Friday

Marker assisted selection

Marker assisted breeding

Marker Assisted Selection

"QTL analysis has produced great advances in plant breeding"

Rex Bernardo *Crop Sci*. **48:**1649–1664 (2008).

"at least 10,000 marker-trait associations in different plant species have been reported"

"exploiting the QTL that have been mapped has not been routinely done"

$$R = ih\sigma_g$$

The breeders' equation

$$R = h^2S.$$

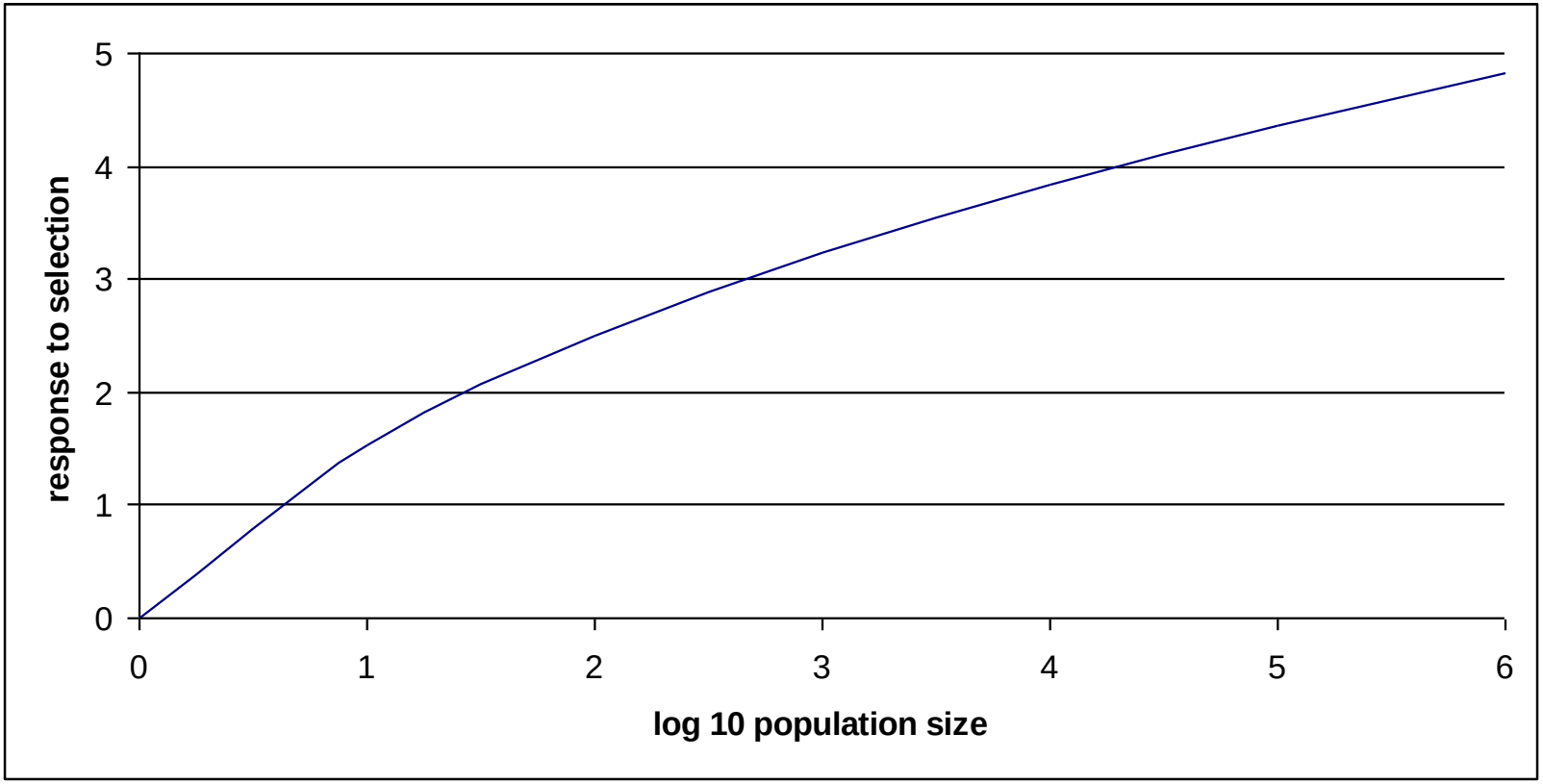
$$R = ih\sigma_g$$

R	response to selection
h^2	heritability – really just a regression coefficient
S	selection differential
i	standardised selection differential
σ_g	additive genetic variance

Ways to increase response to selection:

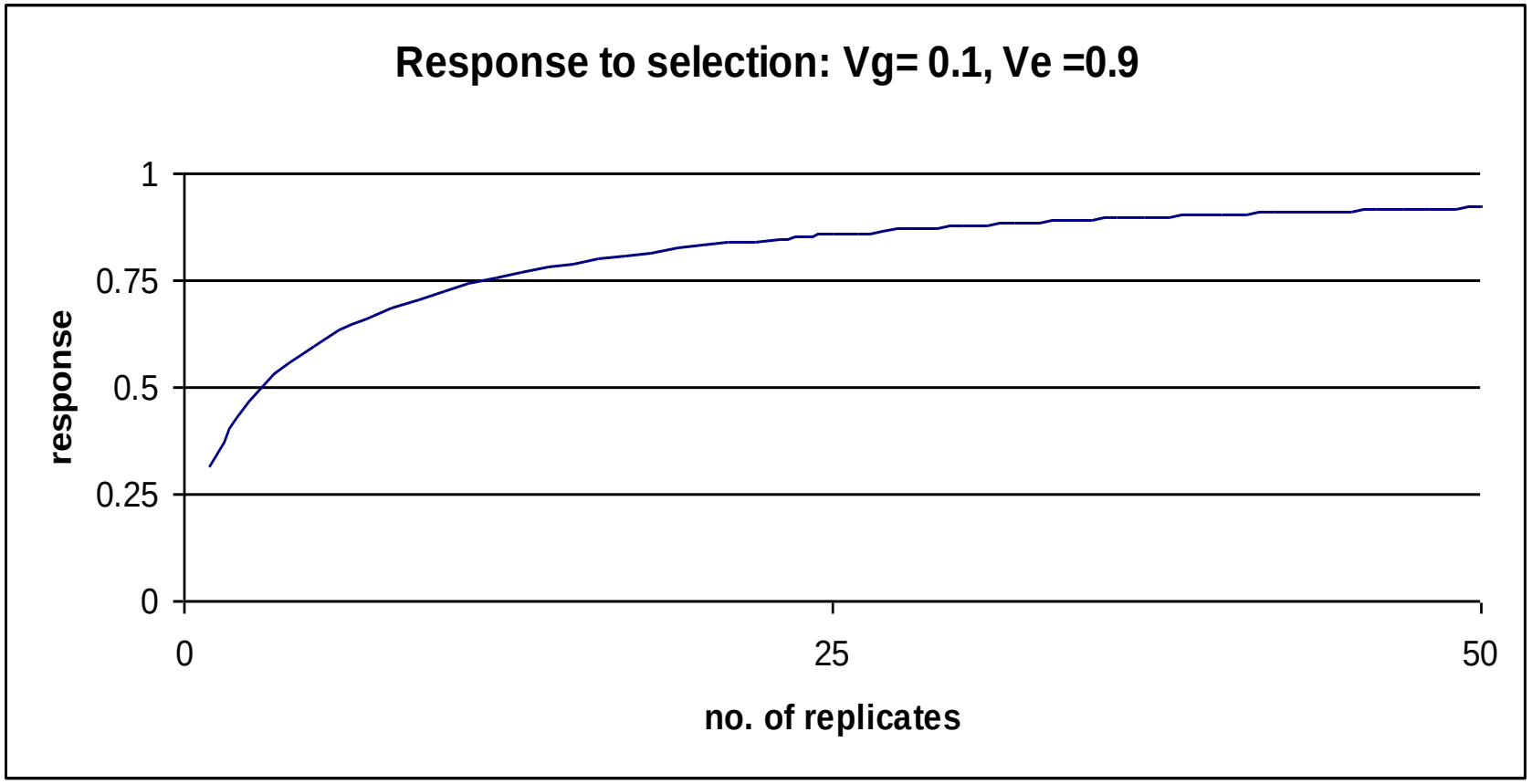
target	method
increase i	test more lines
reduce time	out of season
increase h	test more plots per line
	indirect selection (markers or traits)
reduce costs	indirect selection (eg grain quality)
increase σ_g	wide crosses / mutation

Size isn't everything #1: increasing i is not cost effective.



10	→ 100	increases response by	63%
100	→ 1,000		29%
1,000	→ 10,000		19%

Size isn't everything #2: increasing scale is not cost effective.



Double the speed, double the response

Methods:

1. Out of season nurseries.

2. Make crosses in advance of results.

    Breeders already do this, implicitly.

    Explicit schemes: accelerated recurrent selection

3. Marker assisted breeding

    Not for polygenic traits so far.

    Genomic selection?

Marker Assisted Selection

select on phenotype alone $R = ih_p^2\sigma_p$

select on markers alone $R = ir_g h_m h_p \sigma_p$

For MAS to give a greater response than phenotypic selection

$$ir_g h_m h_p \sigma_p > ih_p^2 \sigma_p$$

$$r_g h_m > h_p$$

but since $h_m^2 = 1$ (assuming no genotype errors)

$$r_g > h_p$$

$$r_g^2 > h_p^2$$

The genetic correlation coefficient squared between marker index and genotype must be higher that the heritability of the phenotype.

When does selection on markers work well?

$$r_g^2 > h_p^2$$

Classic MAS selects for specific tagged loci. Good if:

- most Vg controlled by a very small number of tagged QTL
- low heritability
- trait expensive to score
- trait scored post reproduction
- quicker

GS is more flexible:

- no requirement for large gene effects
- no requirement to tag individual QTL
- but requires many cheap markers

Marker Assisted Selection

Index selection

Combine markers and phenotype information.

“molecular score”

Still need accurate assessment of markers or can make things worse.

“genomic selection.” The next thing.

MAS on quantitative traits

Lande & Thompson 1990.

Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits.

Benchmark treatment of MAS for quantitative traits in the context of quantitative genetics and selection theory.

Proposed method never caught on:

- marker density?
- problems in selecting associated markers?

Marker Selection: “The winner’s curse.”

(The Beavis effect in linkage analysis.)

With multiple QTL of small effect, some get lucky and are detected.

These are genuine QTL, but their effect is overestimated.

E.g.

A mapping experiment with 101 genes & h^2	=	100%
Standardised difference between homozygotes	=	1
Mean of those detected as sig ($p<0.05$)	=	2.49

Genomic selection

Trait effects of all genes or chromosomal positions are estimated simultaneously without significance testing (eliminates bias).

High marker density

Estimate trait effect for every marker or interval

Statistical problem – more markers than individuals

Predicting Unobserved Phenotypes for Complex Traits from Whole-Genome SNP Data *PLoS Genetics* Lee *et al.* 2008

"…correlations between predicted and actual phenotypes are in the range of 0.4 to 0.9. The prediction of unobserved phenotypes from high-density SNP data and appropriate statistical methodology is feasible and can be applied in human medicine, forensics, or artificial breeding programs."

Genomic Selection

Proposed 2001:

Meuwissen et al. 2001 Genetics. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

Trait effects of all genes or chromosomal positions are estimated simultaneously without significance testing so there is no bias.

- Requires high marker density.
- Statistical problem: more markers than individuals.
- Estimate a trait effect for every marker or interval.

How to do it.

Training set (population):

Markers

Phenotypes.

Regress phenotypes on markers in the training set

Use regression equations to predict phenotypes from markers in novel germplasm.

Select, cross and repeat

After a few generations, derive a new training set and start again

How many depends on LD, population structure, h^2.

The calibration phase

questions	variables
What population to use?	LD
How many individuals?	h^2
How many markers?	Allele frequencies
	recalibration interval
	breeding methods

Extreme examples illustrate problems:

Use lines from one cross to predict in another?

p(locus segregating in both crosses) is ≤ ¼

Prediction within a single cross: many linked loci

Some loci linked in dispersion, some in repulsion. Calibration will work on the net effect. Validity after recombination?

Calibration: statistical methods

BLUP on trait (selection index):

Predicts performance on individuals with no trait data from genetic relationship with individuals with trait data.

Used in animal breeding for decades (using pedigree relationships)

Ridge Regression (BLUP on markers):

Add a common penalty to each marker to reduce its effect.

By reducing the influence of every marker, all markers can be fitted.

Bayesian & other methods:

All methods predict on the basis of kinship to some extent

Predictions from kinship are never better than the best observation:

Suppose $h^2 = 1$ for a polygenic trait.

Predicted breeding value is a weighted mean of the phenotyped relatives:

Source of Prediction	r^2
Progeny from parents	½
gp from grand parents	¼
ggp from great grand parents	⅛

We need methods which escape the gravitational pull of kinship.

This is not just do to with algorithms. Small training sets, low numbers of markers: the kinship signal is the only thing the markers can hook.

Recalibration

As cycles of selection proceed:

allele frequencies change.

recombination acts to reduced LD

selection acts to increase LD

Minimum no. of generations before recalibration will depend on:

initial LD

allele frequencies

intensity of selection

Select within crosses or select between crosses?

Between crosses:

we will be selecting mainly on kinship

nothing wrong with this, but we do it already

Within crosses:

higher LD, fewer markers, smaller training sets.

cannot select on pedigree estimates of kinship

But:

Only ½ Va is available within crosses.

Only ¼ Va is available for GS.

High LD in early generations: how long will the predictions last?

Select within n-way crosses

Available Va rises:

	Vg within crosses
F2 lines	1/2
4-way Xs	3/4
8-way	7/8

Constructed to reduce LD: predictions may last longer.

Takes time.

Breeding wheat / barley with genomic selection. Finding the right balance:

Years	F2	4-way cross	8-way cross
1	Cross + DH	Cross	Cross
2	Bulk	Cross + DH	Cross
3	Trial	Bulk	Cross + DH
4	Trial	Trial	Bulk
5	GS / crosses	Trial	Trial
6	GS / crosses	GS / crosses	Trial
7		GS / crosses	GS / crosses
8			GS / crosses

←	LD: fewer markers	←
→	Diversity: more response	→
→	Time: lower resp/year	→
→	Population size: more power	→

Genomic Selection

Reality 2009:

Hayes et al. 2009 J. Dairy Sci. 92:433-44
Genomic selection in dairy cattle: Progress and challenges

"… at least 2 dairy breeding companies are already marketing bull teams for commercial use based on their GEBV only, at 2 yr of age. This strategy should at least double the rate of genetic gain in the dairy industry."

Existing methods should translate more readily to tree species?

Outbreeders.
Long breeding cycle time: more to gain
Learn from the mistakes of animal breeders.

Wong & Bernardo Genomewide selection in oil palm:
Theor Appl Genet (2008) 116:815–824

Marker Assisted Selection & major genes

Opportunities to save time and money

Location and effect are generally well known.

Opportunities to speed things up, especially in

Marker Assisted Backcrossing

Conclusions

In breeding, speed is more important than size.

Genomic selection will reduce cycle time.

Problems remain but:

For the first time in ~25 years of QTL mapping, it may be possible to incorporate MAS for yield into practical (commercial) breeding programmes.

"The merging of quantitative and population genetics, driven by data generated by large-scale high-throughput genomics platforms, offers new approaches to classical problems in quantitative genetics."

P Visscher 2009 Whole genome approaches to quantitative genetics Genetica 136:351-358

Conclusion

$$R = h^2 S$$