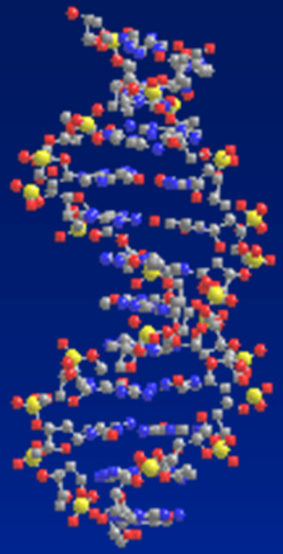
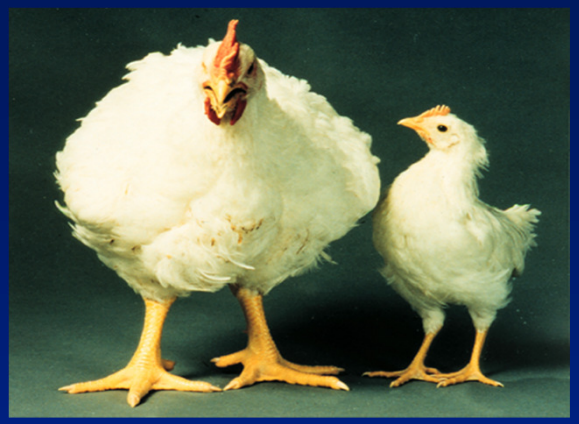
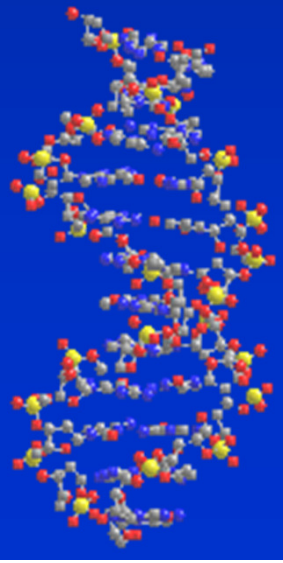




Genomic Selection in the era of Genome sequencing



Genomic selection

- Factors affecting accuracy of genomic selection
- How often to re-estimate the chromosome segment effects?
- Genomic selection with low marker density
- Genomic selection across breeds
- Optimal breeding program design with genomic selection

Accuracy of genomic selection

- Factors affecting accuracy of genomic selection r(GEBV,TBV)
  - Linkage disequilibrium between QTL and markers = density of markers
  - Single markers, haplotypes or IBD
  - Number of records used to estimate chromosome segment effects

Accuracy of genomic selection

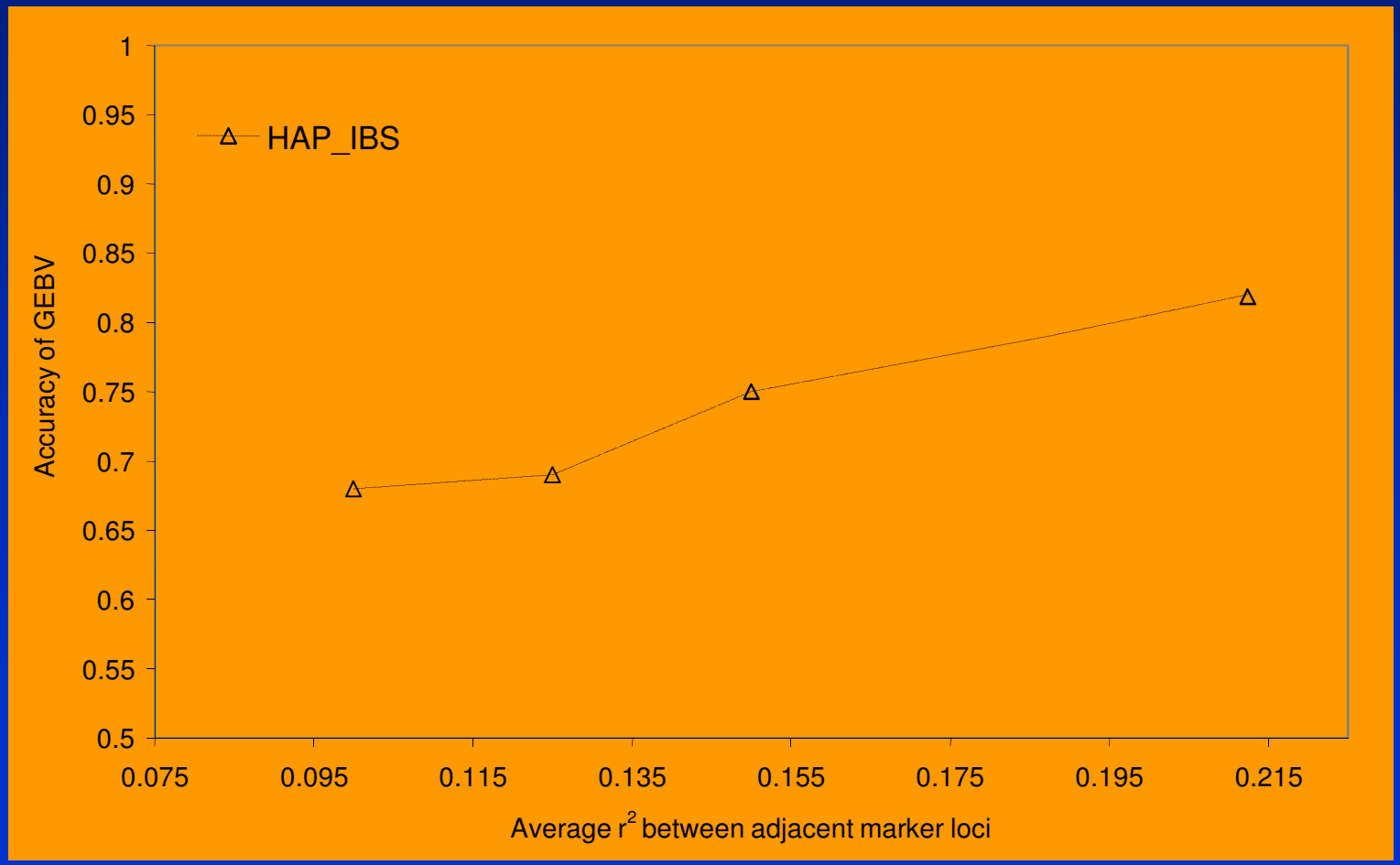
- Factors affecting accuracy of genomic selection r(GEBV,TBV)
  - Linkage disequilibrium between QTL and markers = density of markers
    - Haplotypes or single markers be in sufficient LD with the QTL such that the haplotype or single markers will predict the effects of the QTL across the population.

Accuracy of genomic selection

- Factors affecting accuracy of genomic selection r(GEBV,TBV)
  - Linkage disequilibrium between QTL and markers = density of markers
    - Haplotypes or single markers be in sufficient LD with the QTL such that the haplotype or single markers will predict the effects of the QTL across the population.
    - Calus et al. (2007) used simulation to assess effect of LD between QTL and markers on accuracy of genomic selection

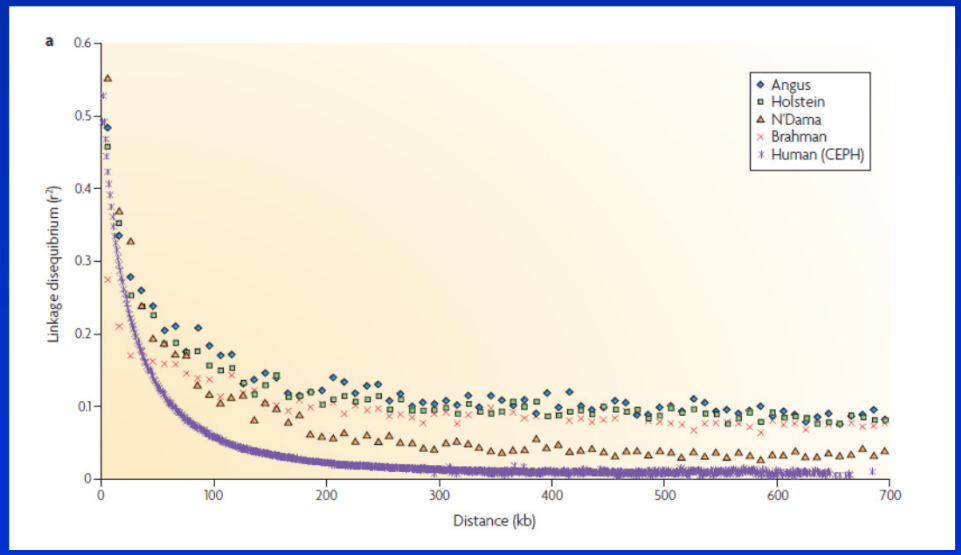
Accuracy of genomic selection

- Effect of LD on accuracy of selection



Accuracy of genomic selection

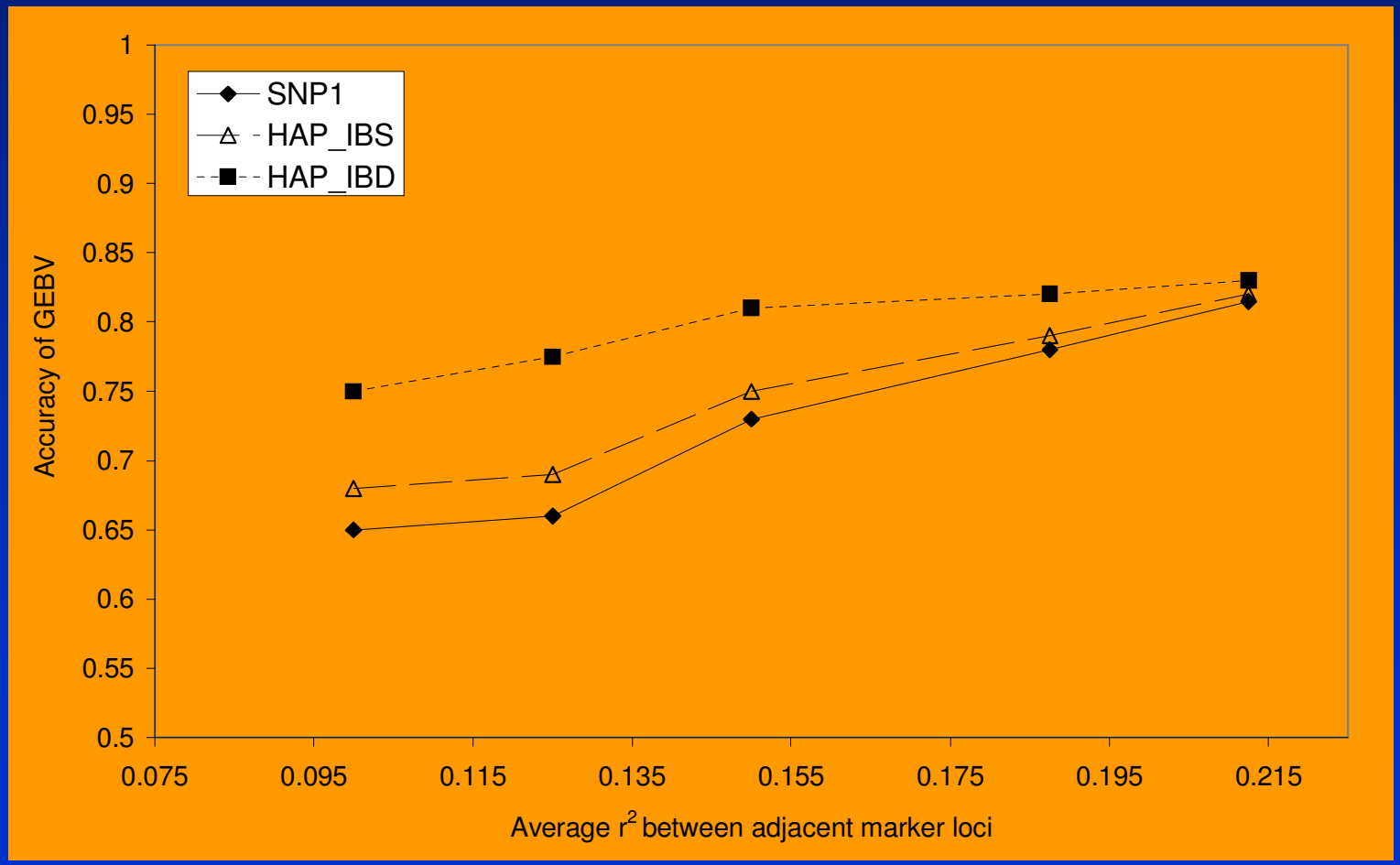
- Factors affecting accuracy of genomic selection r(GEBV,TBV)
  - Linkage disequilibrium between QTL and markers = density of markers
  - In dairy cattle populations, an average r^2 of 0.2 between adjacent markers is only achieved when markers are spaced every 100kb.



Accuracy of genomic selection

- Comparing the accuracy of genomic selection with
  - IBD approach
  - haplotypes
  - single markers
  - Calus et al (2007) used simulated data

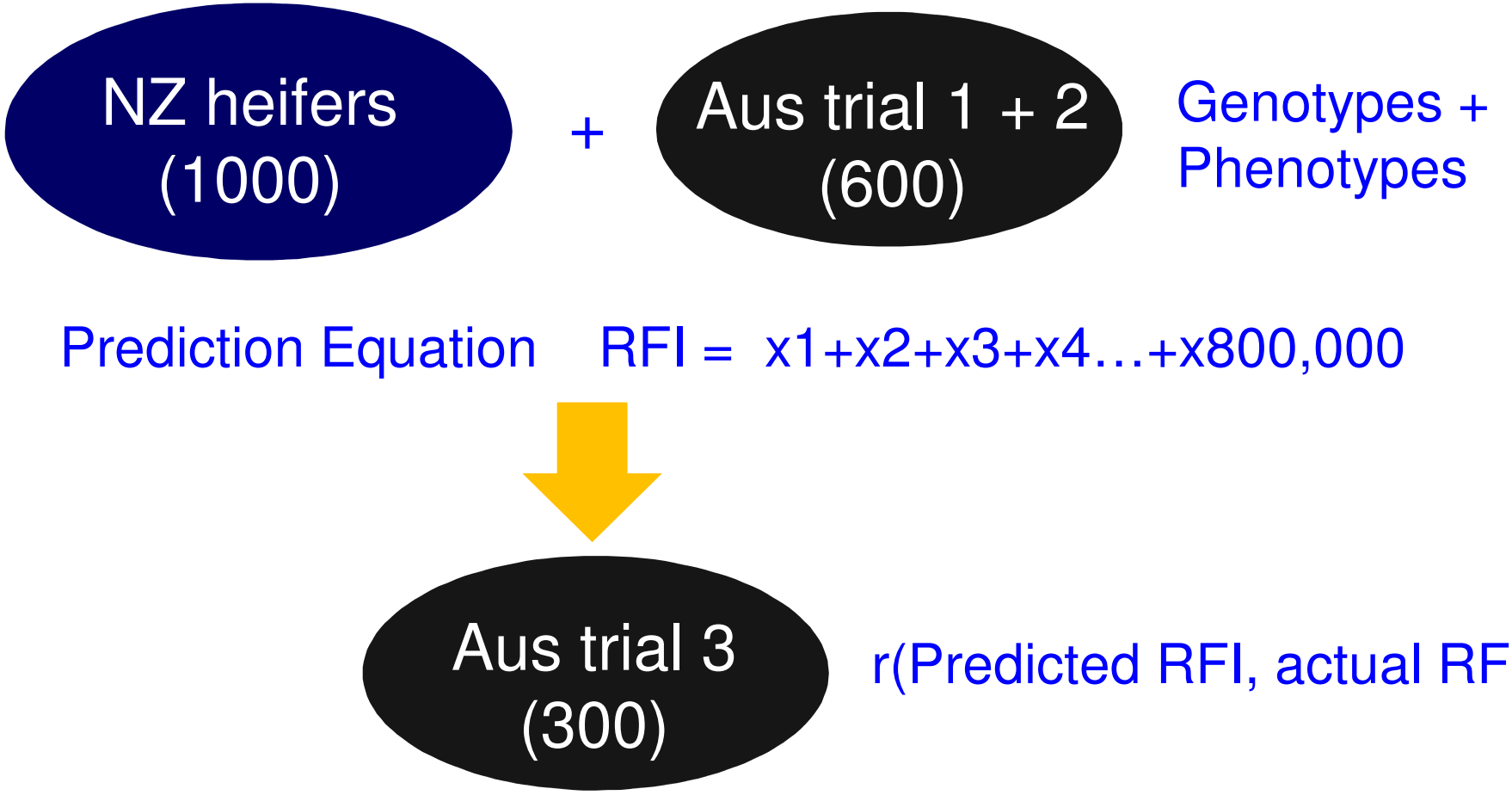
Accuracy of genomic selection



Genomic Predictions Residual Feed Intake

- Collaboration DPI Vic, Livestock Improvement Corporation and Dairy NZ (Richard Spelman, Kevin MacDonald, et al.)
- 1000 heifers each
- Genotyped 800,000 SNPs (Illumina Bovine HD)

Genomic predictions



Genomic Predictions Residual Feed Intake

- To derive prediction equation
- GBLUP -> all markers have small, non zero effect
- BayesR -> proportion of markers have zero effect, rest have small to moderate effects

Accuracy GEBV Residual Feed Intake

Trait	Marker Panel	GBLUP	BayesR
Liveweight	50K	0.35	0.35
	800K	0.38	0.40
Residual Feed Intake	50K	0.29	0.39
	800K	0.29	0.41

Accuracy of genomic selection

- Number of records used to estimate chromosome segment effects
  - Chromosome segment effects g_i estimated in a reference population
  - How big does this reference population need to be?
  - Meuwissen et al. (2001) evaluated accuracy using LS, BLUP, BayesB using 500, 1000 or 2000 records in the reference population

Accuracy of genomic selection

- Number of records used to estimate chromosome segment effects

	No. of phenotypic records		
	500	1000	2200
Least squares	0.124	0.204	0.318
Best linear unbiased prediction (BLUP)	0.579	0.659	0.732
BayesB	0.708	0.787	0.848

Accuracy of genomic selection

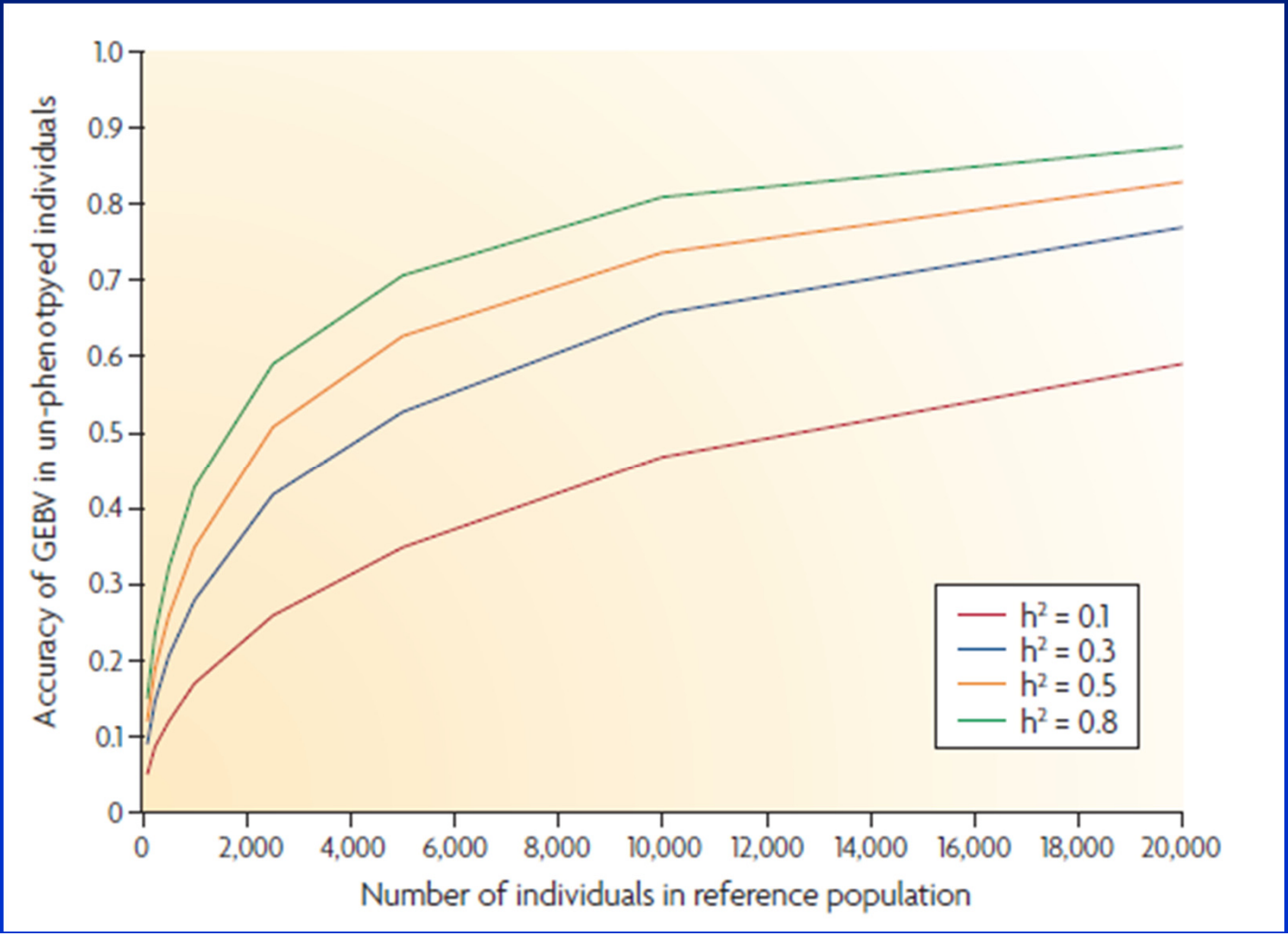
- Number of records used to estimate chromosome segment effects

	No. of phenotypic records		
	500	1000	2200
Least squares	0.124	0.204	0.318
Best linear unbiased prediction (BLUP)	0.579	0.659	0.732
BayesB	0.708	0.787	0.848

h^2=0.5

Accuracy of genomic selection

- Number of records used to estimate chromosome segment effects



Genomic selection

- Factors affecting accuracy of genomic selection
- How often to re-estimate the chromosome segment effects?
- Genomic selection with low marker density
- Genomic selection across breeds
- Optimal breeding program design with genomic selection

Genomic selection

- How often to re-estimate the chromosome segment effects?
  - If the markers used in genomic selection were actually the underlying mutations causing the QTL effects, the estimation of chromosome segment effects could be performed once in the reference population.
  - GEBVs for all subsequent generations could be predicted using these effects.

Genomic selection

- How often to re-estimate the chromosome segment effects?
  - In practise is that there will be markers with low to moderate levels of r^2 with the underlying mutations causing the QTL effect.
  - Do not capture all of QTL variance
  - Over time, recombination between the markers and QTL will reduce the accuracy of the GEBV using chromosome segment effects predicted from the original reference population.
  - We need to re-estimate chromosome segment effects
  - How often?

Genomic selection

- How often to re-estimate the chromosome segment effects?

Table 4.3. 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. From Meuwissen et al. (2001).

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.

Genomic selection

- How often to re-estimate the chromosome segment effects?

Table 4.3. 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. From Meuwissen et al. (2001).

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.

  - Denser markers >> generations between re-estimation of effects

Genomic selection

- However decay of accuracy is dependant on genomic selection method…..

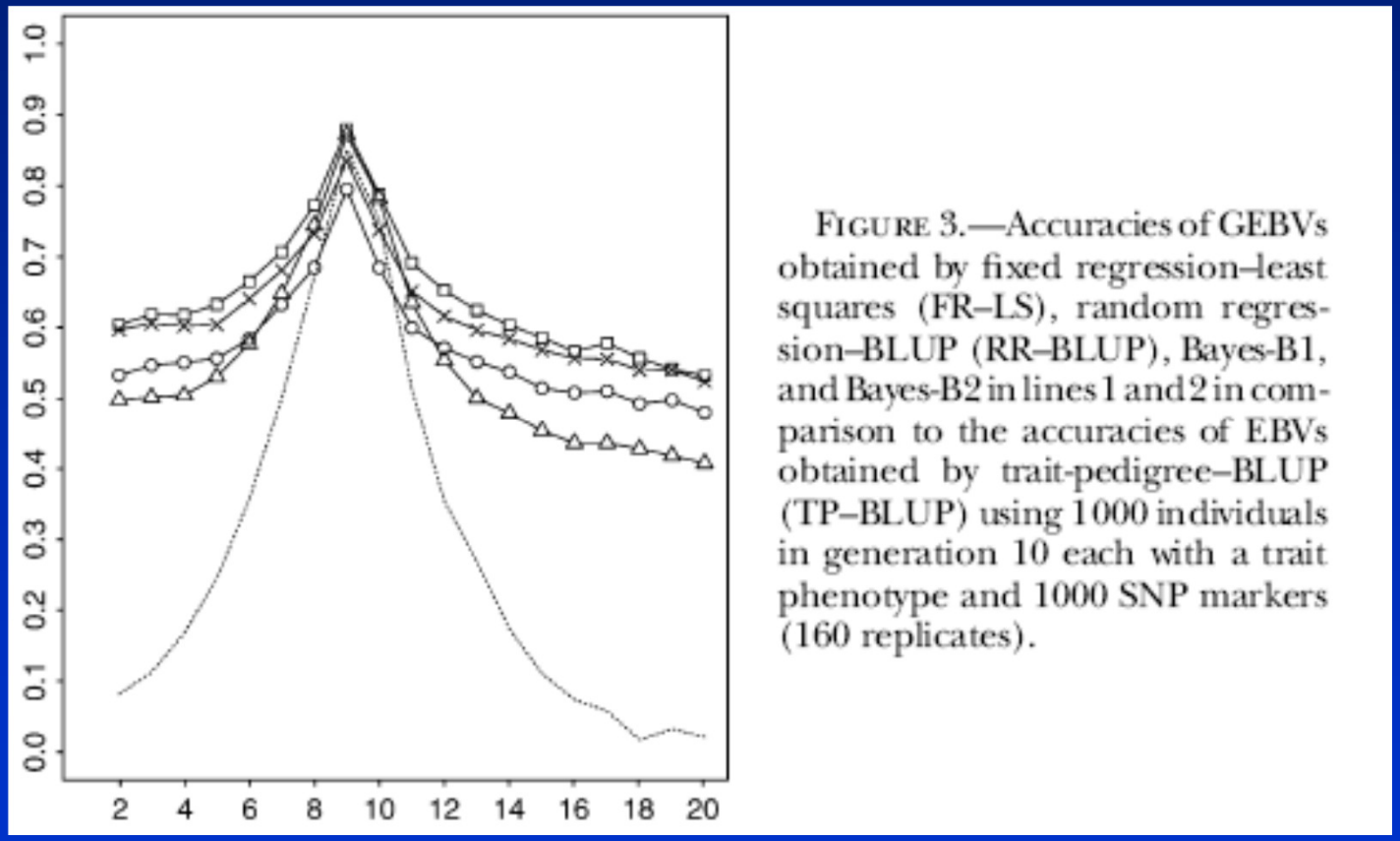


FIGURE 3.—Accuracies of GEBVs obtained by fixed regression–least squares (FR–LS), random regression–BLUP (RR–BLUP), Bayes-B1, and Bayes-B2 in lines 1 and 2 in comparison to the accuracies of EBVs obtained by trait-pedigree–BLUP (TP–BLUP) using 1000 individuals in generation 10 each with a trait phenotype and 1000 SNP markers (160 replicates).

- Habier et al. (Genetics 177:2389)

Genomic selection

- Decay of accuracy actually depends on LD between QTL and SNPs
  - Higher LD slower decay
- ***Genomic selection methods will also pick up pedigree effects if this is not accounted for!!***
  - Eg a rare SNP heterozygous in a sire is a good marker for the family derived from that sire!

Genomic selection

- Decay of accuracy actually depends on LD between QTL and SNPs
  - Higher LD slower decay
- ***Genomic selection methods will also pick up pedigree effects if this is not accounted for!!***
  - Eg a rare SNP heterozygous in a sire is a good marker for the family derived from that sire!
  - BLUP especially bad, as is the same as fitting average relationship matrix derived from markers
    - Eg each segment has same variance

Genomic selection

- Decay of accuracy actually depends on LD between QTL and SNPs
  - Higher LD slower decay
- ***Genomic selection methods will also pick up pedigree effects if this is not accounted for!!***
  - Eg a rare SNP heterozygous in a sire is a good marker for the family derived from that sire!
  - Solutions
    - Fit polygenic effect in model
      - Sample **u** from $N(0,A\sigma^2)$ in Gibbs chain and correct when sampling other effects

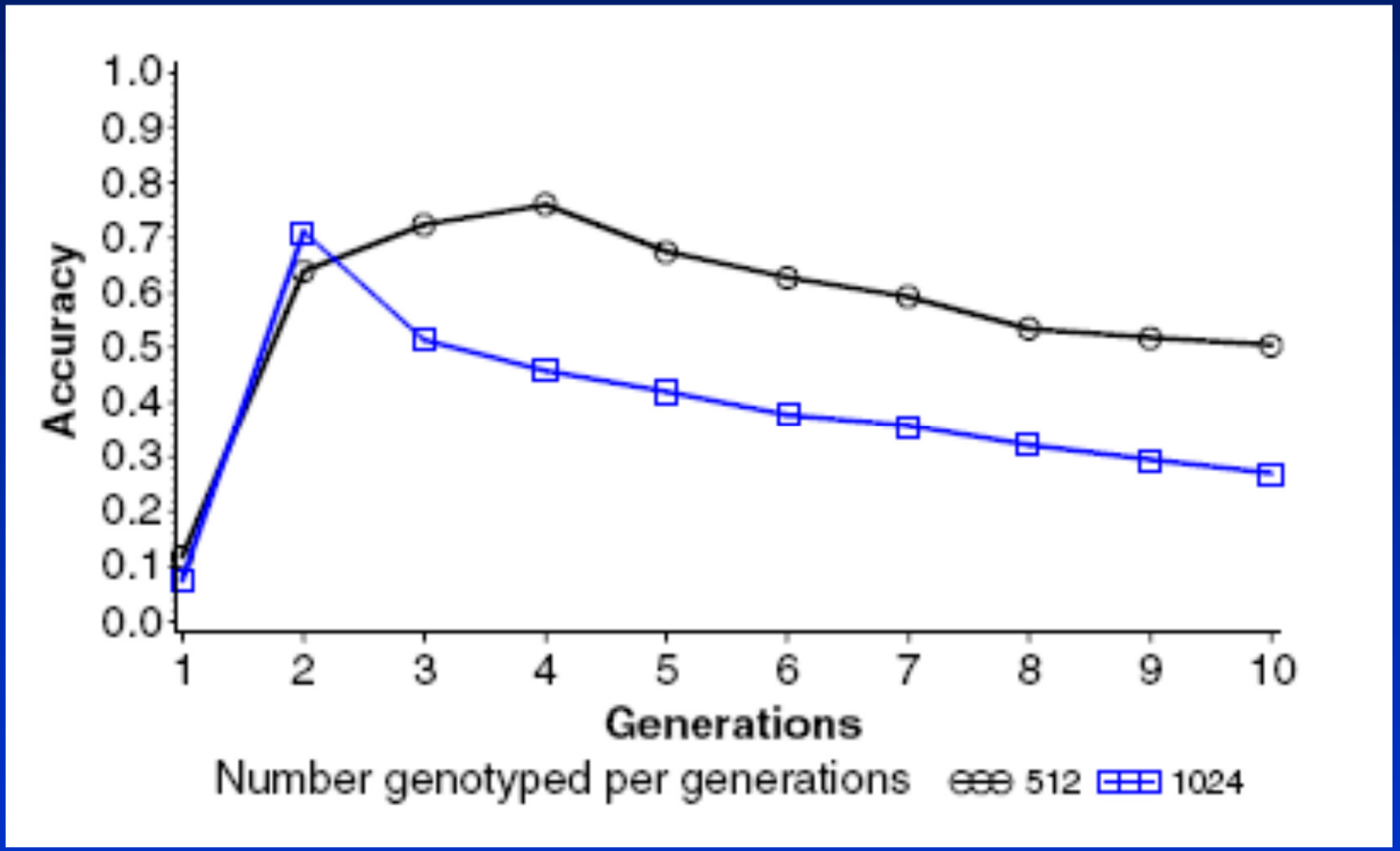
Genomic selection

- Decay of accuracy actually depends on LD between QTL and SNPs
  - Higher LD slower decay
- ***Genomic selection methods will also pick up pedigree effects if this is not accounted for!!***
  - Eg a rare SNP heterozygous in a sire is a good marker for the family derived from that sire!
  - Solutions
    - Fit polygenic effect in model
      - Sample **u** from $N(0,A\sigma^2)$ in Gibbs chain and correct when sampling other effects
    - Use multiple breeds?
      - Must be very close to QTL for SNP to have effect across multiple breeds

Genomic selection

- Decay of accuracy actually depends on LD between QTL and SNPs
  - Higher LD slower decay
- ***Genomic selection methods will also pick up pedigree effects if this is not accounted for!!***
  - Eg a rare SNP heterozygous in a sire is a good marker for the family derived from that sire!
  - Solutions
    - Multiple generations in reference population?

Genomic selection



Two generations in reference pop vs. one generation

(Muir 2008: Journal of Animal Breeding and Genetics 124:342)

Genomic selection

- Factors affecting accuracy of genomic selection
- How often to re-estimate the chromosome segment effects?
- Genomic selection with low marker density
- Genomic selection across breeds
- Optimal breeding program design with genomic selection

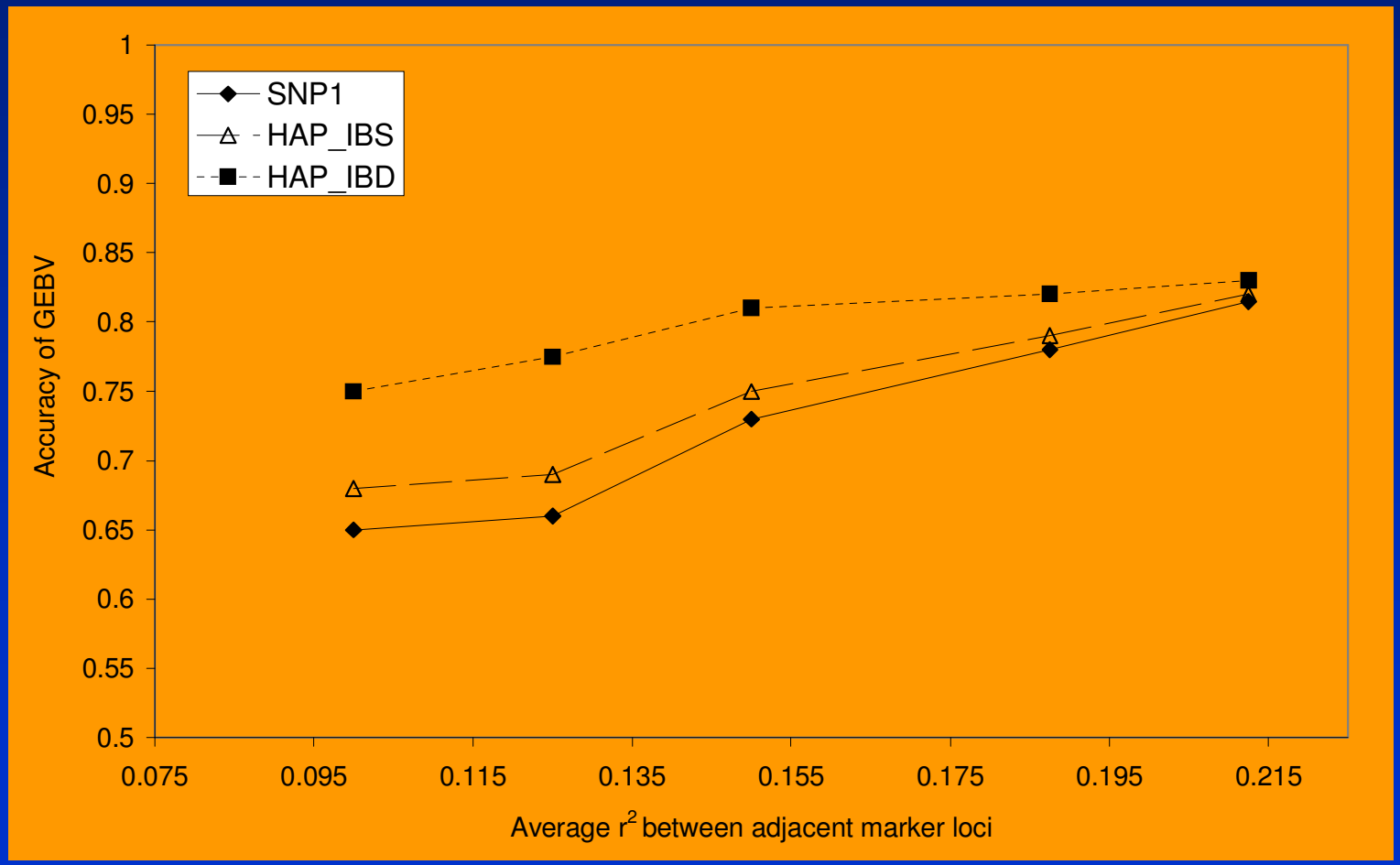
Genomic selection

- Genomic selection with low marker density
  - May not be enough markers across genome to ensure adjacent markers have $r^2>=0.2$.
  - Will not capture all the genetic variance with the markers.

Genomic selection

- Genomic selection with low marker density
  - May not be enough markers across genome to ensure adjacent markers have $r^2>=0.2$.
  - Will not capture all the genetic variance with the markers.
  - Two strategies
    - Exploit linkage as well as linkage disequilibrium by using the IBD approach

Accuracy of genomic selection



Genomic selection

- Genomic selection with low marker density
  - May not be enough markers across genome to ensure adjacent markers have $r^2>=0.2$.
  - Will not capture all the genetic variance with the markers.
  - Two strategies
    - Exploit linkage as well as linkage disequilibrium by using the IBD approach
    - Include a polygenic effect to capture some of the genetic variance not captured by the markers (exploit pedigree)

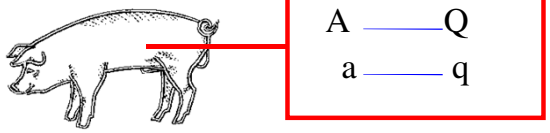
$$\mathbf{GEBV} = \hat{\mathbf{u}} + \sum_{i}^{p} \mathbf{X}_i \hat{\mathbf{g}}_i$$

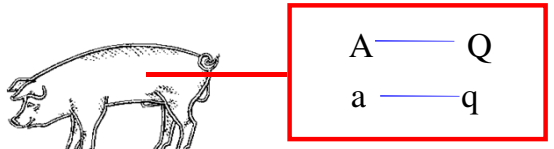
Genomic selection

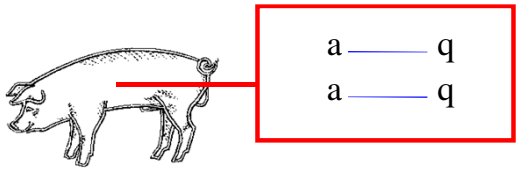
- Factors affecting accuracy of genomic selection
- How often to re-estimate the chromosome segment effects?
- Genomic selection with low marker density
- Genomic selection across breeds
- Optimal breeding program design with genomic selection

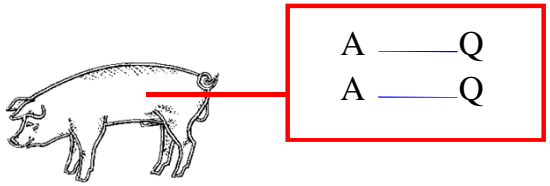
Genomic selection across breeds

- Genomic selection relies on the phase of LD between markers and QTL being the same in the selection candidates as in the reference population.
- However as the two populations diverge, this is less and less likely to be the case
  - especially if the distance between markers and QTL is relatively large.


A Q
a q

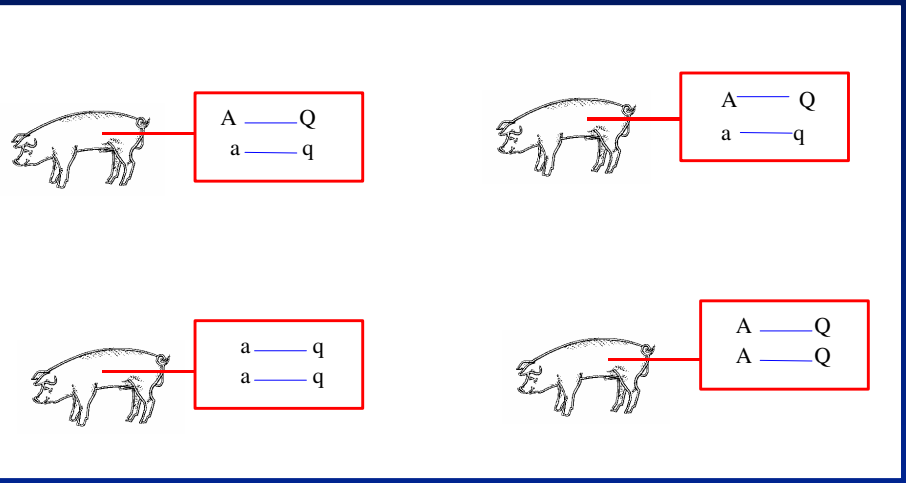

A Q
a q


a q
a q


A Q
A Q

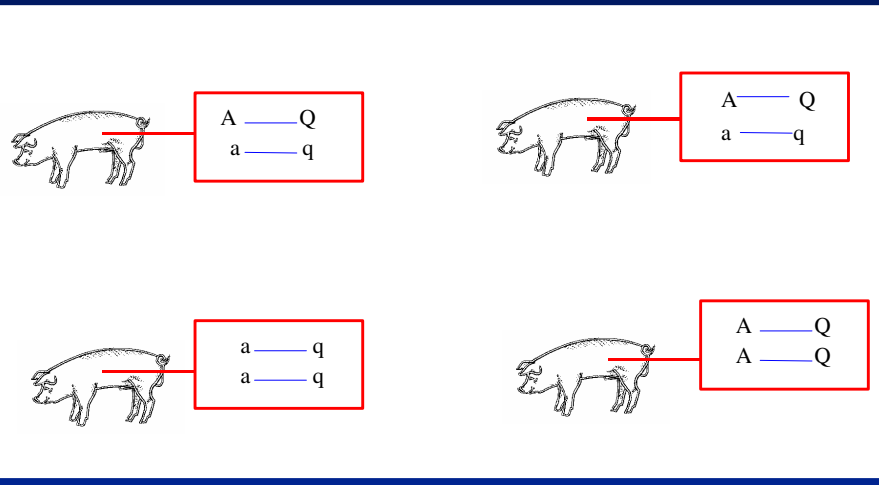
Same Breed

Predict g_i in reference population



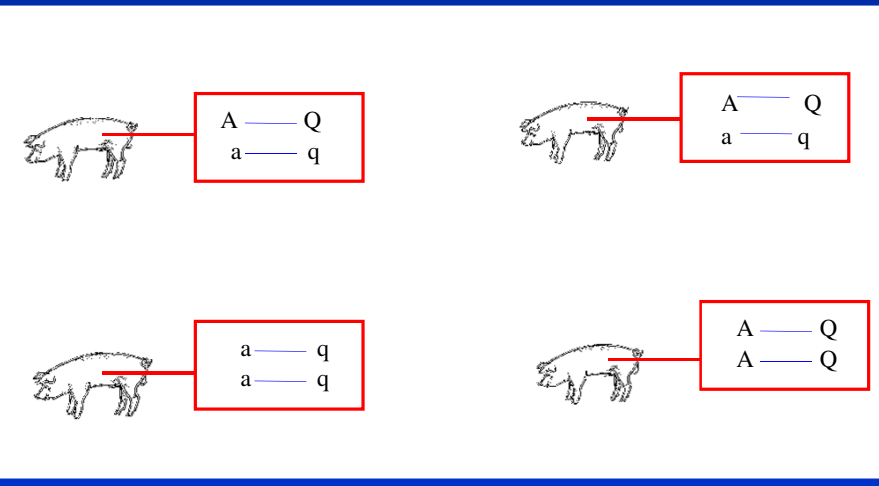
Same Breed

Predict g_i in reference population



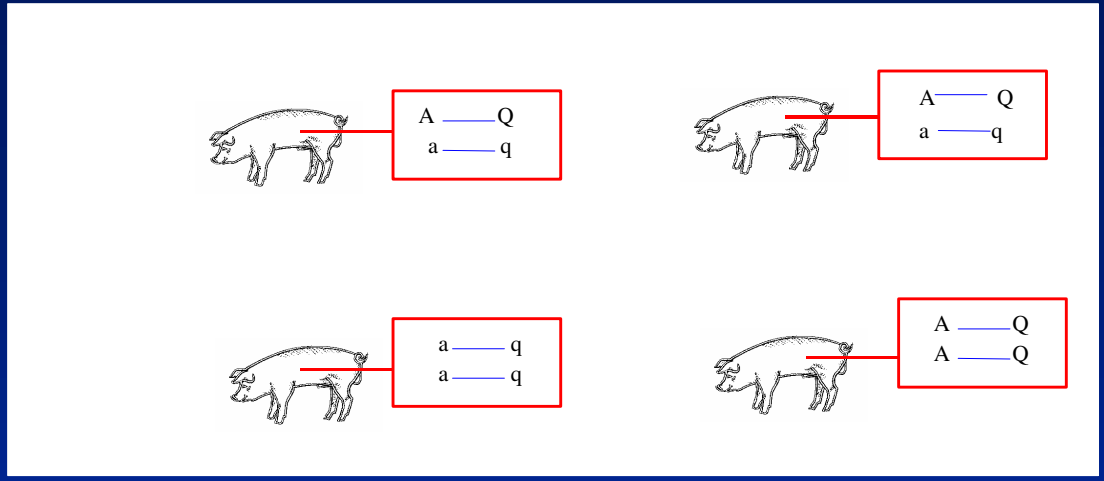
Calculate GEBV in selection candidates

$$\mathbf{GEBV} = \sum_{i}^{p} \mathbf{X}_i \hat{\mathbf{g}}_i$$



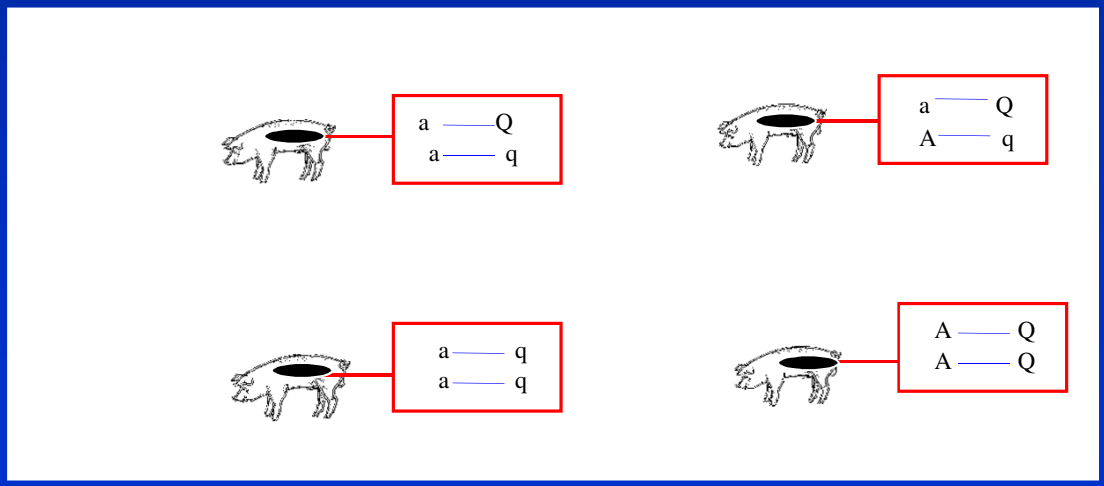
Different Breeds

Predict g_i in reference population



Calculate GEBV in selection candidates

$$\mathbf{GEBV} = \sum_{i}^{p} \mathbf{X}_i \hat{\mathbf{g}}_i$$



Genomic selection across breeds

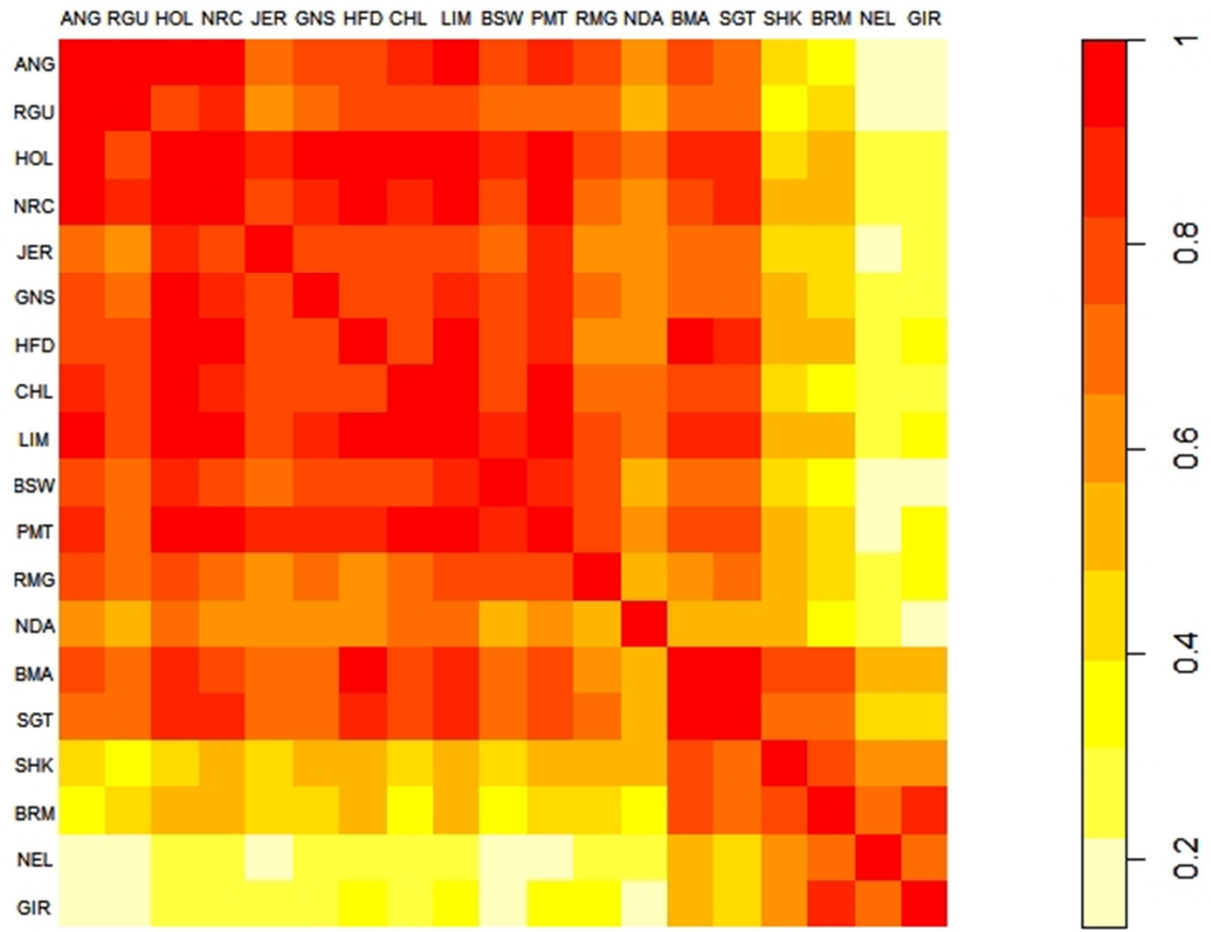
- Use correlation between r in two populations, corr(r1,r2), to assess persistence of LD across populations.
  - Signed r^2 statistic
  - If same sign in different breeds, same marker allele associated with increasing QTL allele

Genomic selection across breeds

- Use correlation between r in two populations, corr(r1,r2), to assess persistence of LD across populations.
  - Signed r^2 statistic
  - If same sign in different breeds, same marker allele associated with increasing QTL allele

- If the chromosome segment effects are estimated in population 1, and GEBVs in that population can be predicted with an accuracy *x1*, then the GEBVs of animals population 2 may be predicted from the chromosome segment effects of population 1 with an accuracy *x2* = *x1**corr(r1,r2)

10 - 50kb

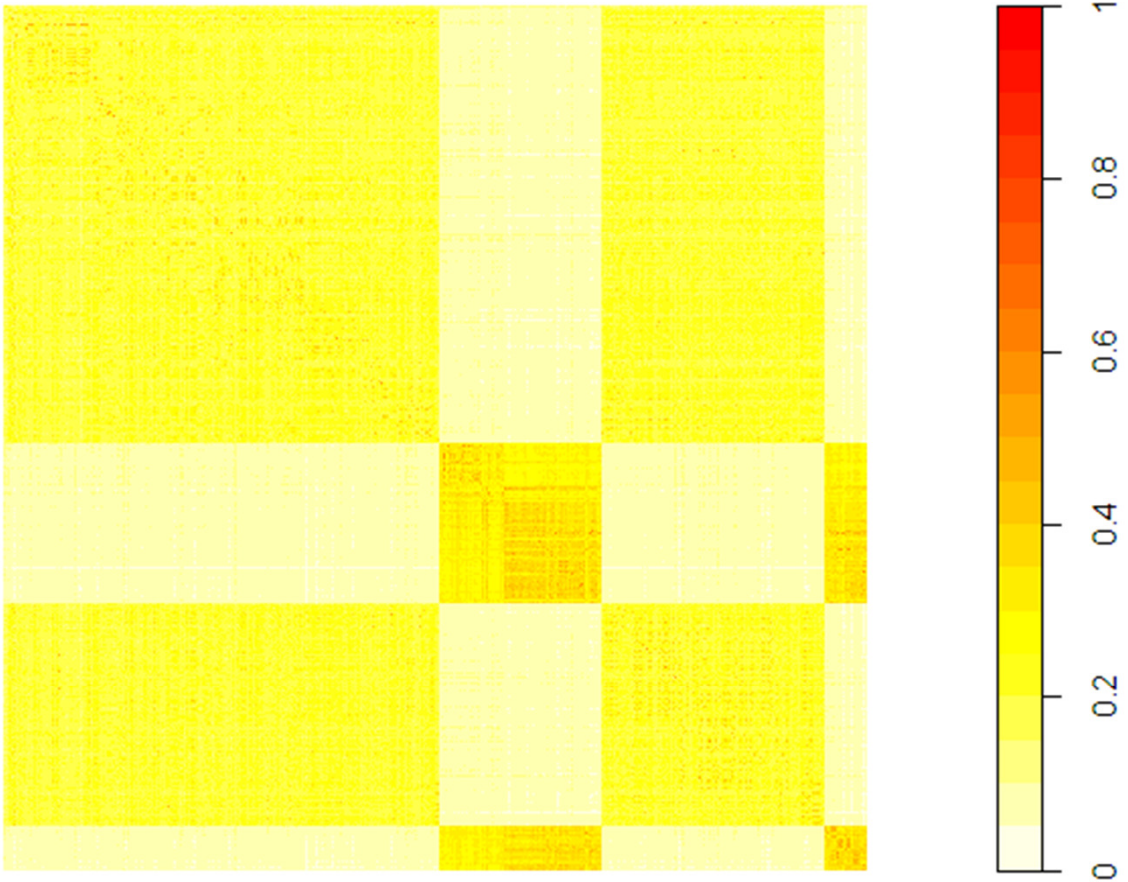

ANG RGU HOL NRC JER GNS HFD CHL LIM BSW PMT RMG NDA BMA SGT SHK BRM NEL GIR
ANG
RGU
HOL
NRC
JER
GNS
HFD
CHL
LIM
BSW
PMT
RMG
NDA
BMA
SGT
SHK
BRM
NEL
GIR
1
0.8
0.6
0.4
0.2

Holstein reference *n = 781*

Jersey reference *n = 287*

Holstein validation *n = 400*

Jersey validation *n = 77*



Genomic selection across breeds

GBLUP

Reference Set	Validation set	Protein	Fat	Milk	Prot%	Fat%
Holstein only	*Holstein*	0.53	0.48	0.64	0.66	0.67
	Jersey	-0.07	-0.02	-0.02	-0.07	0.25
Jersey only	*Holstein*	0.03	-0.01	-0.01	0.03	0.12
	Jersey	0.58	0.45	0.68	0.67	0.78
Holstein and Jersey	*Holstein*	0.53	0.49	0.64	0.66	0.67
	Jersey	0.58	0.46	0.61	0.65	0.79

Genomic selection across breeds

GBLUP

Reference Set	Validation set	Protein	Fat	Milk	Prot%	Fat%
Holstein only	*Holstein*	0.53	0.48	0.64	0.66	0.67
	Jersey	-0.07	-0.02	-0.02	-0.07	0.25
Jersey only	*Holstein*	0.03	-0.01	-0.01	0.03	0.12
	Jersey	0.58	0.45	0.68	0.67	0.78
Holstein and Jersey	*Holstein*	0.53	0.49	0.64	0.66	0.67
	Jersey	0.58	0.46	0.61	0.65	0.79

BAYESA

Reference Set	Validation set	Protein	Fat	Milk	Prot%	Fat%
Holstein only	*Holstein*	0.47	0.44	0.59	0.59	0.71
	Jersey	0.24	0.35	0.37	0.33	0.63
Jersey only	*Holstein*	0.01	0.02	-0.02	0.05	0.17
	Jersey	0.43	0.37	0.59	0.51	0.67
Holstein and Jersey	*Holstein*	0.47	0.44	0.55	0.54	0.69
	Jersey	0.47	0.51	0.58	0.67	0.82

Genomic selection across breeds

- Recently diverged breeds/lines, may be possible to use estimates of SNP effects across lines?
- More distantly related breeds, will need very dense marker maps before implementation?
- Important in multi breed populations
  - eg. beef, sheep, pigs
- Assumes same QTL mutation in both breeds

Genomic selection

- Factors affecting accuracy of genomic selection
- How often to re-estimate the chromosome segment effects?
- Genomic selection with low marker density
- Genomic selection across breeds
- Optimal breeding program design with genomic selection

Optimal breeding program design

- With genomic selection, we can potentially predict GEBV with an accuracy of 0.8 for selection candidates at birth
- How does this change the optimal breeding program design?

Optimal breeding program design

- With genomic selection, we can potentially predict GEBV with an accuracy of 0.8 for selection candidates at birth
- How does this change the optimal breeding program design?

- Breed from animals as early as possible

Optimal breeding program design

- In dairy cattle current structure is
  - Each year select a team of calves to form a progeny test team
  - At two years of age these bulls are mated to random cows from the population
  - At four years of age the daughters of the bulls start lactating

Optimal breeding program design

- In dairy cattle current structure is
  - Each year select a team of calves to form a progeny test team
  - At two years of age these bulls are mated to random cows from the population
  - At four years of age the daughters of the bulls start lactating
  - At five years of age the bulls receive a progeny test "proof" based on the performance of their daughters
  - The bulls are then selected on the basis of these proofs to be "breeding bulls"
    - Semen sold to commercial farmers

Optimal breeding program design

- In dairy cattle with genomic selection..
  - Genotype a large number of bull calves from the population
  - Calculate GEBVs for these calves
    - Accuracy = 0.8 = accuracy of progeny test
  - Select team based on GEBV
  - Sell semen from these bulls as soon as they can produce it

Optimal breeding program design

- In dairy cattle with genomic selection..
  - Genotype a large number of bull calves from the population
  - Calculate GEBVs for these calves
    - Accuracy = 0.8 = accuracy of progeny test
  - Select team based on GEBV
  - Sell semen from these bulls as soon as they can produce it
  - Generation interval reduced from ~4 yrs to ~ 2 yrs
    - $\Delta G = ir\sigma_g/L$
  - Double rate of genetic gain

Optimal breeding program design

- In dairy cattle with genomic selection..
  - Genotype a large number of bull calves from the population
  - Calculate GEBVs for these calves
    - Accuracy = 0.8 = accuracy of progeny test
  - Select team based on GEBV
  - Sell semen from these bulls as soon as they can produce it
  - Generation interval reduced from ~4 yrs to ~ 2 yrs
    - $\Delta G = ir\sigma_g/L$
  - Double rate of genetic gain
  - Save the cost of progeny testing!
    - Reduce costs by 92% (Schaeffer et al. 2006)

Optimal breeding program design

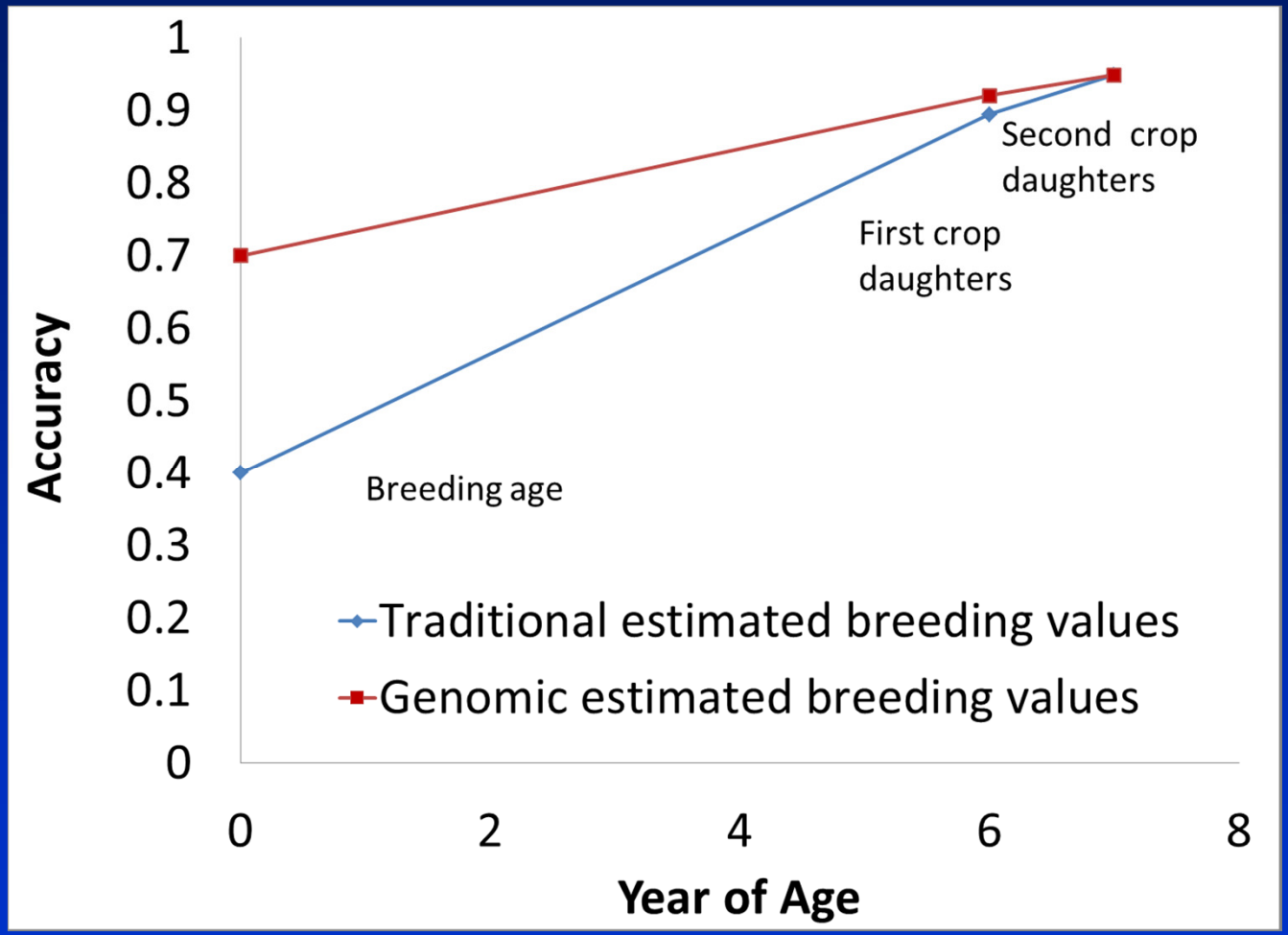
- In pigs
  - Currently EBV for traits like meat quality, sow fertility, disease resistance based on performance of relatives
  - Exploits between family variance, not within
  - Feed conversion efficiency = expensive

Optimal breeding program design

- In pigs with genomic selection
  - Accurate GEBVs for meat quality, sow fertility, disease resistance based on own marker genotype
  - Exploits between and within family variance
  - Feed conversion efficiency GEBV?
  - Will accelerate genetic gain for these traits
  - Reverse declines in meat quality for example

Genomic selection: Dairy cattle

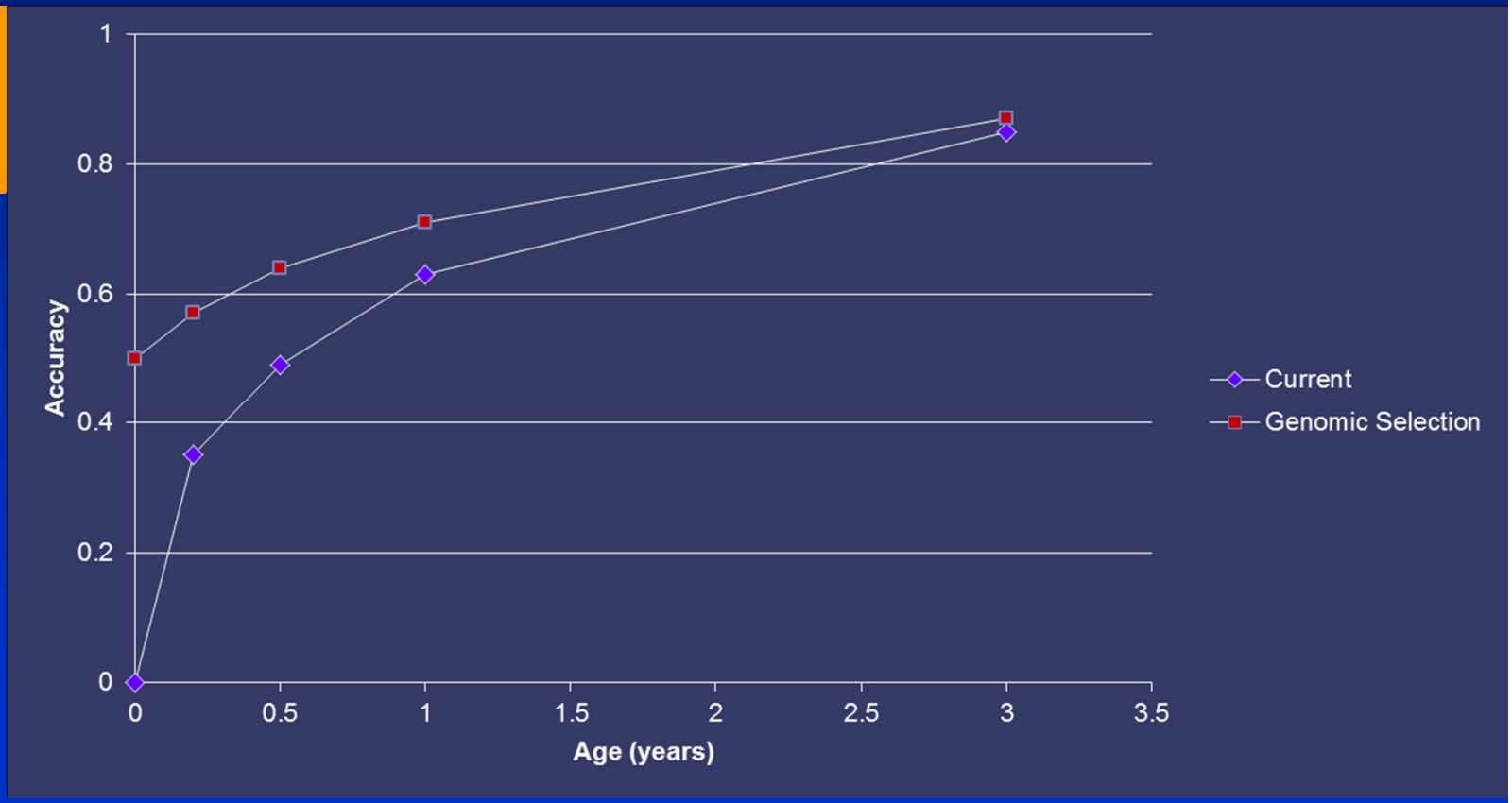
$$\Delta G = \frac{ir\sigma_g}{L}$$



Genomic selection: Meat sheep

But gains to be made by selection for breeding objective traits directly, eg. Lean meat yield vs. scanned eye muscle area

$$\Delta G = \frac{ir\sigma_g}{L}$$



Increased genetic gain from genomic selection

Industry	Potential increase
Dairy Cattle	60-120% (Pryce et al. 2011)
Meat sheep	21% (van der Werf 2011)
Wool sheep	38% (van der Werf 2011)
Beef cattle	29-158% Van Eenennaam 2011
Layers	40% (Dekkers et al 2009)
Broilers	20% (Dekkers et al. 2009)

Optimal breeding program design

- Synergy with reproductive technologies
- If we can predict genetic gain accurately at birth, genetic gain depends on generation interval
- Reproductive technologies to reduce this
  - Juvenile in-vitro embryo transfer?
  - Extreme technologies like in-vitro meosis
- Must manage inbreeding!!

Genomic selection for QTL mapping

- In association studies multiple SNPs pick up the same QTL
  - Problem with positioning QTL
- In genomic selection we fit all QTL simultaneously
- Remove effect of QTL in adjacent marker brackets/adjacent SNPs

Genomic selection

- Accuracy of genomic selection depends on
  - LD between markers and haplotyes
    - r2>=0.2 required to achieve r(GEBV,TBV) = 0.8
  - Number of records used to estimate segment effects

Genomic selection

- Higher marker densities necessary to apply genomic selection across breeds
  - Choose reference populations carefully!
- Number of generations between estimating chromosome segment effects depends on marker density
- Cost effective genomic selection possible?
- May radically alter breeding programs for some species