

# Genomic Selection for Crossbred Performance in Livestock

Jack Dekkers

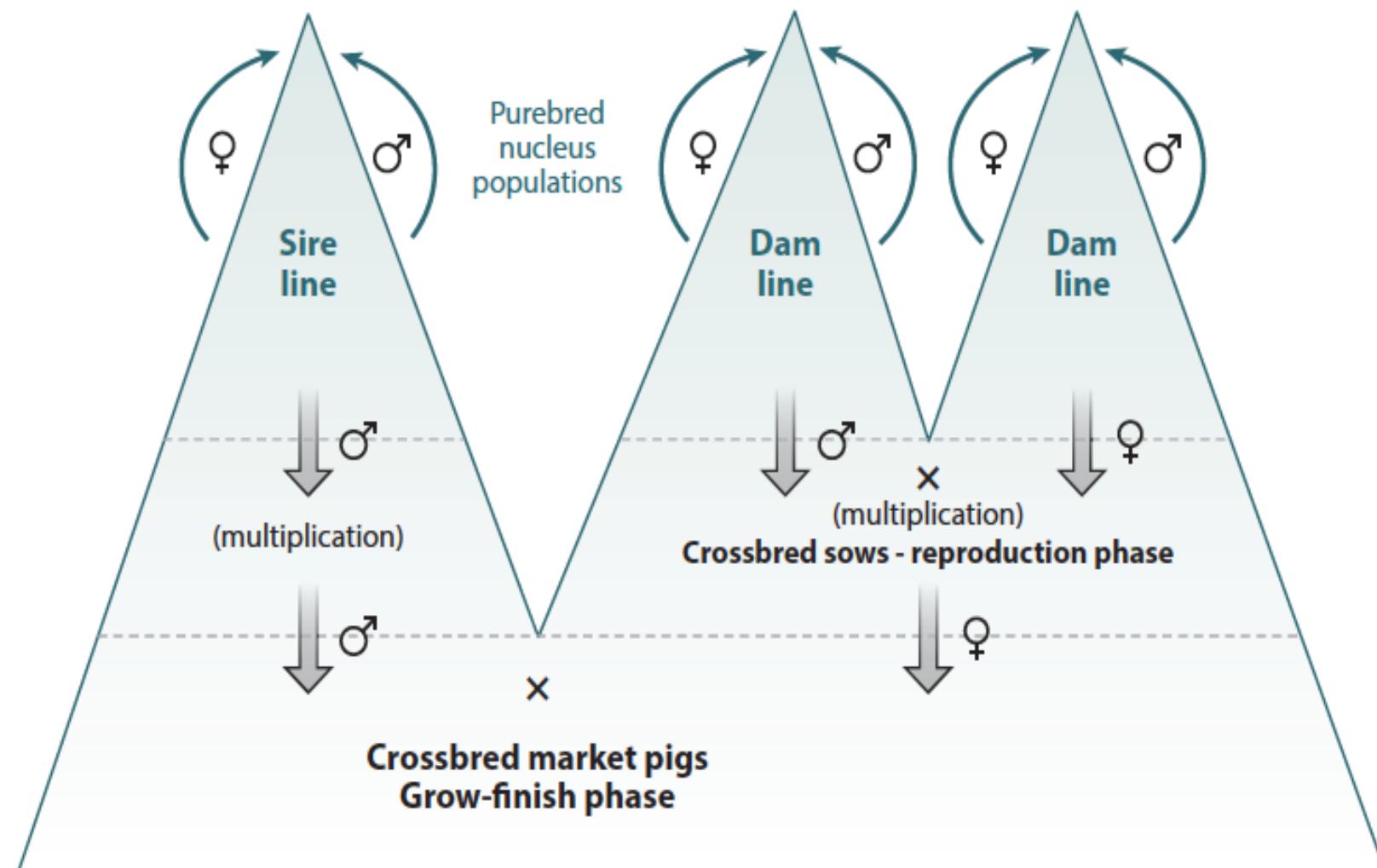
Animal Breeding & Genetics  
Department of Animal Science  
Iowa State University

IOWA STATE  
UNIVERSITY



ANIMAL  
SCIENCE

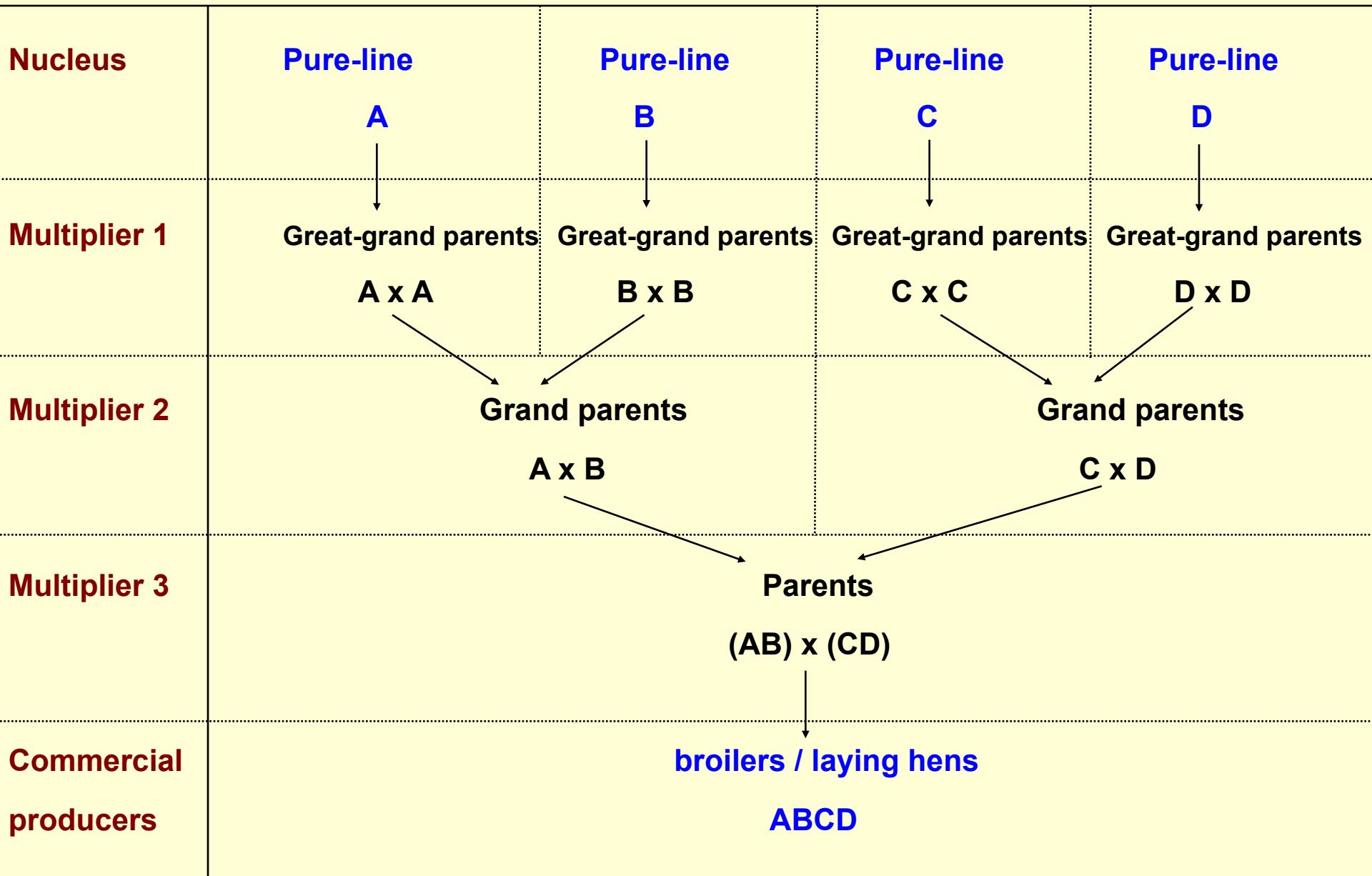
# Typical Breeding Pyramid for Pigs



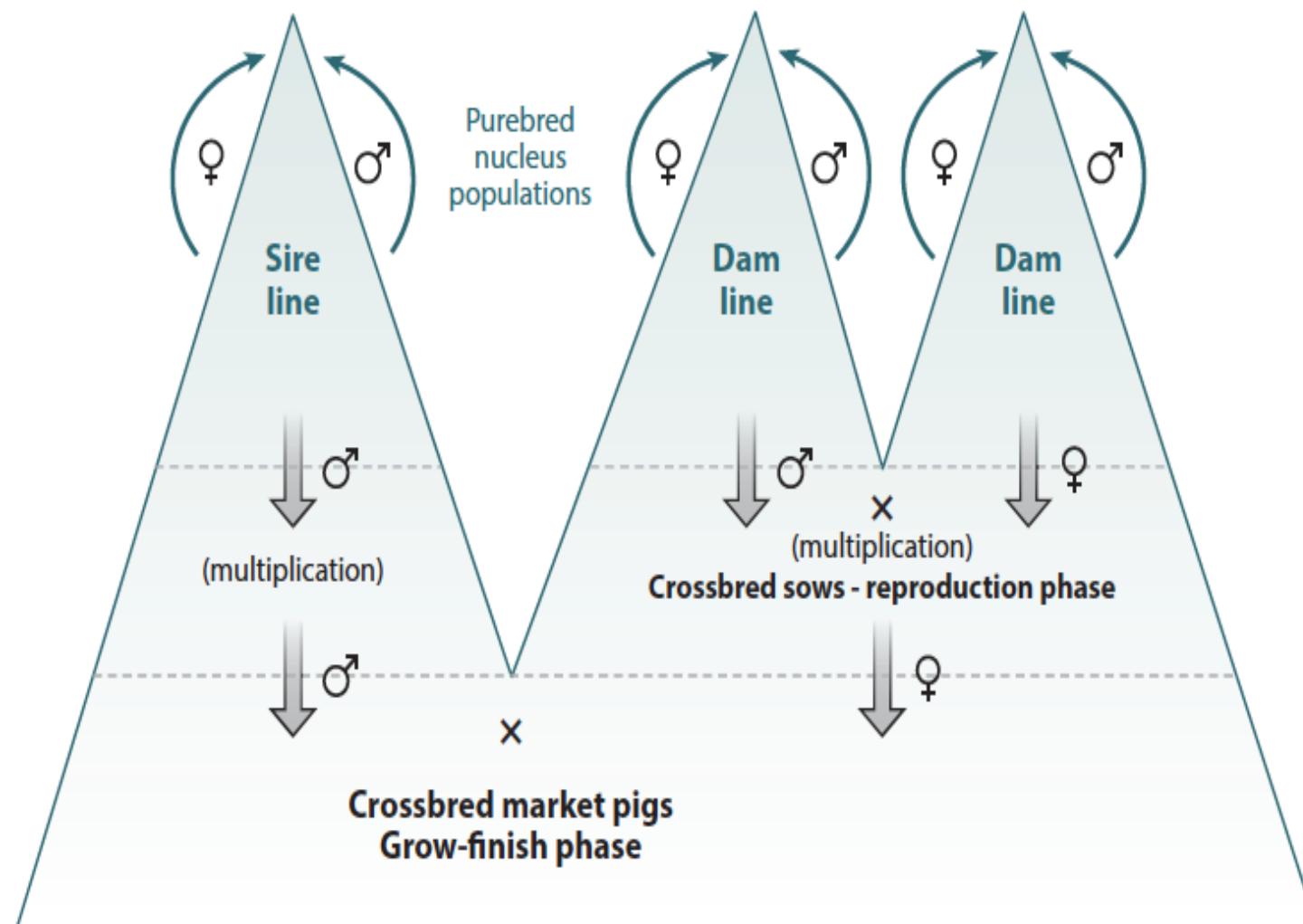
## Reasons for crossbreeding

- Heterosis
- Breed complementarity
- Protecting  $\Delta G$

# General structure of poultry breeding programs



# Typical Breeding Pyramid for Pigs



**Selection on Purebred Performance in High Health Environment**

$$r_{pc} < 1$$

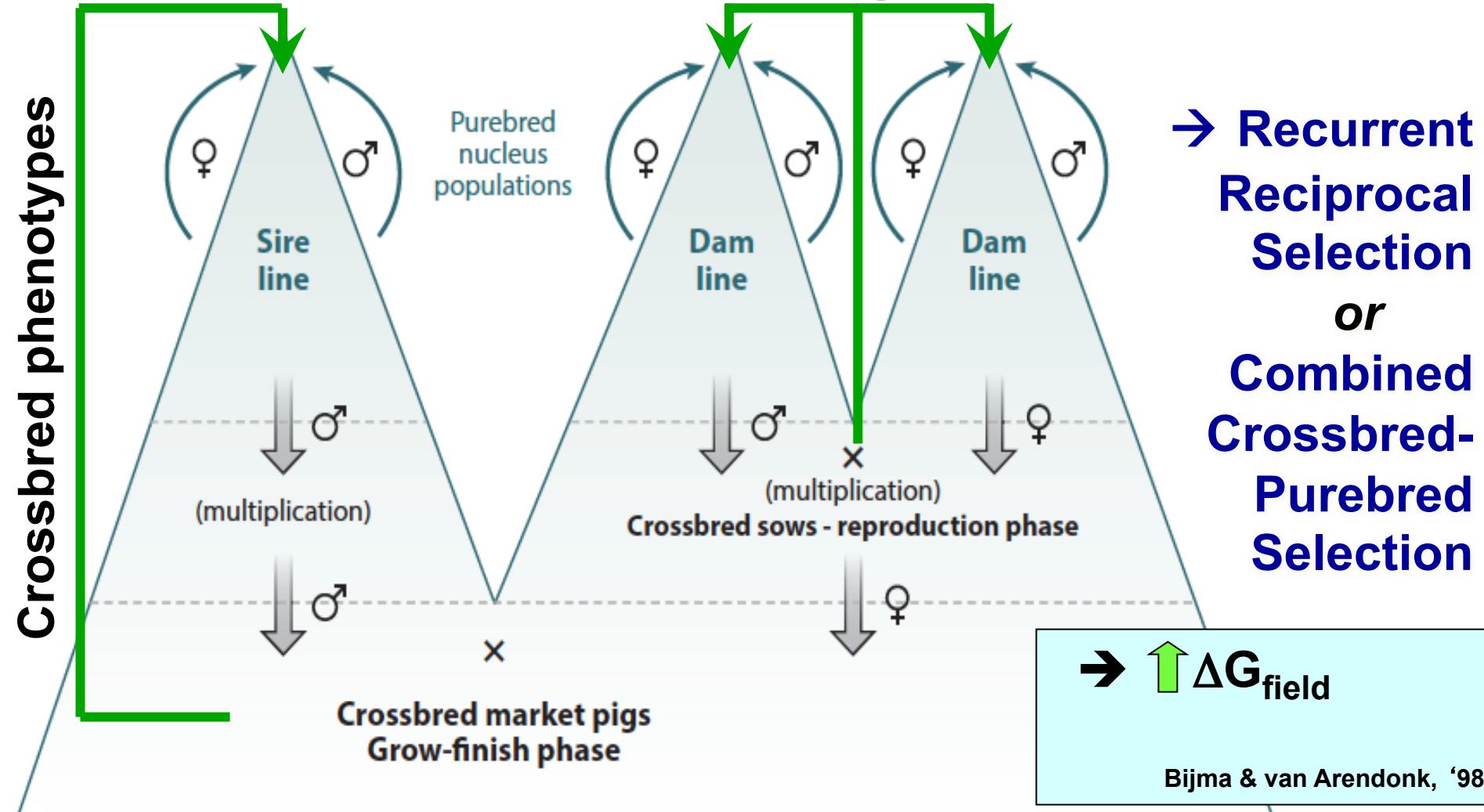
**Crossbred Performance in Field Environment**

## Limitations of purebred selection programs

- limited selection for CB performance in the field
- no selection for traits not recorded in nucleus – disease
- no selection for heterosis

- Non-additive genetics
- G x E

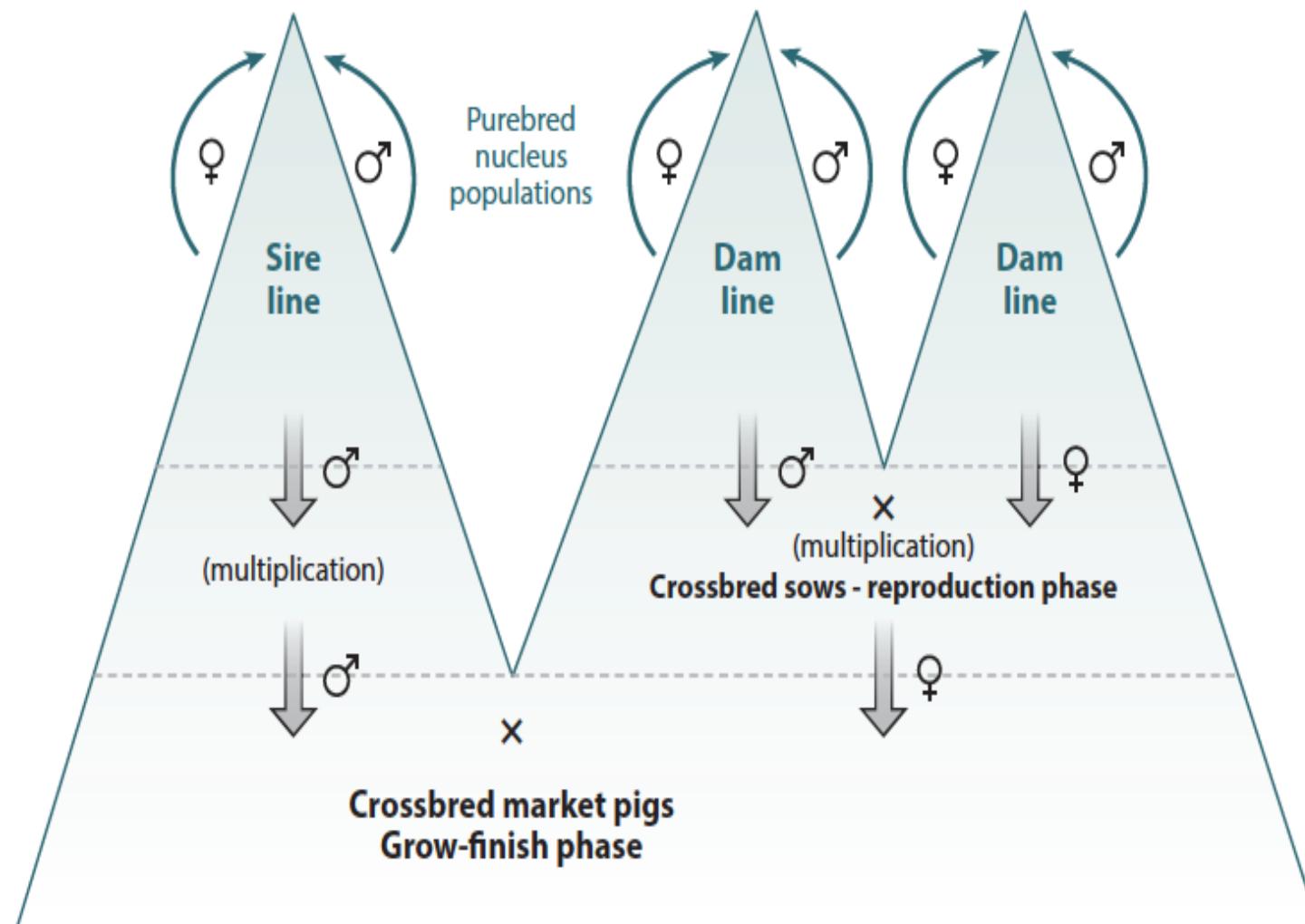
# 'Traditional' Breeding Solution



## Requirements / limitations:

- Costly logistics - Pedigree-based phenotyping in field
- Higher rates of inbreeding - family data vs. own phenotype

# Is there a genomic solution?



**Selection on Purebred Performance in High Health Environment**

$$r_{pc} < 1$$

**Crossbred Performance in Field Environment**

- Non-additive genetics
- G x E

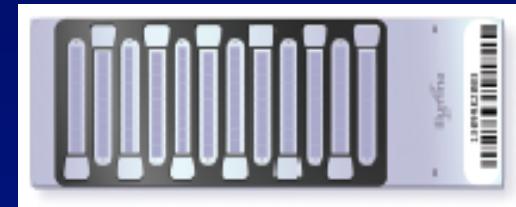
# Genomic selection

Meuwissen et al. 2001 Genetics

## Genetic Evaluation using high-density SNPs

$$y_i = \mu + \sum_{\text{SNP } k} \beta_k X_{ik} + e_i$$

substitution effect  
# allele dosage (0/1/2)



Estimates of SNP effects  $\hat{\beta}_k$

Use to estimate breeding value of new animals based on genotypes alone

$$\text{Genomic EBV} = \sum \hat{\beta}_k X_{ik}$$

Implemented using a variety of Bayesian methods (Bayes-A, -B, -C, C- $\pi$ )  
Or by using genomic vs. pedigree relationships in animal model BLUP (GBLUP)

# Genomic selection

Meuwissen et al. 2001

## Genetic Evaluation using high-density SNPs



Training data



Training data

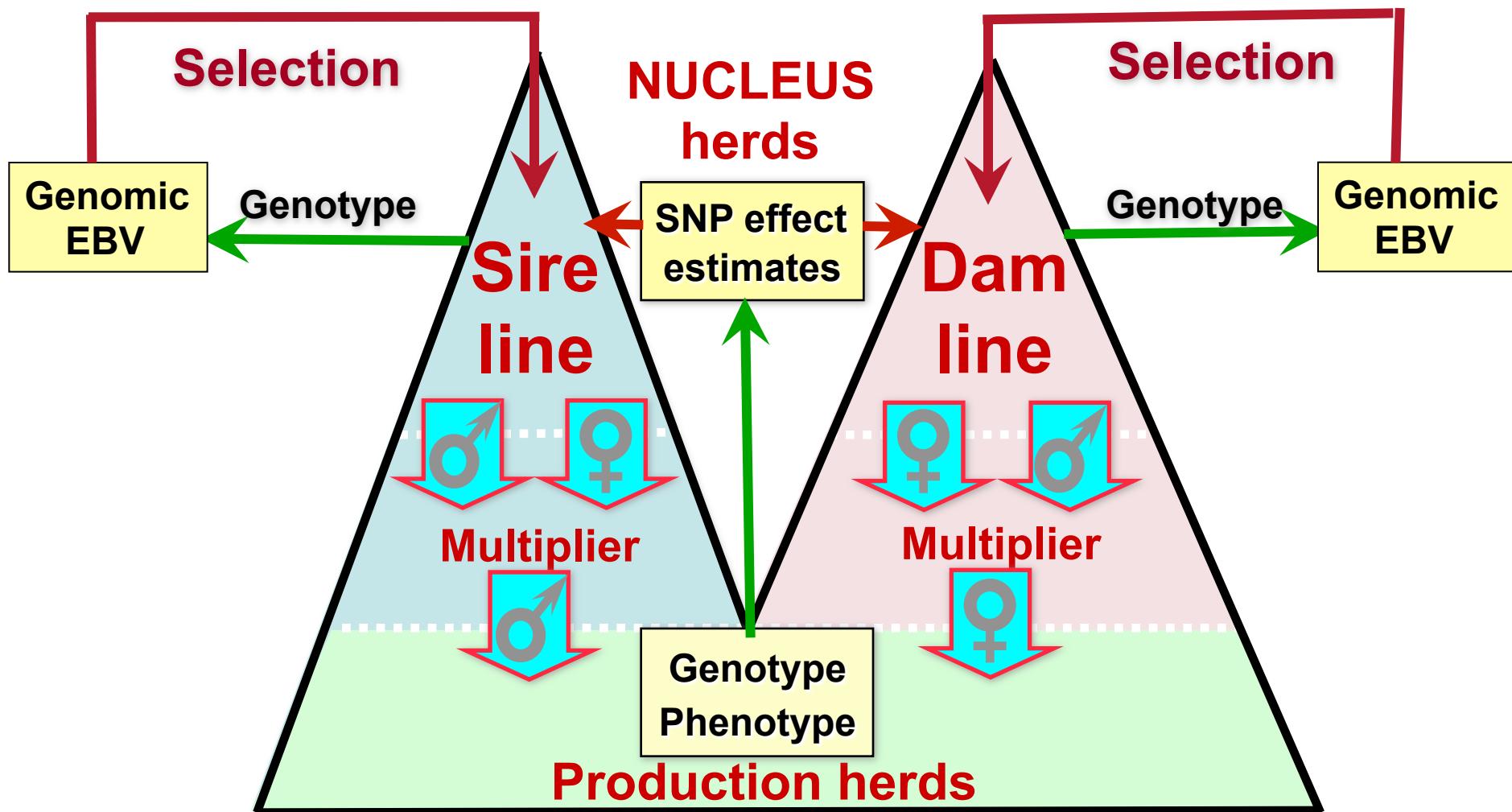
Estimate  
marker  
effects

Predict BV  
from marker  
genotypes at  
early age

Predict BV  
from marker  
genotypes at  
early age

# Genomic Prediction Training on Crossbred data

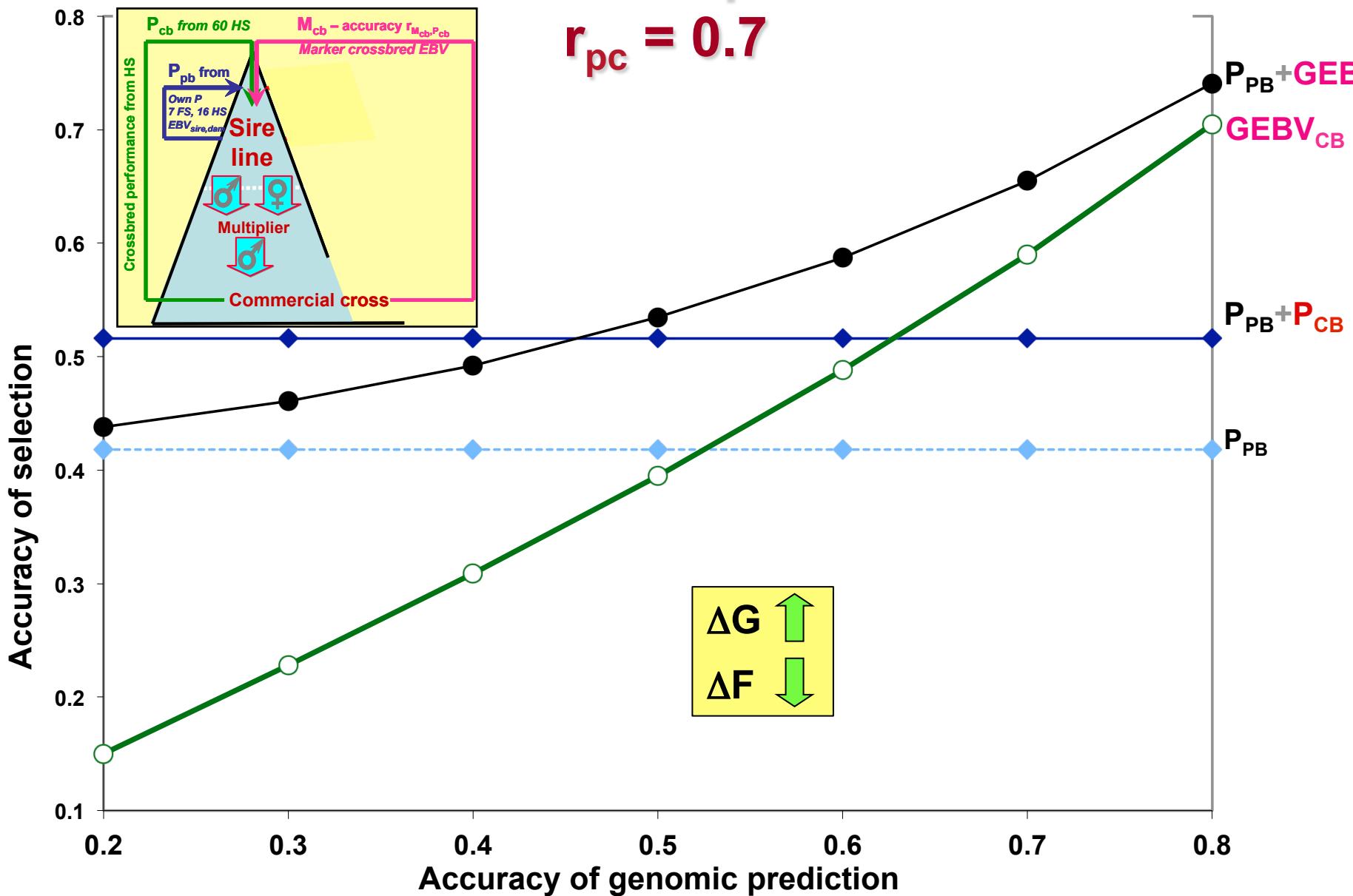
(Dekkers 2007 JAS)



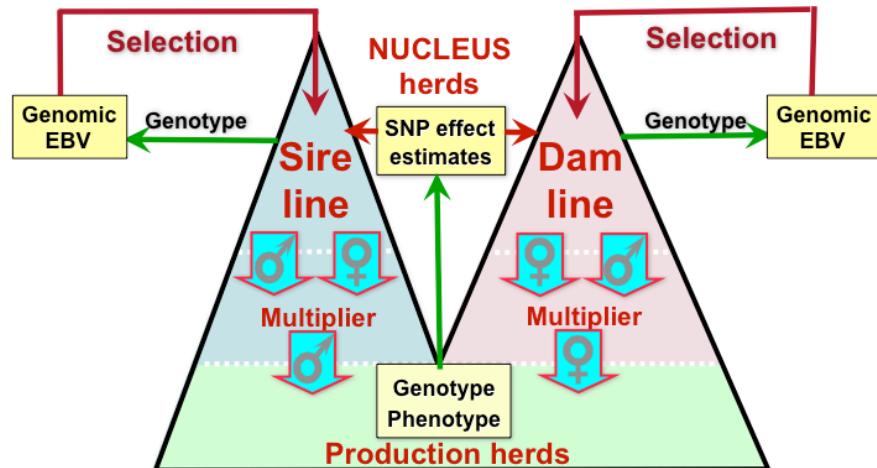
# Accuracy of selection for commercial performance

Dekkers  
JAS, 2007

$$r_{pc} = 0.7$$



# Questions:

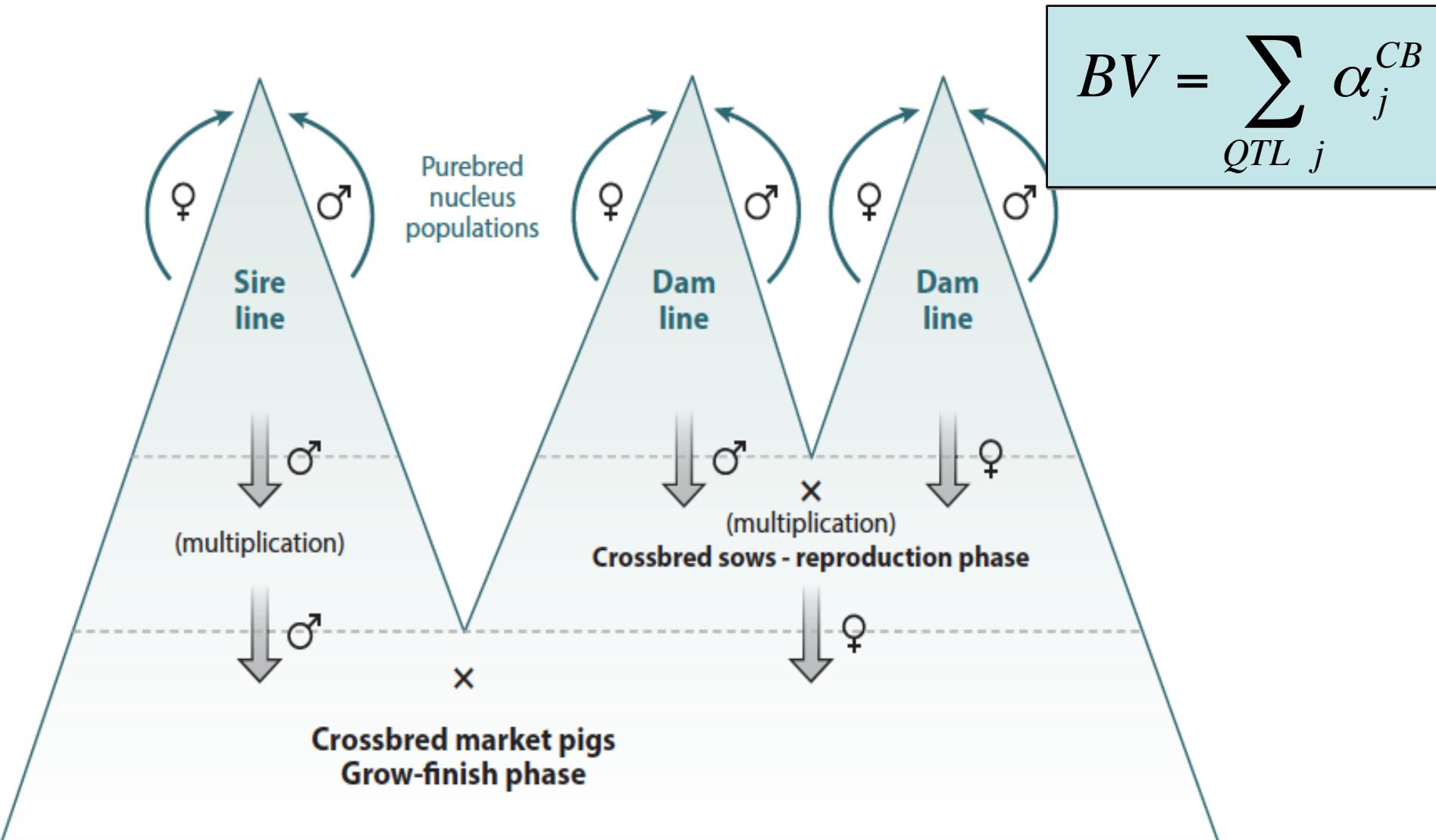


- **What model to use for genomic prediction of purebreds for CB performance?**
  - Additive model      ↔      Dominance model
  - Across-breed effects      ↔      Breed-specific effects
  
- **What training data to use?**
  - Purebred phenotypes      ↔      Crossbred phenotypes
  - Purebred genotypes      ↔      Crossbred genotypes
  
- **How does this depend on the nature of  $r_{pc}$ ?**
  - Dominance      ↔      GxE

# Select on Breeding Value for CB Performance

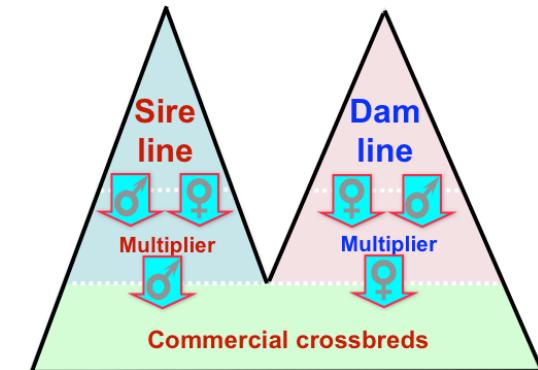
= sum of allele substitution or average allele effects

$$BV = \sum_{QTL\ j} \alpha_j^{CB}$$



# Average Allele Effects

QTL allele	Random allele from mate		Genetic value progeny = <b>Average allele effect</b>
	Q	q	
p <sub>Q</sub>	p <sub>q</sub>		
Q	QQ a	Qq d	$\alpha_Q = p_Q a + p_q d$
q	qQ d	qq -a	$\alpha_q = p_Q d - p_q a$



$$\text{Allele substitution effect} = \alpha = \alpha_Q - \alpha_q = a + (p_q - p_Q)d$$

QTL genotype	Breeding value	Rescaled BV
QQ	$2\alpha_Q = 2p_q \alpha$	$+ \alpha$
Qq	$\alpha_Q + \alpha_q = (p_q - p_Q)\alpha$	0
qq	$2\alpha_q = -2p_Q \alpha$	$- \alpha$

# 2-Breed Cross

PureBreds

**S**  
 $p_S$

**D**  
 $p_D$

Allele Substitution  
effects

PureBred

$$\alpha_{PB} = a + (1 - 2p_S)d$$

$$\alpha_{PB} = a + (1 - 2p_D)d$$

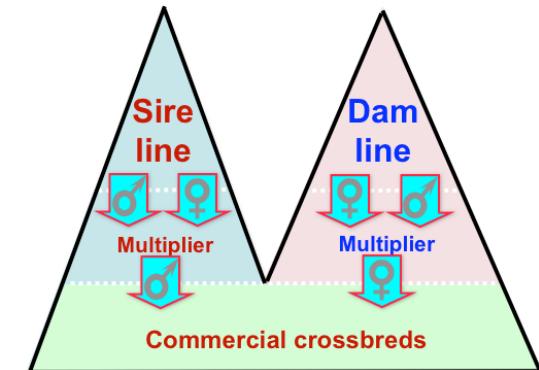
CrossBred

$$\alpha_{CB} = a + (1 - 2p_D)d$$

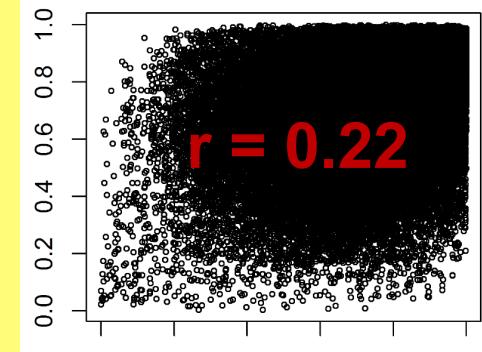
$$\alpha_{CB} = a + (1 - 2p_S)d$$

→  $\alpha_{PB} \neq \alpha_{CB} \rightarrow r_{pc} < 1$

→  $\alpha$ 's change over time



Landrace



Large White

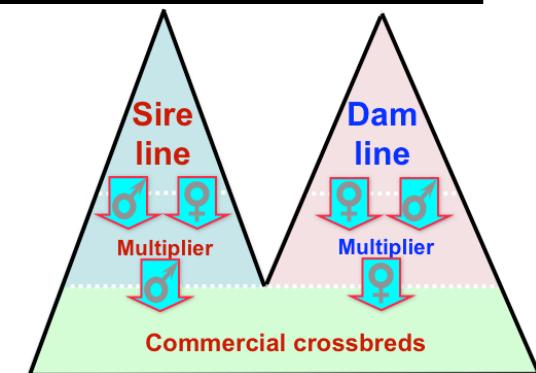
# Effect of Marker-QTL LD

Marker allele	QTL allele $p_Q$ $p_q$	Haplotype frequency	Mate allele		Genetic value of progeny	Average marker allele effects
			$Q$ $p_Q$	$q$ $p_q$		
$M$ $p_M$	$Q$	$p_M p_Q + D$	$QQ$ $a$	$Qq$ $d$	$\alpha_Q = p_Q a + p_q d$	$\beta_M = \frac{(p_M p_Q + D)}{p_M} (\alpha_Q - \alpha_q)$ $\alpha_q = p_Q d - p_q a$
	$q$	$p_M p_q - D$	$qQ$ $d$	$qq$ $-a$	$\alpha_q = p_Q d - p_q a$	
$m$ $p_m$	$Q$	$p_m p_Q - D$	$QQ$ $a$	$Qq$ $d$	$\alpha_Q = p_Q a + p_q d$	$\beta_m = \frac{(p_m p_Q - D)}{p_m} (\alpha_Q - \alpha_q)$ $\alpha_q = p_Q d - p_q a$
	$q$	$p_m p_q + D$	$qQ$ $d$	$qq$ $-a$	$\alpha_q = p_Q d - p_q a$	

$$D = p_{MQ} - p_Q p_M$$

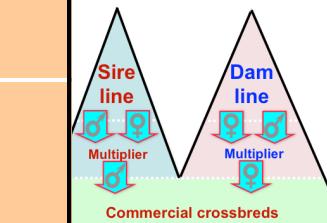
$$\text{Marker allele subst. effect} = \beta = \beta_M - \beta_m = \alpha \frac{D}{p_M p_m}$$

$$\alpha = a + (p_q - p_Q)d$$



# Impact of Marker-QTL LD

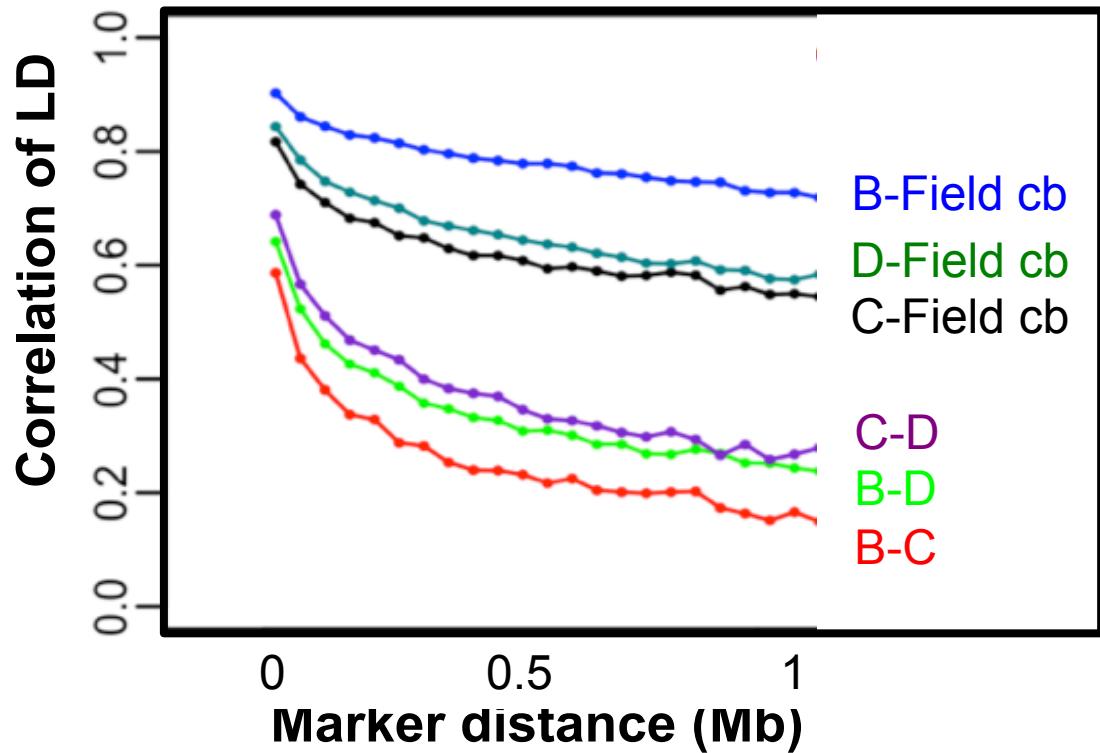
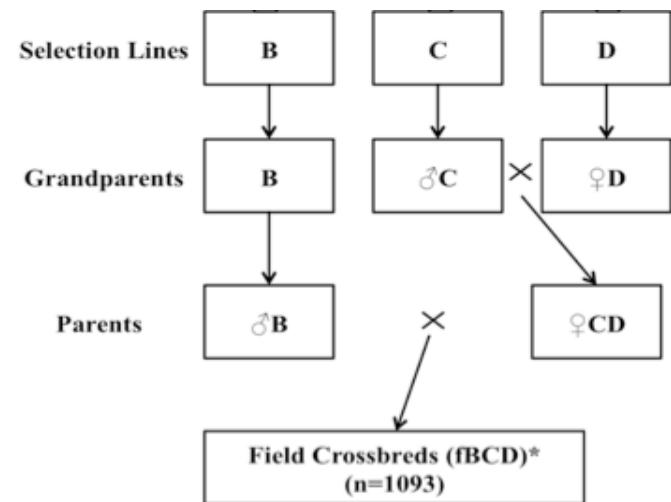
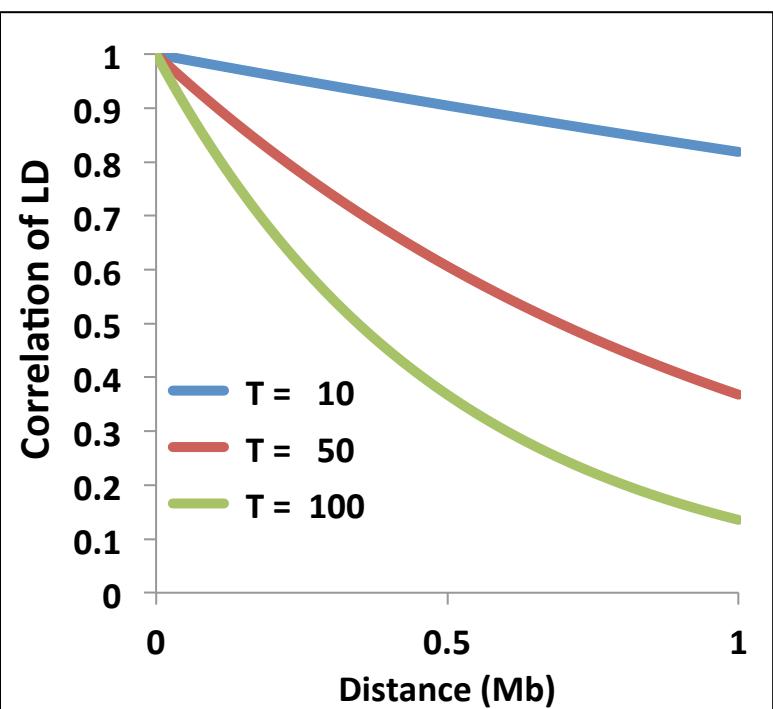
Paternal gamete			Maternal Gamete		D=0.1					
Marker allele	QTL allele	D=0.22 Frequency	M	0.3	m	0.7				
			Q	0.5 0.25	q	0.5 0.05	Q	0.5 0.25	q	0.5 0.45
M 0.6	Q	0.58	QQ	4	Qq	2	QQ	4	Qq	2
	q	0.02	qQ	2	qq	-4	qQ	2	Qq	2
m 0.4	Q	0.02	QQ	4	Qq	2	QQ	4	Qq	2
	q	0.38	qQ	2	qq	-4	qQ	2	qq	-4



Marker allele	Random allele from mate		Genetic value of progeny = Average allele effect	Dam breed Average Allele effect
	M $p_M$	m $p_m$		
M	$g_{MM} = 3.58$	$g_{Mm} = 2.56$	$\beta_M = p_M g_{MM} + p_m g_{Mm} = 2.87$	$\beta_M = 2.60$
m	$g_{mM} = 1.13$	$g_{mm} = -1.62$	$\beta_m = p_M g_{mM} + p_m g_{mm} = -0.80$	$\beta_m = +0.89$

# Consistency of LD between breeds

$$\text{Correlation of LD} = e^{-2cT}$$



# Models for training on CB data

## Standard Additive Model

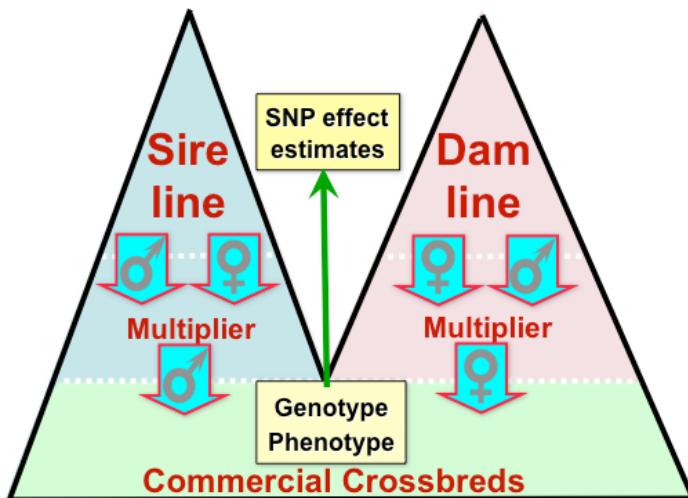
$$y_i^{CB} = \mu + \sum_{SNP \ j} X_{ij} \alpha_j + e_{ij} \quad X_{ij} = \text{genotype code 0/1/2}$$

$$GEBV_s^{CB} = \sum_{SNP \ j} X_{ij} \hat{\alpha}_j$$

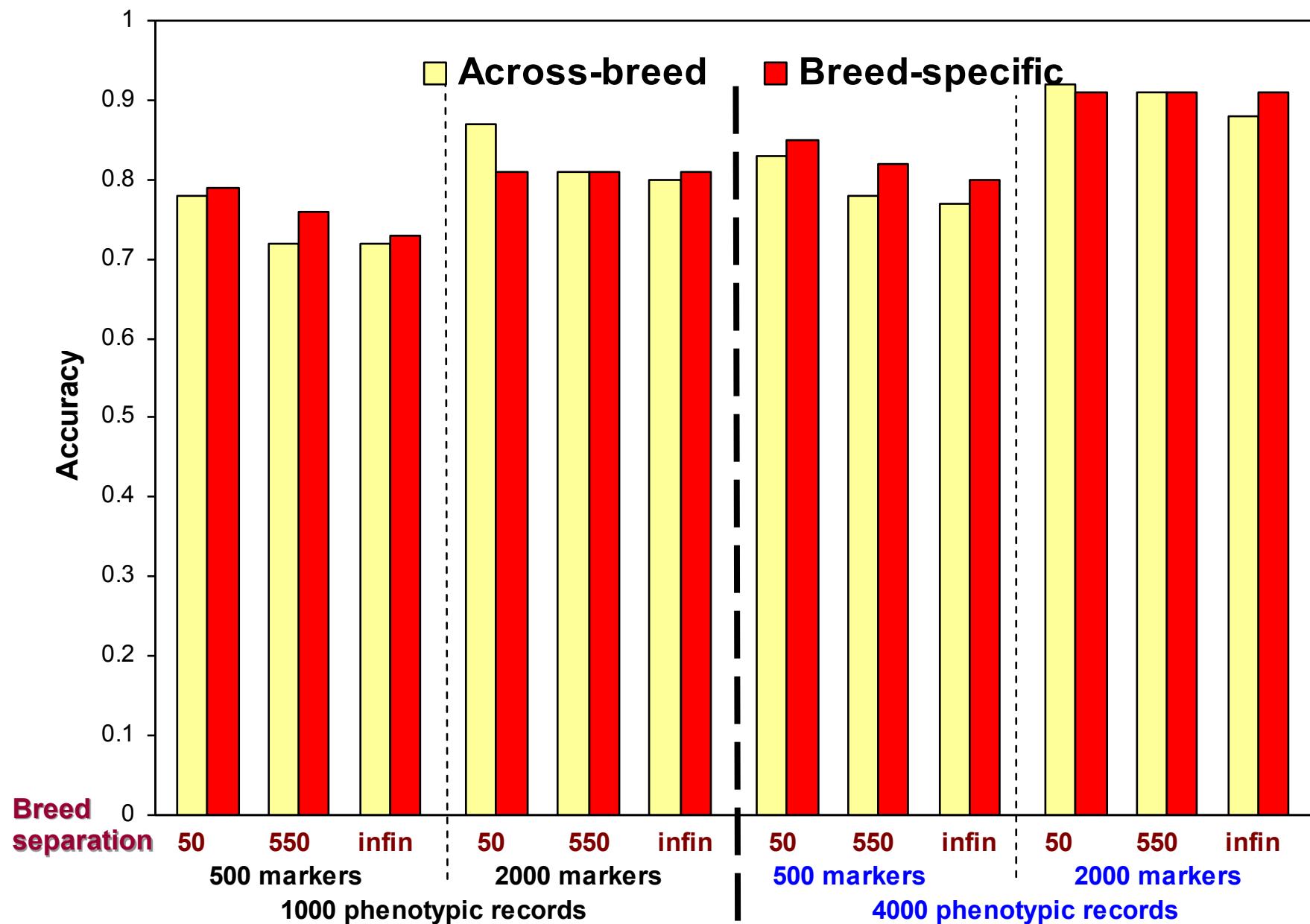
## Breed-specific Additive Model

$$y_i^{CB} = \mu + \sum_{SNP \ j} (X_{ij}^s \alpha_j^s + X_{ij}^d \alpha_j^d) + e_{ij} \quad X_{ij}^s = \text{breed origin allele code 0/1}$$

$$GEBV_i^{CB} = \sum_{SNP \ j} (X_{ij}^s \hat{\alpha}_j^s + X_{ij}^d \hat{\alpha}_j^d)$$



## GSE 2009 Additive QTL model



# Models for training on CB data

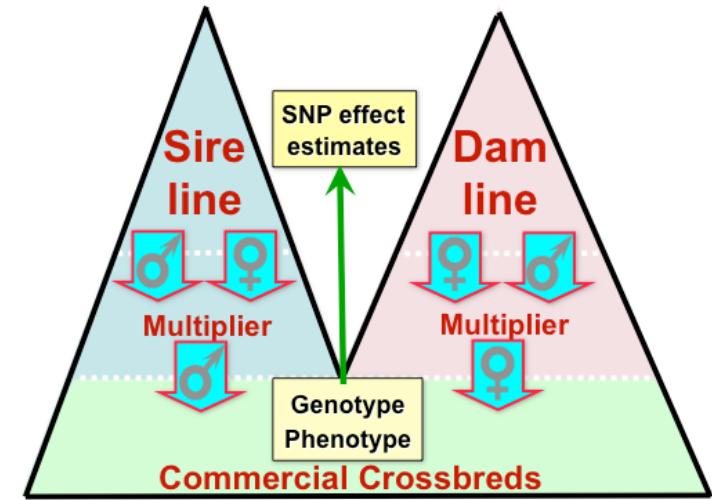
## Standard Additive Model

$$y_i^{CB} = \mu + \sum_{SNP j} X_{ij} \alpha_j + e_{ij} \quad X_{ij} = \text{genotype code 0/1/2}$$

## Breed-specific Additive Model

$$y_i^{CB} = \mu + \sum_{SNP j} (X_{ij}^s \alpha_j^s + X_{ij}^d \alpha_j^d) + e_{ij}$$

$$\alpha = a + (1 - 2p)d$$



## Dominance Model

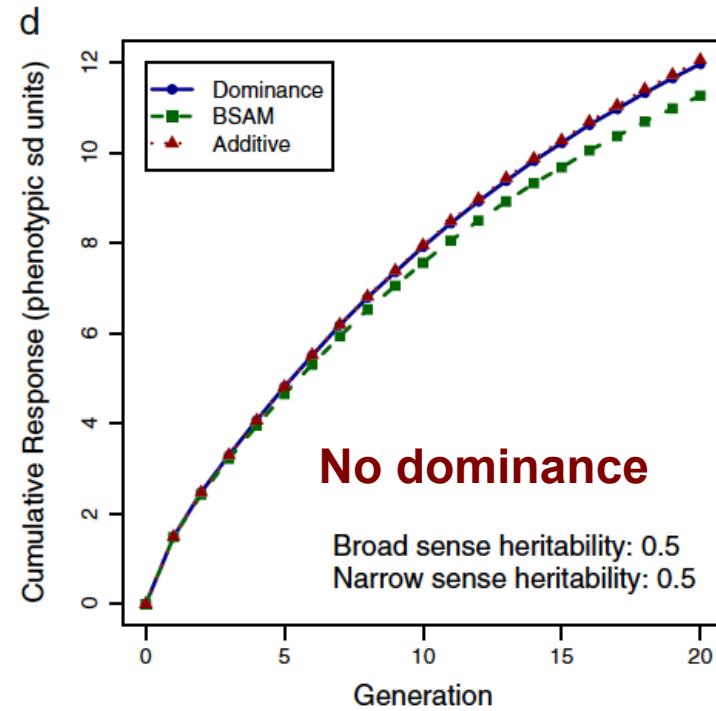
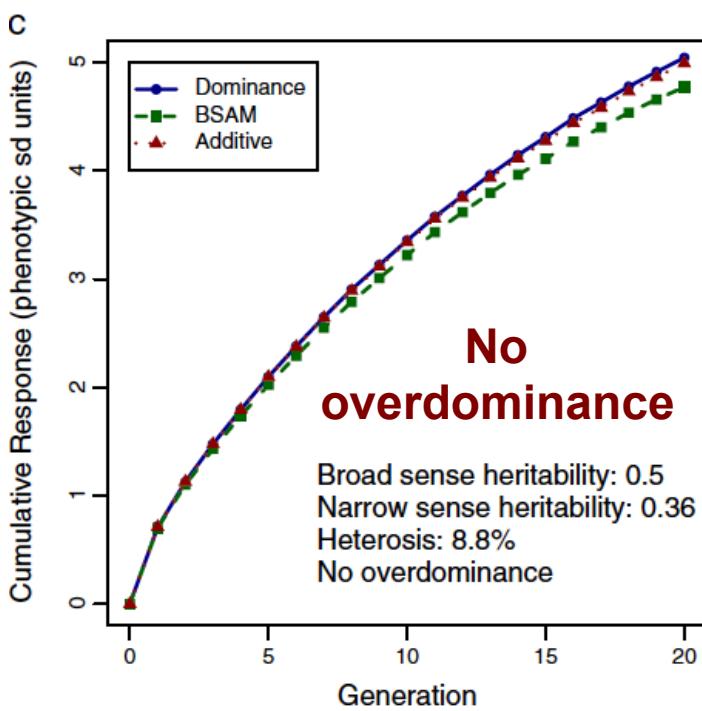
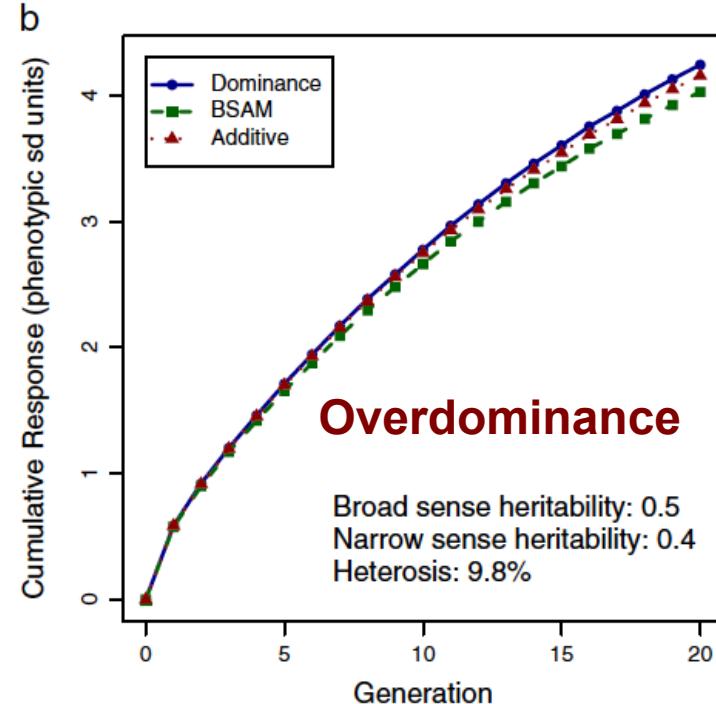
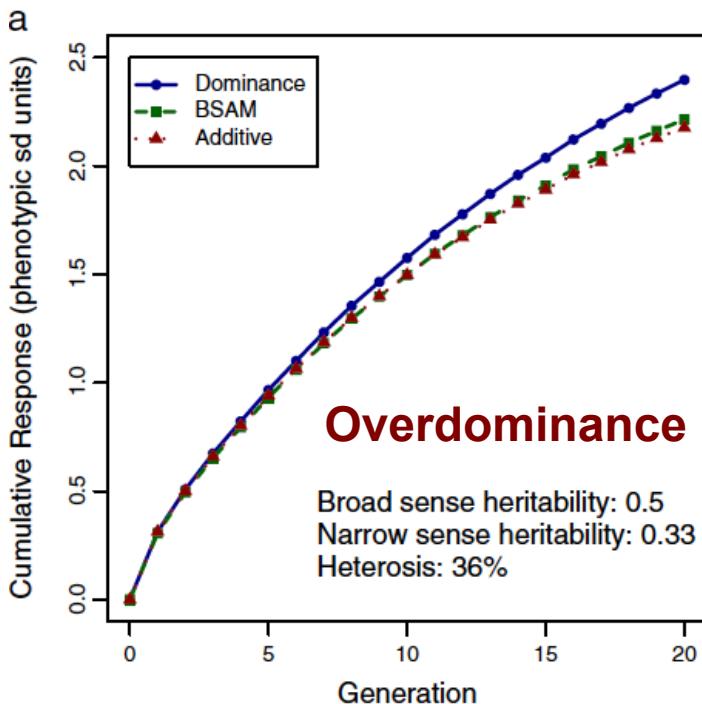
$$y_i^{CB} = \mu + \sum_{SNP j} (X_{ij} a_j + H_{ij} d_j) + e_i \quad H_{ij} = \text{heterozygosity indicator 0/1}$$

$$\hat{\alpha}_s^{CB} = \hat{a}_j + (1 - 2p_j^d)\hat{d}_j$$

$$GEBV_s^{CB} = \sum_{SNP j} X_{ij} \hat{\alpha}_j^{CB}$$

Zeng et al. 2013

10 chromosomes,  
1000 QTL  
10,000 SNPs  
1,000 CB records



# Models for training on CB data

## Dominance Model

$$y_i^{CB} = \mu + \sum_{SNP} j (X_{ij} a_j + H_{ij} d_j) + e_i$$

$H_{ij}$  = het. indicator 0/1

$X_{ij}^{MM}$  = genotype indicator 0/1

## Alternate formulation

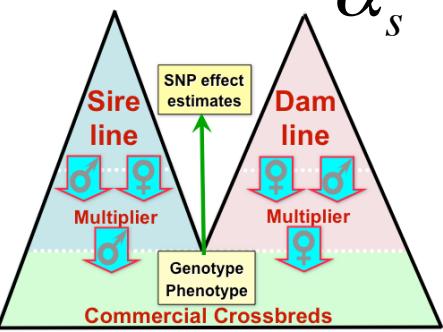
$$y_i^{CB} = \mu + \sum_{SNP} j (X_{ij}^{MM} g_j^{MM} + X_{ij}^{mM} g_j^{mM} + X_{ij}^{mm} g_j^{mm}) + e_i$$

$$\hat{a}_j = \frac{1}{2}(\hat{g}_j^{MM} - \hat{g}_j^{mm})$$

$$\hat{d}_j = \hat{g}_j^{mM} - \frac{1}{2}(\hat{g}_j^{MM} + \hat{g}_j^{mm})$$

$$\hat{\alpha}_s^{CB} = \hat{a}_j + (1 - 2 p_j^d) \hat{d}_j$$

$$GEBV_s^{CB} = \sum_{SNP} j X_{ij} \hat{\alpha}_j^{CB}$$



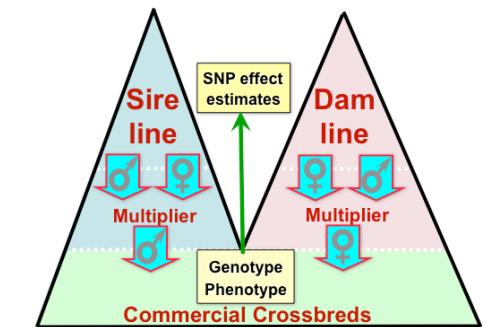
# Models for training on CB data

## Dominance Model

$$y_i^{CB} = \mu + \sum_{SNP} \left( X_{ij} a_j + H_{ij} d_j \right) + e_i$$

## Alternate formulation

$$y_i^{CB} = \mu + \sum_{SNP} \left( X_{ij}^{MM} g_j^{MM} + X_{ij}^{mM} g_j^{mM} + X_{ij}^{mm} g_j^{mm} \right) + e_i$$



$X_{ij}^{MM}$  = genotype ind 0/1

## Breed-specific Dominance Model

$$y_i^{CB} = \mu + \sum_{SNP} \left( X_{ij}^{MM} g_j^{MM} + X_{ij}^{Mm} g_j^{Mm} + X_{ij}^{mM} g_j^{mM} + X_{ij}^{mm} g_j^{mm} \right) + e_i$$

$$\hat{\alpha}_j^{sM} = p_j^d \hat{g}_j^{MM} + q_j^d \hat{g}_j^{Mm}$$

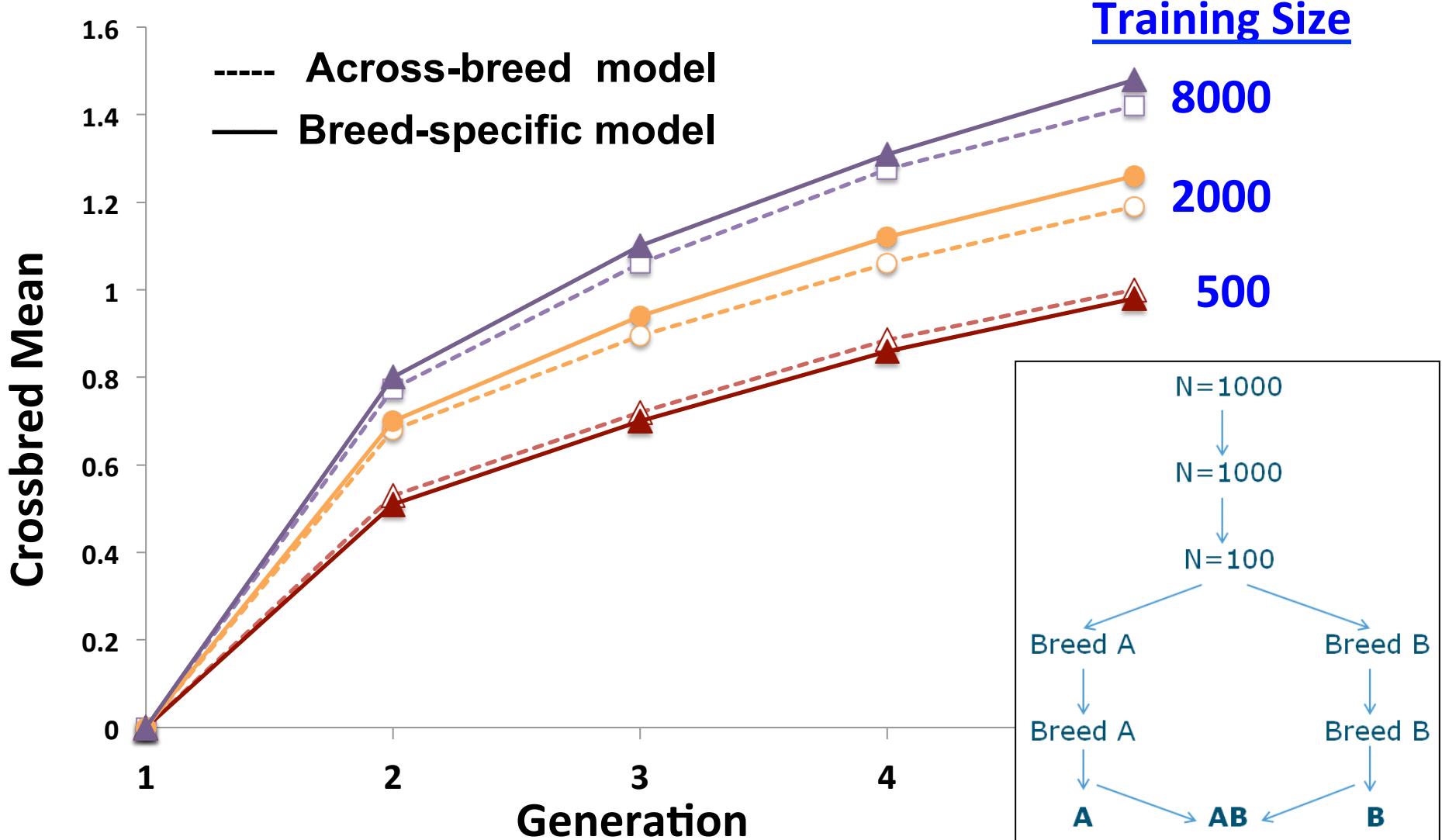
$$\hat{\alpha}_j^{dM} = p_j^s \hat{g}_j^{MM} + q_j^s \hat{g}_j^{Mm}$$

$$\hat{\alpha}_j^{sm} = p_j^d \hat{g}_j^{mM} + q_j^d \hat{g}_j^{mm}$$

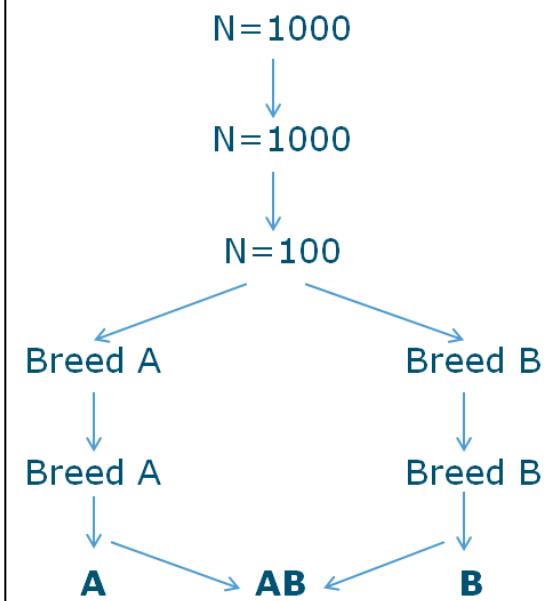
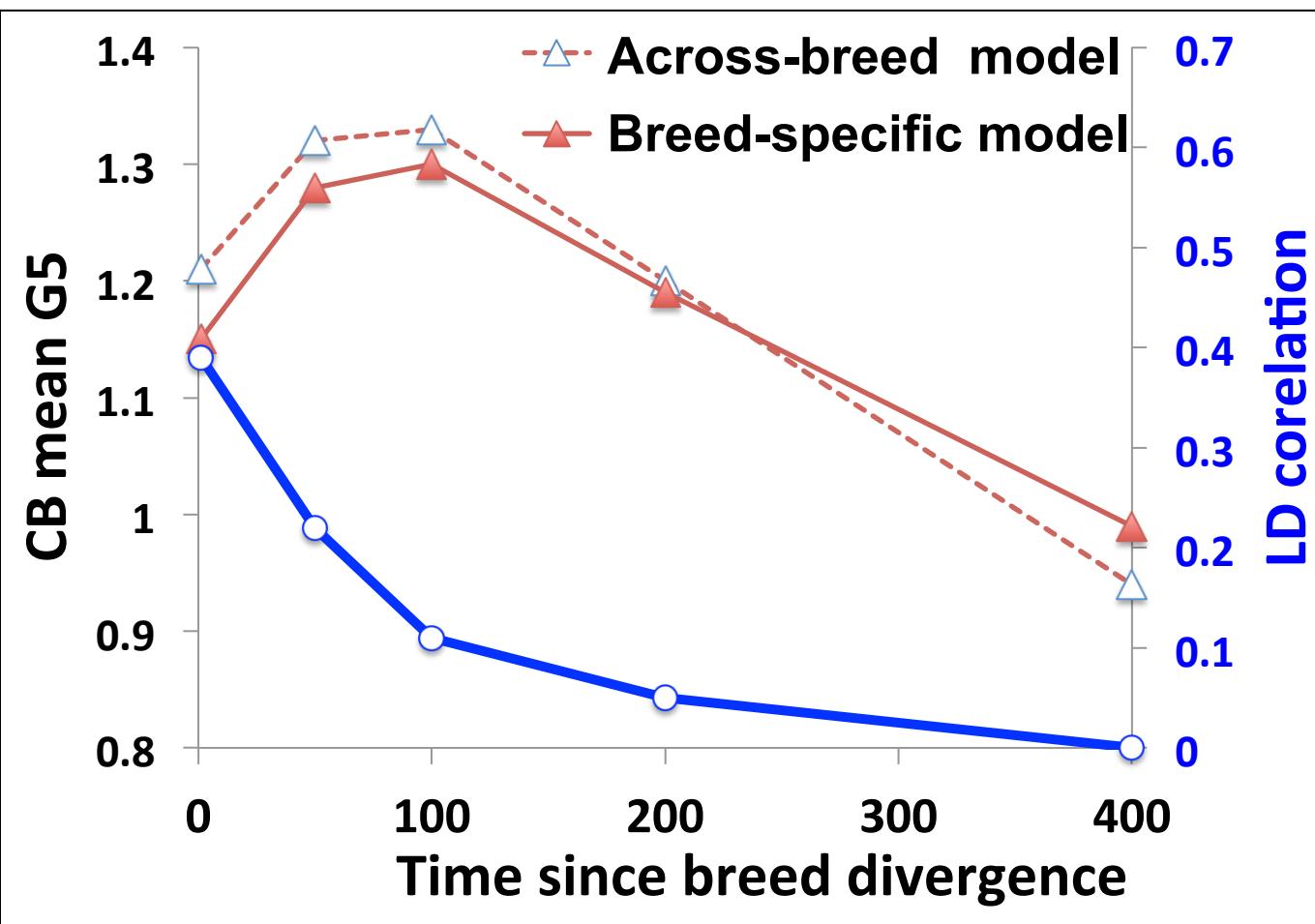
$$\hat{\alpha}_j^{dm} = p_j^s \hat{g}_j^{mM} + q_j^s \hat{g}_j^{mm}$$

$$GEBV_i^{CB} = \sum_{SNP} Average\ Allele\ Effects$$

# Dominance Models for CB training

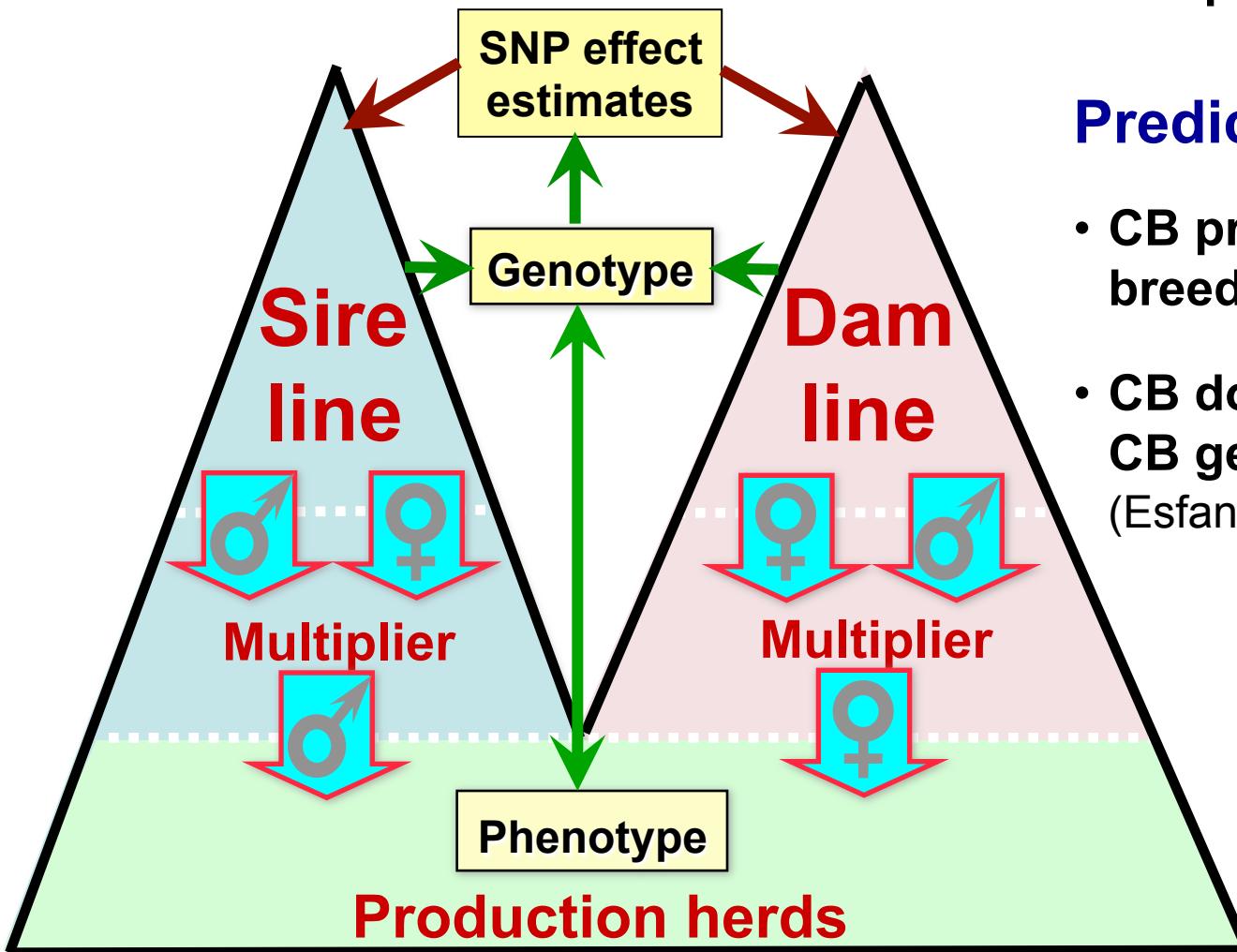


# Impact of time since divergence of breeds



# Other GS training scenarios

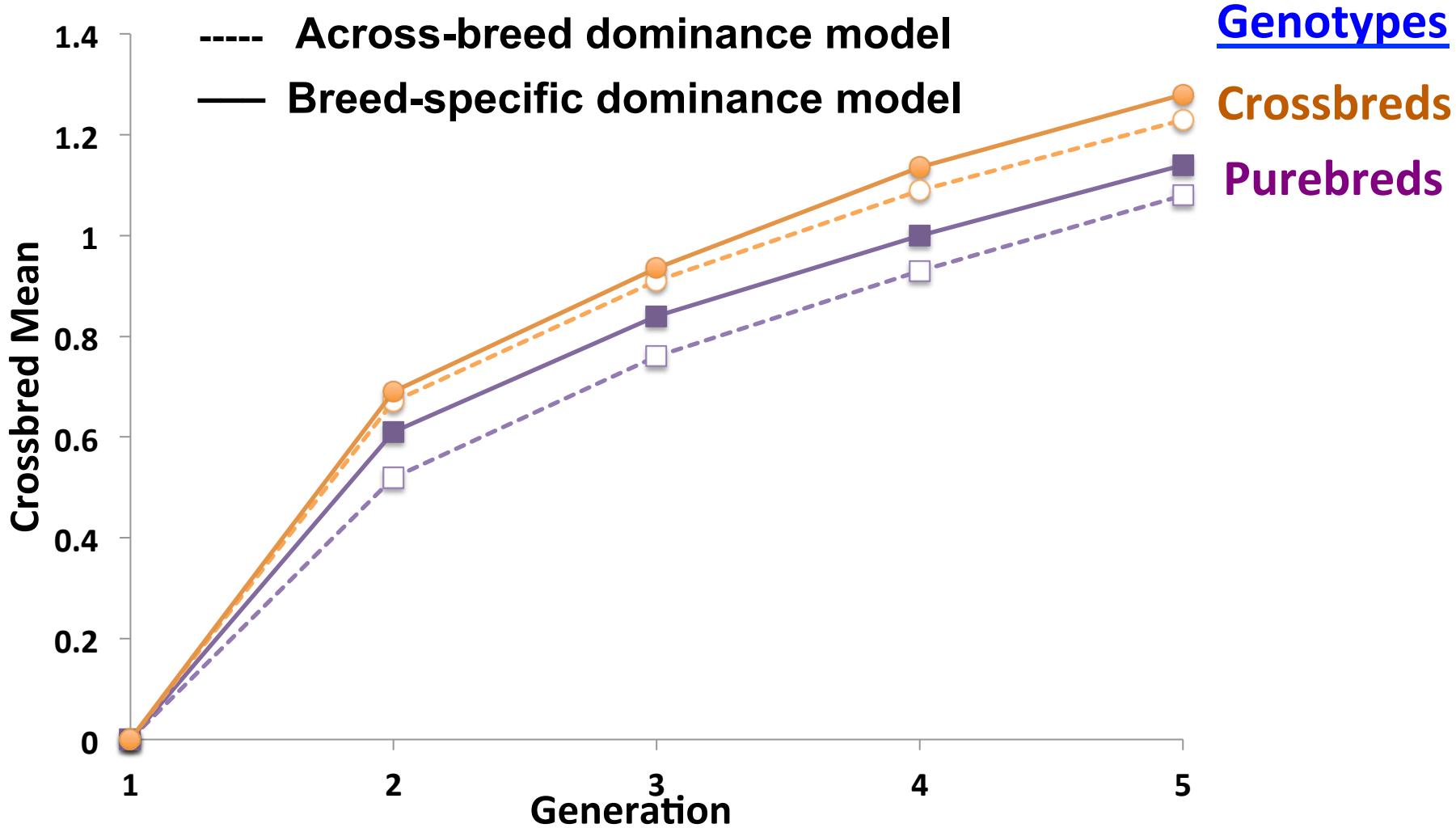
1. Genotype phenotyped CB → train on own phenotype
  - does not require pedigree
2. Genotype PB and train on CB progeny performance
  - requires pedigree



## Prediction models

- CB progeny mean model by breed → breed-specific  $\alpha$
- CB dominance model using CB genotype probabilities  
(Esfandyari et al. 2014)

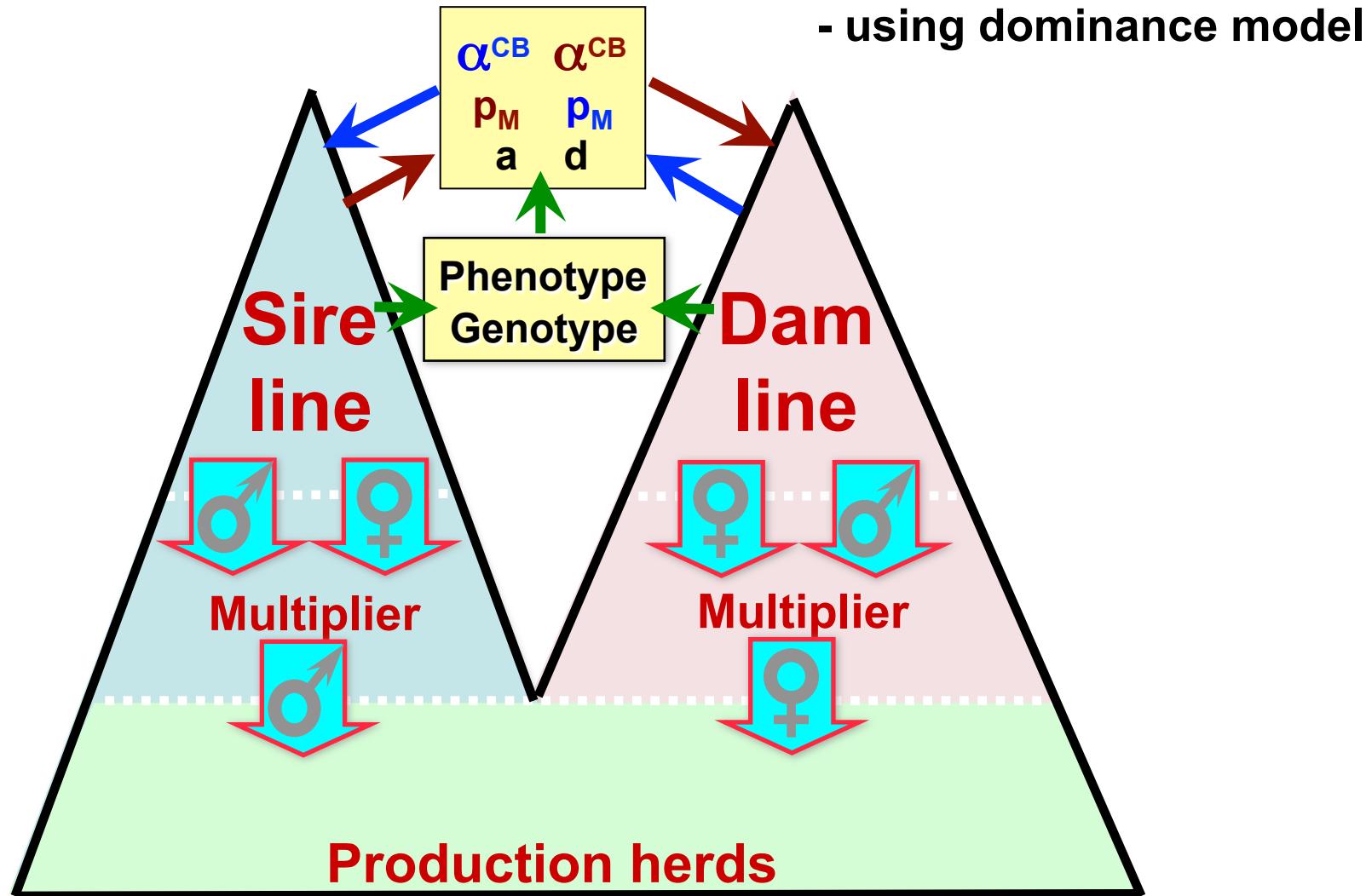
# Training on CB phenotypes



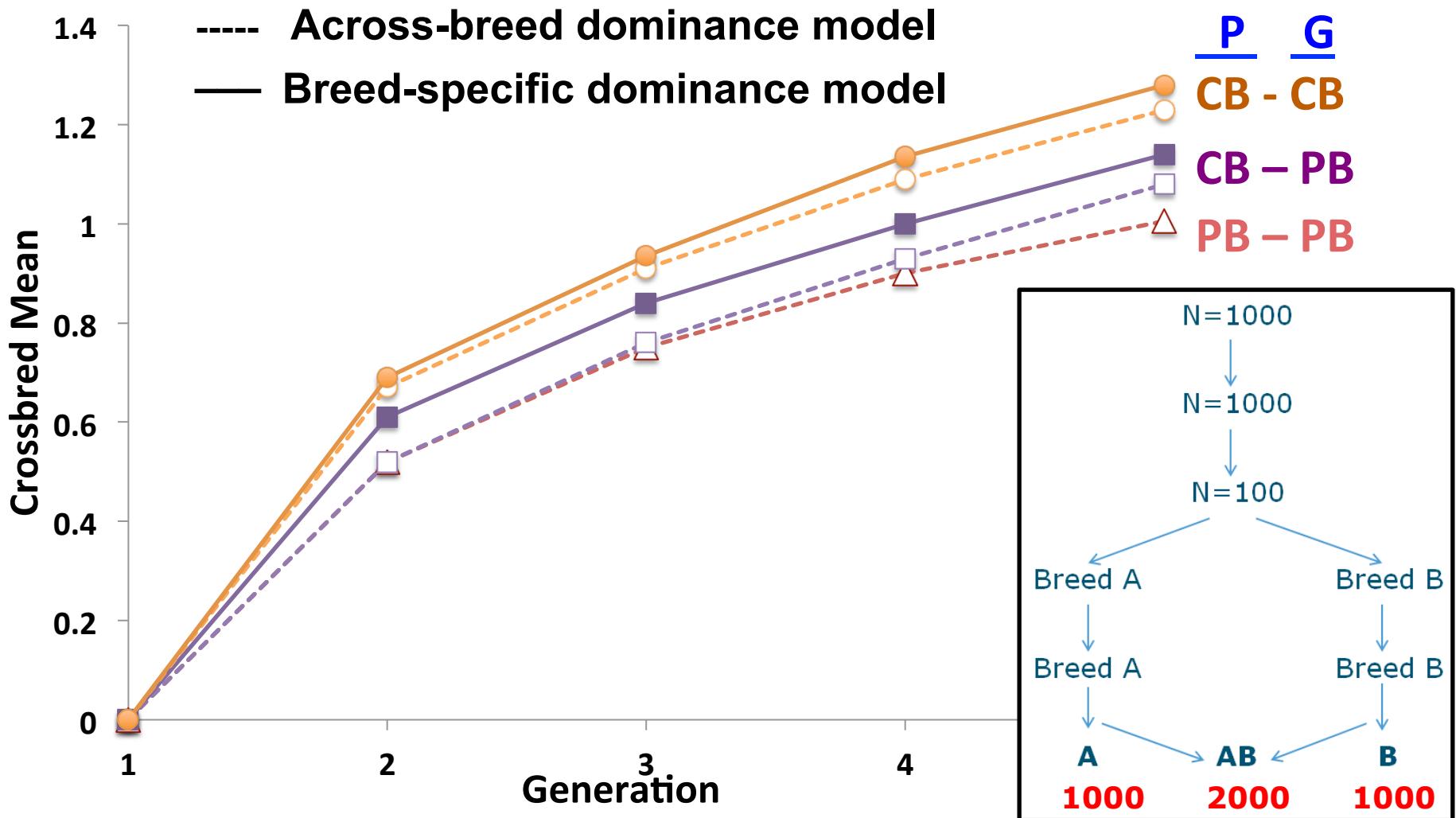
# Other GS training scenarios

1. Genotype phenotyped CB → train on own phenotype
2. Genotype PB and train on CB progeny performance
3. Genotype PB and train on PB performance

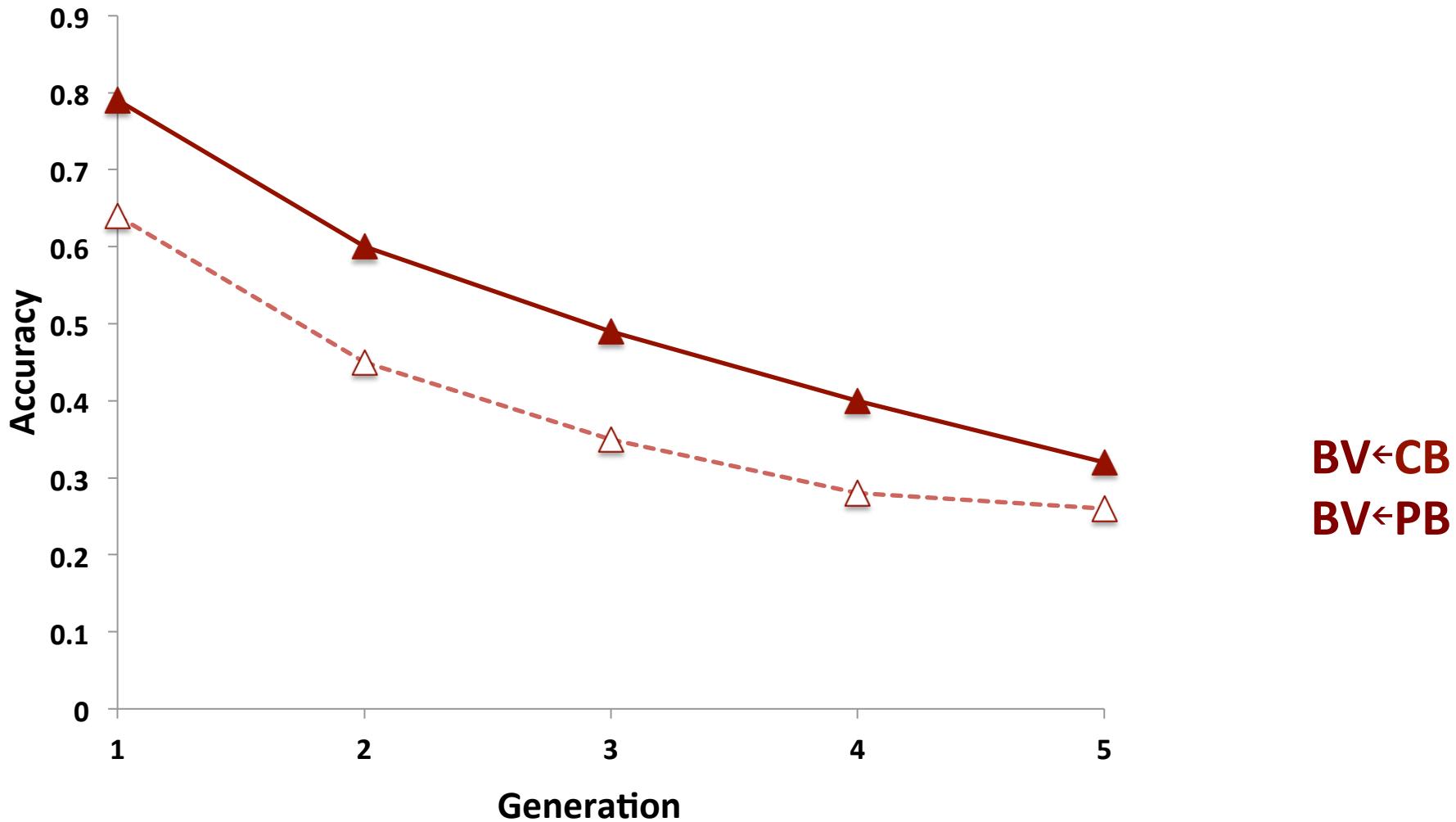
- using dominance model



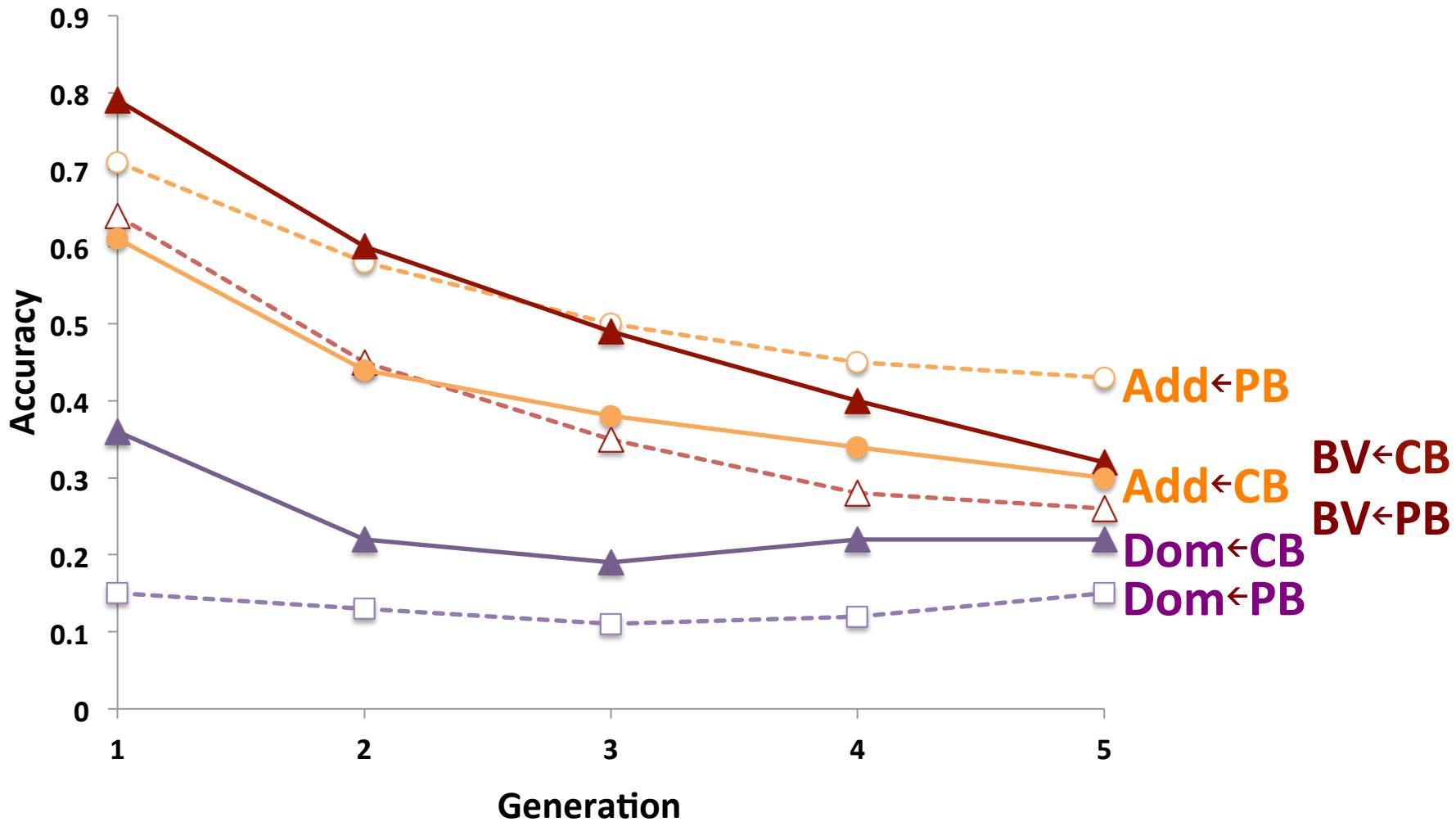
# Training using PB Phenotypes



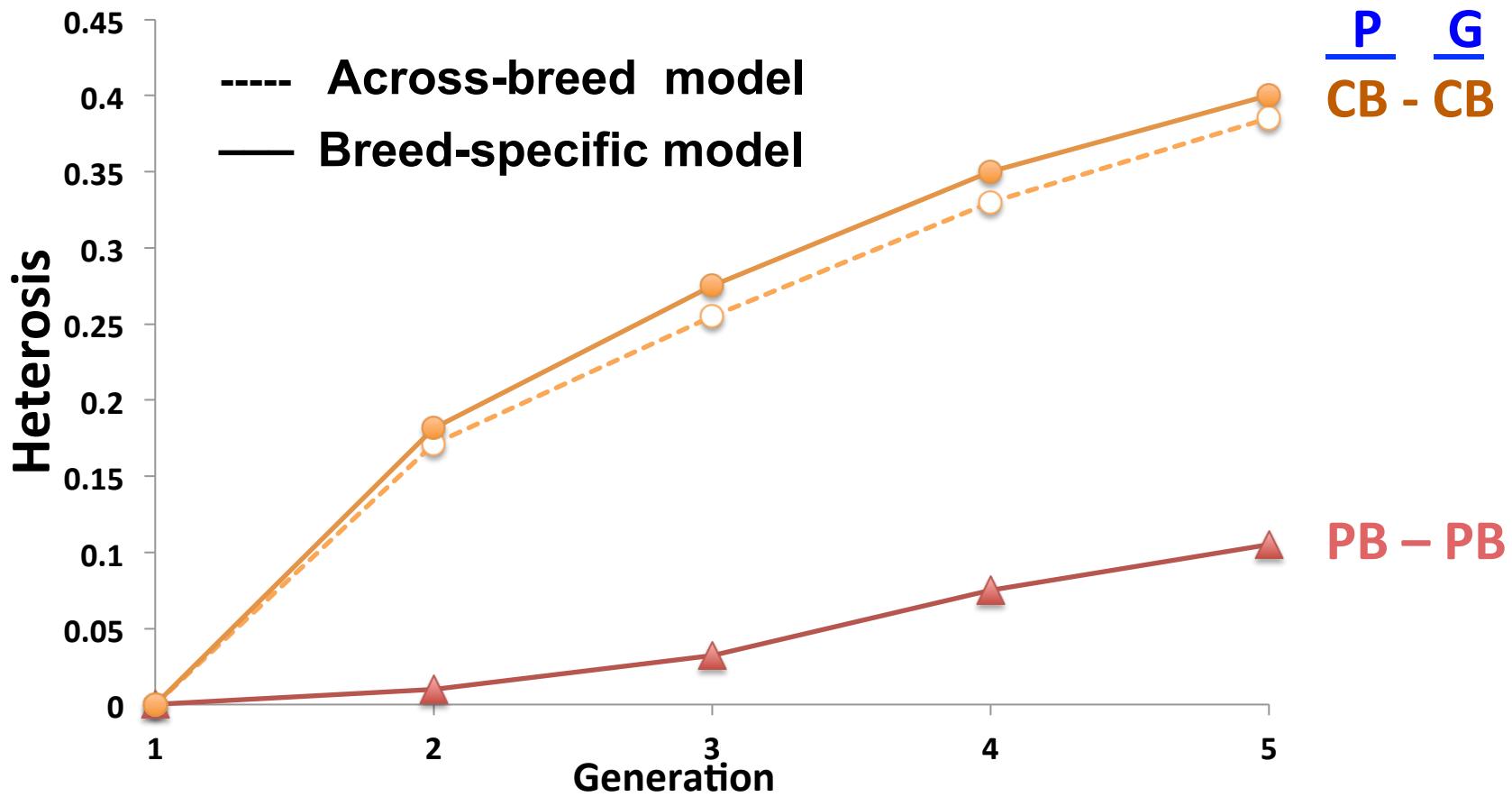
# Accuracy of BV and components



# Accuracy of BV and components

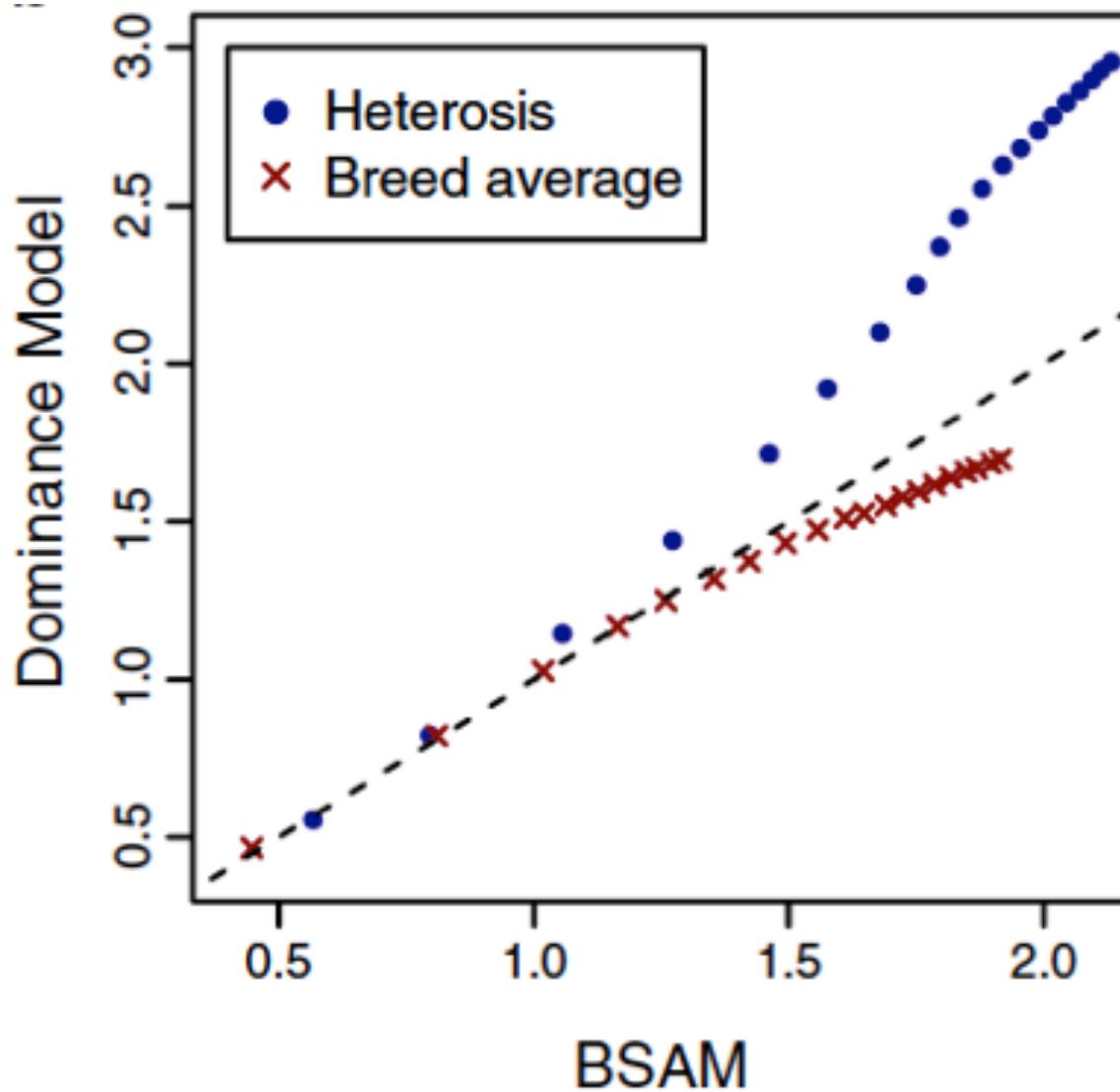


# Heterosis



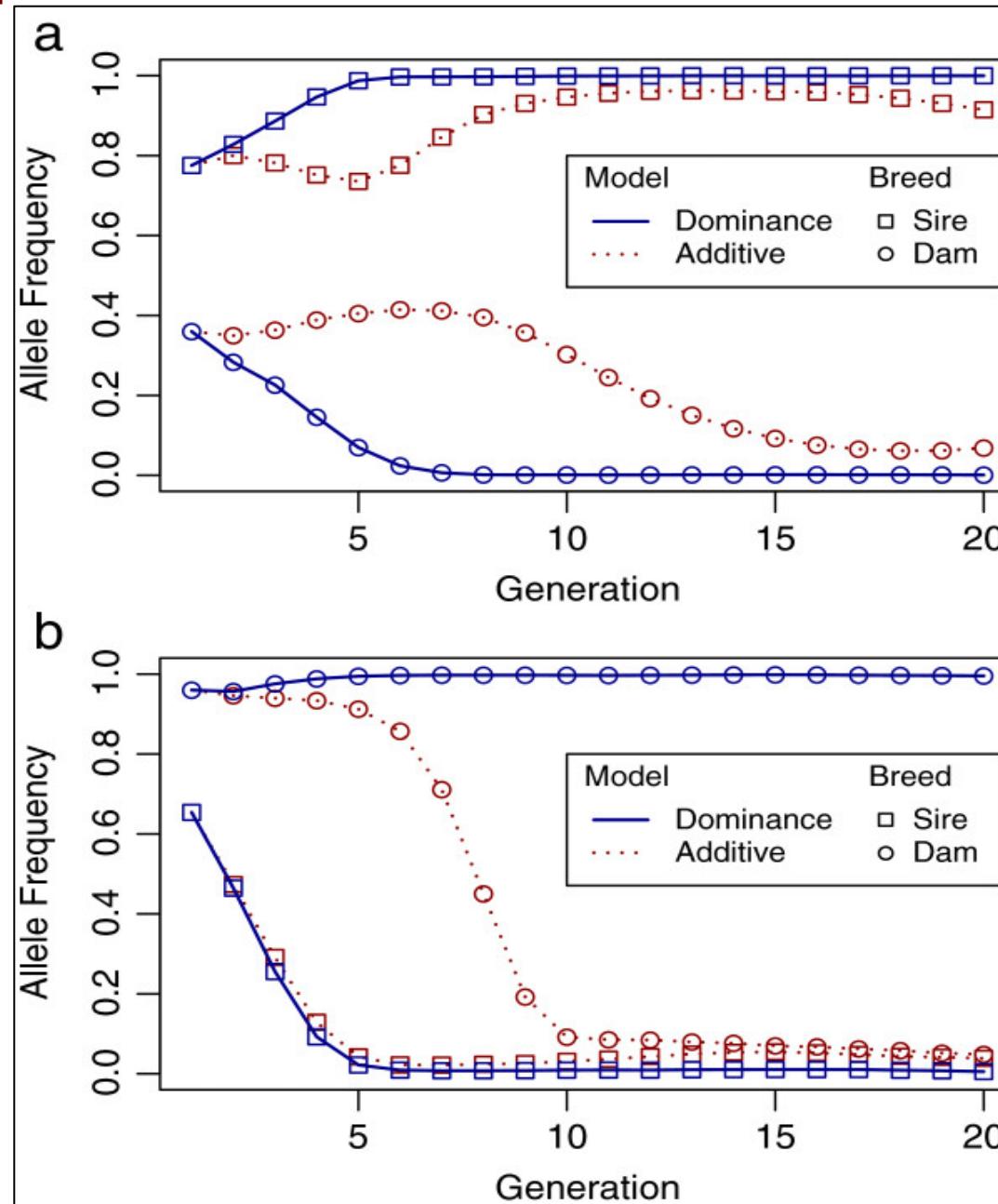
# Impact of Dominance $\leftrightarrow$ Additive model on Heterosis

Zeng et al. 2013



# Allele frequencies for over-dominant QTL

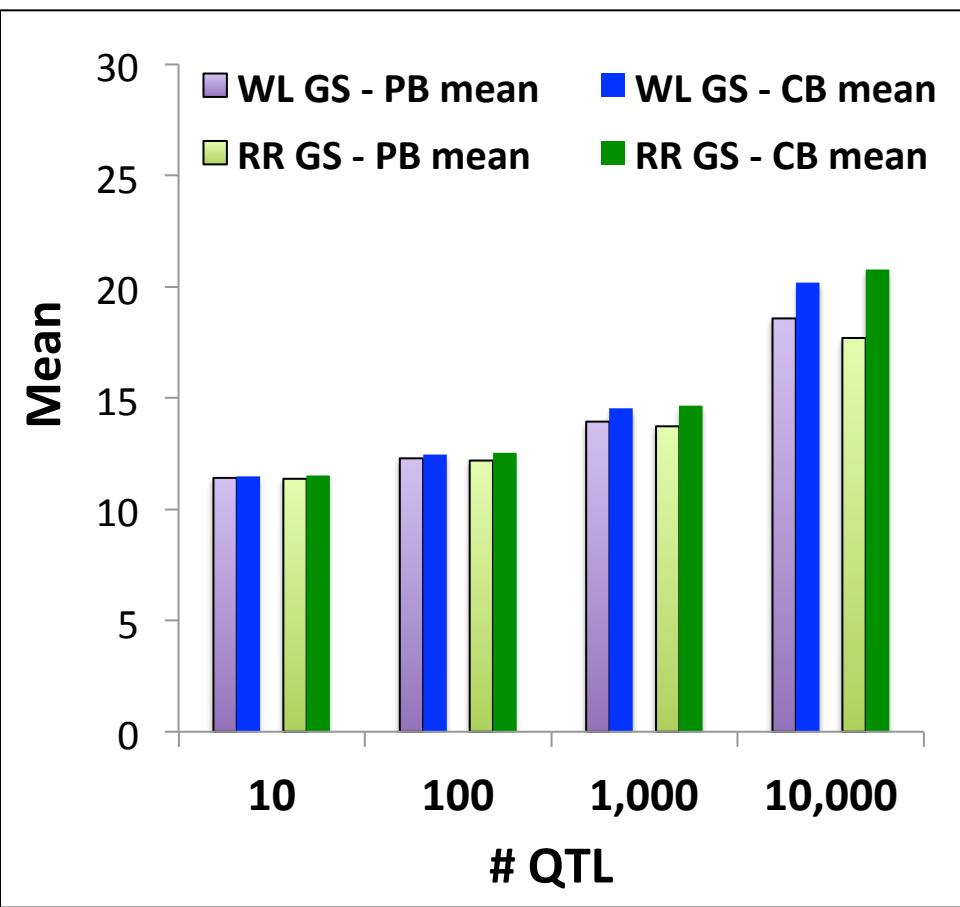
Zeng et al. 2013



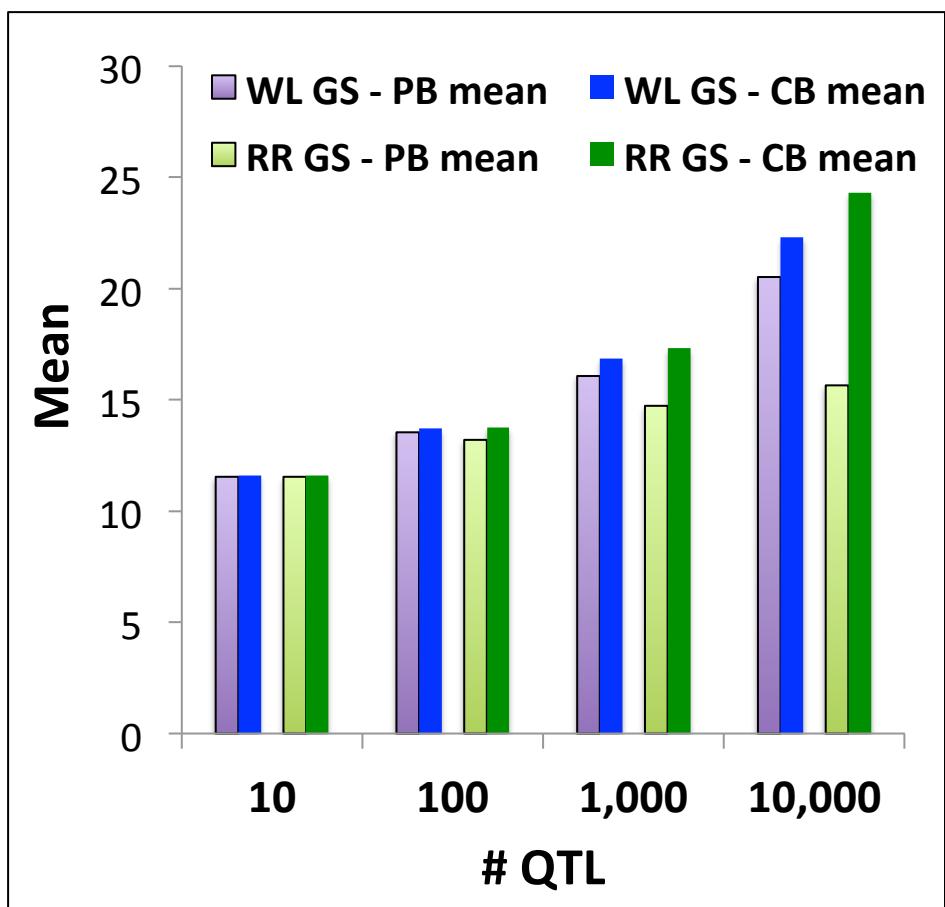
# Heterosis without overdominance?

Kinghorn et al. 2011

After 5 Generations



After 20 Generations



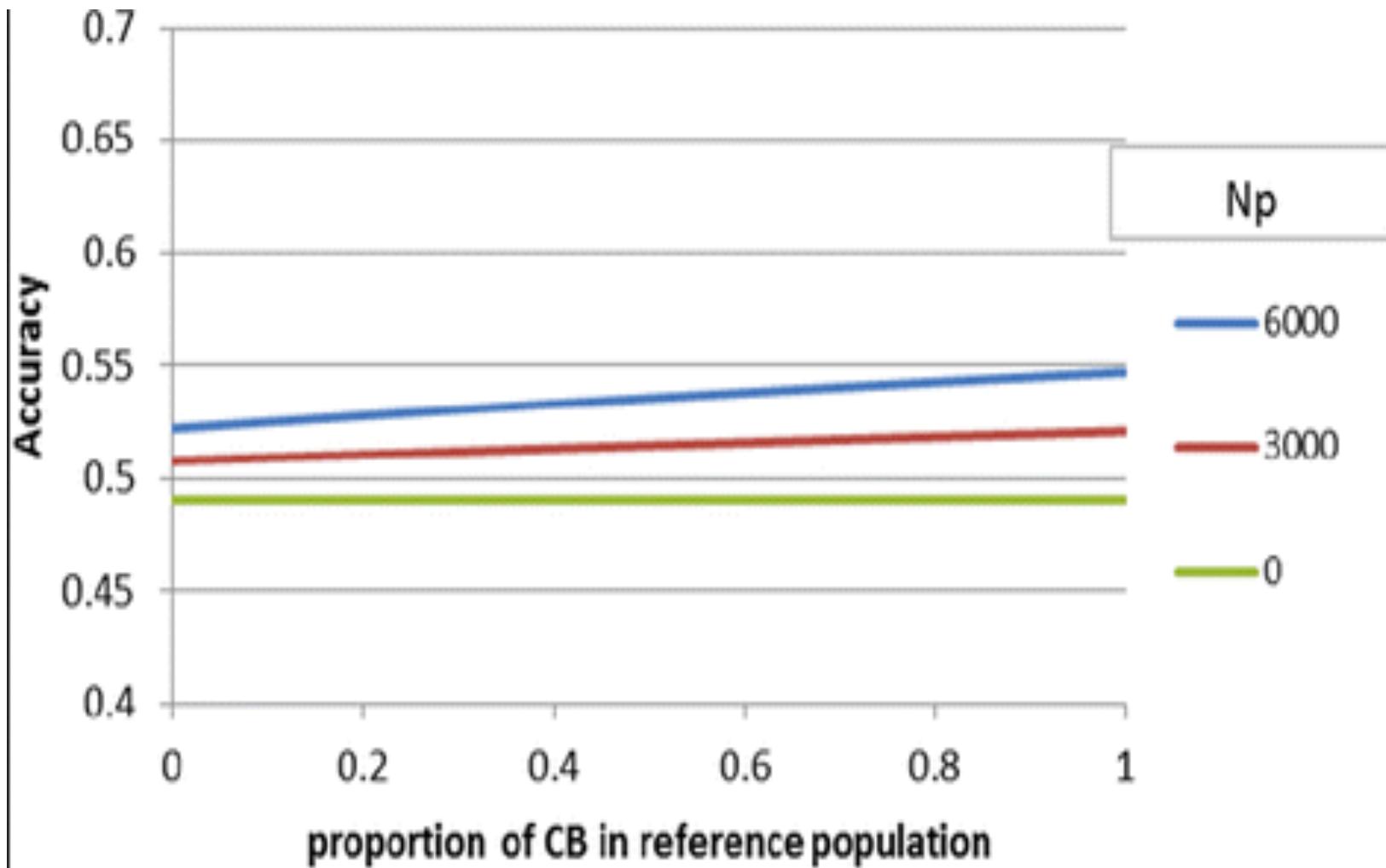
# Phenotypes and models for CB prediction depending on nature of $r_{pc}$

Factors contributing to $r_{pc} < 1$	Phenotypes		Prediction model	
	PB	CB	Across breed dominance model	Breed-specific dominance model
Dominance	v	v	v	v
Epistasis	v	v	v	v
Imprinting	v	v		vv
Marker-QTL LD	v	v		v
GxE		vv	v	v

- Genotype phasing
- Retraining
- Haplotypes

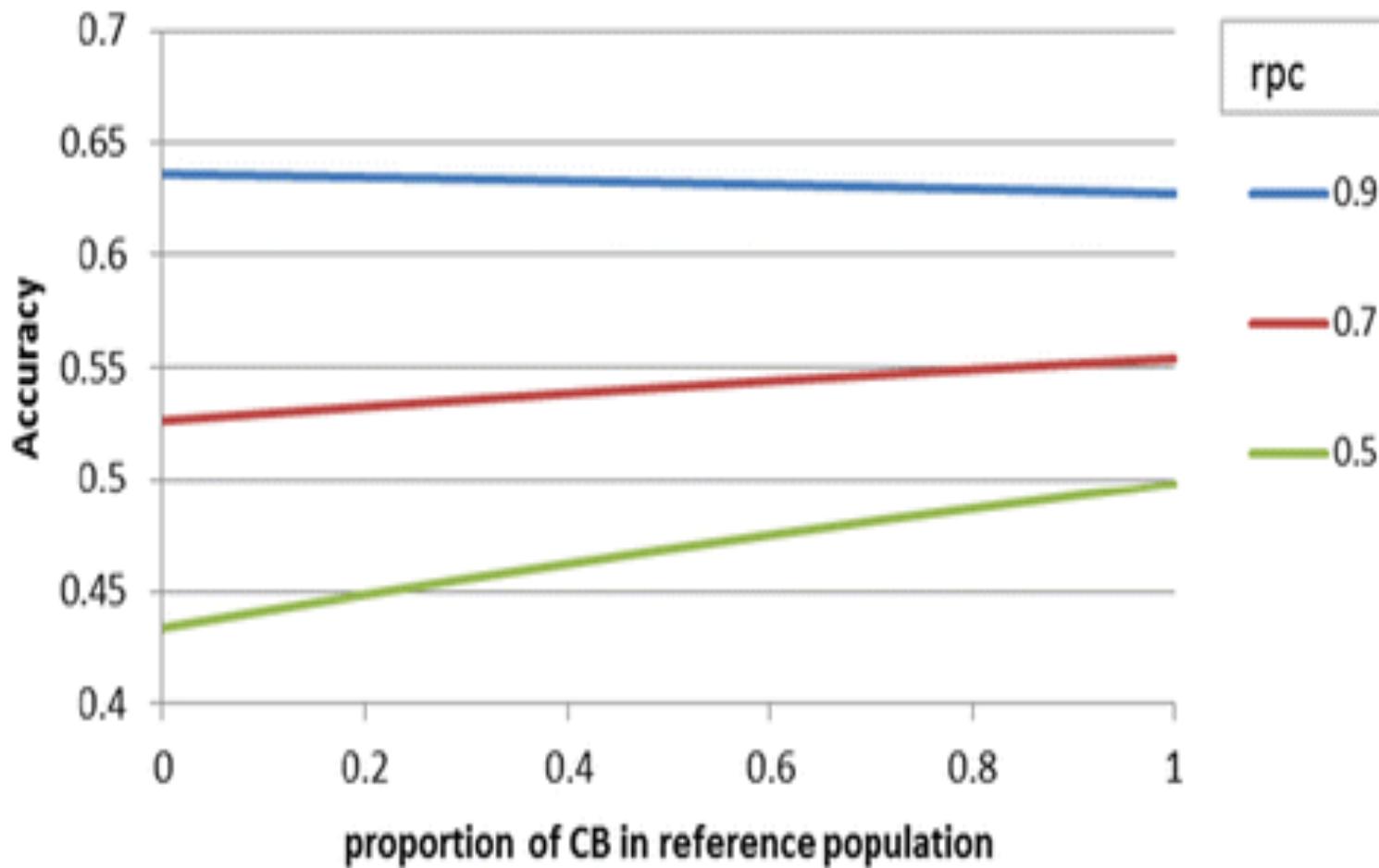
# Design of reference populations for CB selection

## Grevenhof and van der Werf GSE 2015



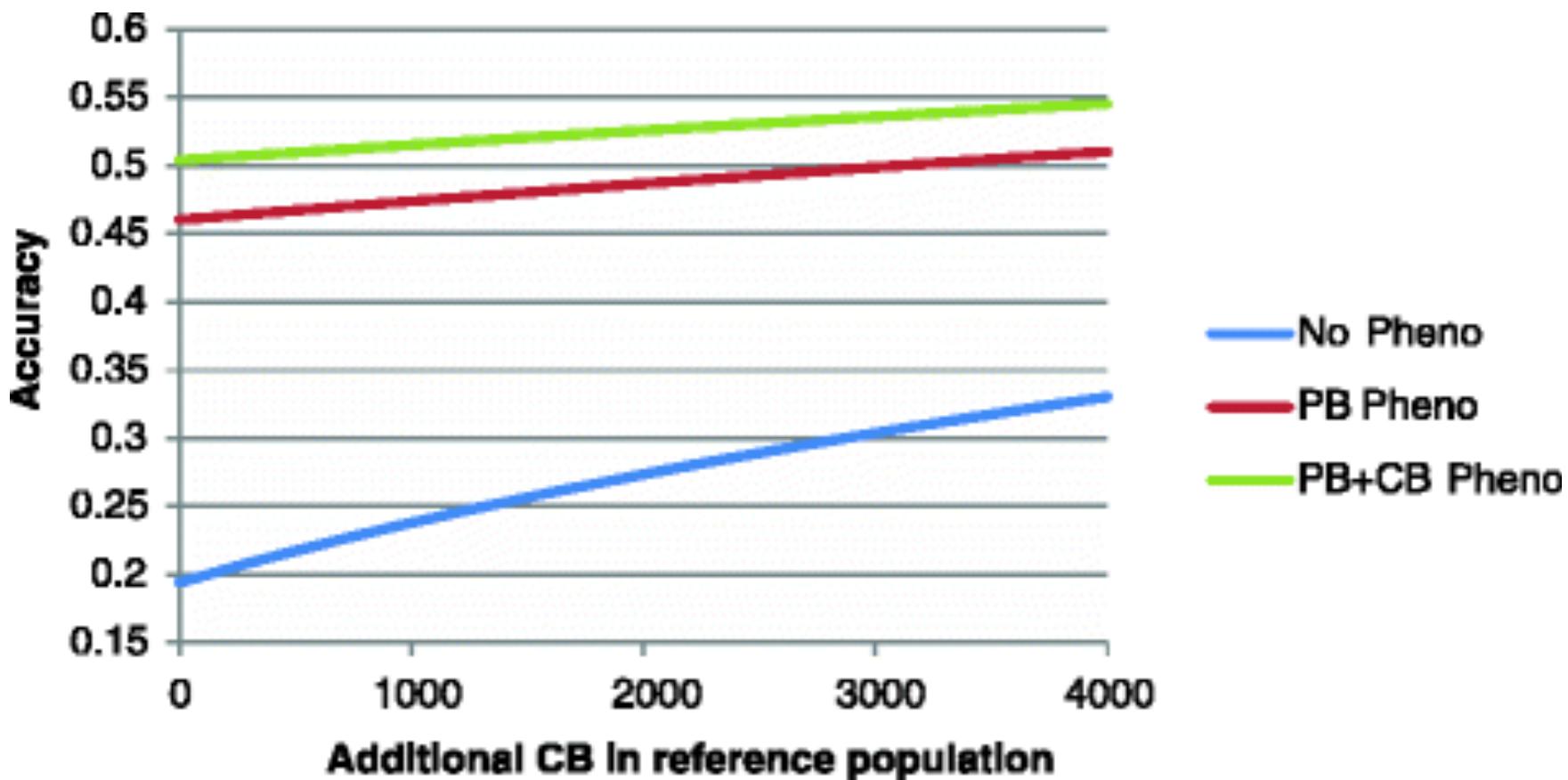
# Design of reference populations for CB selection

## Grevenhof and van der Werf GSE 2015



# Design of reference populations for CB selection

## Grevenhof and van der Werf GSE 2015



# Conclusions

Genomic Selection provides great opportunities for genetic improvement of commercial crossbred performance

- Increase  $\Delta G$  for commercial CB performance
- Reduce  $\Delta F$
- Requirements:
  - Routine phenotyping of crossbreds in field (GxE)
  - Genomic prediction methods that allow for breed-differences in allele frequencies and LD
    - Breed-specific allele models
    - Dominance models