Genome selection: Basics and experience in animal systems

Alison Van Eenennaam, Ph.D.

Animal Genomics and Biotechnology
Cooperative Extension Specialist
Department of Animal Science
University of California, Davis, CA U.S.A.
Ph: (530) 752-7942

alvaneenennaam@ucdavis.edu

UCDAVIS ANIMAL SCIENCE

animalscience.ucdavis.edu/animalbiotech

Phenotyping animals is so much more fun than phenotyping plants…






Relative to plants – animal breeders really have it hard

- Long generation interval
- Cannot self (at least domestic livestock can't!)
- Have limited family sizes (often one offspring/yr)
- Cannot make RILs easily in domestic livestock
- Certainly can't make double haploids (except fish)
- Expensive to phenotype
- Can't measure milk production on bulls
- Can't indiscriminately discard unwanted ones
- Most traits seem to obey infinitesimal model

But the end product is so much more satisfying!

Overview

- Introduction to animal breeding
- Genomic selection
- Practical questions for breeders
  - Dairy Industry as a genomic selection success story
  - Beef Industry as an "opportunity for improvement"
- What does the future hold

Note: I have drawn a lot of my material from published literature and would highly recommend you read the references listed at the bottom of the slide to more fully understand this brief overview of complex concepts.

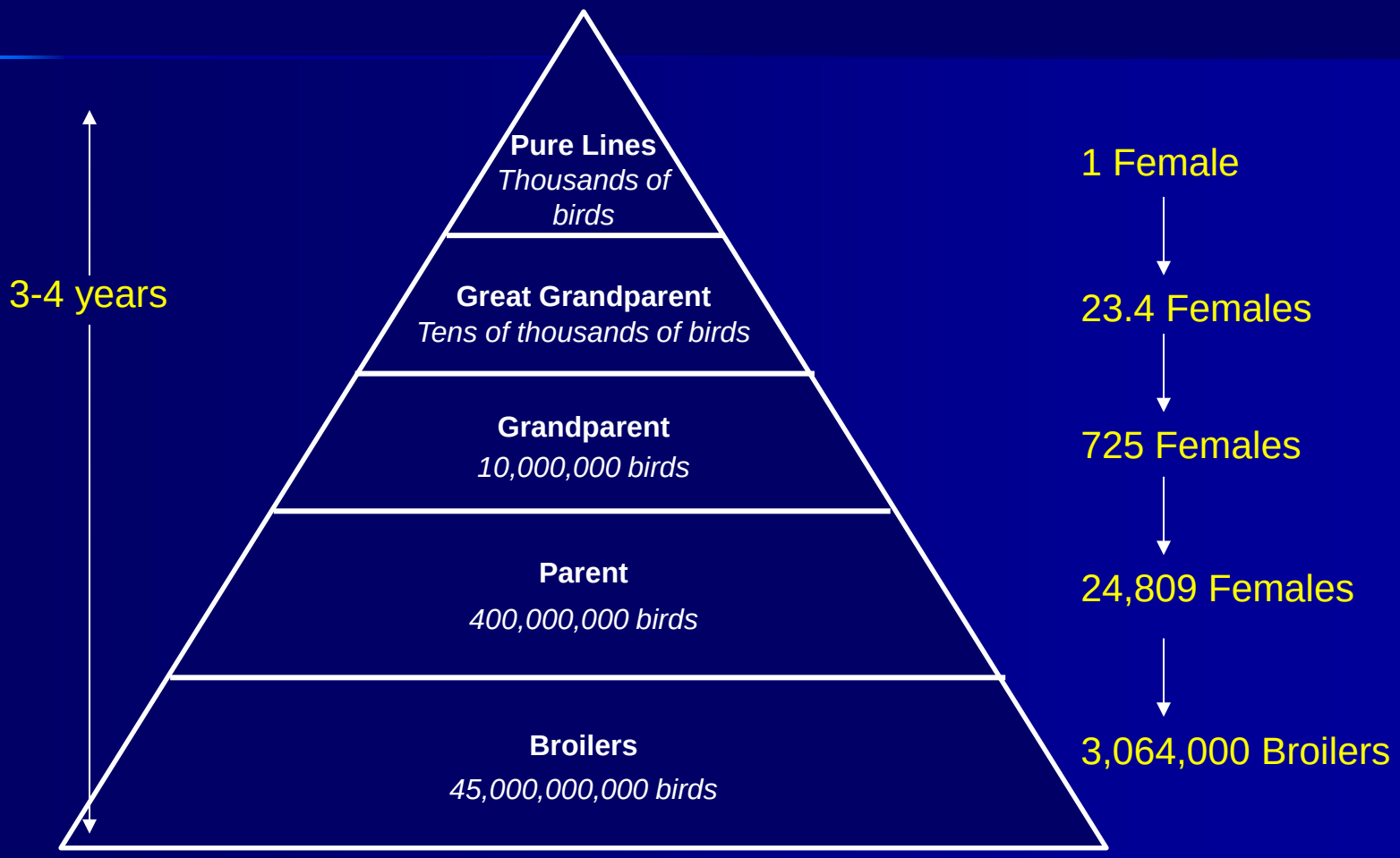
Any technology that can modify components of the breeders equation can accelerate genetic gain

$$\Delta G = \frac{\textit{intensity of selection } X \textit{ accuracy of selection } X \textit{ stand. dev. genetic variation}}{\textit{generation interval}}$$

Animal breeders have been genetically modifying animals for faster growth and improved feed conversion for many years

Havenstein, G., Ferket, P. and Qureshi, M. (2003). Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poultry Science* 82, 1500-1508.

One female broiler hen generates over 3 million direct genetic descendants



Van Eenennaam, A.L., K. A. Weigel, A. E. Young, M. A. Cleveland, and J. C. M. Dekkers 2014. Applied Animal Genomics: Results from the Field. Annual Review of Animal Biosciences. 2:105-139.

MILK

Holstein - Dairy

Angus - Beef

MEAT

EPD = expected progeny difference

ABS Global : Beef

Pasture to Plate™ Genetics

Main | Beef | Dairy | Technical Services | Company Info | News

DREAM ON 29SM0373

CNS DREAM ON L186 ASA 2144976

View Support Photos

NICHOLS LEGACY G151 / NICHOLS BLK DESTINY D12
NICHOLS DEBRA D81
CNS SHEEZA DREAM K107W / SRS FRANCHISE F601
NICHOLS JOLIETTE 107W

2007 GENETIC TYPE SUMMARY

	STA	#HEAD	-2 -1 0 1 2
Stature	-0.4	PE	
Capacity	0.2	PE	
Body Length	0.4	PE	
Muscling	0.6	PE	
Rear Legs	-0.8	PE	
Feet & Pasterns	0.1	PE	
Femininity	0	PE	
Udder Attachment		PE	
Udder Depth		PE	
Teat Size		PE	

Comments: Pedigree Estimate

Profitable and Marketable

- Heterozygous black, DNA tested homozygous polled
- A leader for API (top 2%) combining outstanding calving ease, maternal calving ease, sensible growth and strong carcass traits
- The overall look and kind of his progeny make them extremely popular and easy to sell
- The best return on investment for your semen dollar

AMERICAN SIMMENTAL SIRE SUMMARY FALL 2007

TRAIT	CED	BW	WW	YW	MCE	MILK	MWW	STAY
EPD	+15.7	-0.9	+35.3	+58.6	+13.3	+6.9	+24.5	+16.3
ACC	.88	.95	.92	.90	.77	.80	.81	.15

TRAIT	CW	YG	MARB	BF	REA	SHR
EPD	-10.5	-.01	+.30	+.03	+.11	-.01
ACC	.66	.66	.66	.71	.64	.08

	API	TI
INDEXES	$120	$74

Calving Ease: ★ ★ ★

Carcass Merit:

Individual Performance		
BW	78	
205	886	N/A
365	1,435	N/A
SC	40	12 Mo.
YFS/FS	6	
Weight	2,300	Mat.
Height	58	Mat.
Born: 11/21/01		

Owned by:

- Janssen Farms LLC, , IA
- Chris Schick, , IL
- Parke Livestock Enterprises, , KY

service science success

Before 2010: Marker-assisted selection using 1-100 SNPs

- Meat Tenderness
- Quality Grade (Marbling) ⟶
- Beef Cattle Feed Efficiency
- Meat Yield
- Disease Resistance
- Dairy Form
- Milk and Milk Component Yield

What was wrong with this MAS model?

- Problem with traditional Marker-assisted selection (MAS) approaches is the effect of individual quantitative trait loci (QTL) on complex traits, such as yield, are likely to be small.
- A large number of QTL are necessary to explain the genetic variation in these traits
- The usefulness of information from a sparse marker map in outbreeding species is also limited because the linkage phase between the marker and the QTL must be established for every family

Genomic selection

Alternative is to trace all segments of the genome with markers

- Divide genome into chromosome segments based on marker intervals
- Capture all QTL = all genetic variance
- Marker density must be sufficiently high to ensure that all QTL are in linkage disequilibrium (LD) with a SNP marker

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics, Vol. 157, 1819-1829

Wrong Expert Predictions

I think there's a world market for about five computers.

Thomas J. Watson, chairman of the board of IBM. 1943

There is no reason anyone would want a computer in their home.

Ken Olson, president of Digital Equipment Corp. 1977

The cost for a genome scan (defined as 18 chromosomes* 7 markers (i.e. 126 markers!) * $4/marker) = $504

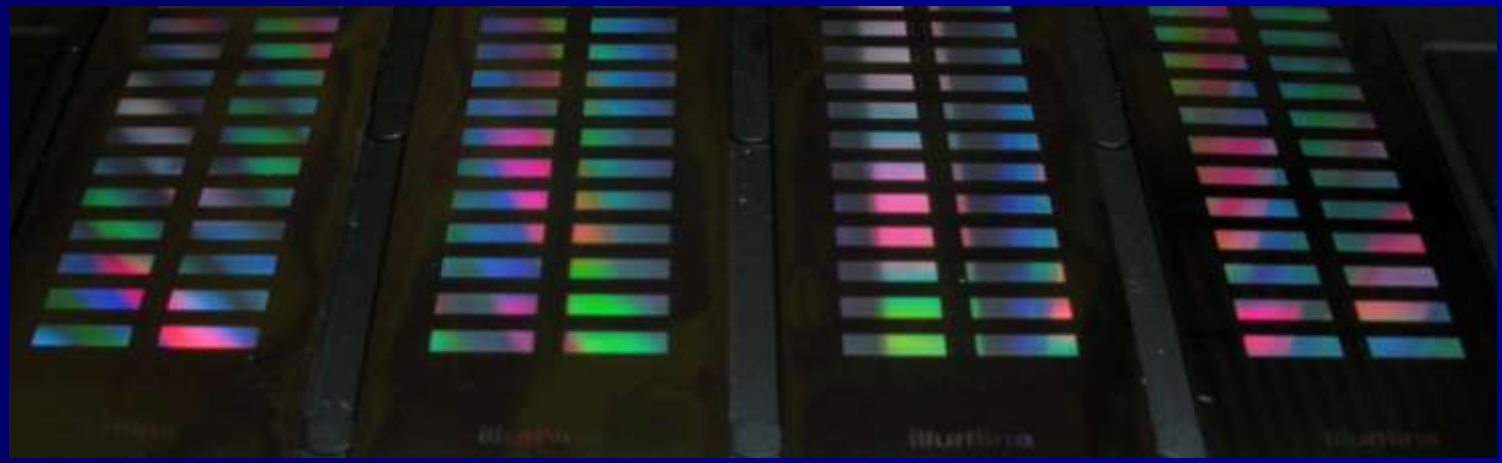
Ben Hayes and Mike Goddard, 2003. Evaluation of marker assisted selection in pig enterprises. Livestock Production Science 81:197-211.

Potential benefits of genomics are greatest for economically-important traits that:

- Are difficult or expensive to measure
- Cannot be measured until late in life or after the animal is dead
- Are not currently selected for because they are not routinely measured and so there are no selection criteria available
- Have low heritability

High-throughput genotyping technology enabled the development of high density "SNP chips"

The sequencing of the bovine genome allowed for the development of a 50,000 SNP chip, then the 800,000 SNP chip; and now whole genome sequence (3 billion)!



We can use these SNP CHIPS for "genomic" selection?

TRAINING POPULATION

1,000s animals
- Phenotypes
- Genotypes

Training = estimate the value of every chromosome fragment contributing variation in a population with phenotypic observations

Prediction = the results of training can then be used to develop prediction equations to predict the merit of new animals (e.g. young bulls)

Practical questions for breeders

- How many phenotypic records are required in the initial experiment (reference population) used to estimate the effect of chromosome segments?

Accuracy of the prediction equation proportional to:

$$\frac{Th^2}{ML}$$

T: total number of records in the training population

h^2: heritability of the trait

L : length of chromosomes (in Morgans)

M: ~ $2N_e$ (effective population size)

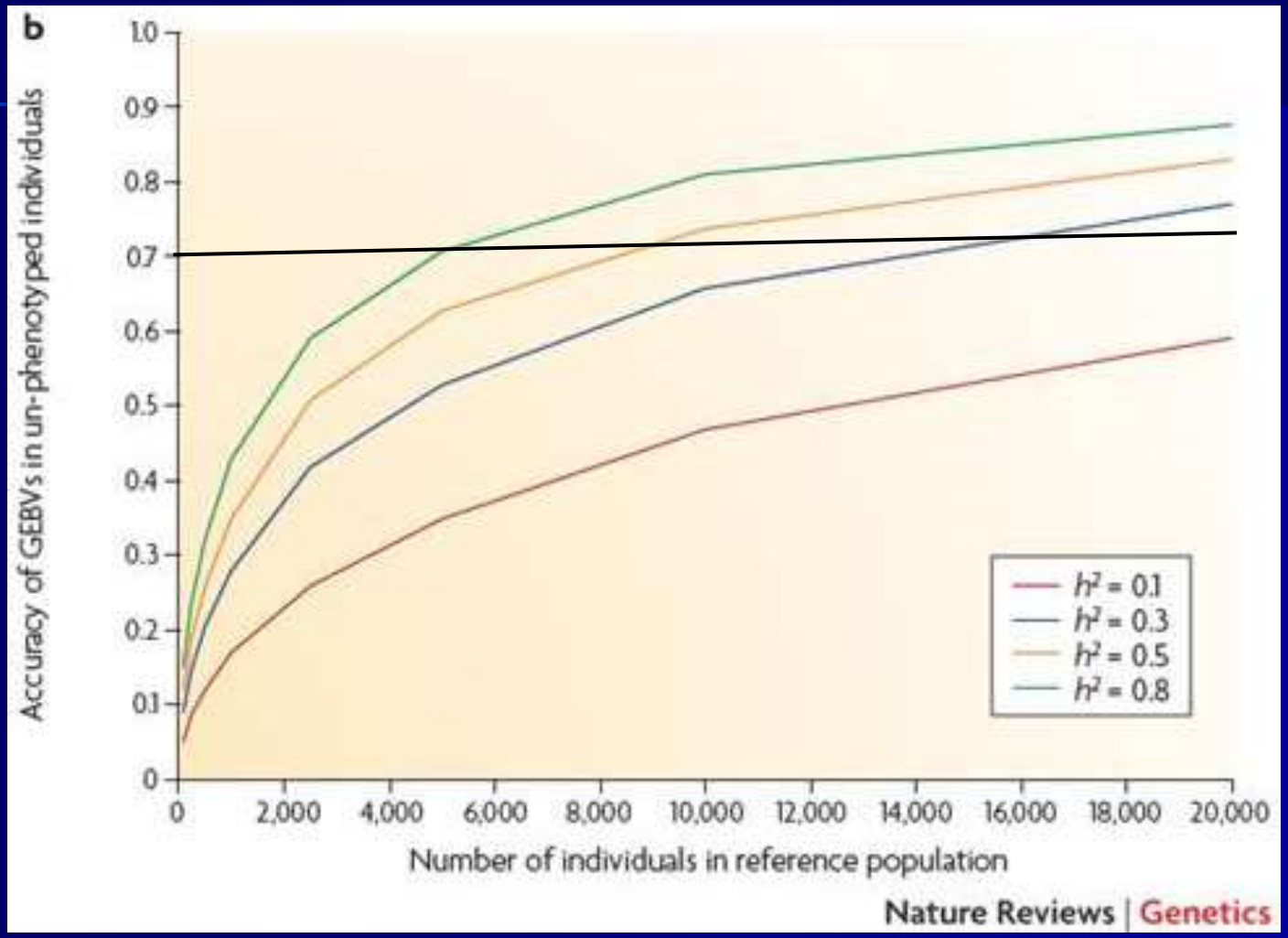
Goddard, M. E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257.

Effective population size estimates for cattle

Breed	N_e	Breed	N_e
Angus	**136**	Brown Swiss	**61**
Charolais	**110**	Guernsey	**76**
Hereford	**97**	Holstein	**99**
Limousin	**174**	Jersey	**73**
Red Angus	**85**	Norwegian Red	**106**
Brahman	**115**	Gir	**133**
Nelore	**86**		
Beef Master	**106**	Merino (sheep)	**~ Big (> 100)**
Santa Gertrudis	**107**		**Ben Hayes (pers. comm.)**

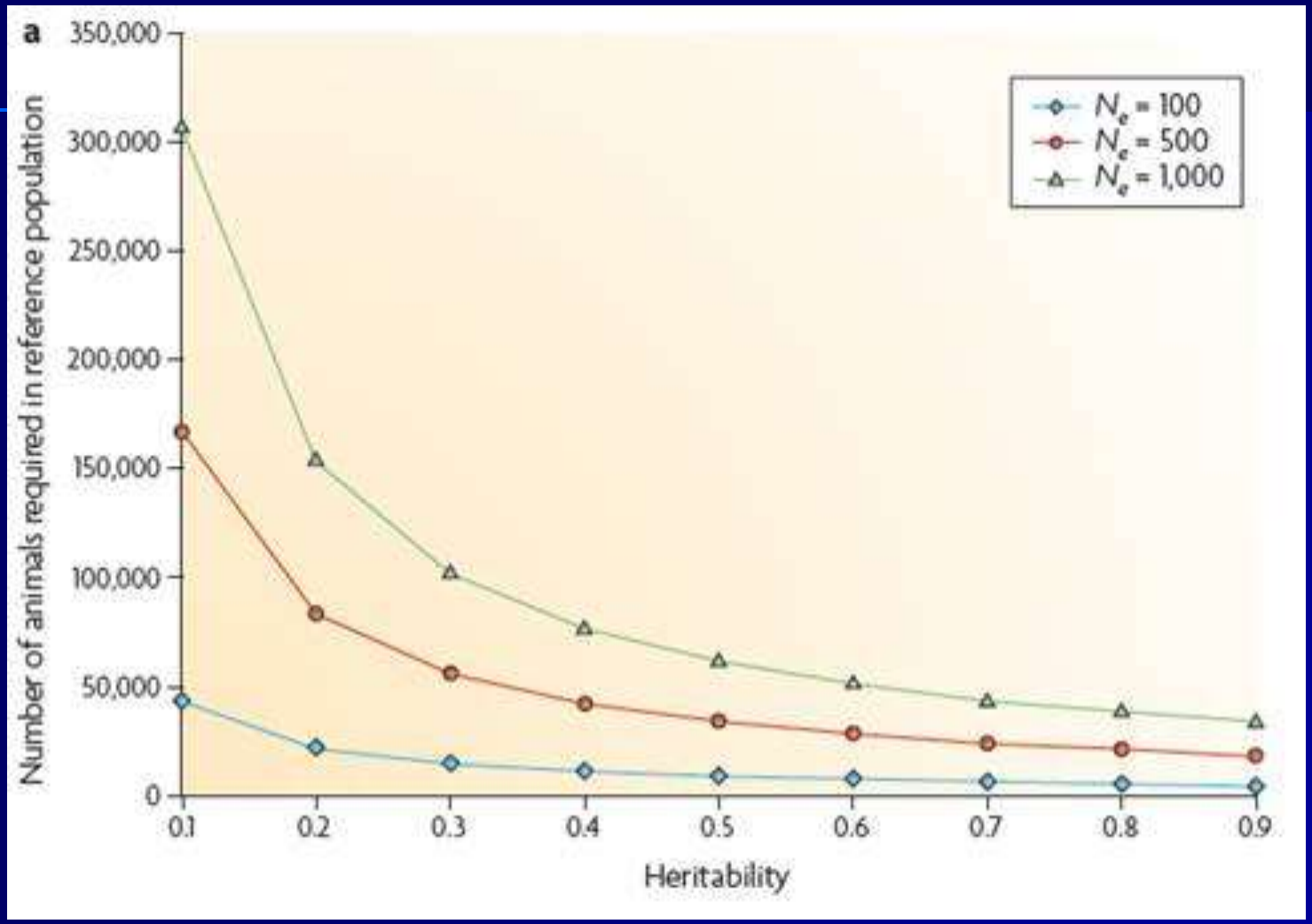
Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. 2009 The Bovine HapMap Consortium. Science 3245: 528-532. Supporting Online Material. Table S1. http://www.sciencemag.org/content/suppl/2009/04/22/324.5926.528.DC1

Effect of number of animals on accuracy of prediction equation (for a N_e of 100)



Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391.

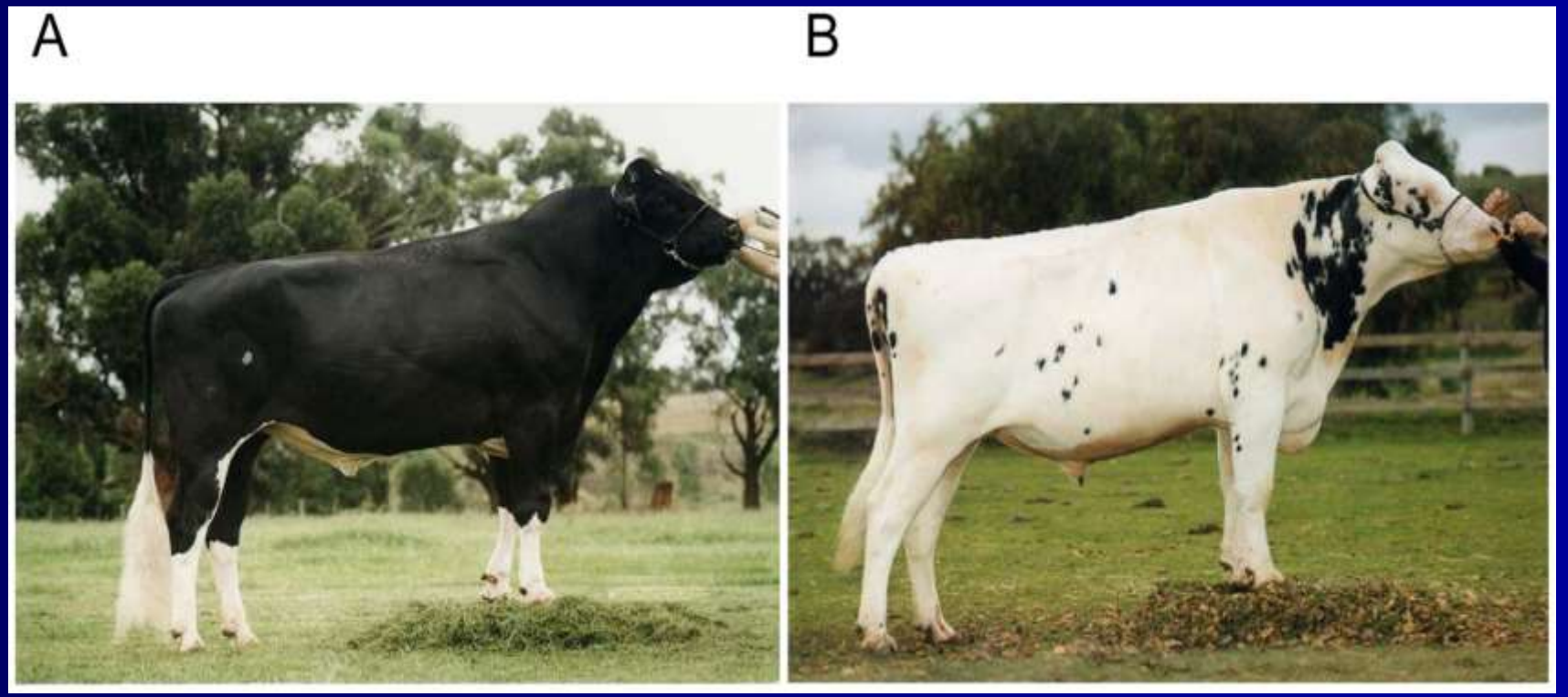
Effect of population size and heritability on the number of animals required in the training population (for an accuracy of 0.7)



Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391.

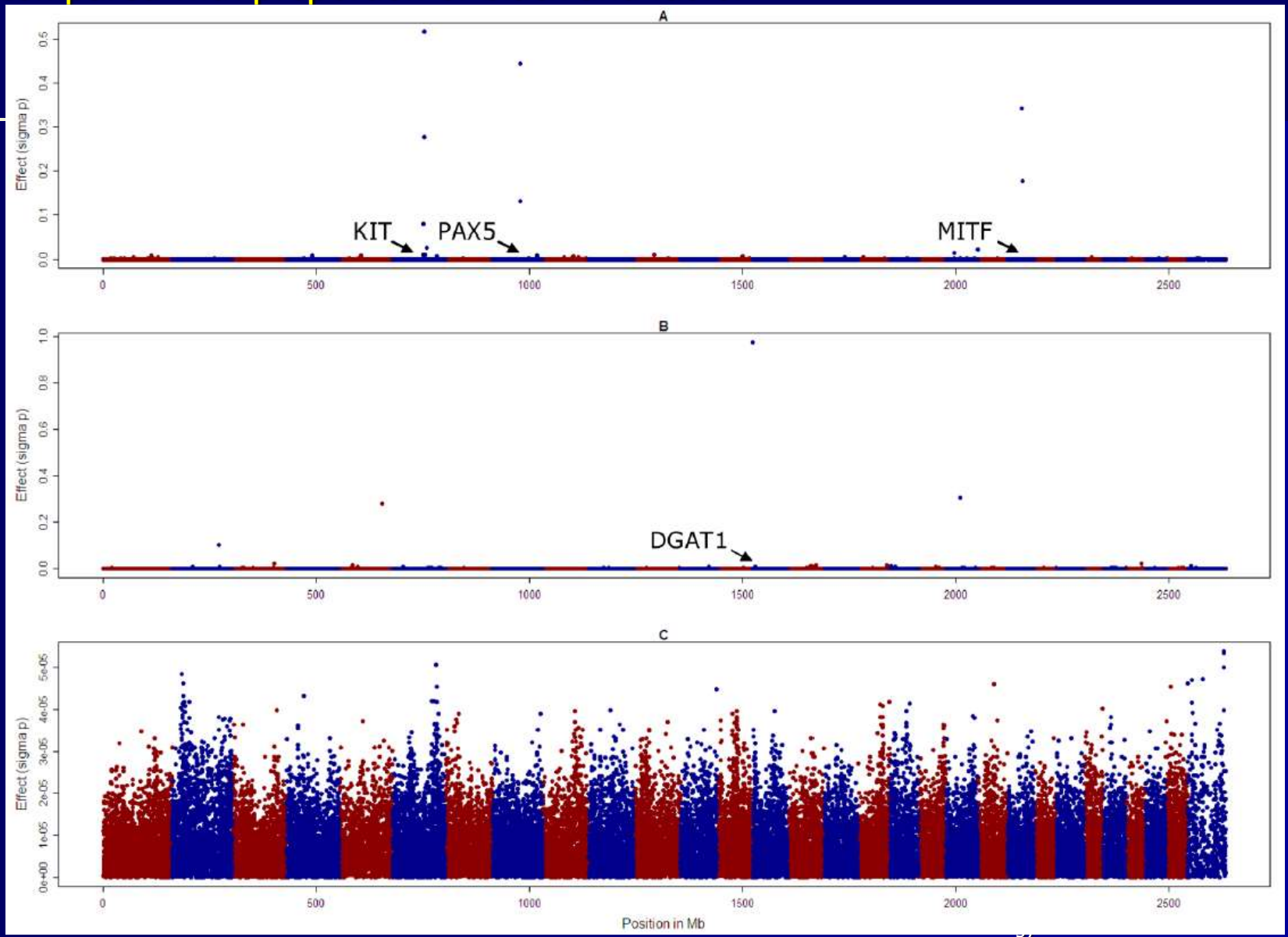
There is also an effect of trait architecture

The accuracy of predicting genetic values is higher for traits with a proportion of large effects (e.g. proportion black and fat percentage) than for a trait with no loci of large effect (e.g. overall type), provided the method of analysis takes advantage of the distribution of loci effects.



Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. Plos Genet 6

Genome-wide SNP effects when all SNPs are fitted simultaneously for three traits in Holstein Friesian cattle. Proportion of black (A), fat% (B), and overall type (C). Note the different scale of the y axis for overall type compared with proportion of black and fat%.



Maybe R.A. Fisher was onto something?

If a nearly infinitesimal model is correct as seems to be the case for most quantitative traits; then large sample sizes will be needed to achieve high accuracy

Number of effects >>> number of records

- When we come to estimate the allelic effects of all of these markers, we are faced with estimating many effects in a data set of limited size, and there are not enough degrees of freedom to fit all marker effects simultaneously
- Need methods that can deal with that

Statistical methods for genomic selection

A number of approaches have been proposed for estimating the single marker or haplotype effects across chromosome segment effects for genomic selection. The key differences between these approaches is the assumption they make about the variances of haplotype or single marker effects across chromosome segments, and whether some proportion of markers have a zero effect (Bayesian approaches).

Practical questions for breeders

- How many phenotypic records are required in the initial experiment (reference population) used to estimate the effect of chromosome segments?
- **How often do we need to re-estimate the chromosome segment effects?**

How often is it necessary to re-estimate the marker effects?

TABLE 5

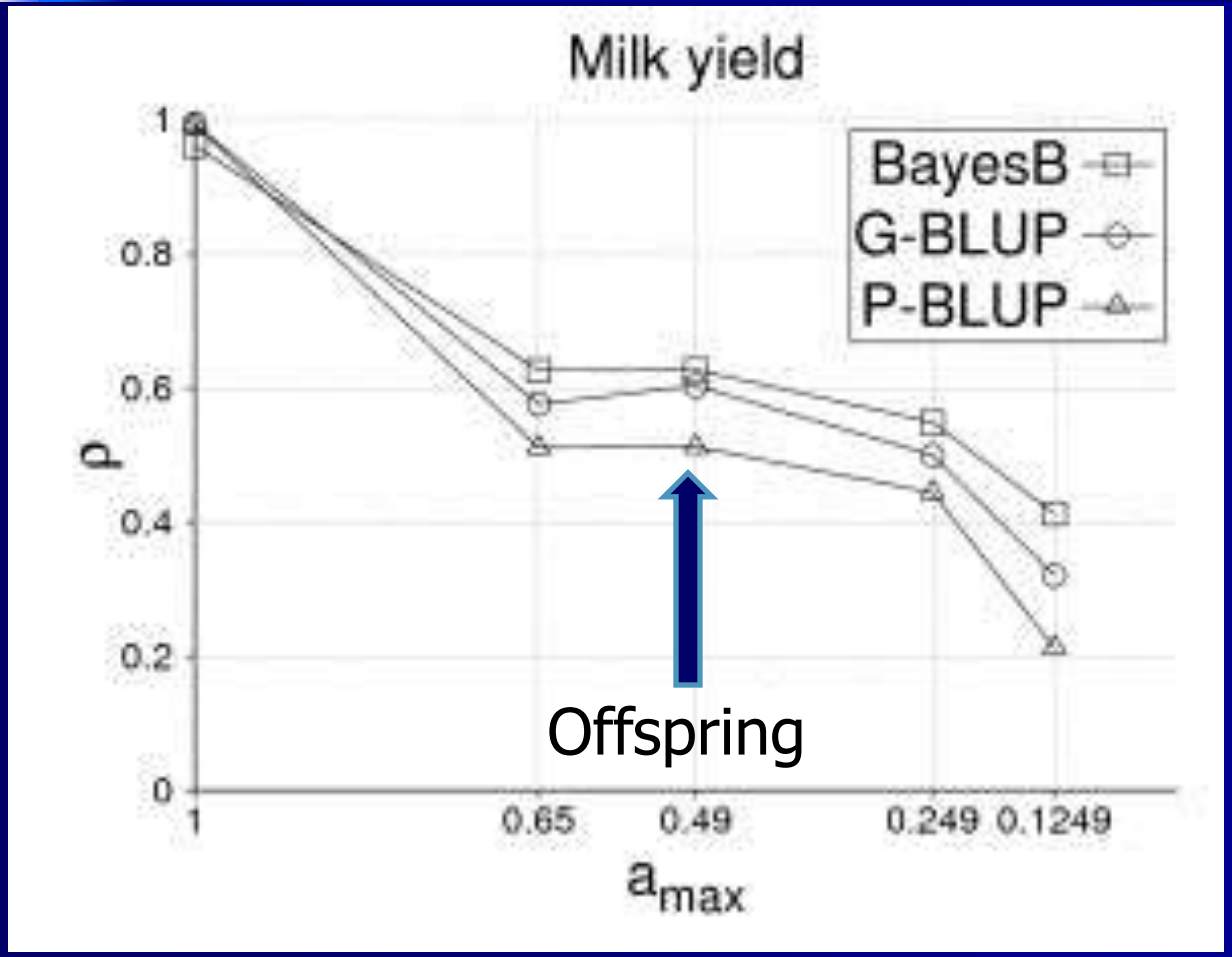
The correlation between estimated and true breeding values in generations 1003–1008, where the estimated breeding values are obtained from the BayesB marker estimates in generations 1001 and 1002

Generation	$r_{TBV;EBV}$
1003	0.848
1004	0.804
1005	0.768
1006	0.758
1007	0.734
1008	0.718

The generations 1004–1008 are obtained in the same way as 1003 from their parental generations.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics, Vol. 157, 1819-1829

Markers can predict family relationships between animals, independently of Linkage Disequilibrium (LD) between the markers and QTL (i.e. due to family relationships or linkage)



Additive-genetic relationships between training and validation animals was found to be a good indicator of accuracy

Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution 42: Article No.: 5

Linkage (i.e. associations exist within families) versus LD

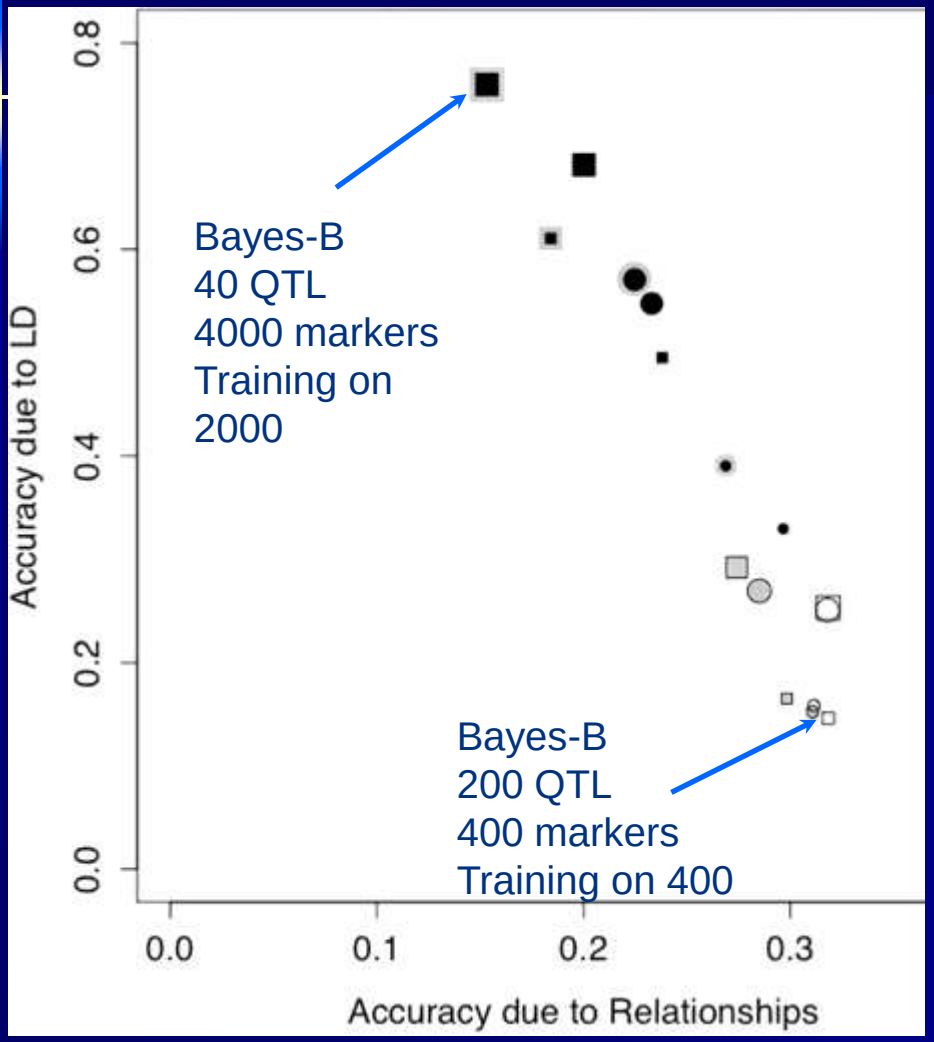
Exploring these two sources of GS accuracy, Habier *et al.* showed that ridge regression (BLUP) is more effective at capturing genetic relationships because it fits more markers into the prediction model.

In contrast, BayesB is more effective at capturing LD between markers and QTL. Because these marker–QTL linkages are tight, recombination does not cause them to decay rapidly, and accuracies from BayesB persist longer than those from ridge regression (BLUP).

Jannink J *et al*. Briefings in Functional Genomics 2010;9:166-177

Decomposition of GS prediction accuracy using the method of Habier et al.*

Jannink J et al. Briefings in Functional Genomics 2010;9:166-177



On a genome comprised of seven chromosomes of 1.5 M each, individuals were generated using a coalescent assuming an effective population size of 100.

- Round and square symbols, ridge regression and Bayes-B, respectively.
- Symbols with gray (inside or around) and without, 40 QTL and 200 QTL, respectively.
- Black and non-black symbols, 4000 and 400 markers, respectively.
- Small and large symbols, training population size of 400 and 2000, respectively

* Habier D, Fernando RL, Dekkers JCM. *The impact of genetic relationship information on genome-assisted breeding values. Genetics 2007;177:2389-97.*

In general accuracy is higher when:

- A large number of animals and high-quality phenotypic records available for training
- Trait is highly heritable
- Small effective population size so small number of chromosome segments to track
- There are genetic relationships (linkage) between training and selection candidates
- Small number of QTL affecting the trait so there is a marker associated with every QTL
- Retrain the prediction equation every generation

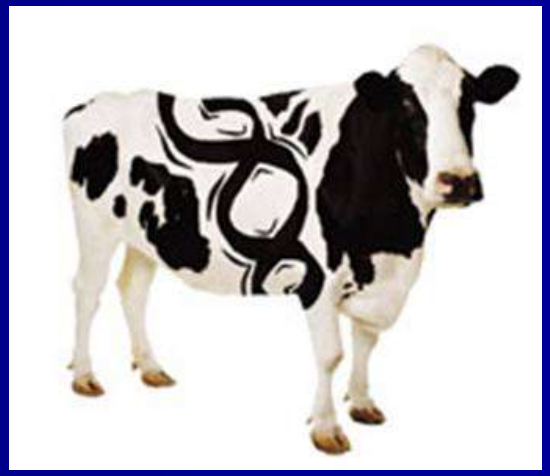
Dairy industry has successfully implemented genomic selection

Validation: Purpose is to estimate the correlation between the prediction and the true genetic merit

r_0

Validation:
New Progeny Tested Bulls

Training 1:
Old Progeny Tested Bulls



Training 2:
Old & New Progeny Tested Bulls

r_1

Application:
New Sire Candidates

Slide courtesy of Marc Thallman, US MARC

Dairy industry suited to WGS

- High use of AI
- Clear selection goal
- One breed used extensively
- Large number of high accuracy A.I. sires for training
- Extensive, uniform collection of data on traits
- Central evaluation (AIPL) receiving genotypes
- Obvious way to increase rate of genetic gain
- AI companies funding the genotyping because they get a clear cost savings in terms of young sire program

Breeding value prediction in Dairy Sires

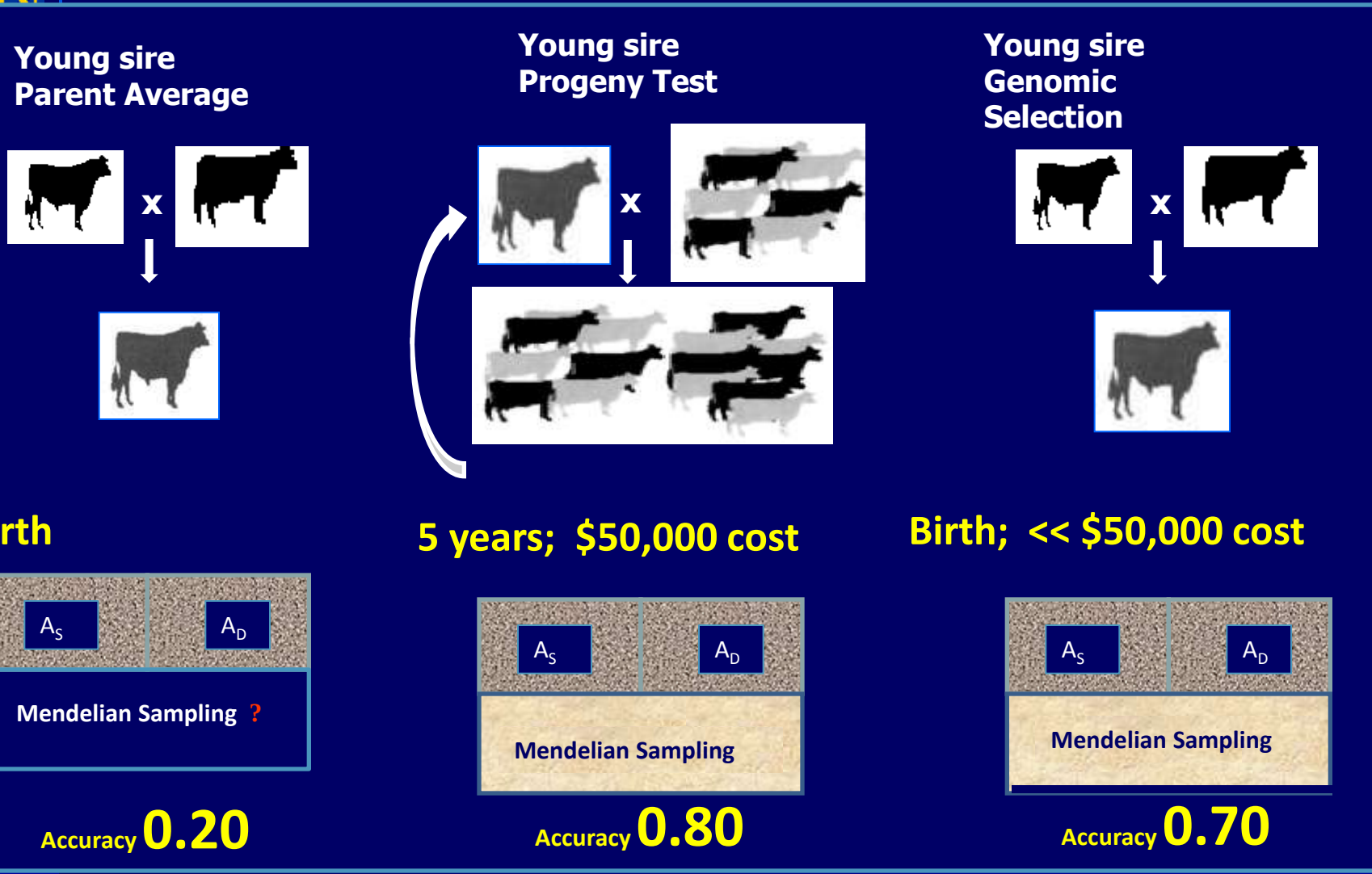


Figure by Gonzalo Rincon

Genomic selection can double rate of genetic gain

Rate of genetic gain ΔG

$\Delta G = (i_m r_m + i_f r_f)/ (L_m + L_f)$ genetic standard deviation/year

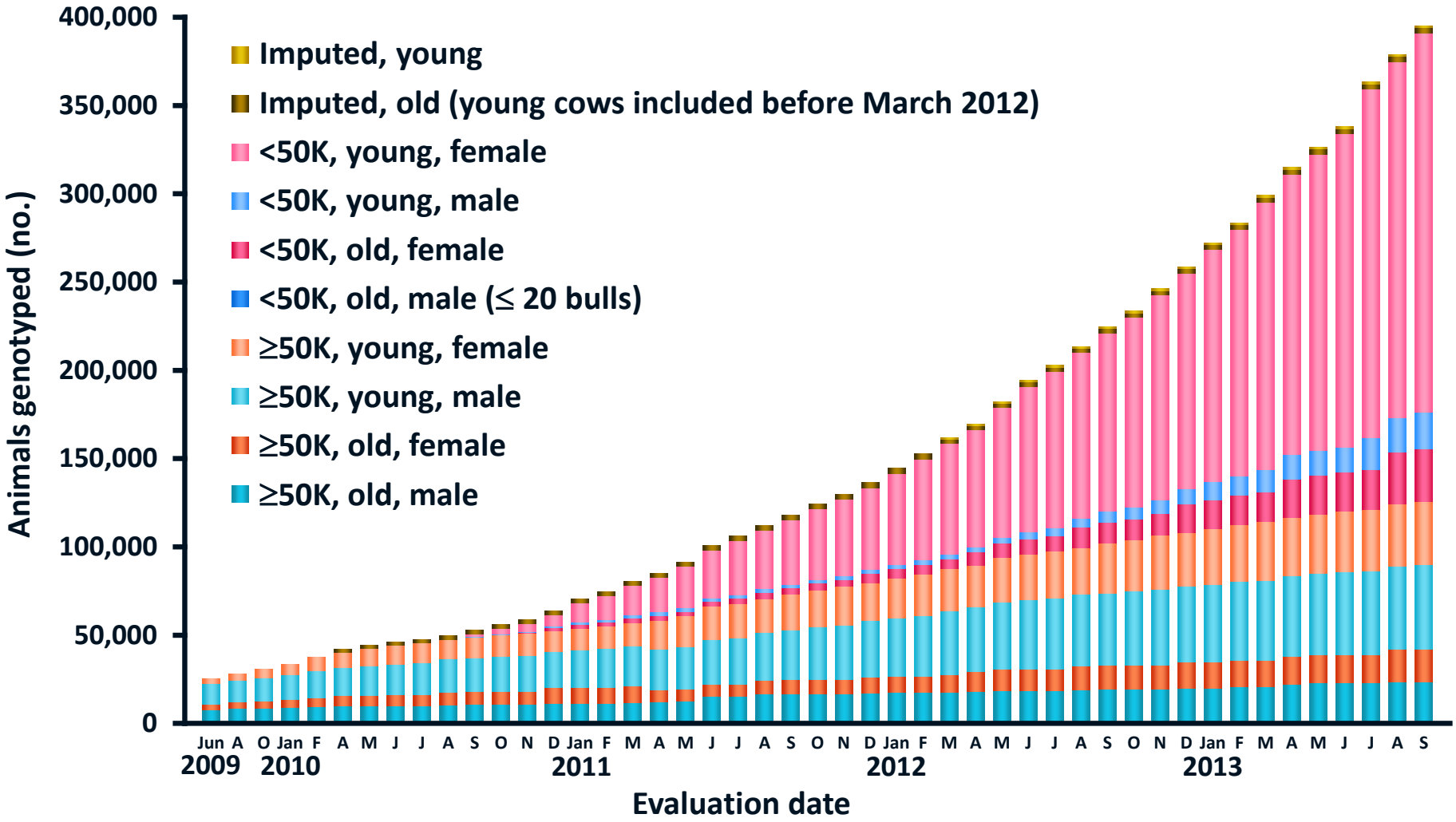
$= (2*0.8 + 0)/ (6+2)$ = 0.2 s.d./year (progeny test)

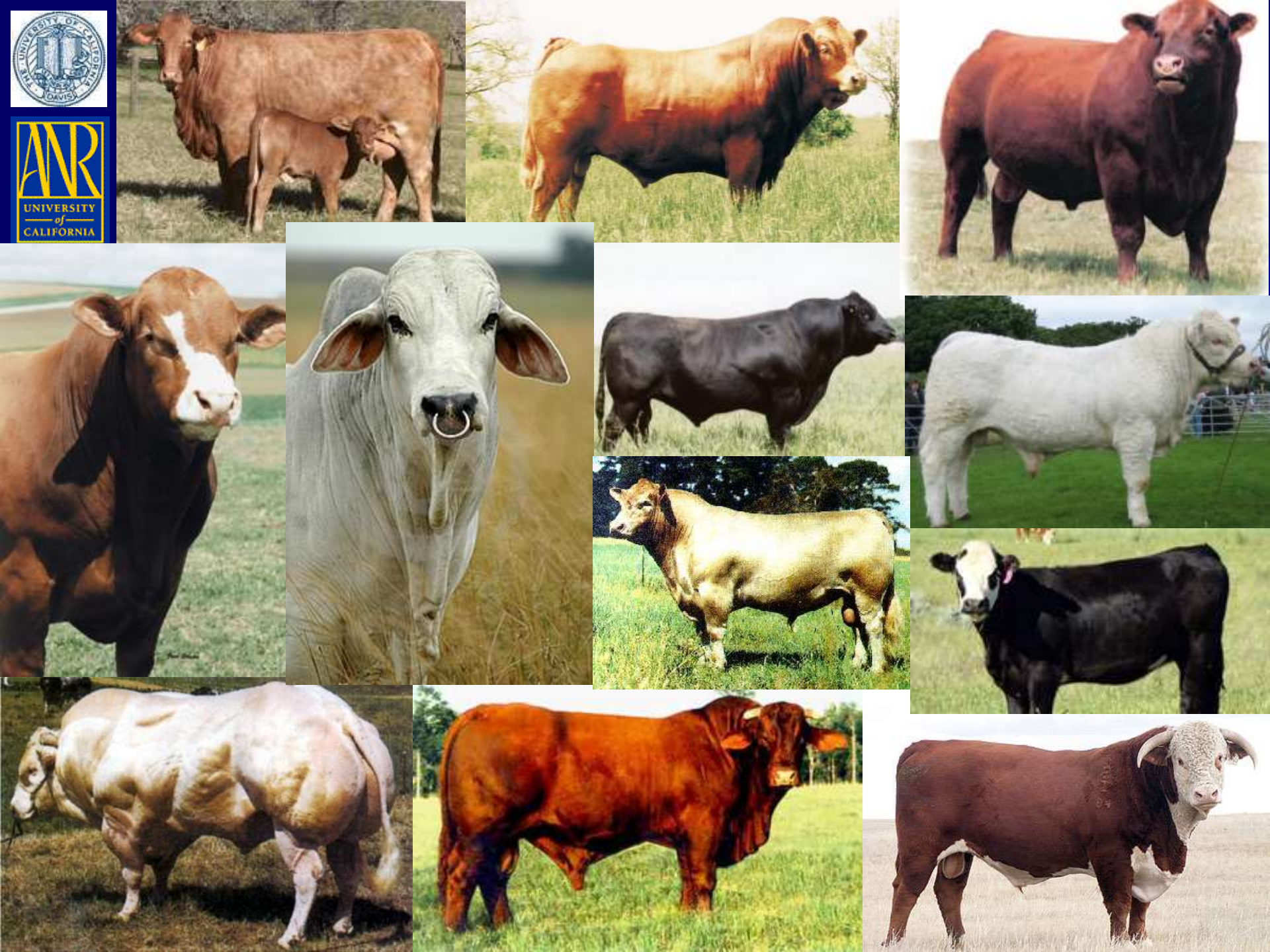
$= (2*0.6 + 0.8*0.6)/ (2+2)$ = 0.42 (genomic selection)

i = intensity of selection
r = accuracy of selection
L = generation interval

Modified from Goddard. (2009) BIF Meeting

> 400,000 Genotypes run in US dairy cattle




UNIVERSITY
of
CALIFORNIA

The Beef Cattle Industry

- Little use of AI
- Relatively few high accuracy sires for training
- Multiple competing selection goals – cow/calf, feedlot, processor – little data/value sharing between sectors
- Few/no records on many economically-relevant traits
- Many breeds, some small with limited resources
- Crossbreeding is important
- No one wants to pay as value is not recovered by breeder

A perfect storm is a confluence of events that drastically aggravates a situation

Practical questions for breeders

- How many phenotypic records are required in the initial experiment (reference population) used to estimate the effect of chromosome segments?
- How often do we need to re-estimate the chromosome segment effects?
- **Does it work across breeds/strain/cultivars?**

NO

at least not with 50K in cattle

Angus predictions (r) are not very accurate in Red Angus (Data provided by Dorian Garrick)

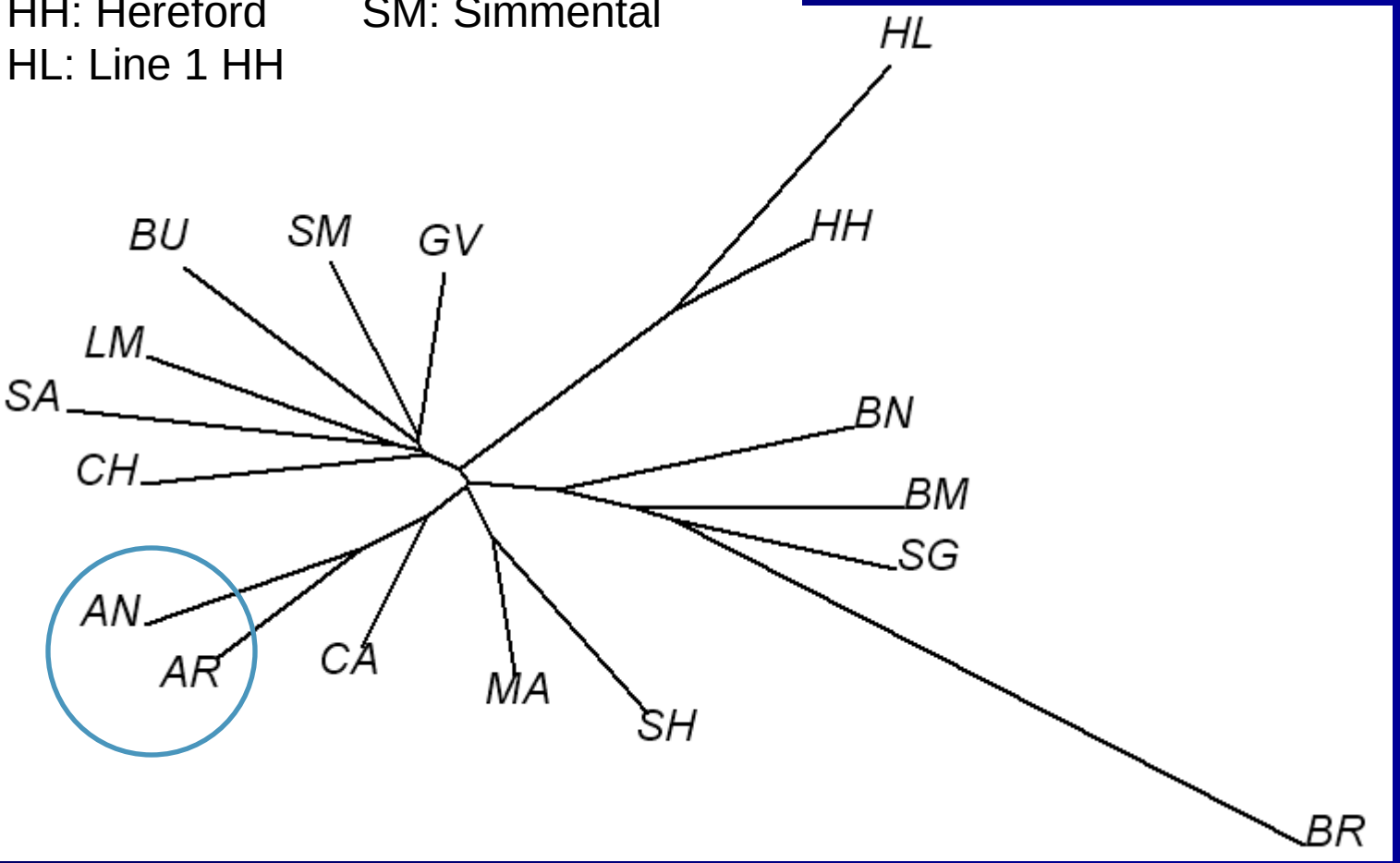
Trait	Trained in Black Angus/Validated in Black Angus	Trained in Black Angus/Validated in Red Angus
BirthWt	0.64	0.27
WeanWt	0.67	0.28
YearlingWt	0.75	0.23
Fat	0.70	0.21
Rib Eye Area	0.75	0.29
Marbling	0.80	0.21
CalvEase (D)	0.69	0.14
CalvEase (M)	0.73	0.18

Angus = ASREML 5-fold validation Red Angus = correlation Training on de-regressed EPDs Saatchi et al (GSE)

AN: Angus
BM: Beefmaster
BN: Brangus
BR: Brahman
BU: Braunvieh
CA: Chiangus
CH: Charolais
HH: Hereford
HL: Line 1 HH
GV: Gelbvieh
LM: Limousin
MA: Maine Anjou
RA: Red Angus
SA: Salers
SG: Santa Gertrudis
SH: Shorthorn
SM: Simmental

Approximate genetic distance between breeds using data from the 2,000 Bull Project.

Larry Keuhn, USDA MARC
http://www.nbcec.org/topics/BeefBreeds.pdf



Realized accuracies (r) resulting from genomic selection prediction equations trained in US beef cattle breeds

Trait	Red Angus (6,412)[b]	Angus (3,500)	Hereford (2,980)	Simmental (2,800)	Limousin (2,400)	Gelbvieh (1,181)
Birth weight	0.75	0.64	0.68	0.65	0.58	0.41
Wean weight	0.67	0.67	0.52	0.52	0.58	0.34
Yearling weight	0.69	0.75	0.60	0.45	0.76	—
Milk	0.51	0.51	0.37	0.34	0.46	0.34
Fat thickness	0.90	0.70	0.48	0.29	—	—
Rib eye area	0.75	0.75	0.49	0.59	0.63	0.48
Marbling	0.85	0.80	0.43	0.63	0.65	0.56
Calving ease direct	0.60	0.69	0.68	0.45	0.52	0.48
Calving ease (maternal)	0.32	0.73	0.51	0.32	0.51	—
Scrotal circumference	—	0.71	0.43	—	0.45	0.50

[a]Data taken from References 29, 30, 131; D. Garrick, unpublished data (personal communication).
[b]Numbers indicate training population. The Red Angus training data set includes some Black Angus cattle that have expected progeny difference in the Red Angus Association.

Van Eenennaam *et al*. 2104. Annual Review Animal Biosciences 2:105-139.

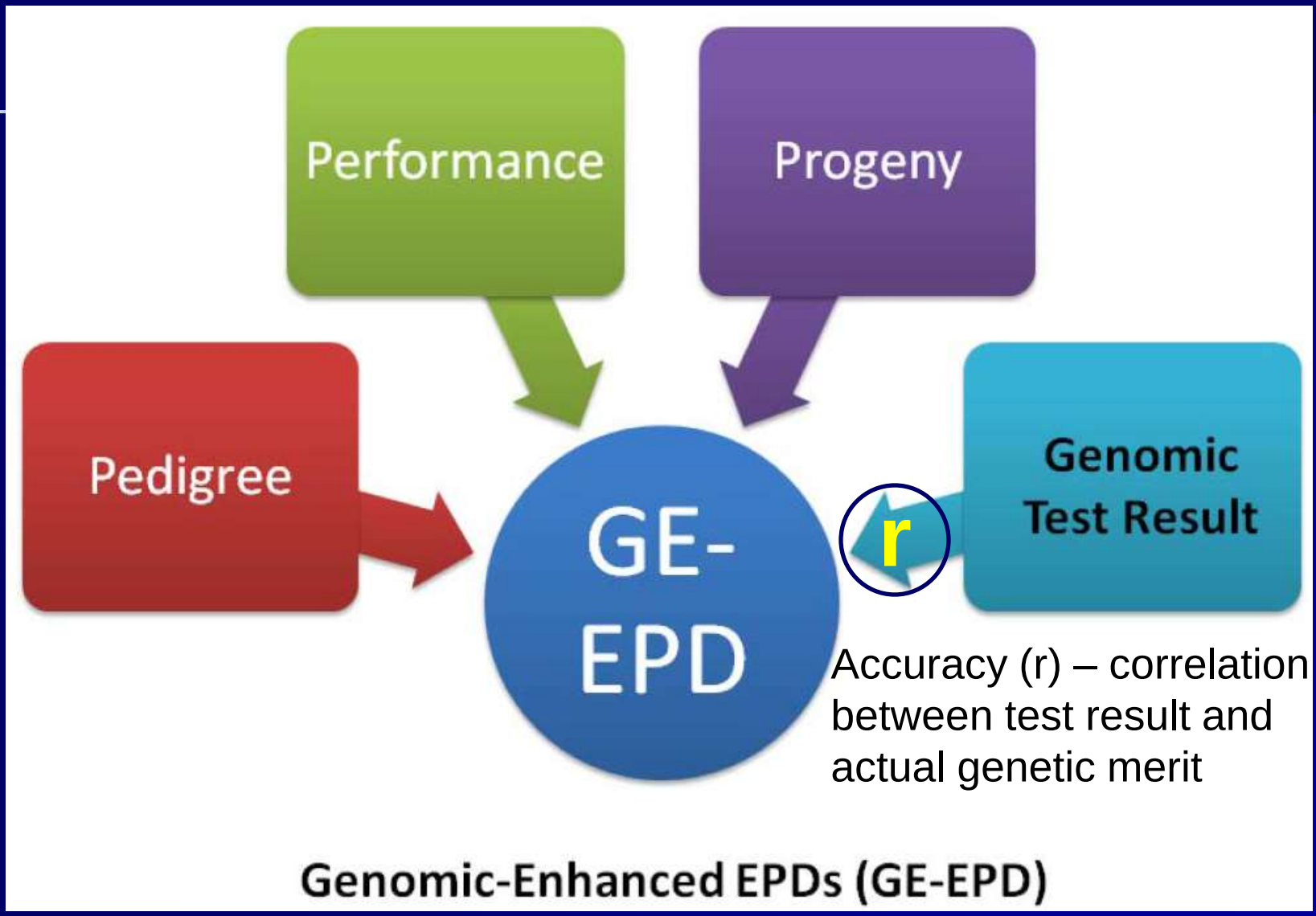
Lead Today with 50K

1. Birth weight
2. Weaning weight
3. Weaning maternal (milk)
4. Calving ease direct
5. Calving ease maternal
6. Marbling
7. Backfat thickness
8. Ribeye area
9. Carcass weight
10. Tenderness
11. Postweaning average daily gain
12. Daily feed intake
13. Feed efficiency (net feed intake)

Pfizer Animal Health
Animal Genetics

50K SNP chip assays
50,000 SNPs spread throughout genome ($75)

Information sources for EPDs – DNA just one source of data for GE-EBV

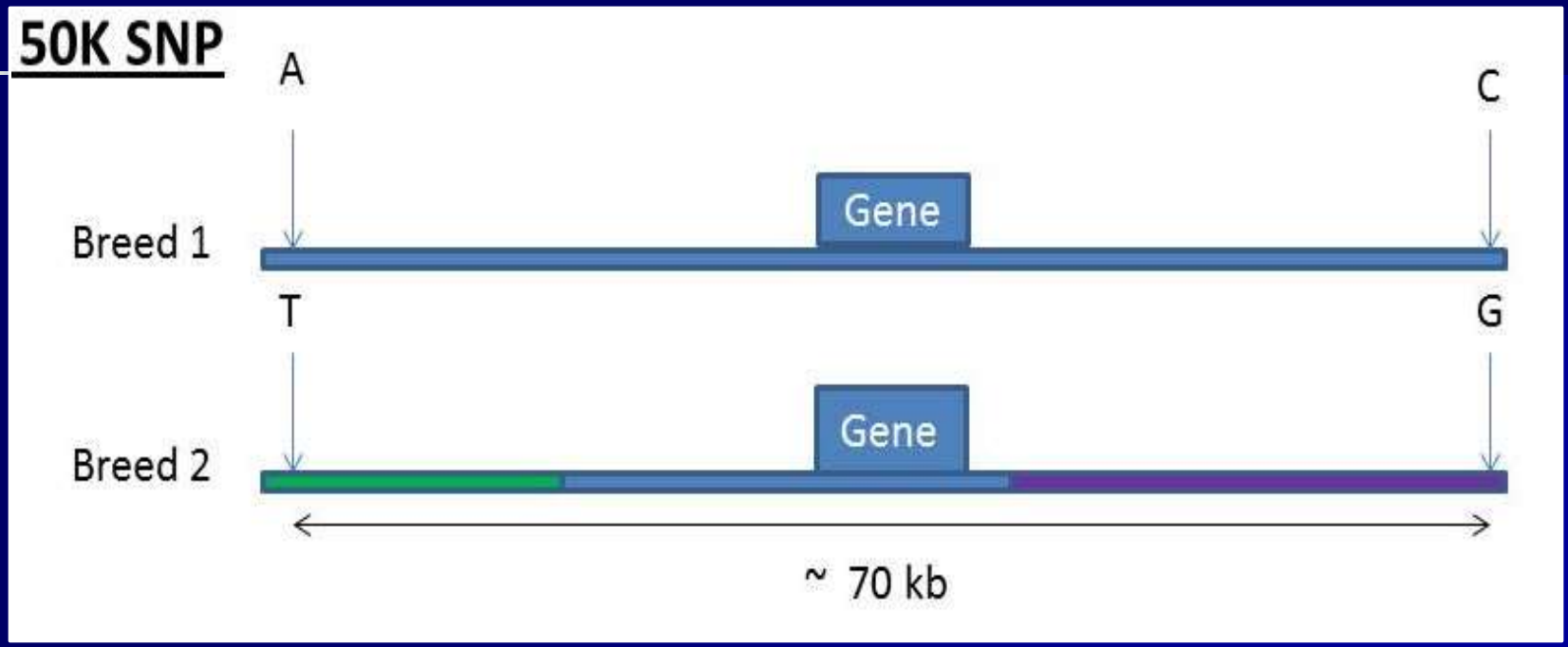


American Angus Association performs weekly evaluations with genomic data – recently updated to include heifer pregnancy

- Association's genetic evaluations, the DNA test results are incorporated into the EPDs using a correlated trait approach.
- The correlations (r) between the HD 50K prediction and the phenotypic data at the Association are updated with each recalibration effort and effectively range from .60 to .70, except for milk (.38) and heifer pregnancy (.49).
- The December 6, 2013, EPD update includes HD 50K predictions from over 51,000 registered Angus animals with genotypes retained at the Association. Results are incorporated into at least 15 EPDs which are then components of the Angus $Value selection index suite.

http://www.angus.org/AGI/GenomicCalibrationRelease.pdf December 2013

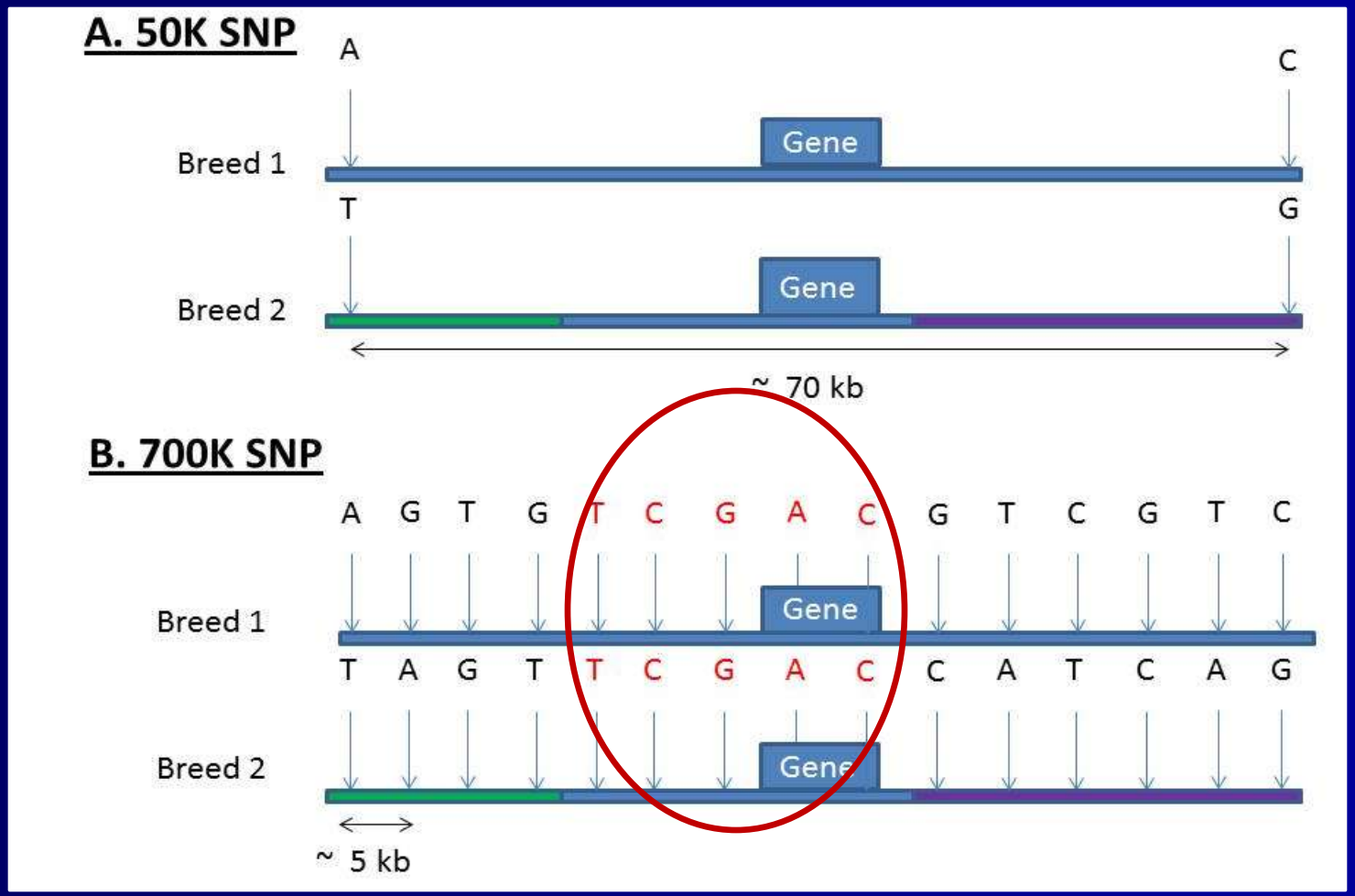
Marker location relative to the gene of interest in two breeds when using the 50K SNP chip assay does not work across breeds

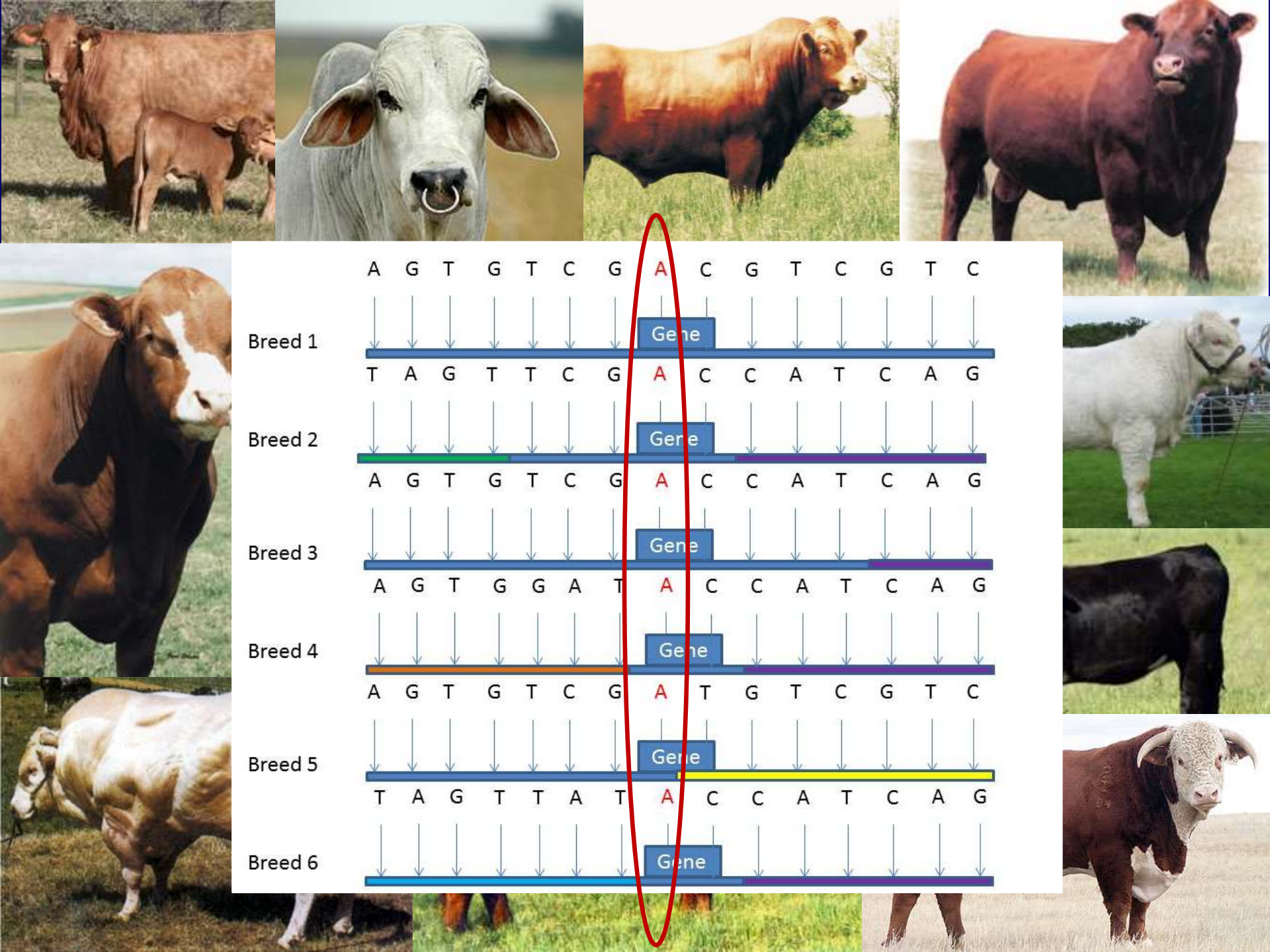


"Our results suggest that the most accurate genomic predictions are achieved when phenotypes from all populations are combined in one training set, while for more diverged populations a higher marker density (in the case of cattle >300,000 SNP) *is required."*

de Roos, A.P.W., B.J. Hayes, and M. E. Goddard. 2009. Reliability of Genomic Predictions Across Multiple Populations. Genetics. 183(4): 1545–1553

Marker location relative to the gene of interest in two breeds when using the (A) 50K SNP chip assay (markers spaced at ~ 70 kb intervals), or (B) the high density 700 K SNP chip assay (markers spaced at ~ 5 kb intervals)




A G T G T C G A C G T C G T C
Breed 1
Gene
T A G T T C G A C C A T C A G
Breed 2
Gene
A G T G T C G A C C A T C A G
Breed 3
Gene
A G T G G A T A C C A T C A G
Breed 4
Gene
A G T G T C G A T G T C G T C
Breed 5
Gene
T A G T T A T A C C A T C A G
Breed 6
Gene

Practical questions for breeders

- How many phenotypic records are required in the initial experiment (reference population) used to estimate the effect of chromosome segments?
- How often do we need to re-estimate the chromosome segment effects?
- Does it work across breeds/strain/cultivars?
- **How many markers do you need – 1, 384, 10K, 50K, 770K, whole genome?**
- **What about less expensive reduced panels e.g. 100 SNP panels– can they work?**

The Power of the IGENITY® profile for Angus

384 SNP

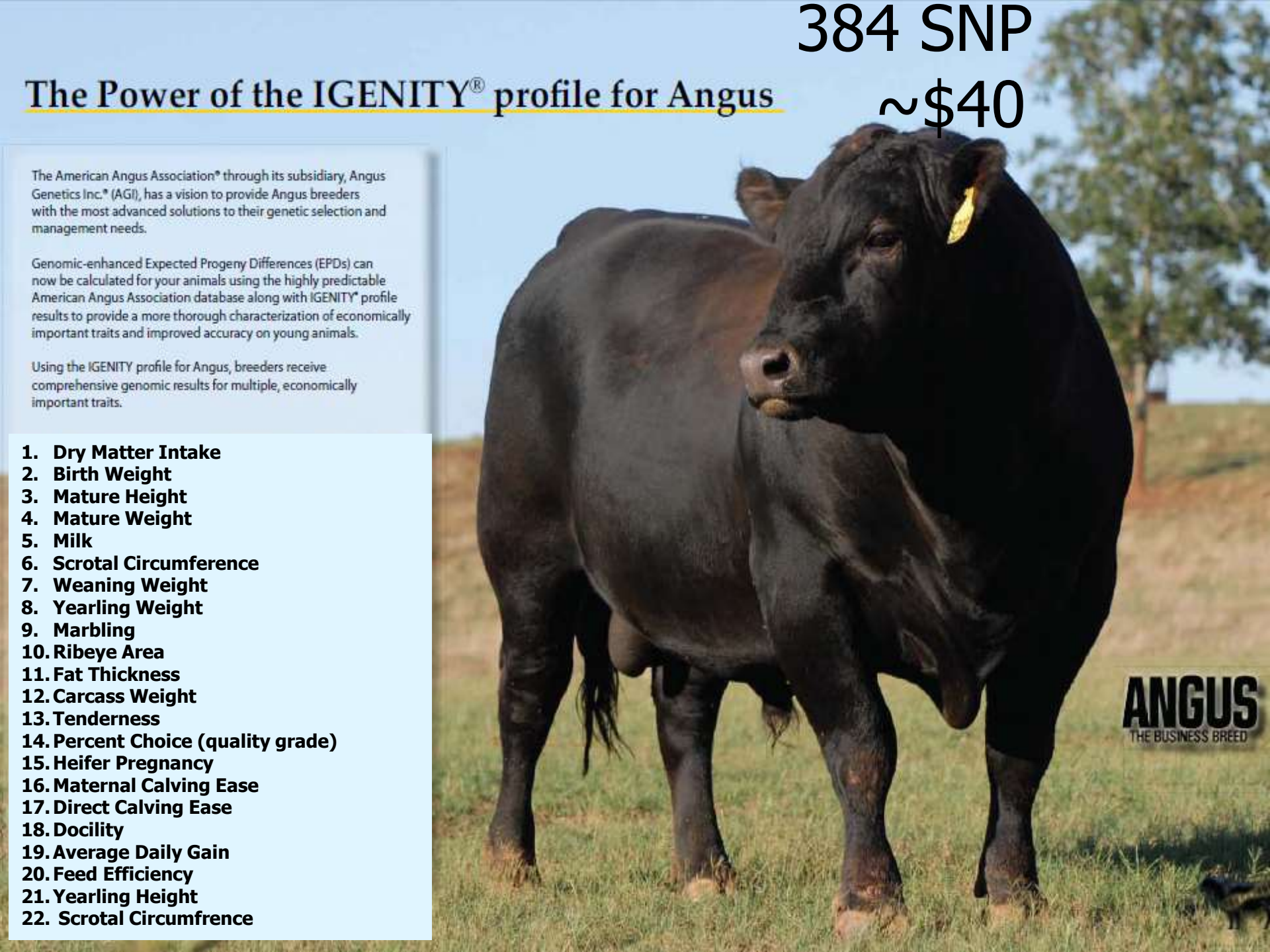
~$40

The American Angus Association® through its subsidiary, Angus Genetics Inc.® (AGI), has a vision to provide Angus breeders with the most advanced solutions to their genetic selection and management needs.

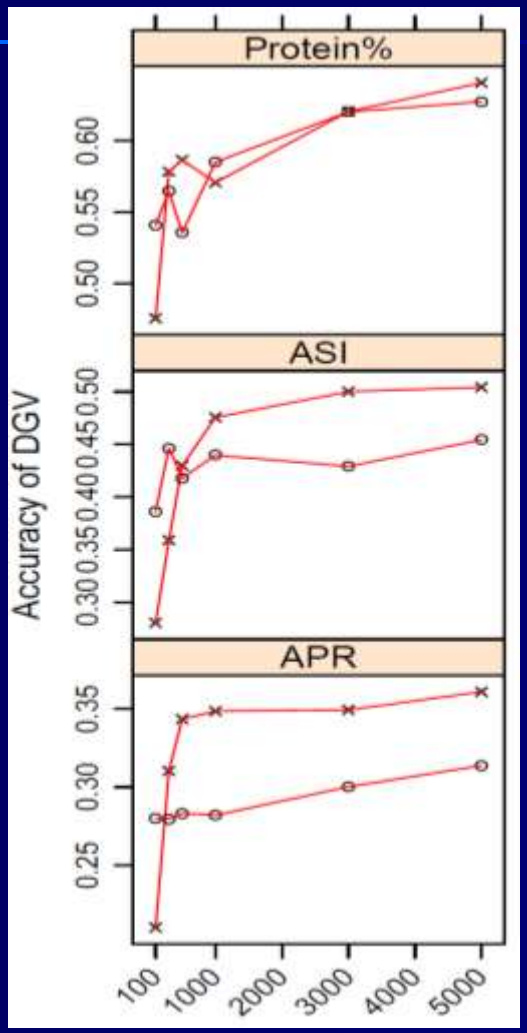
Genomic-enhanced Expected Progeny Differences (EPDs) can now be calculated for your animals using the highly predictable American Angus Association database along with IGENITY® profile results to provide a more thorough characterization of economically important traits and improved accuracy on young animals.

Using the IGENITY profile for Angus, breeders receive comprehensive genomic results for multiple, economically important traits.

1. **Dry Matter Intake**
2. **Birth Weight**
3. **Mature Height**
4. **Mature Weight**
5. **Milk**
6. **Scrotal Circumference**
7. **Weaning Weight**
8. **Yearling Weight**
9. **Marbling**
10. **Ribeye Area**
11. **Fat Thickness**
12. **Carcass Weight**
13. **Tenderness**
14. **Percent Choice (quality grade)**
15. **Heifer Pregnancy**
16. **Maternal Calving Ease**
17. **Direct Calving Ease**
18. **Docility**
19. **Average Daily Gain**
20. **Feed Efficiency**
21. **Yearling Height**
22. **Scrotal Circumfrence**



Reduced SNP panels: Accuracy of direct genomic value (DGV) of dairy bulls using subsets of 5,000 or less of best SNP for each trait



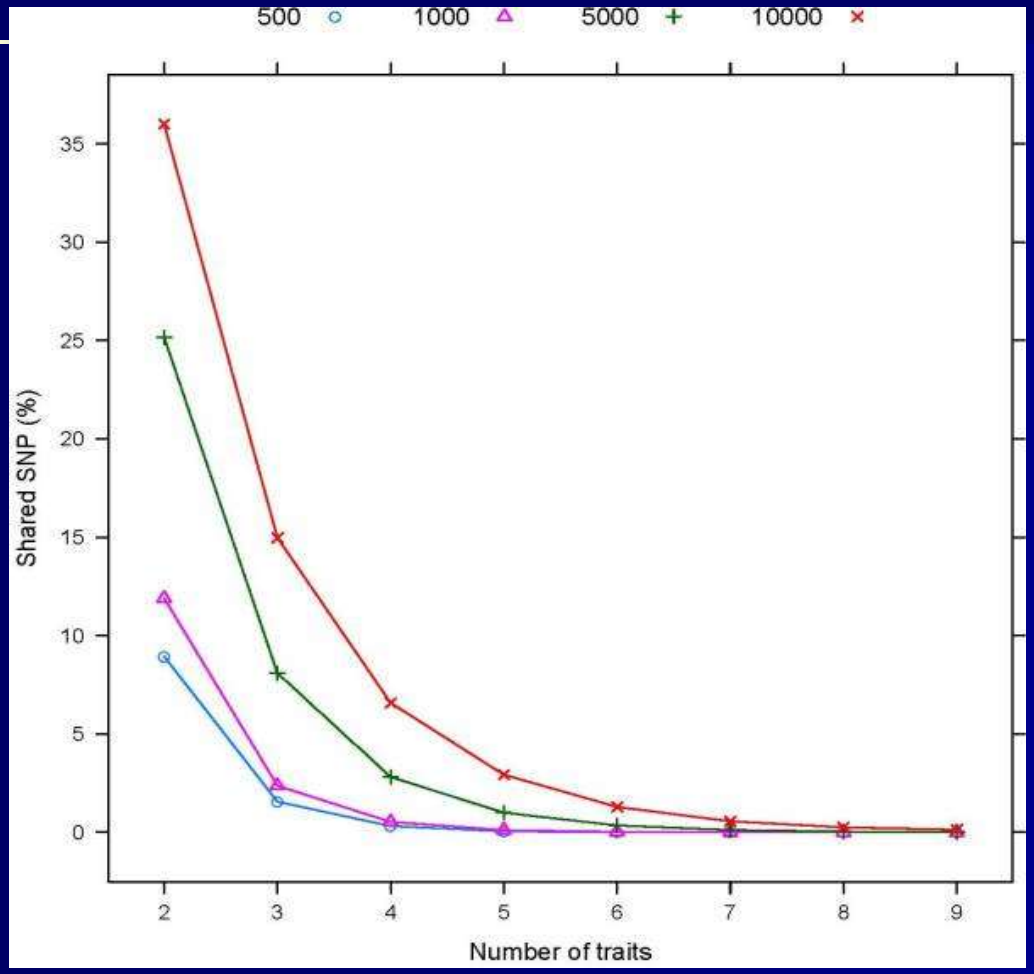
Traits :
Protein %
ASI (Australian Selection Index)
APR (Australian Profit Rank)

Predictions based on <1,000 SNP panels were very sensitive to the selection method and tended to be low accuracy

Moser, G., M. S. Khatkar, B. J. Hayes, and H. W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genetics Selection Evolution 42.

Reduced SNP panels: Percentage of the highest ranked SNP that are shared between sets of traits*

for subsets including 500, 1,000, 5,000 or 10,000 SNP



* **Dairy traits included**:

1. Protein
2. Protein %
3. Survival
4. Fat %
5. Milk
6. Overall Type
7. APR (Australian Profit Rank)
8. ASI (Australian Selection Index)
9. Fat

Few SNPs were in common between the trait-specific subsets

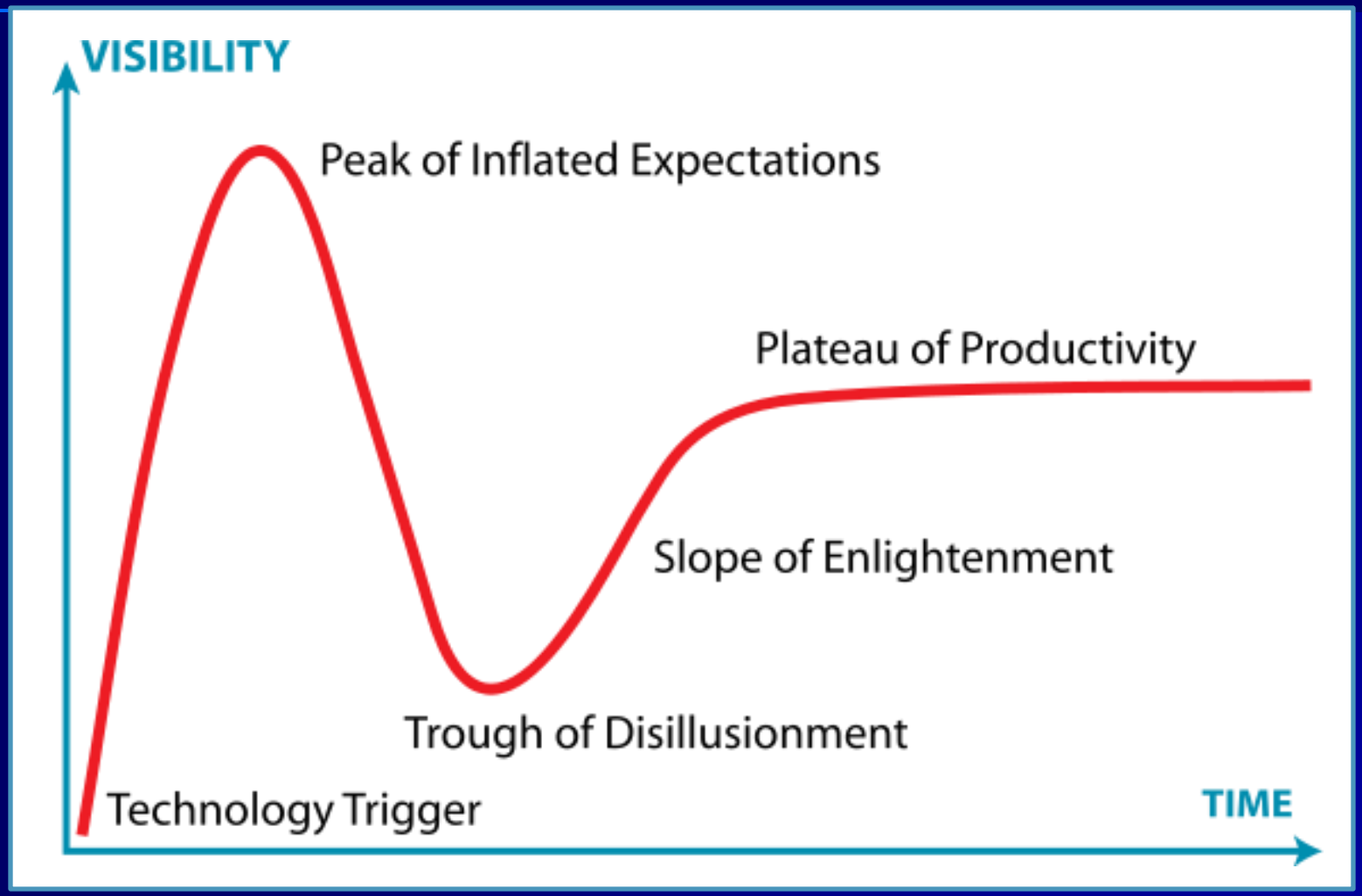
Moser, G., M. S. Khatkar, B. J. Hayes, and H. W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genetics Selection Evolution 42.

Summary of what the literature is telling us about genomic selection

- Prediction equations derived in one breed do not predict accurate GEBVs when applied to other breeds
- Combining breeds into one large multi-breed reference population may give reasonable accuracies in purebreds
- To find markers that are in LD with QTL across diverged breeds, such as Holstein, Jersey, and Angus, will require high density SNPs (>300,000 informative markers in cattle)
- If markers are picking up family relationships (linkage), then the accuracy of marker-based selection will decay over generations within a breed/line/cultivar
- Few of the "best" markers for one trait are common to another so "reduced panels" will need to be designed for imputation not single trait selection


The Future
NEXT EXIT

Hype cycle: the over-enthusiasm or "hype" and subsequent disappointment that typically happens with the introduction of new technologies



It may be necessary to go to whole genome resequencing – select with the causative SNPs (rather than LD)?

- Cost is likely to get to as low as $1000/animal
- Whole genome data >40% more accurate than dense SNP chips
- Need to use Bayesian approaches to estimate SNP effects
- Predictions remained accurate in populations 10 generations removed from the reference population

Meuwissen, T. and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 183:623-631.

Practical questions for breeders – some still unanswered!

- How many phenotypic records are required in the initial experiment (reference population) used to estimate the effect of chromosome segments?
- How often do we need to re-estimate the chromosome segment effects?
- Does it work across breeds/strain/cultivars?
- How many markers do you need – reduced panel (eg. 3K), 50K, 770K, whole genome?
- **How much can you afford to pay? (and who pays)**
- **Does this technology change optimal breeding program design? Absolutely need a multi-trait $selection index based on breeding objective.**


Questions?