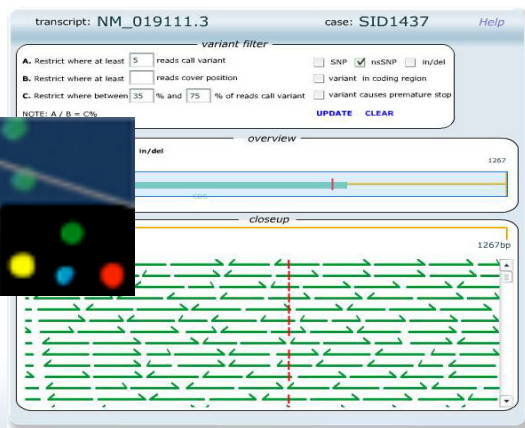
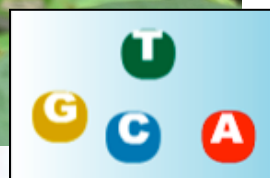


TRANSLATING DISCOVERIES FOR SOYBEAN BREEDING

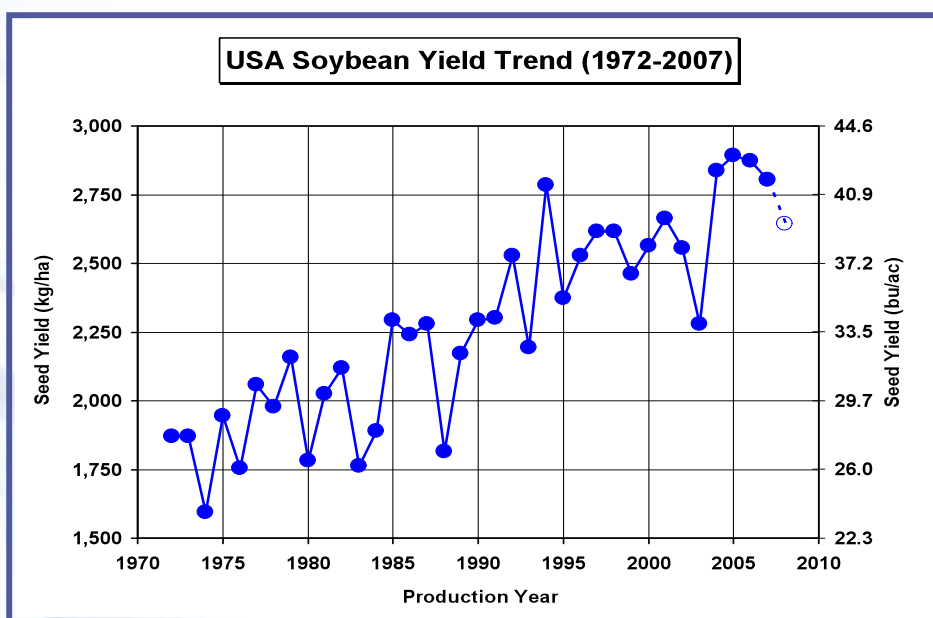
Soybean Breeders and Physiologists Workshop
February 21-23, 2011
St. Louis, Missouri



Bill Beavis

IOWA STATE UNIVERSITY
Department of Agronomy
Crop, Soil, and Environmental Sciences

40 Years of Successful Soybean Breeding



Is this rate sufficient to meet the needs of nine billion people ?

Jim Specht: 0.35 bu/ac per year

An Indictment of Public Plant Breeders

- The vast majority QTL reside in journals rather than in improved cultivars.

-Rex Bernardo, 2009

Why have we focused on identifying QTL and not crop improvement?

- **“We already know how to use biological markers for crop improvement”**

- Single gene traits - MABC
- Oligogenic traits - gene pyramiding and F2 enrichment
- Polygenic traits - MARS and Genomic Selection

- **“Role for the commercial sector”**

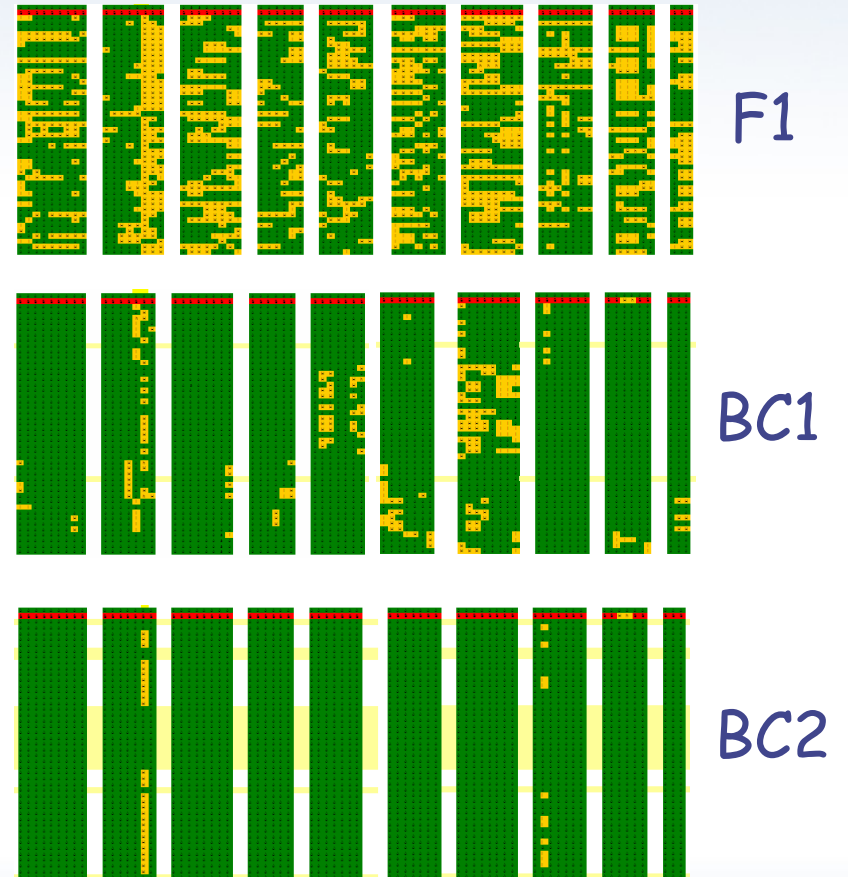
“We have no idea whether our MAB methods are optimal. Will public plant breeders take up the challenge of developing and evaluating MAB methods?” - Ted Crosbie, 2010

Marker Assisted Backcrossing

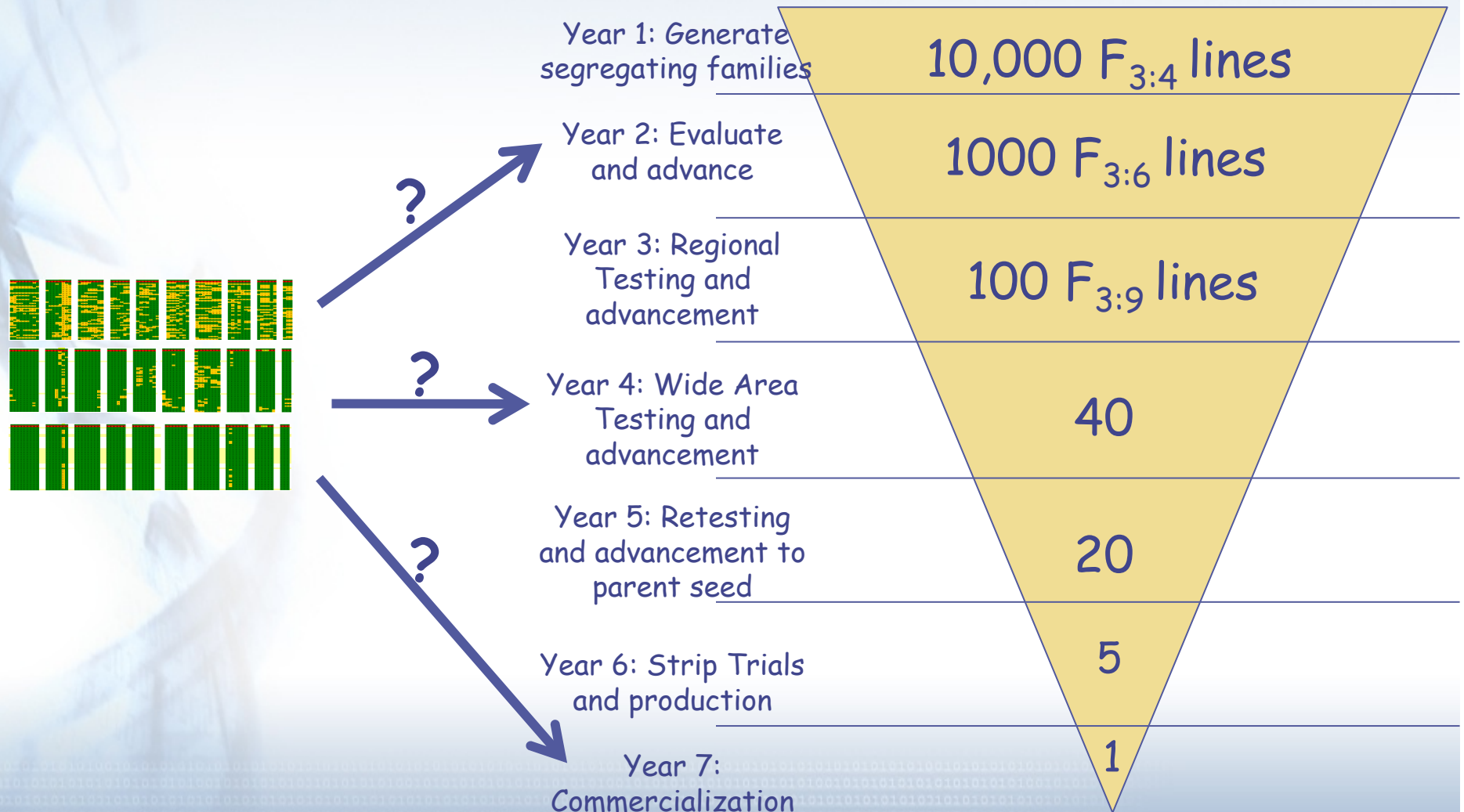
Criterion: Relative Efficiency

$$\Delta G_m / \Delta G_p = (r_m / r_p) (y_p / y_m)$$

- All Transgenics
- Helguera et al, 2003
- Benchimol et al, 2005



Is MABC optimized in the context of a breeding program?



Optimization: What does it mean?

- Is the relative efficiency as measured by a ratio of ΔG for pairs of *ad hoc* (trial and error) methods the best way to approach the question of optimizing the breeding process?
 - Should a point estimate of r , i.e., h^2 be used?
 - What about the likelihood of success?
 - What about the cost of increasing likelihood?
- Is there a better (more objective) process for including sets of criteria that need to be considered in the decision?

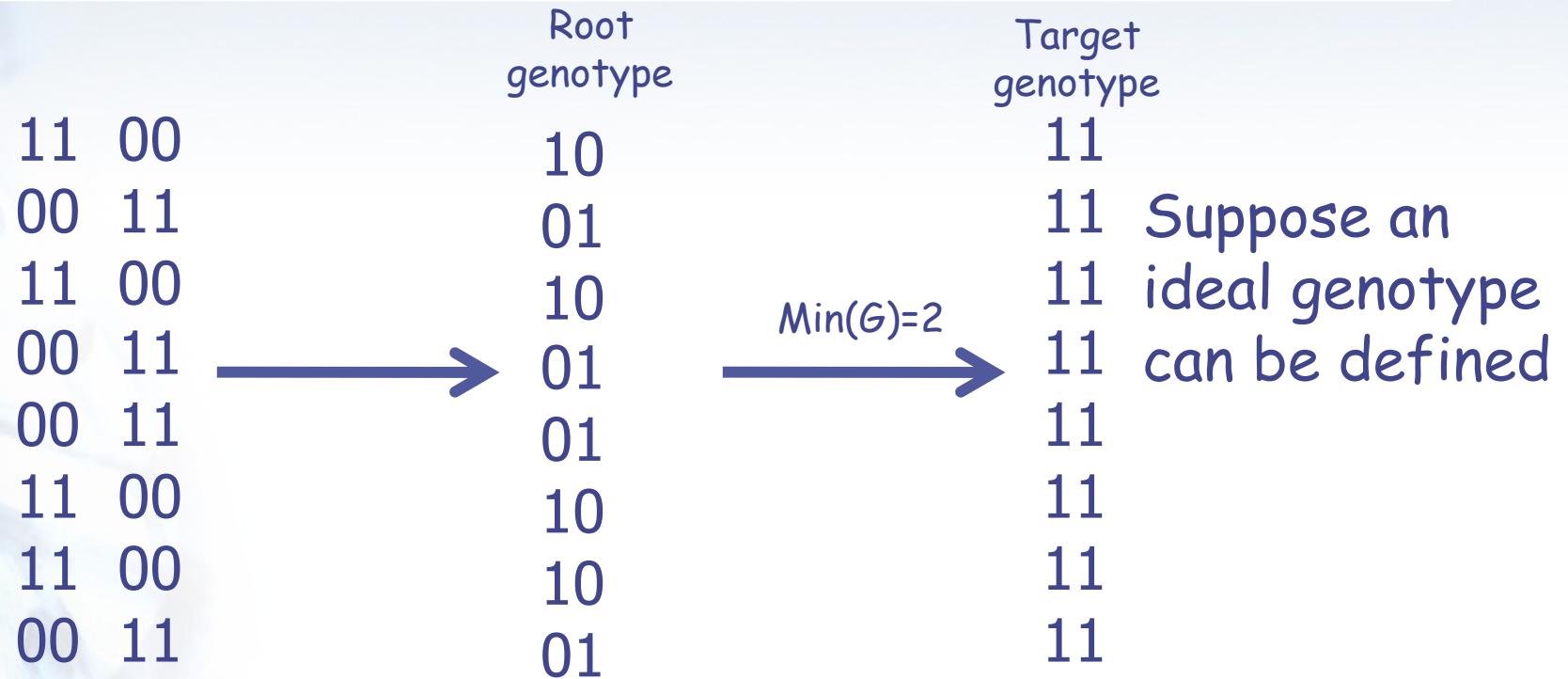
Operations Research: Definition and Purpose

- Hypothesis driven research to find a best (optimal) set of operations (activities) needed for a desired outcome.
- OR consists of the following steps:
 - Define the problem (breeding objective)
 - Model the process
 - Simulate solutions based on the model
 - Identify optimal sets of operations
 - Refine the model

Modeling: Definition and Purpose

- Natural or artificial processes are modeled for purposes of predicting outcomes.
 - In plant breeding, simulation models could be used to choose among proposed breeding methods because experimental evaluation of breeding methods is time and resource limited.

Is there an optimal breeding strategy to stack genes?



Bonnett et al, 2005
Kuchel et al, 2005*
Ye et al, 2007*

A more difficult example.

								Root genotype	Target genotype
11	00	00	00	00	00	00	00	10	11
00	11	00	00	00	00	00	00	10	11
00	00	11	00	00	00	00	00	01	11
00	00	00	11	00	00	00	00	→ 10 →	11
00	00	00	00	11	00	00	00	10	11
00	00	00	00	00	11	00	00	01	11
00	00	00	00	00	00	11	00	10	11
00	00	00	00	00	00	00	11	01	11

Consider all possible breeding pedigrees (binary trees). Find the pedigree that will minimize the number of progeny that need to be genotyped, assuming recombination between adjacent loci = 0.2

Servin et al, 2004

Comparison of sample sizes (N) and number of Generations (G) required to be 99% sure of obtaining the target genotype

	Root genotype	Target genotype	G	N _p	N _{mars}
	10	11	5	4415	----
	10	11	6	2741	----
	01	11	7	2421	7560
Min(G)=3 →	10 →	11	8	2183	3440
	10	11	9	1394	1710
	01	11	10	-----	1100
	10	11	11	-----	880
	01	11			

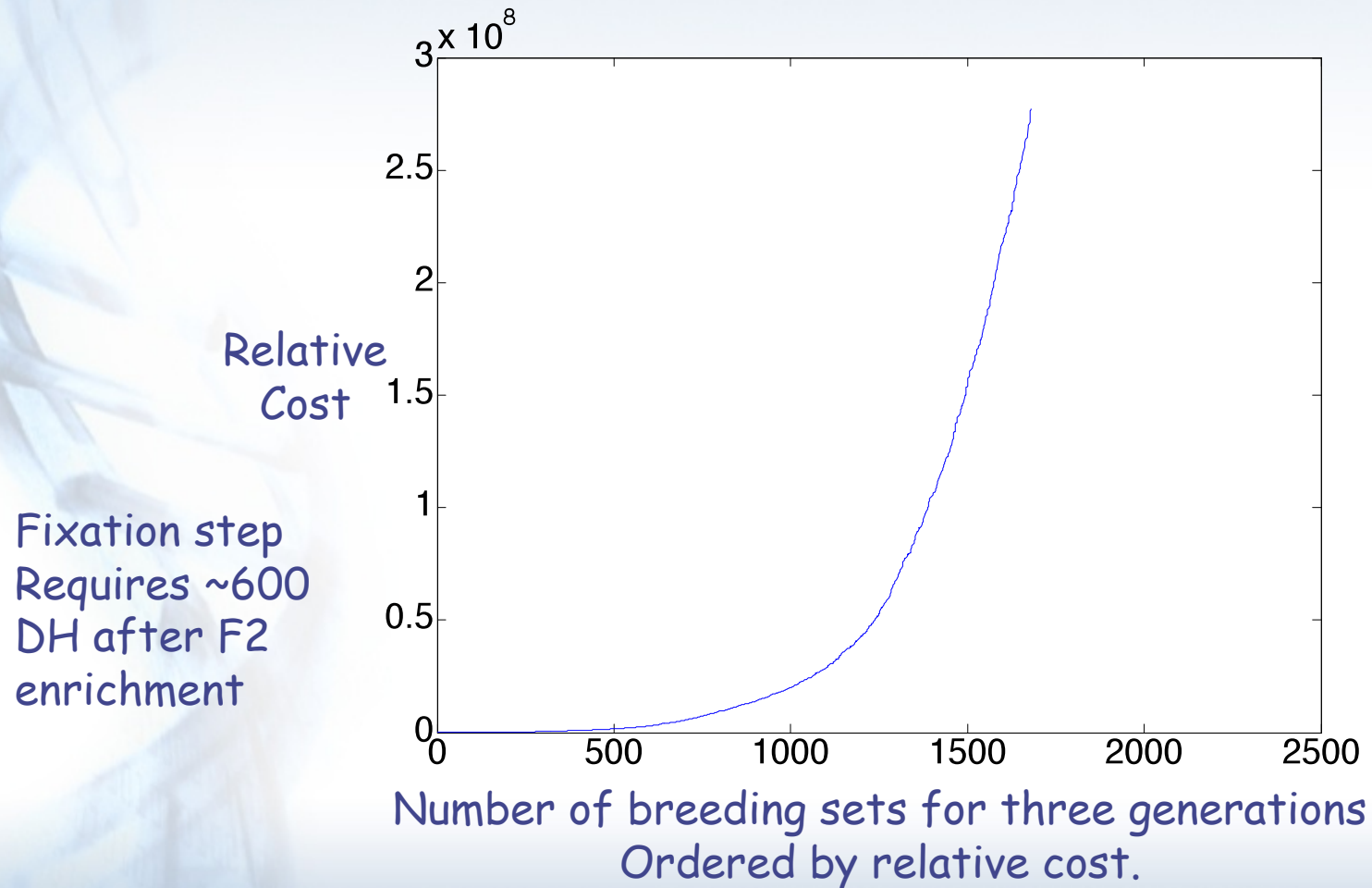
Servin et al, 2004

Consider 12 independent loci with desirable alleles in 8 parents?

1	0	1	0	0	1	0	0
0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	1	1	0	1	1
0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	1	0	1	0	0	0
0	0	1	0	0	0	0	1
0	0	0	0	1	0	1	0
1	0	0	1	0	1	0	0
0	1	0	0	0	1	0	0

One possible solution:
3x4->9 5x8->10
2x4->11 9x10->12
11x12->root genotype


“Set covering” to identify breeding crosses needed to assemble the ‘root genotype’ in three generations.



Wang and Beavis, 2008 unpublished

Stack genes and maintain genetic variability in the breeding population

1	2	3	4	5	6	7	8	9	10
0	0	0	0	1	0	0	0	0	1
C	A	A	C	B	C	C	A	A	A
1	0	1	0	0	0	0	1	0	1
B	C	B	A	C	B	A	A	B	B
1	1	0	1	0	1	0	1	0	0
C	A	B	A	C	B	A	A	C	B
0	0	1	0	1	1	0	0	0	0
B	C	C	C	A	B	C	B	C	C
0	0	0	0	1	0	0	0	1	1
C	A	A	B	A	B	C	A	B	B

Integer Programming

 Min(G)=2

T1	T2	T3	T4
1	1	1	1
A	A	B	C
1	1	1	1
B	C	A	B
1	1	1	1
B	A	C	C
1	1	1	1
C	B	A	A
1	1	1	1
B	A	A	C

Computation time: 0.03 seconds

For a 30x30 where $\frac{1}{2}$ of the loci need to be fixed and $\frac{1}{2}$ of loci need to maintain 4 alleles
 Min(G) = 8 and Computation time: 0.1 seconds

Unresolved Issue: Genomic Selection vs. Gene Stacking

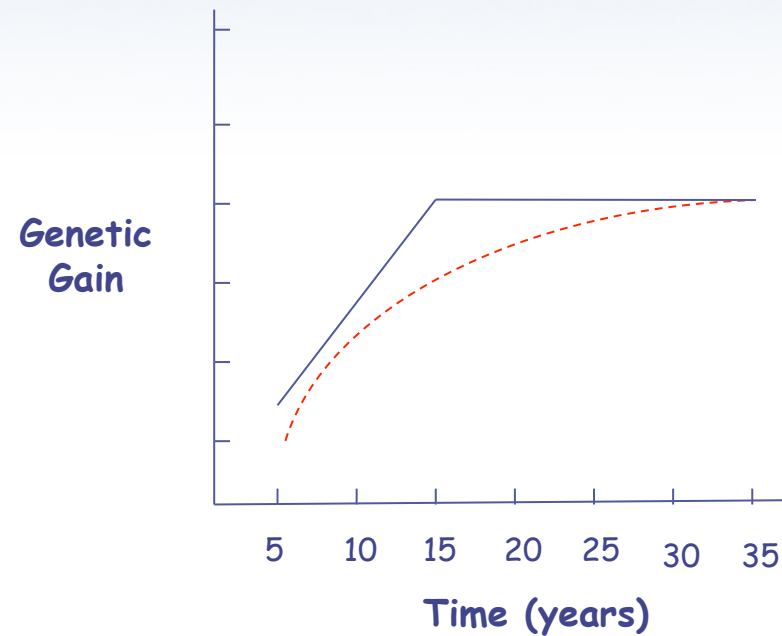
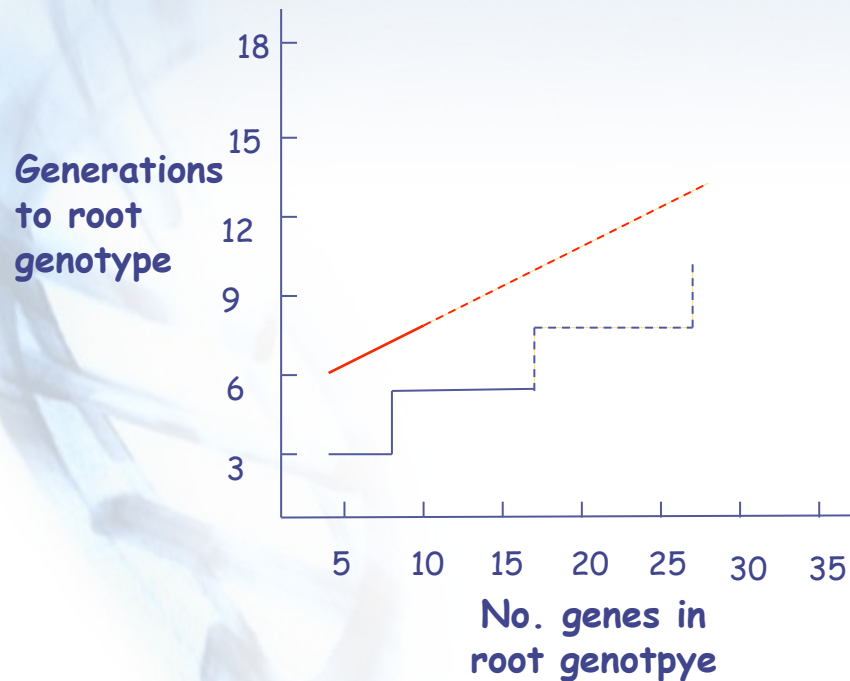
N_{qtl}	$\Delta G_{gs}/\Delta G_{mars}$	G	N_p	N_{mars}
20	1.13	5	4415	----
40	1.15	6	2741	----
100	1.18	7	2421	7560
		8	2183	3440
		9	1394	1710
		10	-----	1100
		11	-----	880

Bernardo and Yu, 2007

Servin et al, 2004

- Conventional wisdom: better to use gene stacking for a few genes and GS for many genes (Heffner et al, 2009).
- Is this a testable hypothesis? By what criteria?

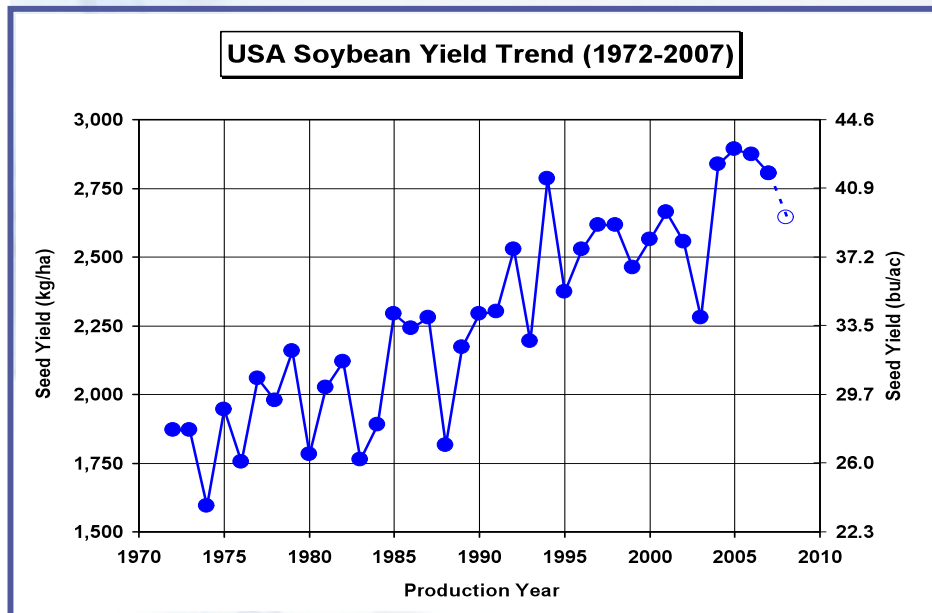
Hypothesis: Gene Stacking is more effective than **GS**



Assuming a target genotype can be defined

If the target genotype has to be determined using experimental methods, then **GS** will be more effective because experimental methods are underpowered and biased (Heffner et al, 2009)

Summary



- MAB has not been solved.
- Methods from optimization research will contribute objective criteria for optimizing MAB.

Acknowledgements

- Iowa State University
 - Kendall Lamkey
 - Thomas Lubberstedt
 - **Lizhi Wang**
- Academic Plant Breeders:
 - Rex Bernardo
 - Brian Diers
 - Dick Johnson
 - Jim Specht
- USDA-ARS Ames
 - **Jode Edwards**
 - Candy Gardner
 - Randy Shoemaker
- Pioneer/DuPont:
 - Mark Cooper
 - Antoni Rafalski
- Syngenta
 - Venkata Kishore
 - Ray Riley
- Research supported by Departments of Agronomy and IMSE and by the Plant Sciences Institute at ISU

A Deo lumen, ab amicis auxilium