Predicting genetic changes with genomic information

Julius van der Werf

une
University of New England

SHEEP CRC
An Australian Government Initiative
CRC AUSTRALIA

Overview

- Genomic information can be used to achieve more genetic change – can we predict effect on genetic change, and inbreeding
- Combine genotype information with information from phenotypes and pedigree – 'additional value' of a genomic test
- How to best use genetic testing in breeding programs

Genomic breeding values

Good for:

Hard to measure, late in life traits HTML

- Lean meat yield, meat quality
- Reproductive Rate
- Adult Weight

But how does it change selection response?

- Overall
- For each trait

Potential benefits of GS - some principles

% increase in EBV accuracy (male 1yo) and genetic gain

	$h^2 = 0.1 = r^2$		$h^2 = 0.3 = r^2$	
Trait Measurability	%Δ Acc	%Δ Gain	%Δ Acc	%Δ Gain
< 1 year, both sexes	15	7	7	7
> 1 year, both sexes	68	19	59	37
>1 year, females only	119	27	112	52
on Corr. Trait, $r_g = 0.9$	20	12	20	26
on Corr. Trait, $r_g = 0.5$	67	50	76	86

→ More increase inaccuracy when there is limited information

Potential benefits of GS - some principles

% increase in EBV accuracy (male 1yo) and genetic gain

	$h^2 = 0.1 = r^2$		$h^2 = 0.3 = r^2$	
Trait Measurability	%Δ Acc	%Δ Gain	%Δ Acc	%Δ Gain
< 1 year, both sexes	15	7	7	7
> 1 year, both sexes	68	19	59	37
>1 year, females only	119	27	112	52
on Corr. Trait, $r_g = 0.9$	20	12	20	26
on Corr. Trait, $r_g = 0.5$	67	50	76	86

Note that: gain = (accuracy * selection intensity/generation interval) * genetic SD

These effects even higher in selected populations (need to account for Bulmer effect)

How is additional response calculated?

- Selection index theory
  - Index weights for various information sources
  - Accuracies of EBV and GEBV -overall index and per trait-
  - Response -overall index and per trait-
    - Some traits benefit more than other from GS

Some definitions

Information used to estimate it

- TBV — True breeding value
- EBV — Estimated breeding value — Phenotype + Pedigree
- GBV — Genomically estimated BV — DNA test
- GEBV — Genomically enhanced estimated BV — Phenotype + Pedigree + DNA test

Selection Index Approach

Genomic Selection: Predict TBV with Accuracy = x

→ GS explains x^2 % of V_A → $V_{qtl} = x^2 V_A$

Lande and Thompson, 1990 Genetics:

$V_{polygenic} = (1-x^2)V_A$

Models

Without GS: $V_{Pheno} = V_{AddGen} + V_{error}$

With GS: $V_{Pheno} = V_{polygenic} + V_{qtl} + V_{error}$

Predict from phenotypes pedigree

Predict from DNA markers

Selection Index *some formal stuff*

for reference only

First summarize some definitions

$I = b_1x_1 + b_2x_2 + \ldots\ldots + b_nx_n$

$var(X)=P$

Index:
X_i = selection criteria
b_i = index weight

Single trait breeding objective:

$H = A$ (breeding value)

Breeding Objective

$Cov(X,A) = G$ (a vector with ST objective)

Optimal weights are $b = P^{-1}G$

Calculate optimal index weights given the breeding objective

$Var(I) = var(b'X) = b'var(X)b = b'Pb = \sigma_I^2$

$Var(H) = \sigma_a^2$

Variance of index = covariance between index and breeding objective

$Cov(I,H) = cov(b'X,A) = b'cov(X,A) = b'G = b'Pb = \sigma_I^2$

$b = P^{-1}G \rightarrow Pb = G$

Accuracy of selection index (single trait)

r_{IA} = correlation between Index (=EBV) and A

$$= \frac{cov(I,A)}{\sigma_I \sigma_A} = \frac{\sigma^2_I}{\sigma_I \sigma_A} = \sigma_I / \sigma_A = \sqrt{(b'Pb/ \sigma_a^2)}$$

Because cov(I,A) = var(I)

Selection Index = Best Linear Prediction BLP

Index (I) is best estimate of breeding value: I = E(A|X) = cov(X,A)/var(X)

Same as BLUP, but without fixed effects.

$Var(I) = var(EBV) = r_{IA}^2 .var(BV) = r_{IA}^2 \sigma_a^2$

r_{IA}^2 also known as reliability

Selection Index Approach with genomic info

use info on various information sources: *below for one trait only*

Variance-covariance of information sources
P-matrix

covariance with TBV
G-matrix

Without GS:

Own perf.	Vp						Va
Sire		Vp					Va/2
Dam			Vp				Va/2
FullSibs	etc			{t-(1-t)/n}Vp			Va/2
HalfSibs					{t-(1-t)/n}Vp		Va/4
Progeny						{t-(1-t)/n}Vp	Va/2

With GS:

Own perf.	Vp-Vq					0	Va-Vq
Sire		Vp-Vq				0	(Va-Vq)/2
Dam			Vp-Vq			0	(Va-Vq)/2
FullSibs	etc			{t-(1-t)/n}(Vp-Vq)		0	(Va-Vq)/2
HalfSibs					{t-(1-t)/n}(Vp-Vq)	0	(Va-Vq)/4
Progeny						{t-(1-t)/n}(Vp-Vq) 0	(Va-Vq)/2
QTL	0	0	0	0	0	0 Vq	Vq

Selection Index Approach with genomic info

Pseudo BLUP: Genomic Breeding value is an additional trait with $h^2 = 1$

Without GS:

Own perf.	Vp						Va
Sire		Vp					Va/2
Dam			Vp				Va/2
FullSibs	etc			{t+(1-t)/n}Vp			Va/2
HalfSibs					{t+(1-t)/n}Vp		Va/4
Progeny						{t+(1-t)/n}Vp	Va/2

With GS:

Own perf.	Vp						Vq	Va
Sire		Vp					Vq/2	(Va)/2
Dam			Vp				Vq/2	(Va)/2
FullSibs	etc			{t+(1-t)/n}(Vp)			Vq/2	(Va)/2
HalfSibs					{t+(1-t)/n}(Vp)		Vq/4	(Va)/4
Progeny						{t+(1-t)/n}(Vp)	Vq/2	(Va)/2
QTL	Vq	Vq/2	Vq/2	Vq/2	Vq/4	Vq/2	Vq	Vq

P-matrix G-matrix

Selection index: example of 2 approaches

own phenotype + GBV

$h^2 = 0.5$

GBV reliability "x^2" = 0.5

	P	G (,BV)		b	varIndex	acc
Phenotype	1	0.5	a	0.5000	0.2500	0.7071

	P		G				
Phenotype	1	0.25	0.5	a	0.3333	0.3333	0.8165
mBV	0.25	0.25	0.25	q	0.6667		

	P		G				
Corrected Phenotype	0.75	0	0.25	u	0.3333	0.3333	0.8165
mBV	0	0.25	0.25	q	1.0000		

Note weights on QTL info

MBV = GBV = "QTL"

Selection index: example of 2 approaches

- information from relatives

	P			G	b	VarIndex	accuracy
ownPoly	0.75	0	0.125	0.25	0.3143	0.3429	0.8281
Own GBV	0	0.25	0	0.25	1.0000		
sirepoly	0.125	0	0.75	0.125	0.1143		

Selection index: example of 2 approaches

- information from relatives

	P				G	b	VarIndex	accuracy
ownPoly	0.75	0	0.125	0	0.25	0.3143	0.3429	0.8281
ownGBV	0	0.25	0	0.125	0.25	1		
sirepoly	0.125	0	0.75	0	0.125	0.1143		
sireMBV	0	0.125	0	0.25	0.125	0		

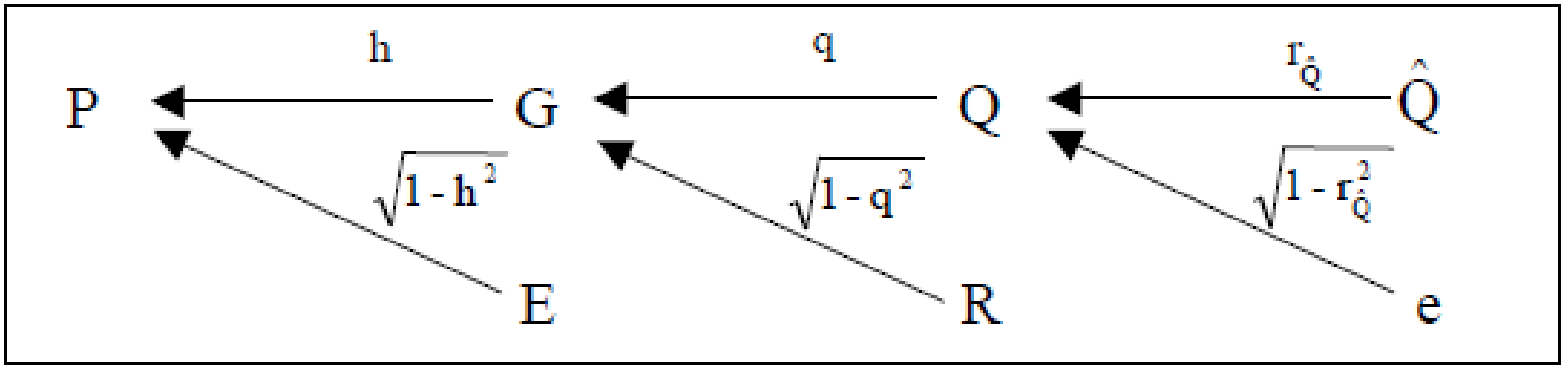
	P			G	b	varIndex	acc
ownPoly	0.75	0	0.125	0.25	0.3182	0.3409	0.8257
ownGBV	0	0.25	0.125	0.25	0.9545		
Sirepheno	0.125	0.125	1	0.25	0.0909		

Top: sire is genotyped, bottom: sire not genotyped, not same accuracy

Conclusion: Relatives info needs to be ‘corrected for markers’

Path coefficient method following

Dekkers Dec 2007 JABG



P = Phenotype

G = Breeding Value

Q = BV component associated with markers (=GBV)

Q_{hat} = estimate of Q

Accuracy GBV = "x" = $q.r_{Qhat}$

Phenotypic correlation: $r_{P,Qhat} = h.x$

Genetic correlation $r_{G,Qhat} = x$

Conclusion: single trait

- Can include GBV as a correlated trait
  - And use standard software for selection index

- r_g = accuracy, same as 'x'
- r_p = h.x
- econ value for GBV = 0

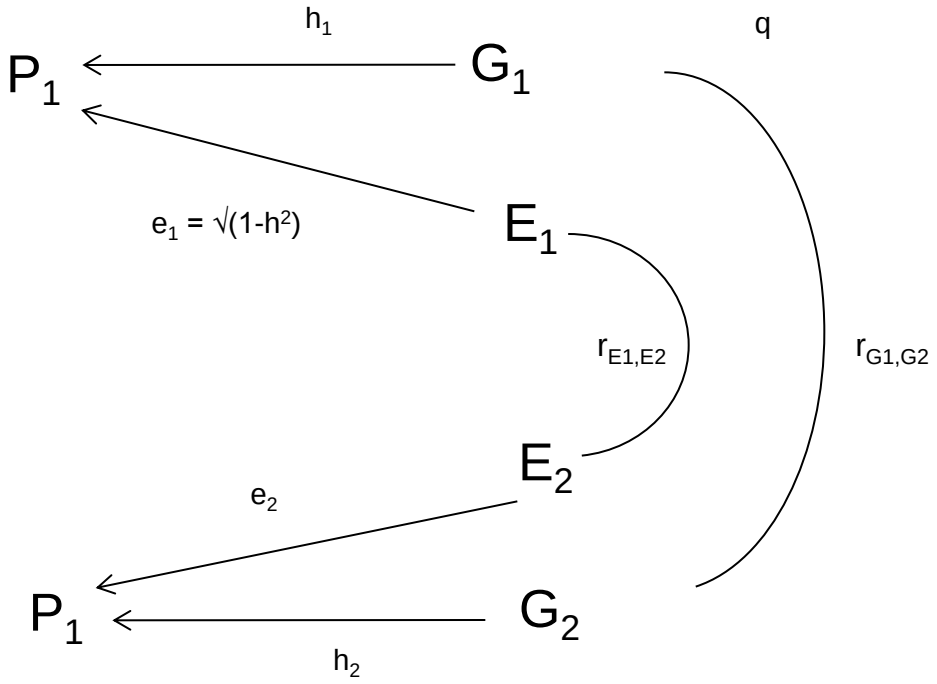
- This is equivalent to treating it as an extra info source in a single trait multiple info sources approach:

  EBV = f(own perf, dam, sire, sibs, progeny, GBV)

Extension to multiple traits

- Some traits may have GBV, others may not
- Need correlations....
  - between GBV and other trait phenotypes
  - between GBV and other trait genotypes
  - between different GBVs
  - These can be predicted from genetic correlations between traits, only when assuming infinitesimal model

Path coefficient method *Dekkers Dec 2007 JABG*



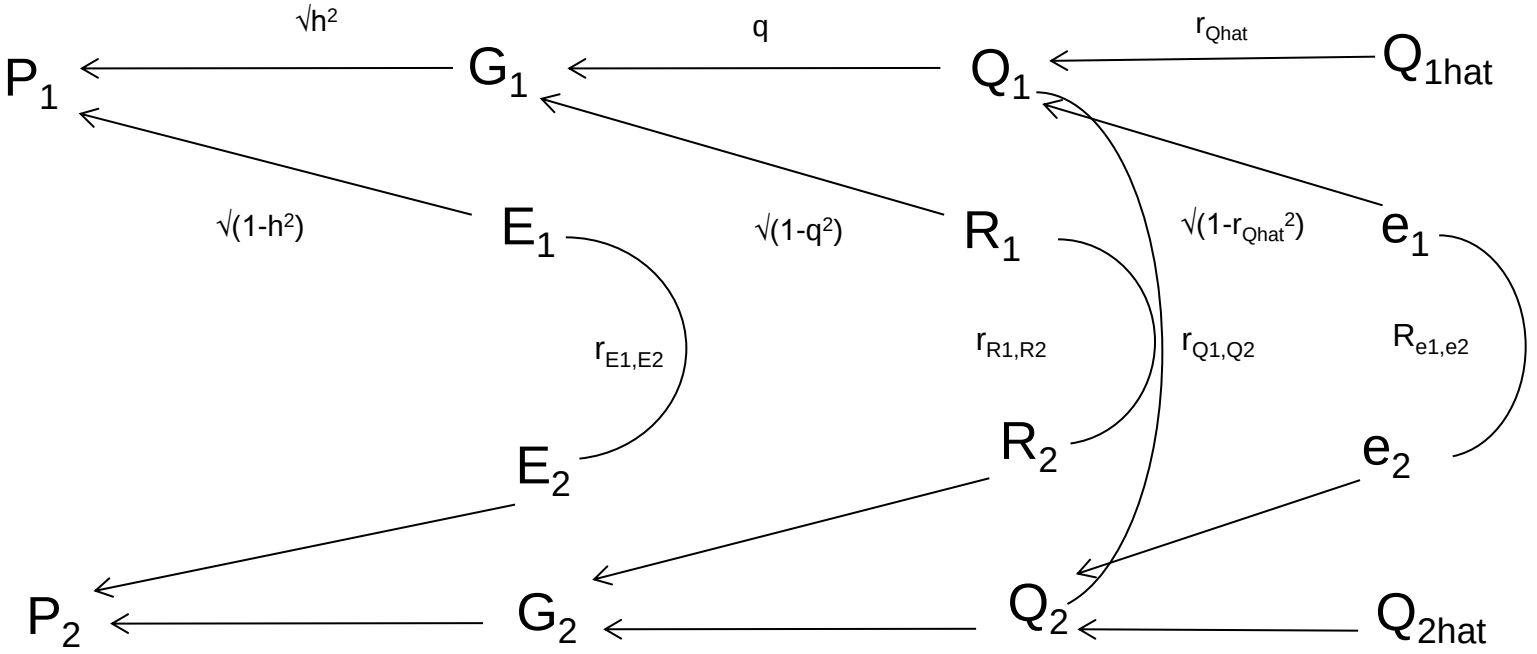
$r_{P1G2} = h_1 r_{G1,G2}$

$r_{p1p1} = h_1\ h_2\ r_{G1,G2} + e_1\ e_2\ r_{E1,E2}$

$e_i = \mathrm{sqrt}(1-h^2_i)$

Path coefficient method following

Dekkers Dec 2007 JABG



$r_{Gi,Qhatj} = r_{Q1hat}.r_{Q1,Q2}$

$r_{Pi,Qhatj} = h_i r_{Q1hat}.r_{Q1,Q2}$

$r_{Qhati\ ,Qhatj} = r_{Q1hat}.r_{Q2hat}.r_{Q1,Q2}$

Summary

To predict accuracy of GEBV we can use selection index approach

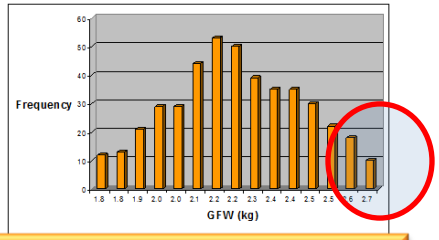
- Either: GBV + polygenic (no correlation)
- Or: GBV + P, correlation is r^2
- The latter is easier: Genomic BV as a correlated trait.

r_g = accuracy of GBV = 'x'
r_p = h.x
Econ value of GBV = 0

How much genetic change?

Breeder's Equation

$$\text{Selection Response} = i\, r_{IA} \sigma_A$$



Selection intensity

Selection accuracy

Genetic variation

How much genetic change?

Selection Differential

$$S = i\ r_{IA}\sigma_A$$

Mean of selected parents

Response per year

- Response per generation

$$S = i\, r_{IA} \sigma_A$$

$$\tfrac{1}{2} S_{sires} + \tfrac{1}{2} S_{dams}$$

Superiority of parents averaged over males and females

- Response per year

$$\frac{\tfrac{1}{2} S_{sires} + \tfrac{1}{2} S_{dams}}{\tfrac{1}{2} L_{sires} + \tfrac{1}{2} L_{dams}} = \frac{S_{sires} + S_{dams}}{L_{sires} + L_{dams}}$$

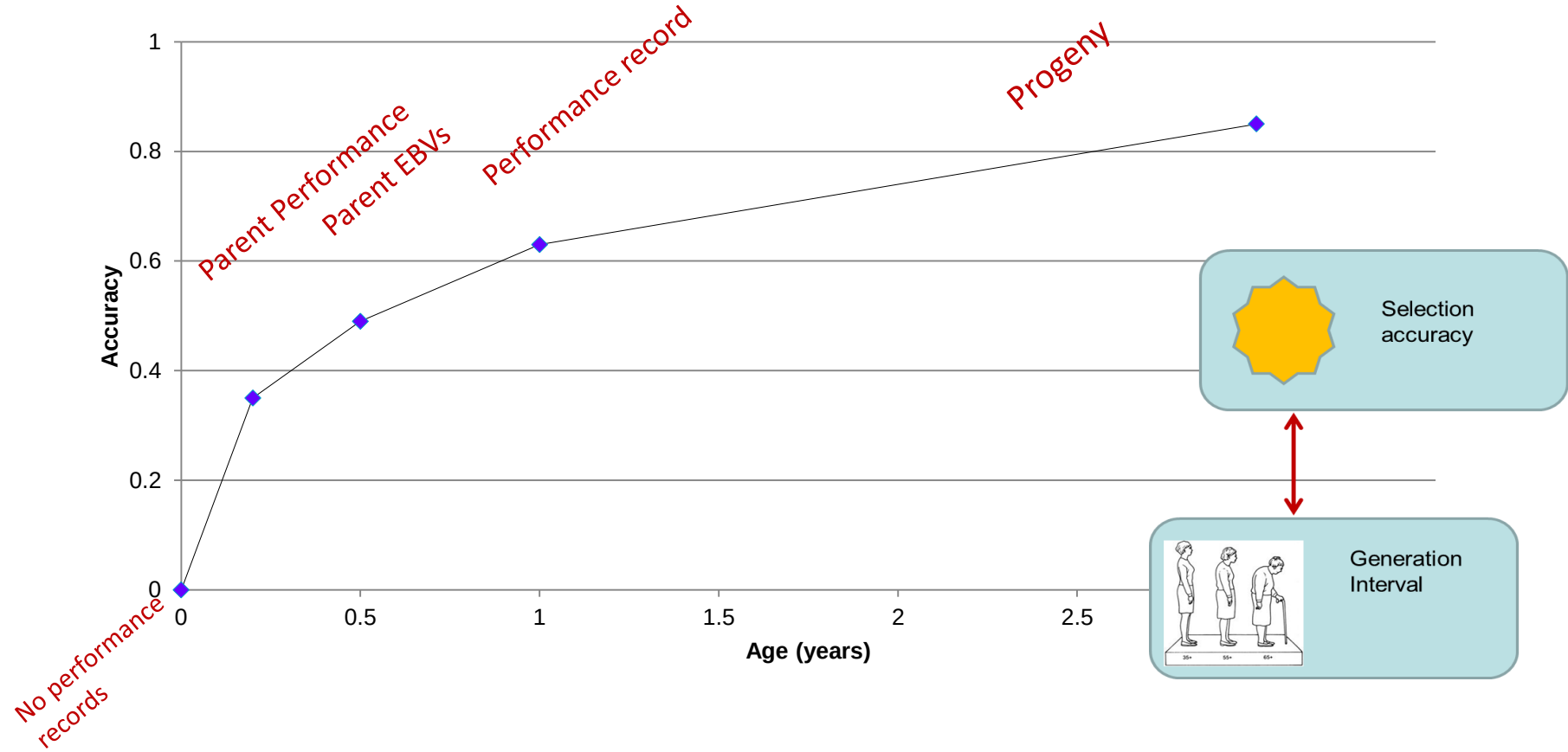
Generation interval (in years) averaged over males and females

Generation interval is average age of sires (dams) when their progeny are born

$$R = \frac{i_m r_m + i_f r_f}{L_m + L_f} \sigma_A$$

Accuracy of predicting a breeding value (r)

- increases as an animal gets older

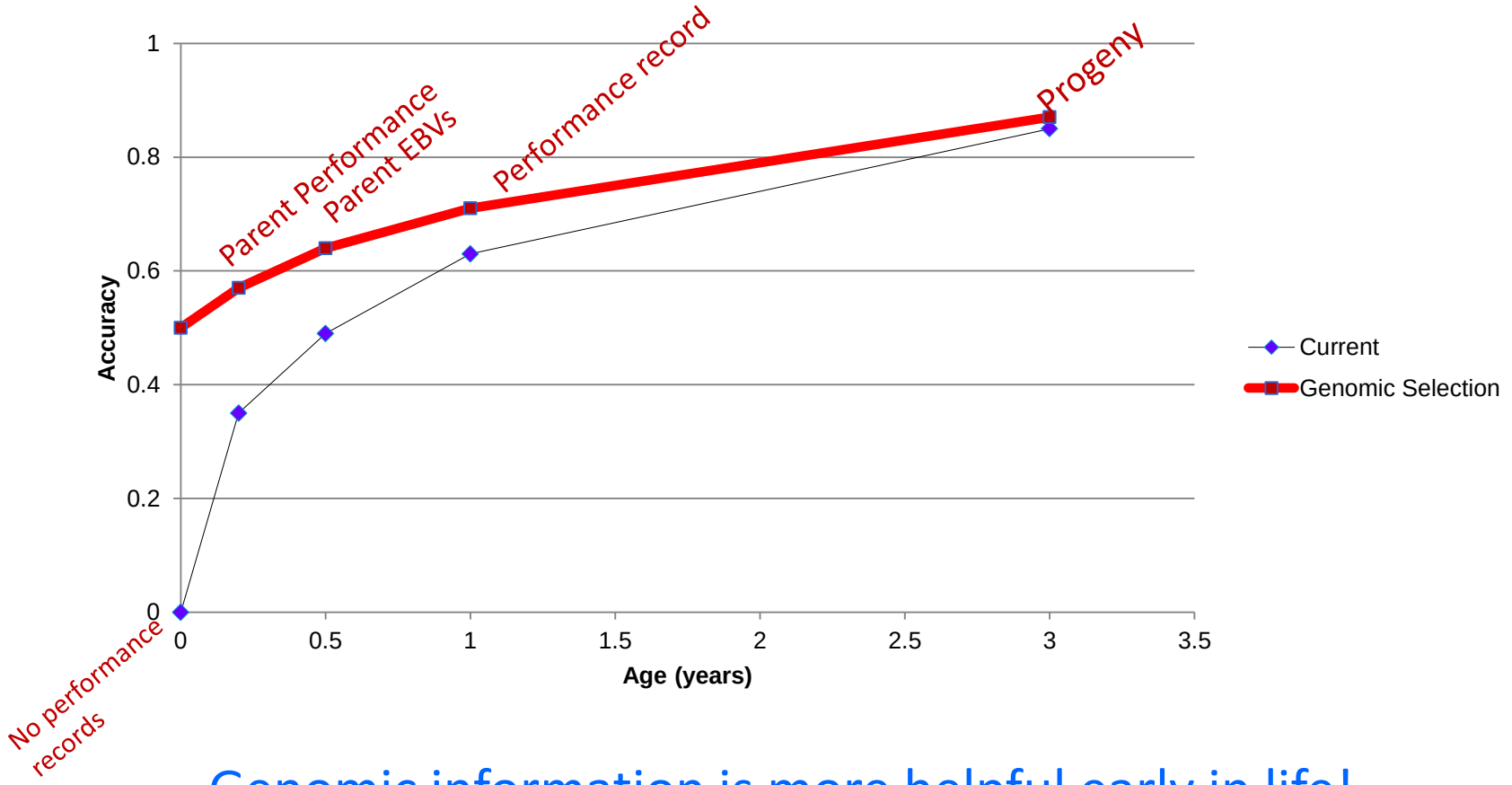


Assumed heritability = 25%

Need to balance accuracy and generation interval!

Accuracy of predicting a breeding value

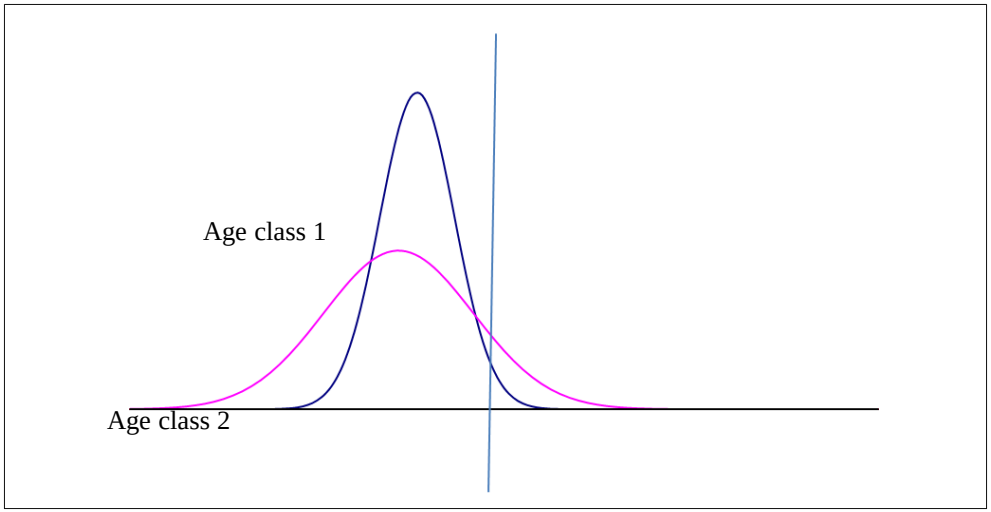
- increases as an animal gets older -

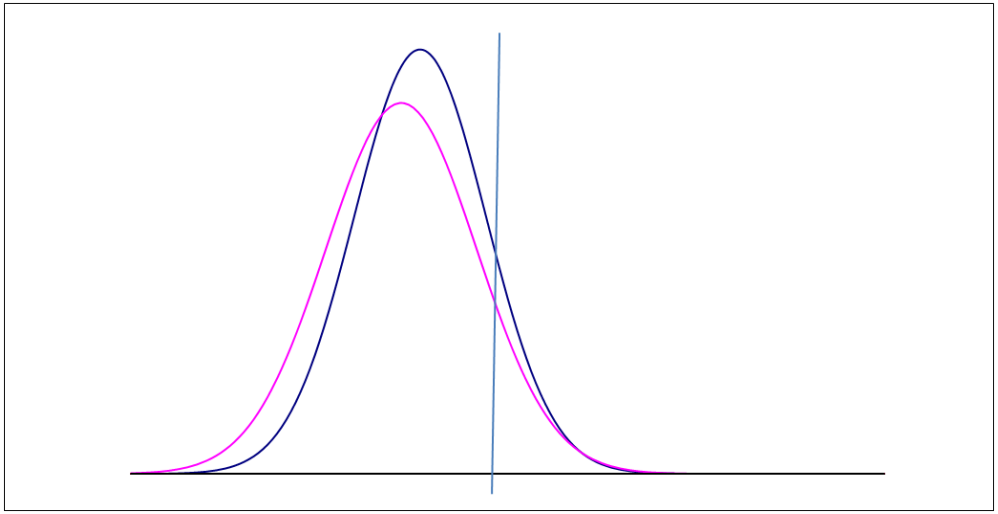


Genomic information is more helpful early in life!

Assumed heritability = 25%; Accuracy of genomic test = 50%

Optimizing age structure





Accuracy changes with age class !

Without genomic selection

ageclass	N in group	mean	SD	Nr Selected
1	50	10.20	0.4	2.7
2	50	10.00	0.8	7.3

Accuracy

With genomic selection

ageclass	N in group	mean	SD	Nr Selected
1	50	10.20	0.7	5.4
2	50	10.00	0.8	4.6

Potential benefits of GS - some cases

% increase in EBV accuracy (male 1yo) and genetic gain

	$h^2 = 0.1 = r^2$		$h^2 = 0.3 = r^2$	
Trait Measurability	%Δ Acc	%Δ Gain	%Δ Acc	%Δ Gain
< 1 year, both sexes	15	7	7	7
> 1 year, both sexes	68	19	59	37
>1 year, females only	119	27	112	52
on Corr. Trait, $r_g = 0.9$	20	12	20	26
on Corr. Trait, $r_g = 0.5$	67	50	76	86

These effects underestimated due to not accounting for Bulmer effect

Effect of GS on genetic change

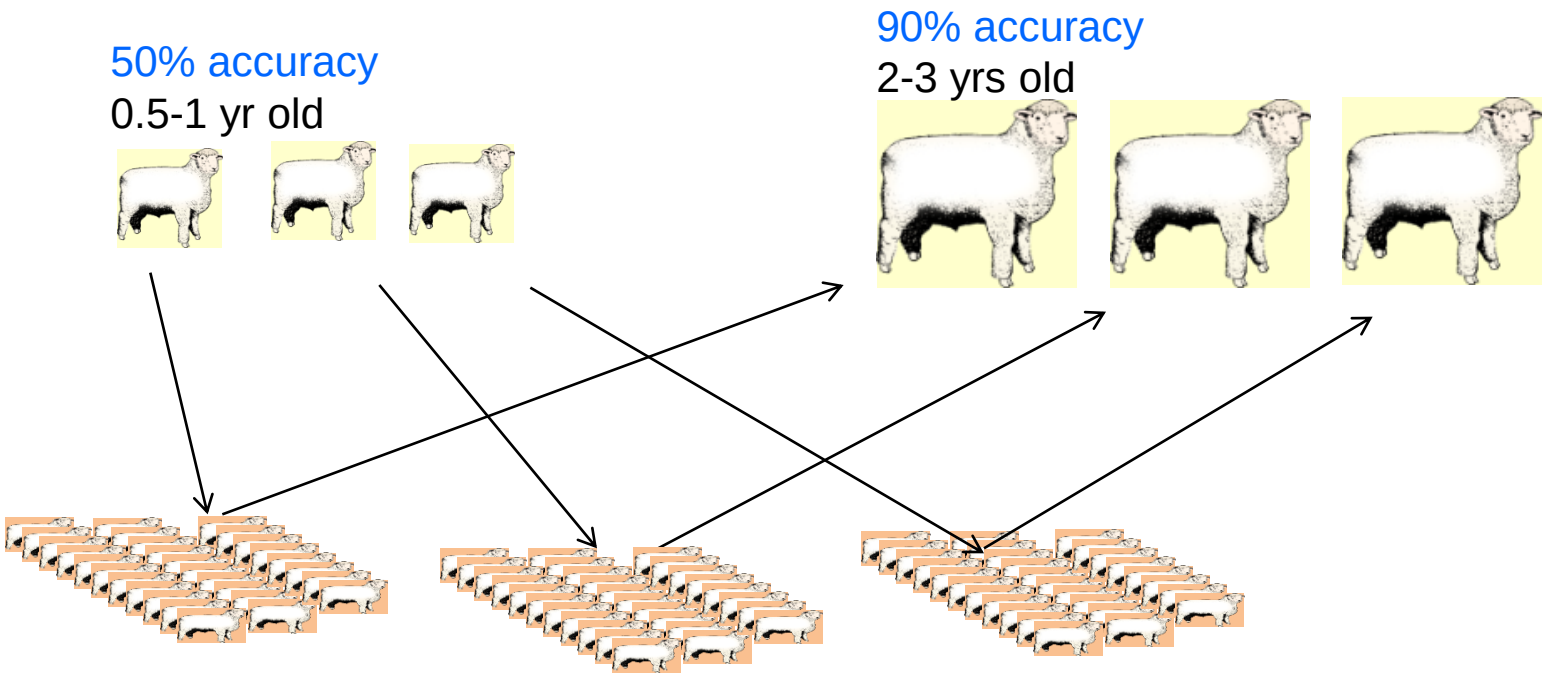
- More accuracy of GEBV,
  - esp younger animals
  - Depends on trait measurability (early/late trat, sex limited)
- More response due to higher accuracy
- ……Or lower generation interval

- E.g. dairy: accuracy of GBV is lower than of progeny test, but generation interval can be much reduced

Benefits across Species

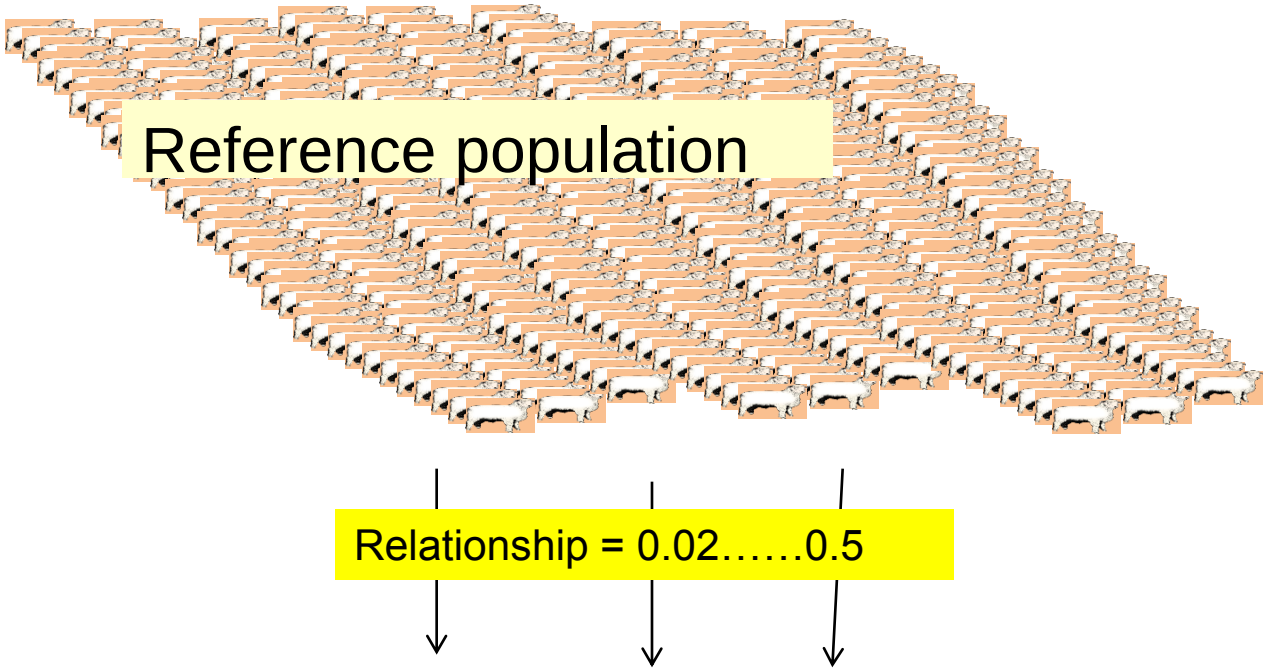
	% extra gain	impact
• Early trait	small	small
		accuracy/ gen int
• Late Trait	moderate	gen int/acc
• Sex limited trait		
– females only, late	very large	gen int
– Males only early	small to modest	acc/gen int

Compare: Progeny Testing

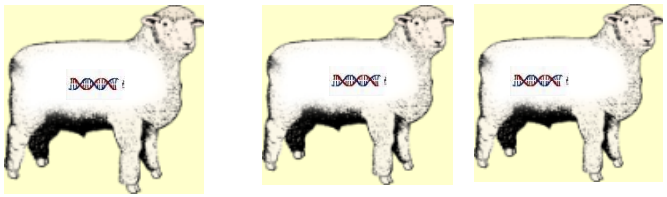


Each progeny group only informs one sire

Genomic Testing



70% accuracy
0.5-1 yrs old

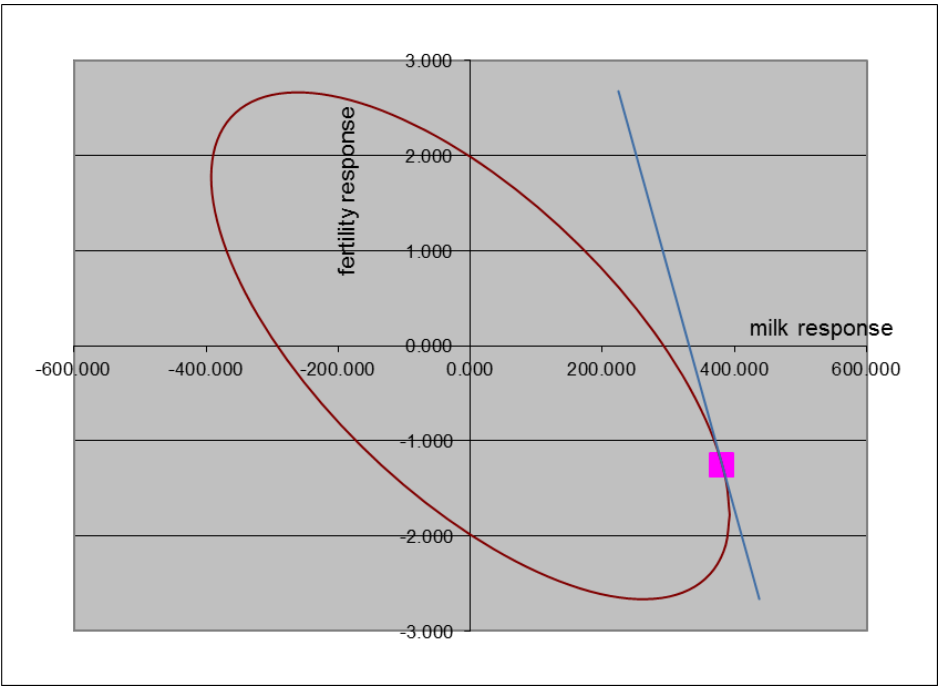


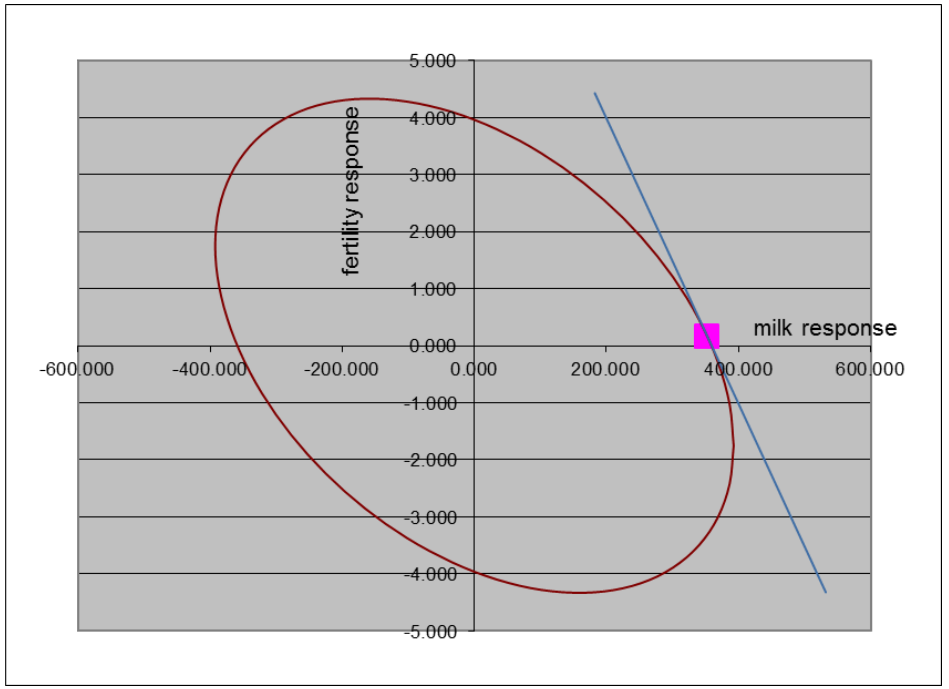
One large reference population informs all young rams

EBV accuracy increased at young age

Genomic Selection also changes the balance between traits

Selection for milk Yield and Fertility





	economic weights		*progeny measured*		*response (4 yrs)*	
	milk	*fertility*	*milk*	*fertility*	*milk*	*feed*
left	0.2	**8**	50	10	381	-1.25
right	0.2	**8**	50	50	352	0.17

Shifting the *trait* balance with genomic selection

	Current Selection	
	Accuracy	Response
Weight kg	0.71	0.79
Dressing %	0.26	0.23
Saleable meat yield %	0.33	0.29
Overall Merit $Index	0.58	2.03

Shifting the *trait* balance with genomic selection

	Current Selection		Genomic Selection		Difference
	Accuracy	Response	Accuracy	Response	
Weight kg	0.71	0.79	0.75	0.76	-4%
Dressing %	0.26	0.23	0.59	0.42	83%
Saleable meat yield %	0.33	0.29	0.60	0.46	59%
Overall Merit $Index	0.58	2.03	0.69	2.43	20%

Note: not only more gain overall, but shift to HTML traits

Effect of GS on genetic change

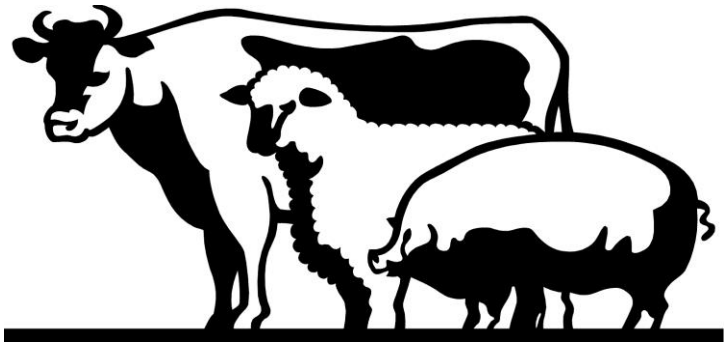
With multiple trait selection:

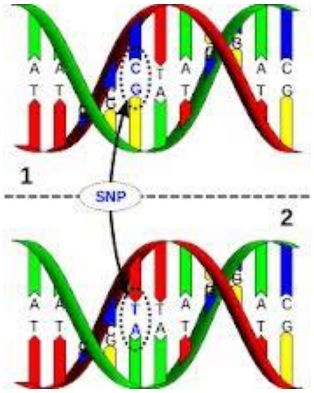
The accuracy of hard to measure trait will improve more, and as a result, these traits get more pushed,

possibly at the expense of the easy to measure traits

(but these were ‘overemphasized’ response before GS)

Genomic information and inbreeding



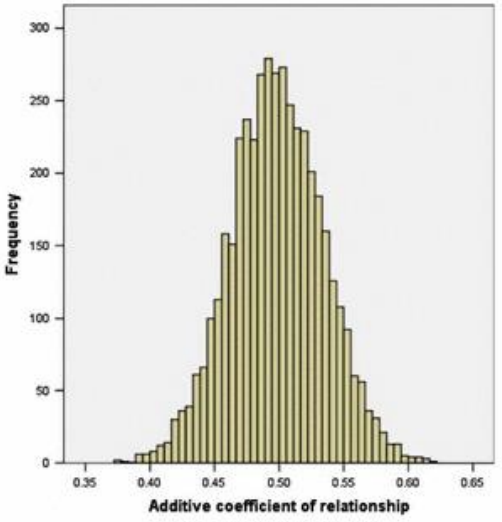


Relationships between Individuals

- Estimated using:
  - Expected relatedness from PEDIGREE
  - ‘Observed’ relatedness from proportion of genome shared
    - Thousands of genetic markers (SNPs)

Estimates of relationship using genotypes:

- The expectations A
- Replaced by the estimated G
  - Genomic relationship matrix



  - Still half mum and half dad
    - But which half?
    - Variation around the expectation?

There is variation in actual relationship, e.g. 0.40-0.50 in FS

We can see this with genomic relationships

What information is used in BVs?

Clark et al, 2013 GSE

- Va= ¼ sire + ¼ dam + ½ MS

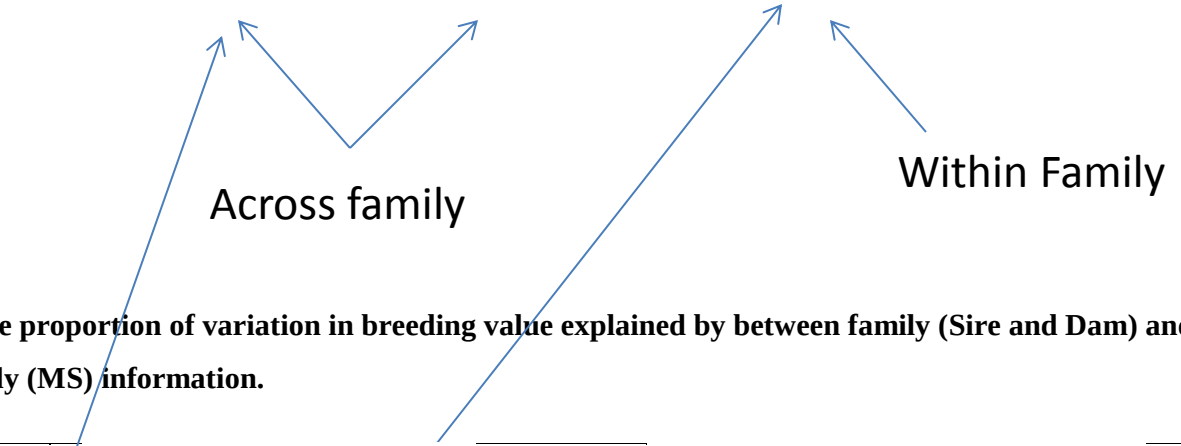


Table 2- The proportion of variation in breeding value explained by between family (Sire and Dam) and within family (MS) information.

	NZ dairy bulls					Australian dairy bulls				
	BV	**Sire**	**Dam**	**MS+e**	***Prop. of PT***	**BV**	**Sire**	**Dam**	**MS+e**	***Prop. of PT***
Parent Average		0.56	0.44	0.001	0.001	**PA EBV**	0.44	0.52	0.04	0.05
Genomic BV		0.43	0.26	0.31	0.56	**GEBV**	0.33	0.37	0.30	0.36
Progeny Test		0.21	0.31	0.48	1.0	**PT**	0.16	0.32	0.52	1.0

Correlation of breeding values and co-selection of relatives

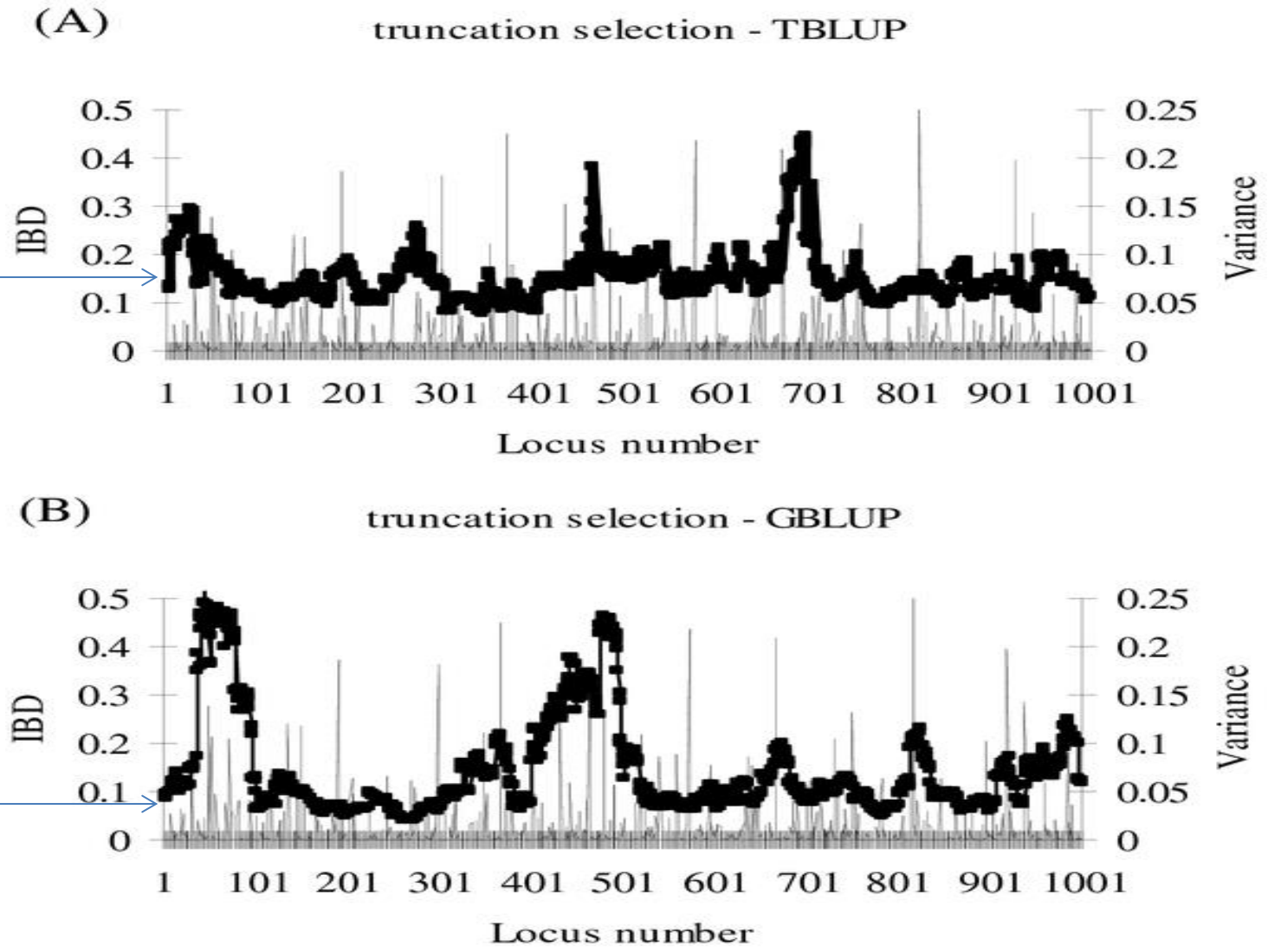
Breeding value type	Half sib correlation	Full Sib correlation	Accuracy
Parent Average	0.55	1.0	0.45
Genomic BV	0.50	0.85	0.57
Progeny Test	0.26	0.53	1.0

Full Sibs
- share the same Parent average BV (½ sire ½ dam)
- no longer the case with genomics

Half Sibs
- Share different PA breeding values
- Small advantage of using G to restrict inbreeding

Truncation selection on breeding values estimated using TBLUP or GBLUP

(Sonesson, Woolliams, Meuwissen, 2012)



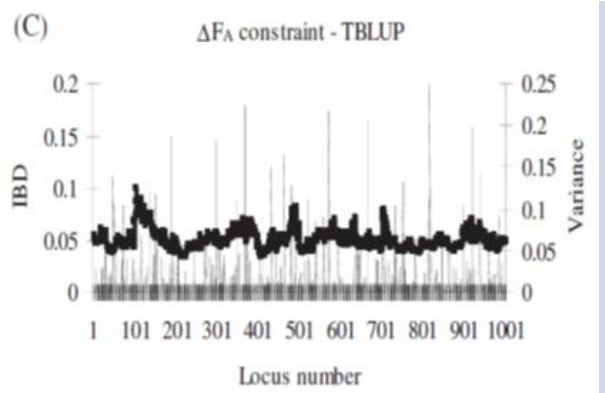
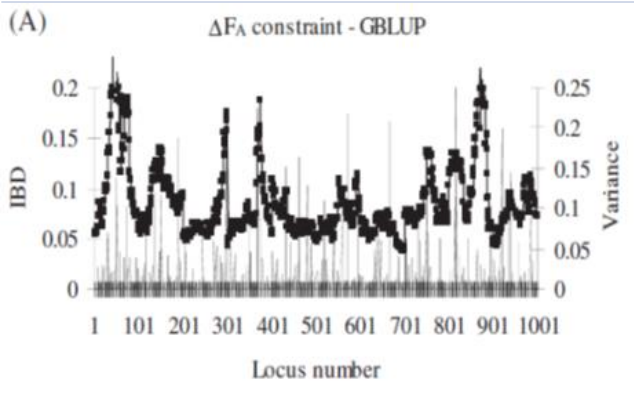
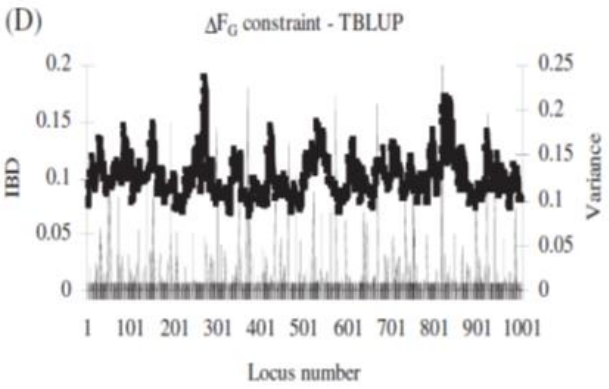
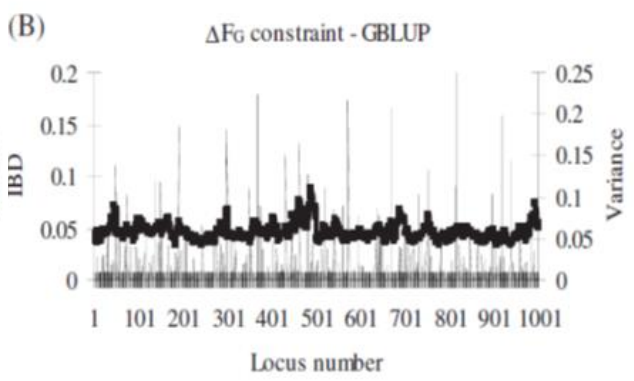
Higher →

Lower →

---- Selecting 100 sires and 100 dams from 3000 cand. ---- After 10 generations

Constraining Inbreeding:

Pedigree or Genomics, Optimal contributions

Selection on Constraint	BLUP	GBLUP
ΔF_A		
ΔF_G		

Genomic selection and inbreeding

- Effect on IBD and variance at loci is different between GBLUP and BLUP
- GBLUP could give more ‘local inbreeding’
- Can constrain inbreeding: better use G when applying GBLUP and use A when applying BLUP

- When constraining with G, can exploit some more variation within family (useful for large FS families)