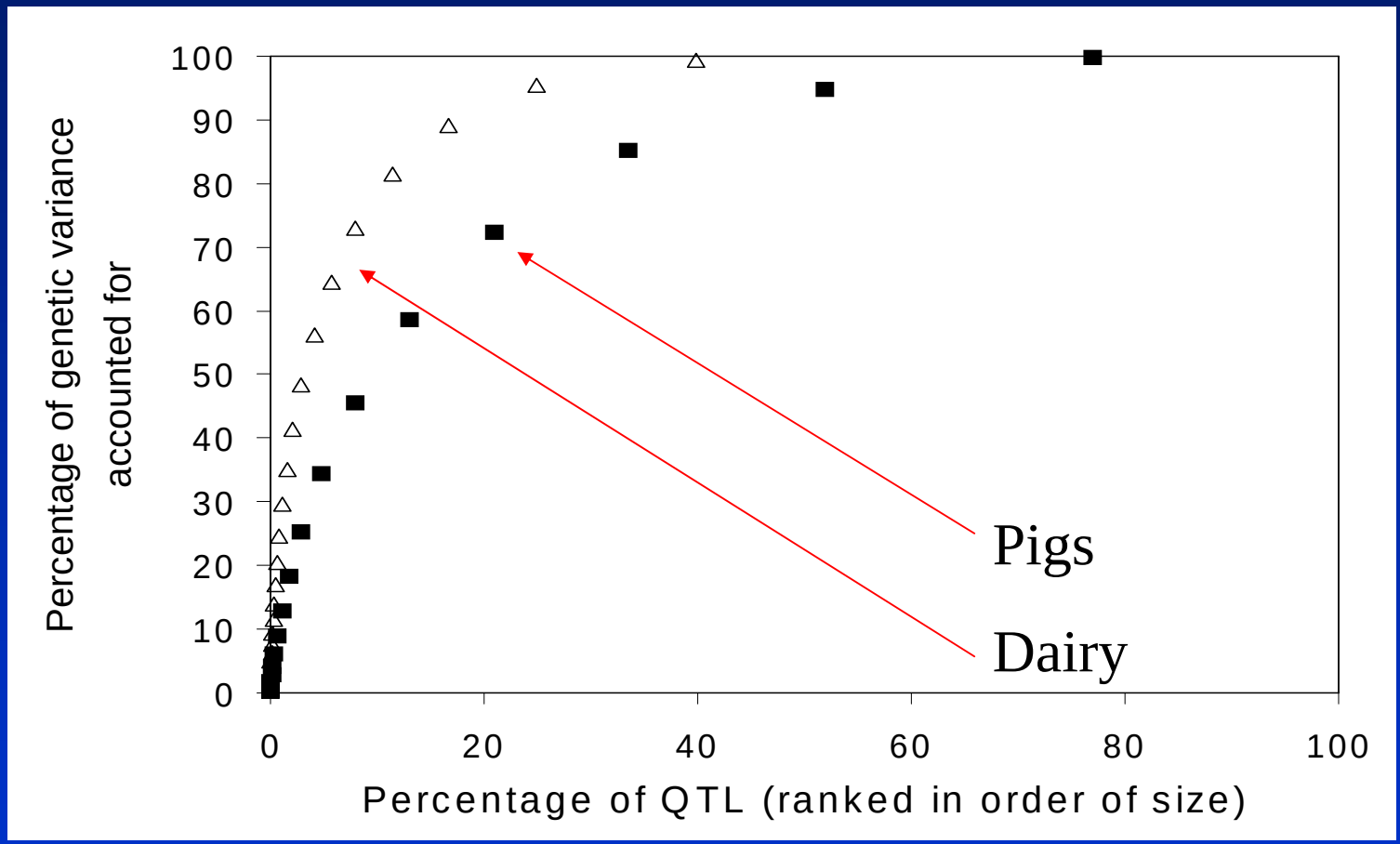
Marker Assisted Selection using LD

- LD-MAS with a single marker
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

How many QTL to use in LD-MAS

- Advantage of MAS over non-MAS approximately proportional to proportion of total genetic variance explained by QTL
- Estimates of number of QTL per trait between 100 and 200
- Do we need to track all these with markers?

How many QTL to use in LD-MAS



How many QTL to use in LD-MAS

- If we use 10-20 QTL per trait in our LD-MAS program, we will exploit ~ 50% of the genetic variance.
- Assumes we have perfect knowledge of the QTL alleles.
- The proportion of genetic variance captured at each QTL in LD-MAS depends on the extent of linkage disequilibrium between the marker and the QTL.

How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X_1}g_1 + \mathbf{X_2}g_2 + \mathbf{e}$$

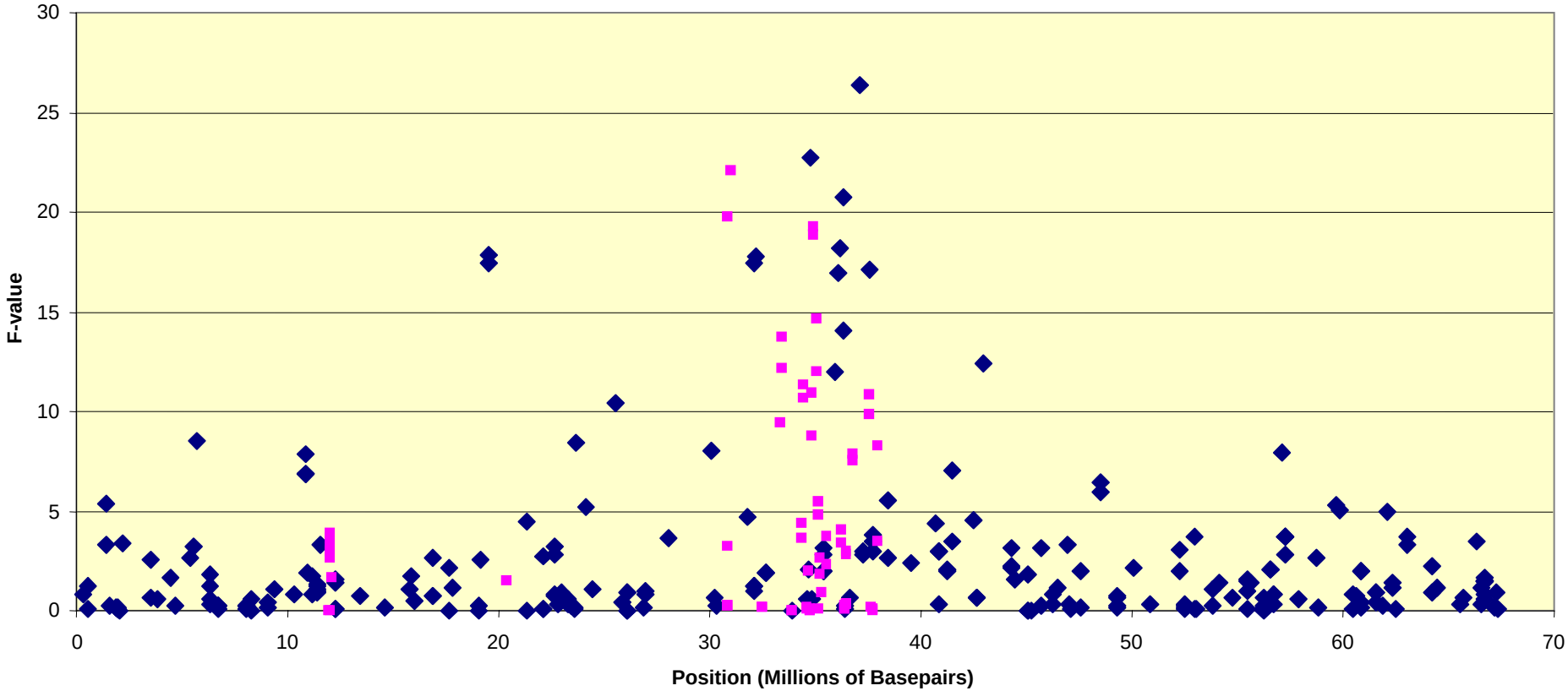
How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers

$$\begin{bmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1_n'1_n & 1_n'X_1 & 1_n'X_2 & 1_n'Z \\ X_1'1_n & X_1'X_1 & X_1'X_2 & X_1'Z \\ X_2'1_n & X_2'X_1 & X_2'X_2 & X_2'Z \\ Z'1_n & Z'X_1 & Z'X_2 & Z'Z + A^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} 1_n'y \\ X_1'y \\ X_2'y \\ Z'y \end{bmatrix}$$

How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers
- Accounts for the fact that some SNPs may be picking up the same QTL


30
25
20
15
10
5
0
F-value
0
10
20
30
40
50
60
70
Position (Millions of Basepairs)

LD-MAS with single markers

- Predict breeding values using marker information:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}_1 \hat{g}_1 + \mathbf{X}_2 \hat{g}_2 + ...$$

How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers (random?)

$$\begin{bmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X_1} & \mathbf{1_n'X_2} & \mathbf{1_n'Z} \\ \mathbf{X_1'1_n} & \mathbf{X_1'X_1 + I\lambda_1} & \mathbf{X_1'X_2} & \mathbf{X_1'Z} \\ \mathbf{X_2'1_n} & \mathbf{X_2'X_1} & \mathbf{X_2'X_2 + I\lambda_2} & \mathbf{X_2'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'X_1} & \mathbf{Z'X_2} & \mathbf{Z'Z + A^{-1}\lambda} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X_1'y} \\ \mathbf{X_2'y} \\ \mathbf{Z'y} \end{bmatrix}$$

- Use variance component estimation to get SNP effects

Marker Assisted Selection using LD

- LD-MAS with a single marker
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term

Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term
- This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.

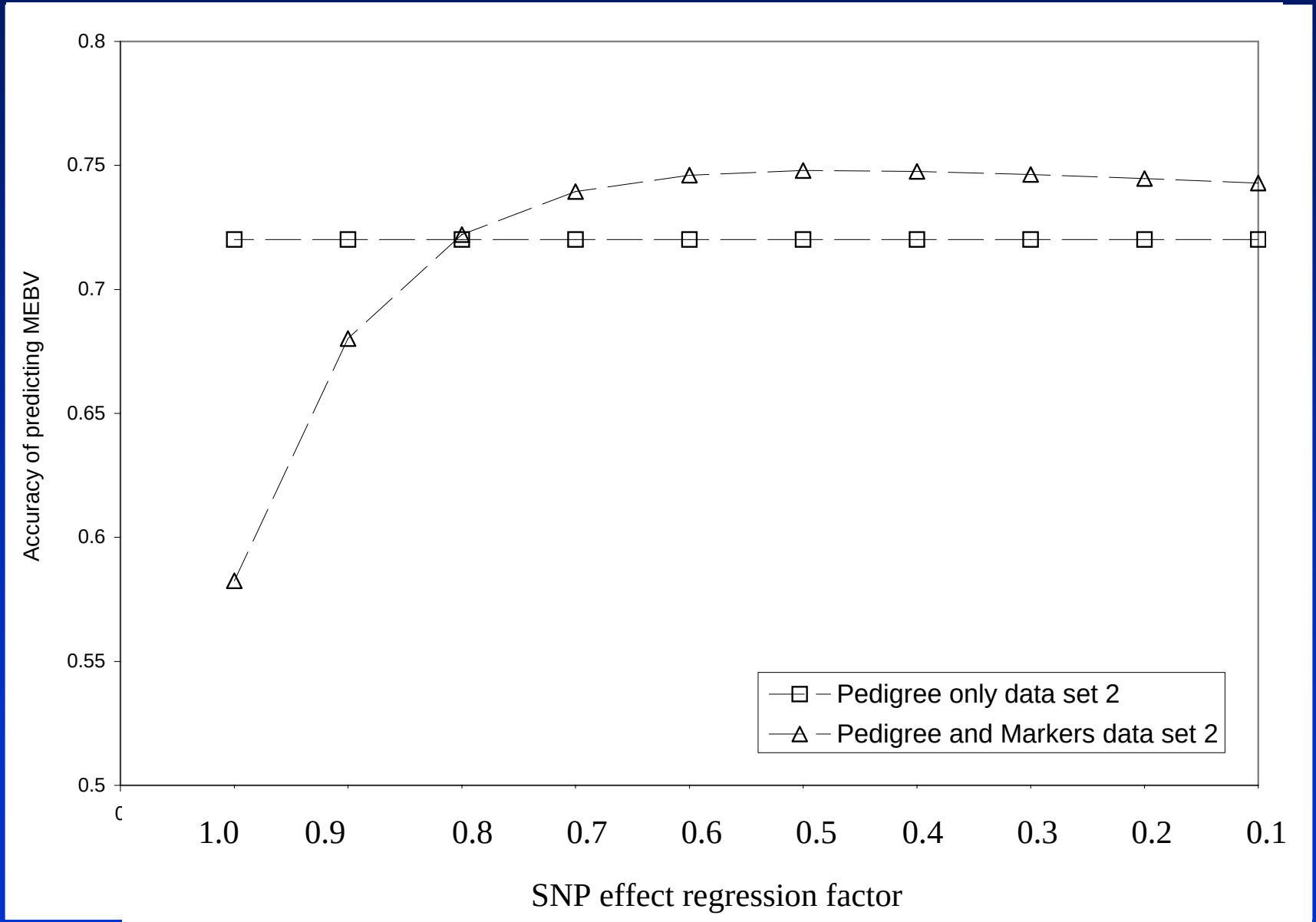
Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term
- This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.
- If the QTL effect is over-estimated, the advantage of MAS can be eroded substantially (eg LD-MAS with a single marker)

Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term
- This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.
- If the QTL effect is over-estimated, the advantage of MAS can be eroded substantially (eg LD-MAS with a single marker)
- Must regress QTL effects prior to use in MAS

Accounting for bias in QTL effects

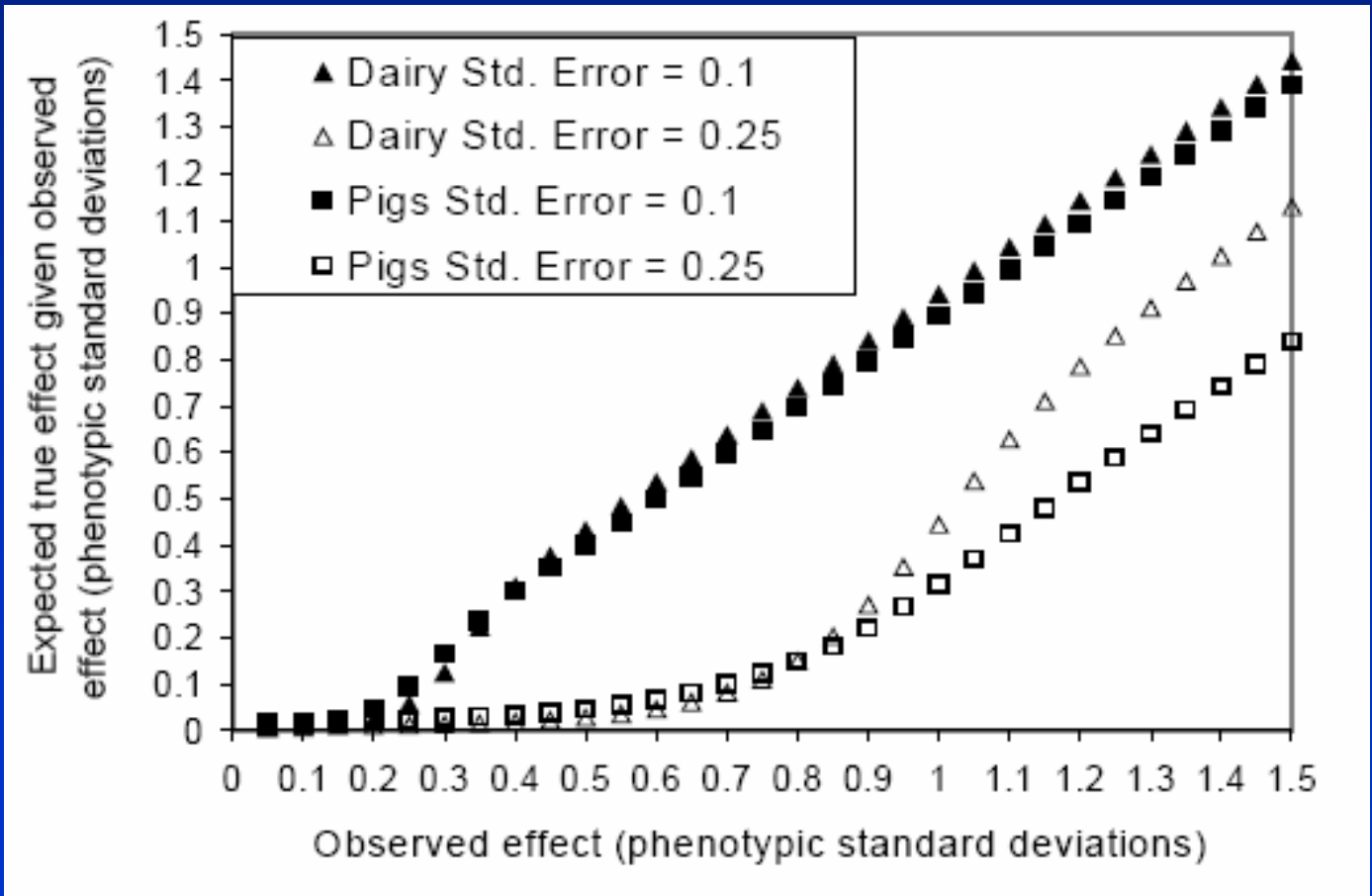


Accounting for bias in QTL effects

- Options for estimating unbiased estimates of QTL effect
  - Best method is to estimate QTL effects in a population which is completely independent of the sample used in the original genome scan where the QTL were first detected.
  - This will also validate that the markers are not an artefact of the statistical model used in the genome scan or some unaccounted for population stratification.
  - But maybe too expensive
  - Use prior knowledge of distribution of QTL effects to regress effects
  - Cross validation

Accounting for bias in QTL effects

- Use prior knowledge of distribution of QTL effects to regress effects
- Then for a given size of experiment and estimated size of effect, we can calculate the true effect



Accounting for bias in QTL effects

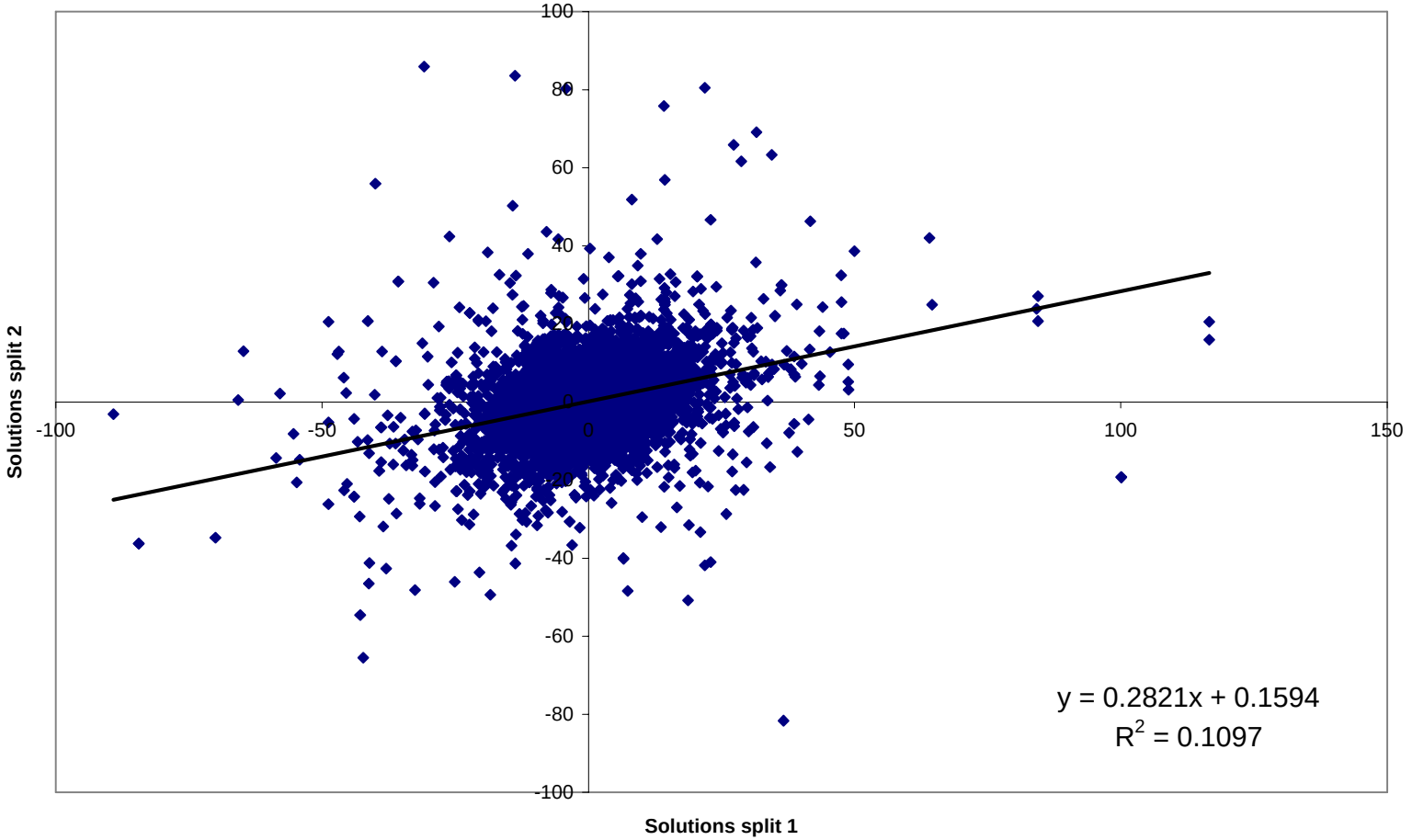
- Use prior knowledge of distribution of QTL effects to regress effects
- Then for a given size of experiment and estimated size of effect, we can calculate the true effect
- See Weller et al. 2005 for distributions of QTL effects across traits

Accounting for bias in QTL effects

- Cross validation
  - split data set in two
  - regress solutions from data set two on data set one to get b_{x1x2}
  - then the regression of the true effects of the SNPs on the solutions from the full data set is
    - $b_{u,xt} = 2b_{x1x2}/(1+b_{x1x2})$

Accounting for bias in QTL effects

- Cross validation



Accounting for bias in QTL effects

- Cross validation
  - split data set in two
  - regress solutions from data set two on data set one to get b_{x1x2}
  - then the regression of the true effects of the SNPs on the solutions from the full data set is
    - $b_{u,xt} = 2b_{x1x2}/(1+b_{x1x2})$
    - $= 0.44$

Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

LD-MAS with haplotypes

- Model:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \mathbf{X}\hat{\mathbf{g}}$$

- **g** is a vector of haplotype effects, eg.

Haplotype	Effect
1	0.2
2	-0.12
3	-0.11
4	0.21

LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes
    - Number of haplotype effects to estimate
    - Number of phenotypic records
    - Accuracy of inferring haplotypes

LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes
    - Number of haplotype effects to estimate
    - Number of phenotypic records
    - Accuracy of inferring haplotypes

LD-MAS with haplotypes

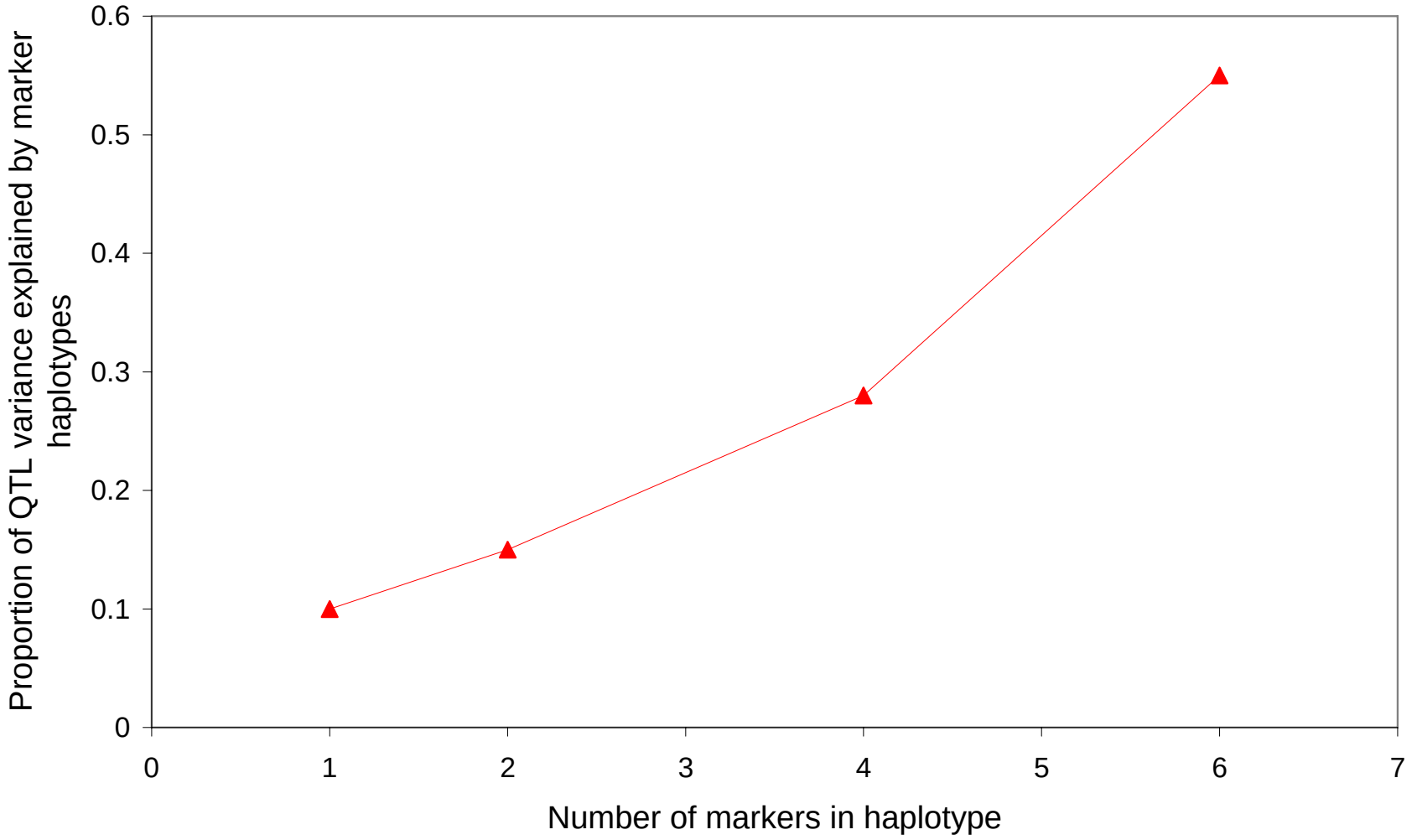
- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes

$$r^2(h,q) = \sum_{i=1}^{n} \frac{D_i^2}{p_i} \Big/ q_1 q_2$$

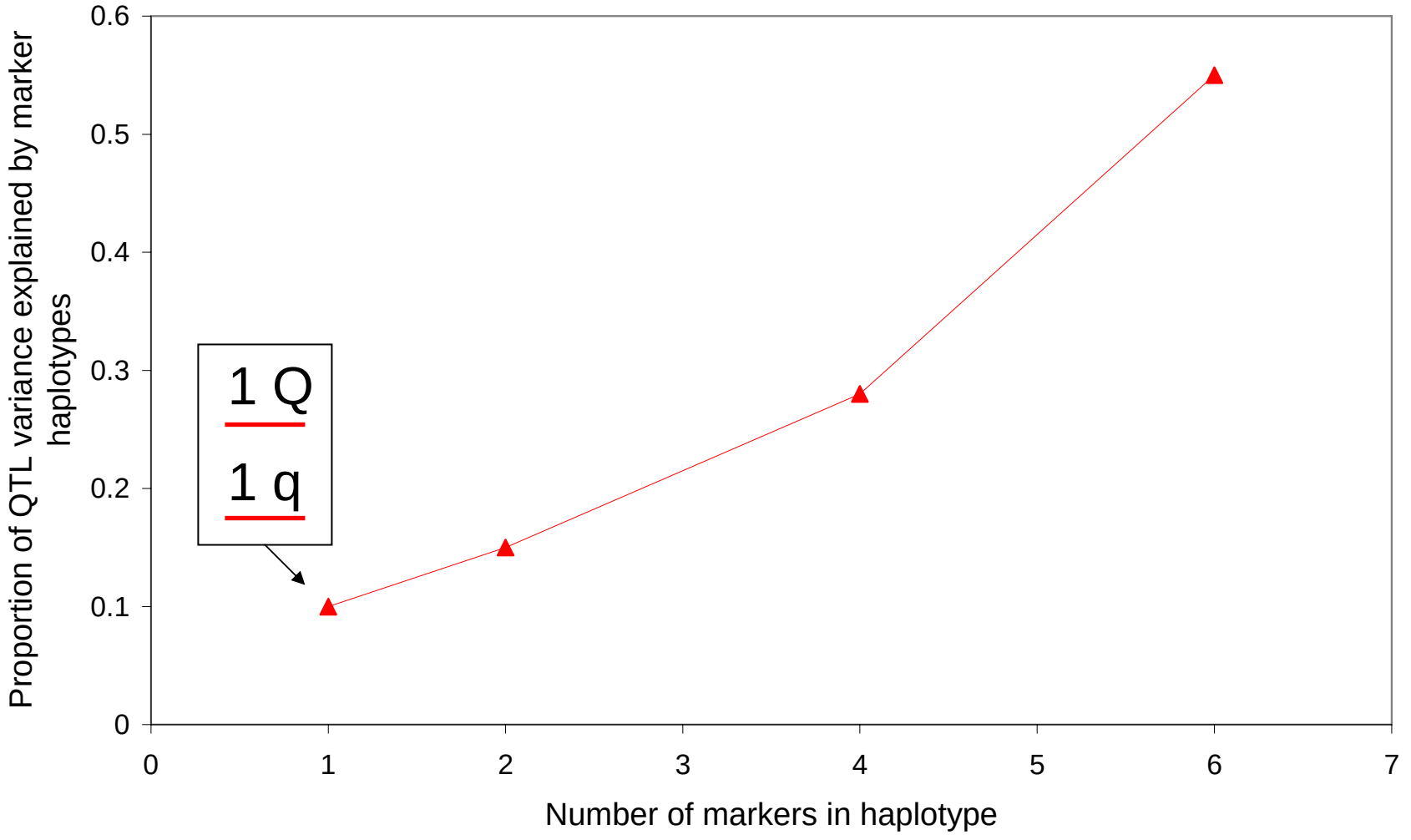
LD-MAS with haplotypes

- Example:
- 10 000 SNPs genotyped in 379 Angus animals
- Select a SNP from the 10 000 at random to be a “QTL”
  - determine
    - Nearest marker
    - 2, 4 or 6 marker haplotypes
      - Haplotypes estimated using PHASE program (Stephens et al. 2001)
  - This takes into account LD structure in the cattle populations
- Calculate the proportion of QTL variance explained by the marker haplotypes.

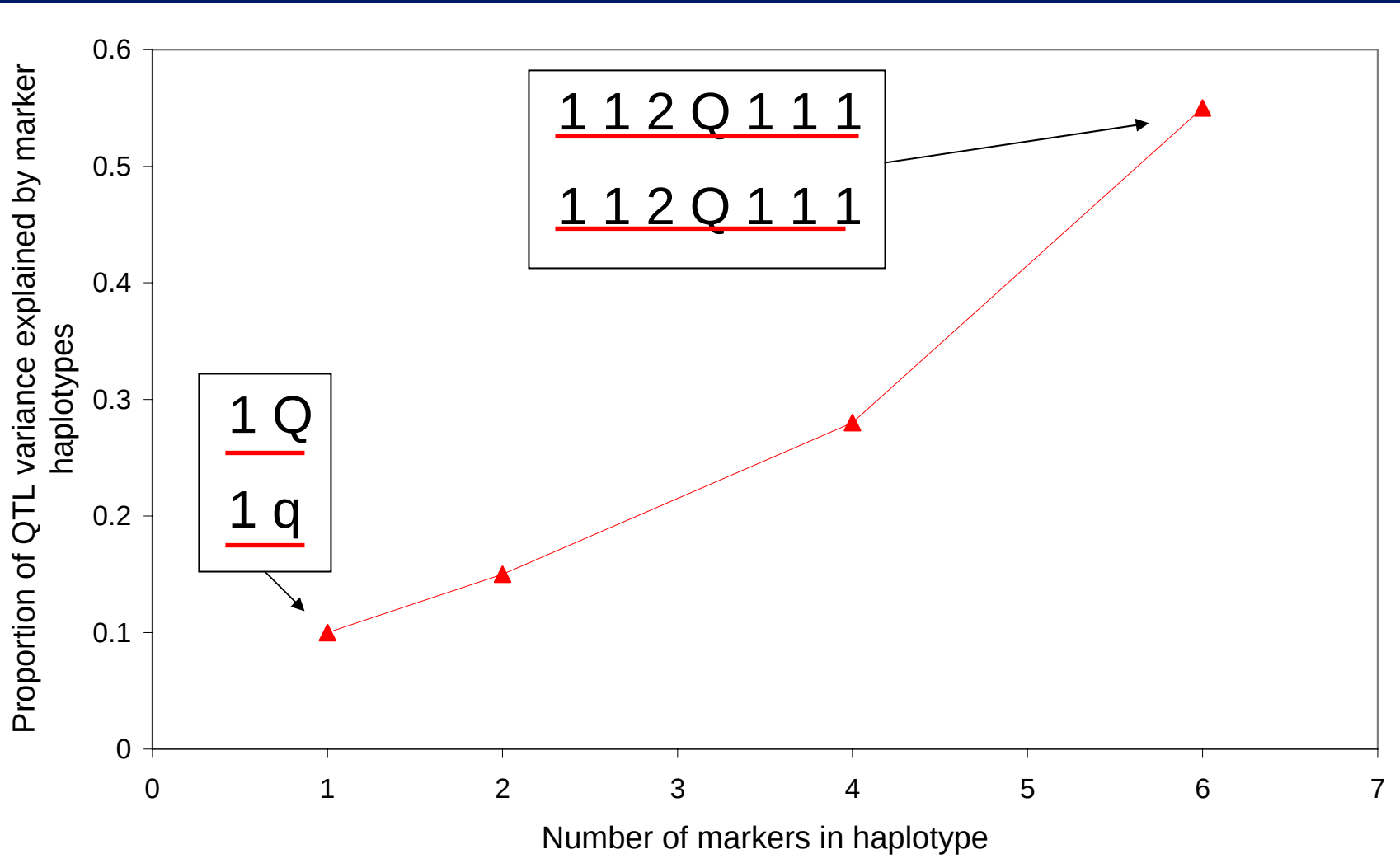
Results


Proportion of QTL variance explained by marker haplotypes
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
Number of markers in haplotype

Results


Proportion of QTL variance explained by marker haplotypes
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
Number of markers in haplotype
1 Q
1 q

Results


Proportion of QTL variance explained by marker haplotypes
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
Number of markers in haplotype
1 1 2 Q 1 1 1
1 1 2 Q 1 1 1
1 Q
1 q

LD-MAS with haplotypes

- Example:

	Proportion of QTL variance explained	Maximum number of haplotypes	Observed number of haplotypes
Nearest marker	0.10	2	2
Best marker	0.20	2	2
2 Marker haplotypes	0.15	4	3.4
4 Marker haplotypes	0.28	16	9.4
6 Marker haplotypes	0.55	64	20.8

LD-MAS with haplotypes

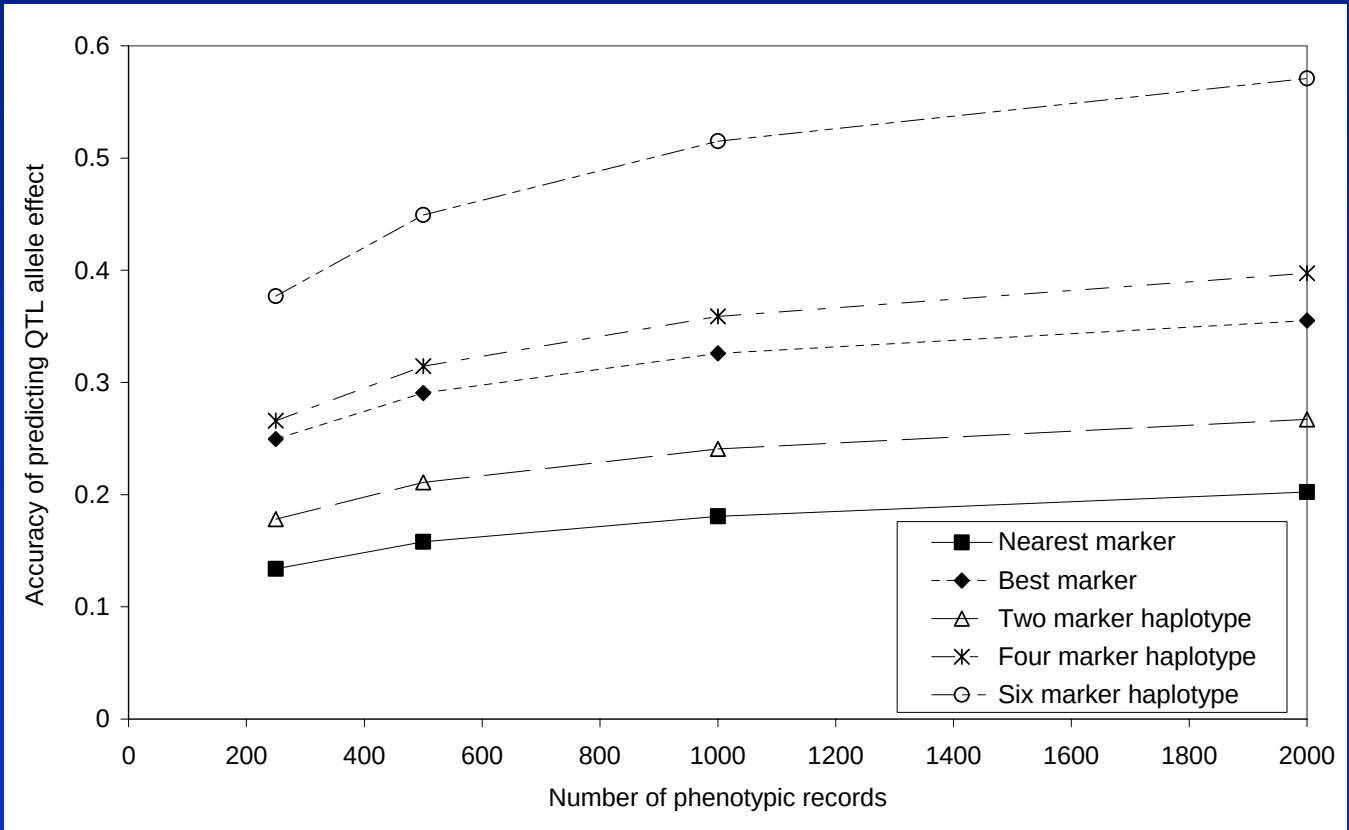
- Accuracy of estimating QTL allele effects from haplotypes:

$$r(q,\hat{h}) = \sqrt{r(h,q)^2 \sum_{i=1}^{n} \frac{p_i^2}{p_i + \lambda / T}}$$

$$\lambda = \sigma_e^2 / \sigma_h^2$$

LD-MAS with haplotypes

- Accuracy of estimating QTL allele effects from haplotypes:



LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes
    - Number of haplotype effects to estimate
    - Number of phenotypic records
    - Accuracy of inferring haplotypes??

Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

LD-MAS with the IBD approach

- MEBVs:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \hat{\mathbf{v}}$$

LD-MAS with the IBD approach

- MEBVs:

$$\mathbf{MEBV} = \hat{\mathbf{u}} + \hat{\mathbf{v}}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'Z} & \mathbf{1_n'W} \\ \mathbf{Z'1_n} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda_1 & \mathbf{Z'W} \\ \mathbf{W'1_n} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix}$$

- Where W is a matrix allocating records to QTL allele effects

LD-MAS with the IBD approach

- Has the potential to be most accurate method for LD-MAS because can capture linkage information as well
  - Particularly with sub-optimal markers densities

Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

Gene Assisted Selection

- Greatest increases in response (not limited by LD)
- Simplest, cheapest to implement in breeding program
  - No need to establish phase within families
  - Cost of discovery very high
  - Number of examples now (Dekkers 2004)
  - May become apparent that mode of inheritance is not additive
  - Eg. IGF2 mutation in pigs is imprinted (only expressed if mutated allele from father)

Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information

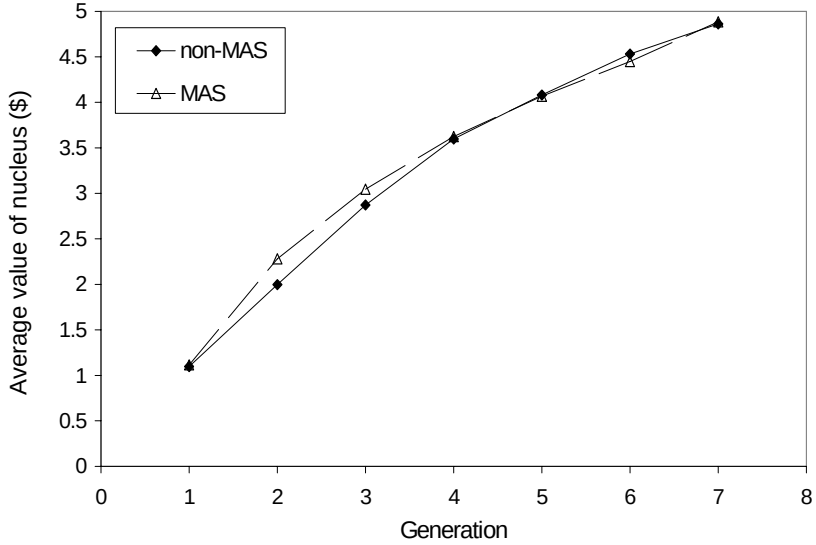
Optimising the breeding scheme with MAS

- Which traits
- Age at selection?

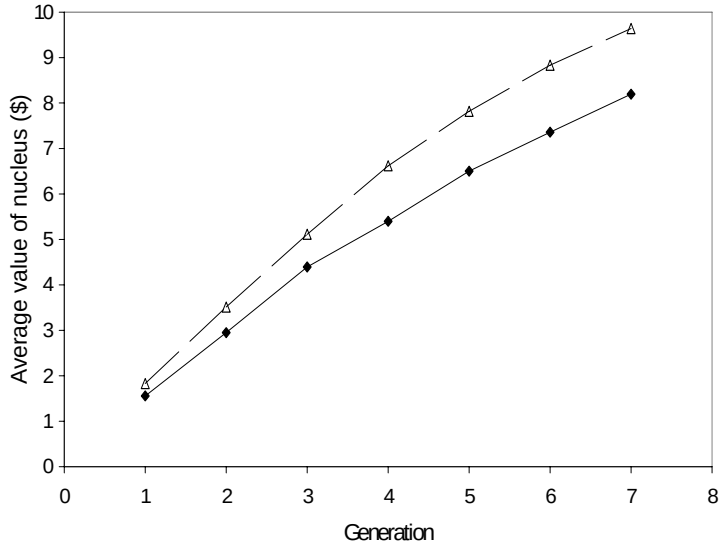
Optimising the breeding scheme with MAS

- Expected response from MAS
  - Traits measured on both sexes before selection << traits measured on one sex before selection << traits measured after selection << traits measured on relatives

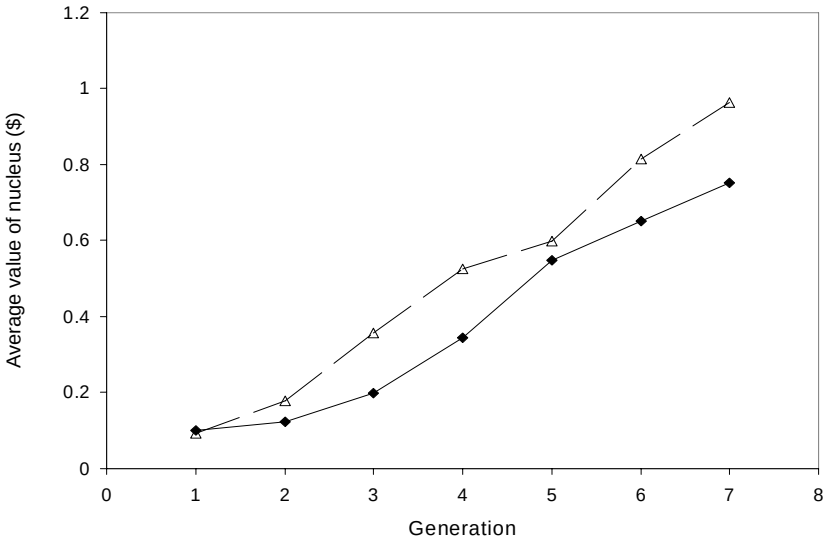
Traits measured before selection	*Traits measured on one sex before selection*	*Traits measured after selection*	*Traits measured on relatives*
Growth Fatness	Feed intake Milk production	Pigs born alive Fertility Disease resistance (cattle)	Carcass quality Disease resistance (fish)



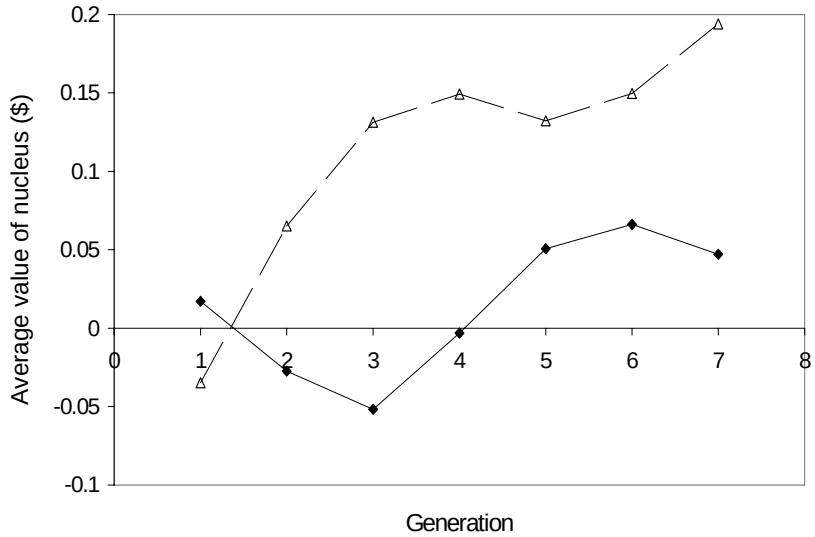
GI



NFI



PBA



MQI

Optimising the breeding scheme with MAS

- Which traits
- Age at selection
  - $G = ir\sigma_g / L$
    - where G = genetic gain
    - i is the intensity of selection
    - r is the accuracy of selection
    - σ_g is the genetic standard deviation and
    - L is the generation length

Optimising the breeding scheme with MAS

- Age at selection
  - We have already discussed improving r
  - What about L?
- Accuracy of traditional EBVs increase as animal ages and it and its relatives acquire phenotypic data.
- But animals can be typed for markers at any age
- Gain in accuracy from markers greatest at young age.
- So if selection optimised, marker data should lead to a decrease in generation length
- Eg. in dairy cattle selected for milk production, MAS leads to greater gains if selection of yearling bulls and cows is practiced than if a traditional progeny testing system is adhered to
- Reproductive technologies?

Take home points

- Markers in LD with QTL relatively easy to use in breeding programs
- Using haplotypes may improve accuracy?
- IBD approach allows linkage information to be used as well
- Response: Traits measured on both sexes before selection << traits measured on one sex before selection << traits measured after selection << traits measured on relatives
- Optimal use of marker information with selection at younger ages