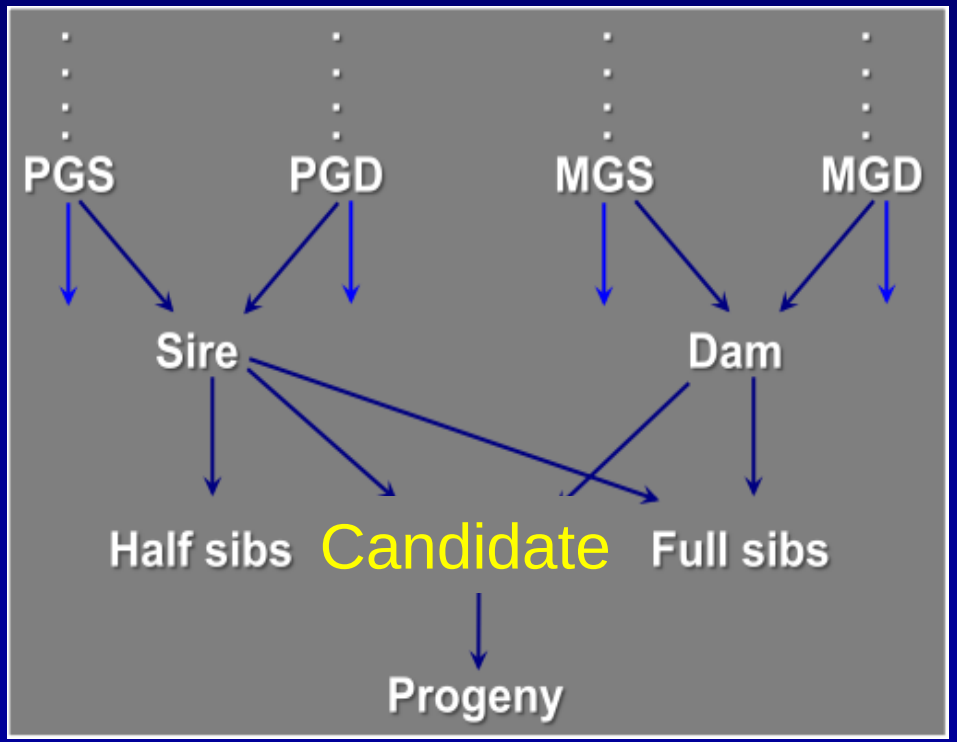
Reference Population Design
Retraining
etc.

Jack Dekkers

The Promise of Genomic Selection

Reduce requirement to get phenotypes on selection candidates and on close relatives in order to estimate EBV

Traditional EBV



Genomic Selection

How to build a Reference Population for a Closed Breeding Population?

Accuracy of GEBV is greater if

- More individuals are genotyped and phenotyped
- Heritability of phenotype is higher

  → genotype individuals that have high EBV accuracy

  + use deregressed EBV
- Selection candidates are more related to training data

  → include parental generation in training

- How important is it to achieve a 'target' accuracy of genomics right from the start?
  - You're going to need to retrain anyway?
  - Build-up training data during GS implementation

To Retrain or Not to Retrain

Results from Stochastic Simulation

Hong-hua Zhao, Jennifer Young,

David Habier, Rohan Fernando, Jack Dekkers

(unpublished)

Genus PIC Hy-Line MONSANTO Choice Genetics MAP YOUR SUCCESS

ANIMAL SCIENCE

CIAG

IOWA STATE UNIVERSITY

Animal Breeding & Genetics

Response from Genomic Selection - Simulation

Generation

0

20 chr of 150 cM

100,000 SNPs (freq. = ½ , LE)

Random mating, N_e =100

LD generated by drift and mutation

1000

Allocate 100 loci with MAF>0.1 as QTL and 2,000 as SNPs

1001

Expand pop.size to 1,000 – phenotype - h^2=0.3

Estimate marker effects by Bayes-B

Mate random 20 males to random 60 females

1002

Select 20/240 males 60/240 females

1012

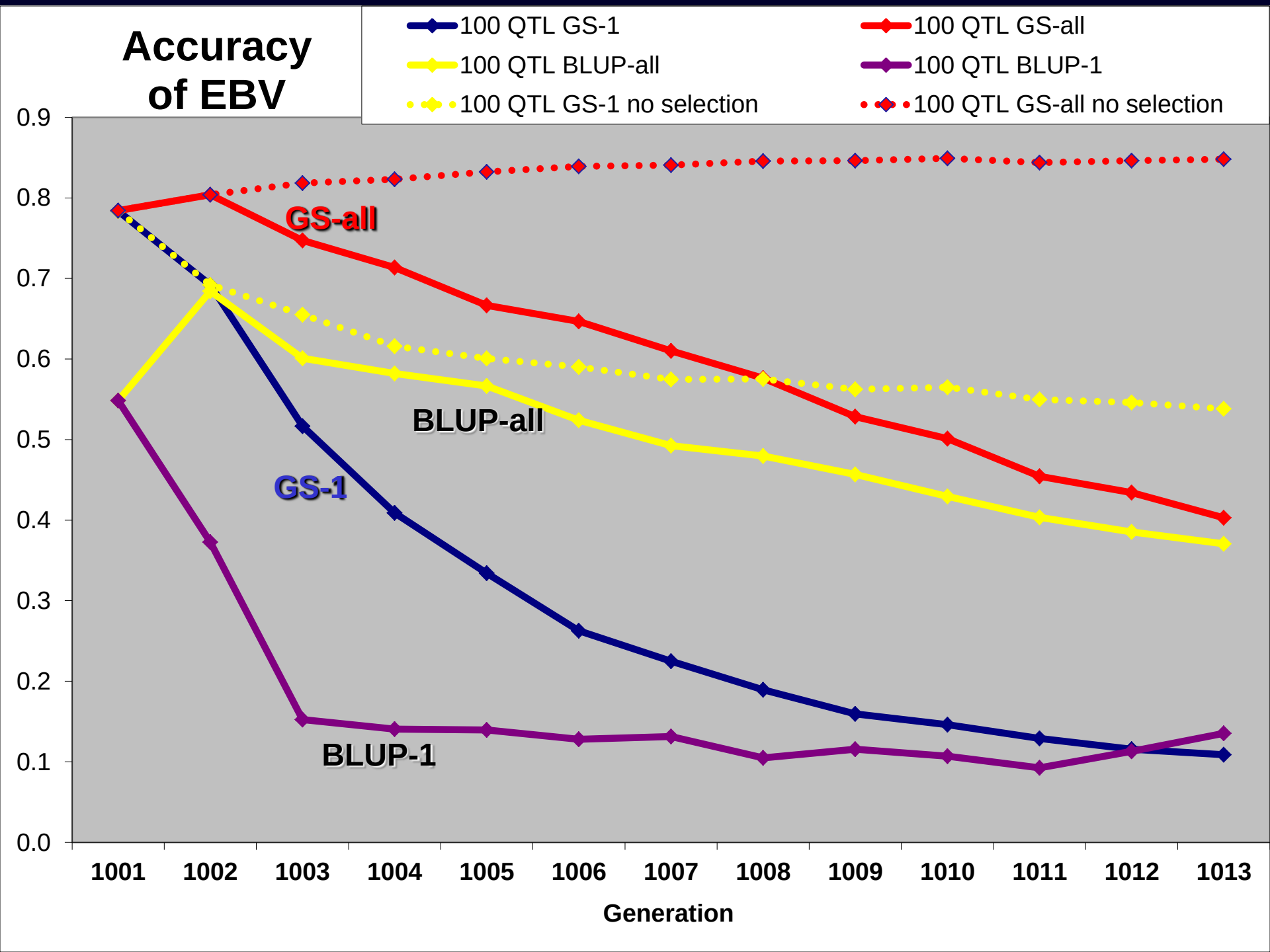
Select

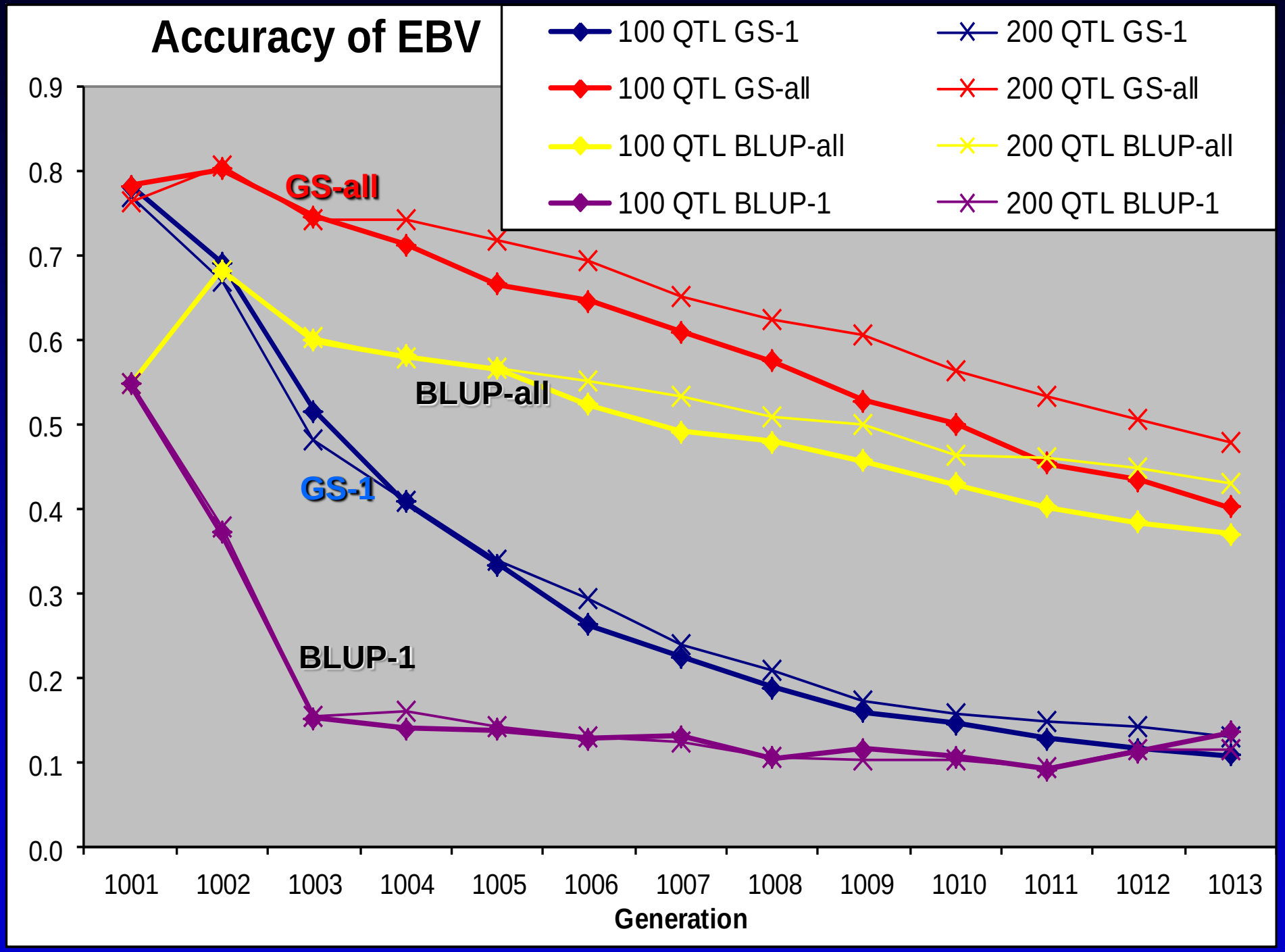
Strategies

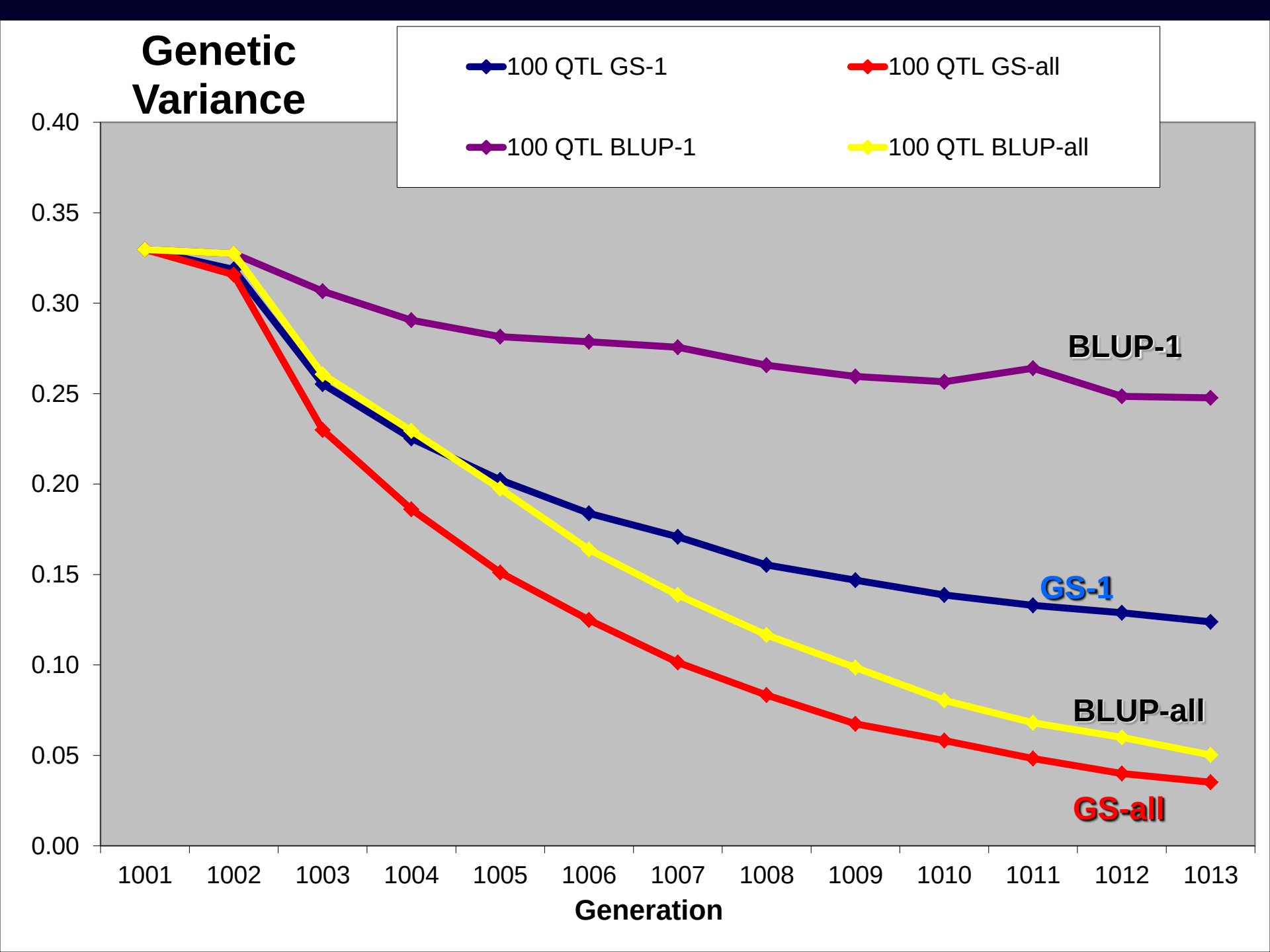
BLUP-1 = PBLUP – last phenotypes collected in G1001

BLUP-all= PBLUP – continuous phenotyping

GS-1 = Bayes-B GS – without retraining – no P after G1001


Accuracy of EBV
100 QTL GS-1
100 QTL GS-all
100 QTL BLUP-all
100 QTL BLUP-1
100 QTL GS-1 no selection
100 QTL GS-all no selection
GS-all
BLUP-all
GS-1
BLUP-1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
Generation

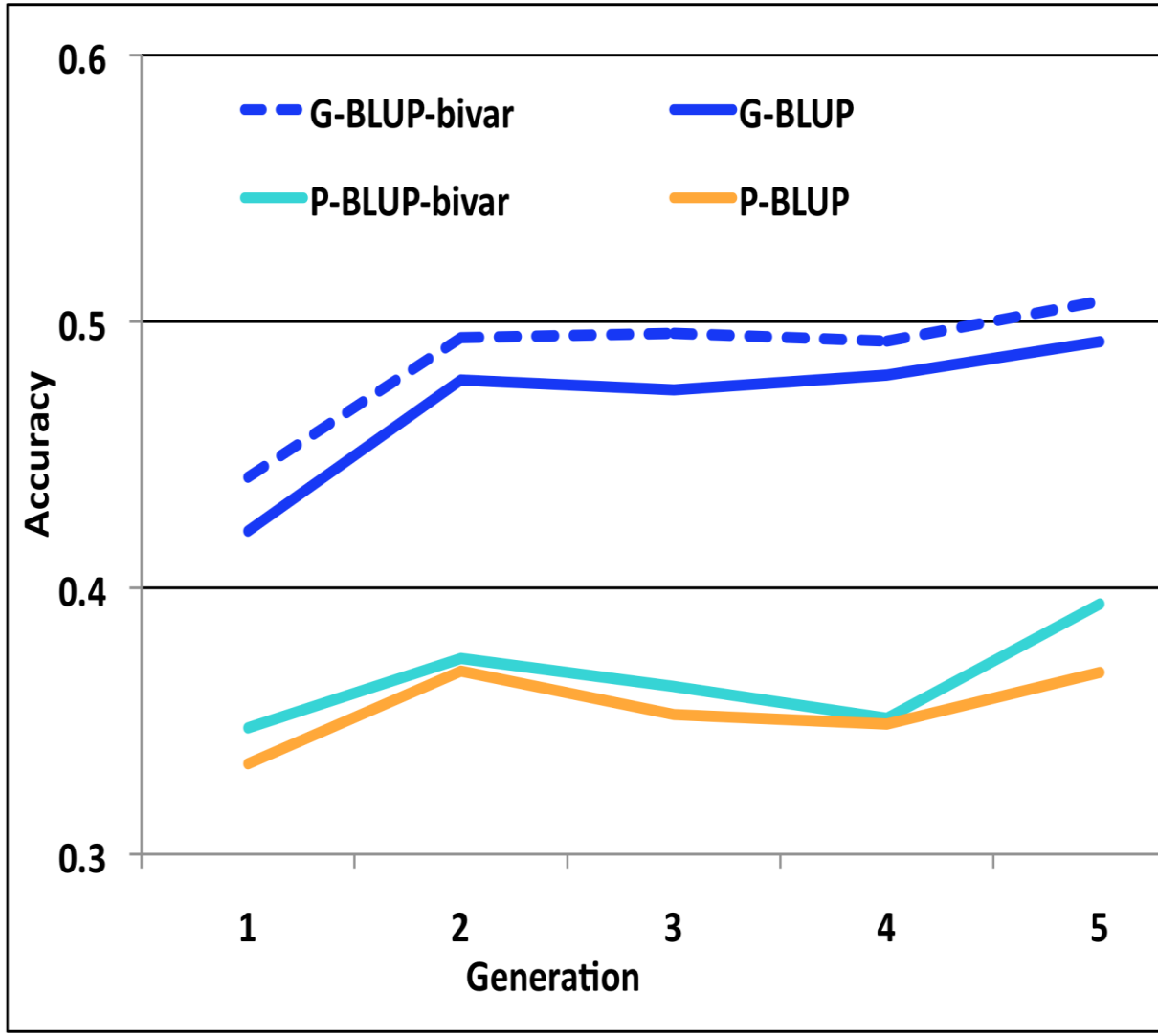

Accuracy of EBV
100 QTL GS-1
200 QTL GS-1
100 QTL GS-all
200 QTL GS-all
100 QTL BLUP-all
200 QTL BLUP-all
100 QTL BLUP-1
200 QTL BLUP-1
GS-all
BLUP-all
GS-1
BLUP-1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
Generation


Genetic Variance
100 QTL GS-1
100 QTL GS-all
100 QTL BLUP-1
100 QTL BLUP-all
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
BLUP-1
GS-1
BLUP-all
GS-all
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
Generation

Genomic Selection Training in a Layer Breeding Population Wolc et al. GSE, 2011

Accuracies averaged over 16 traits

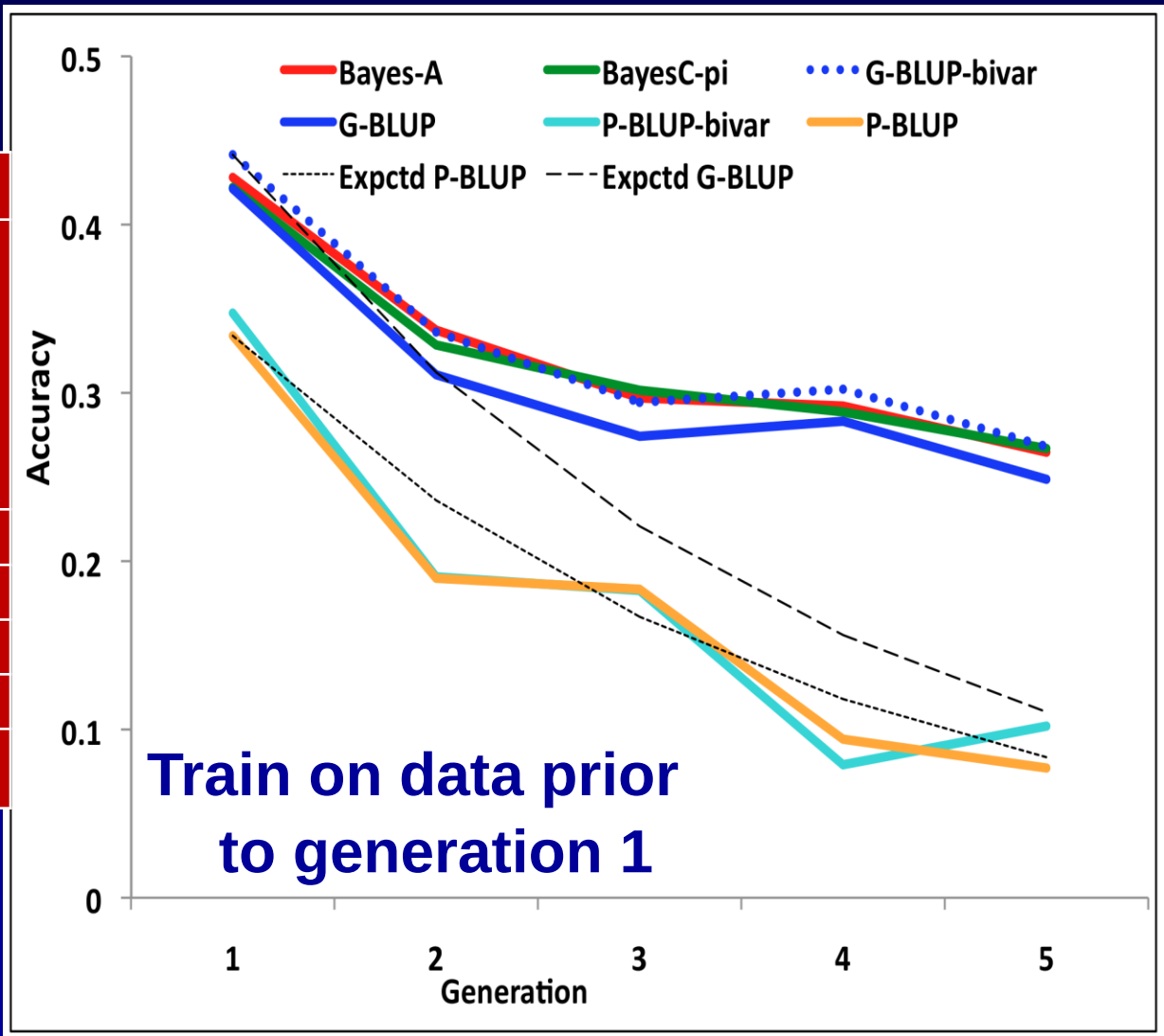
Size of training and validation data				
Ge-ne-ra-tion	Training data: # geno-typed	# with own record	# progeny with genotyped parents Early	Vali-dation data size
<1	777	295	2443	322
2	1215	618	4892	295
3	1628	913	7562	357
4	2108	1273	9319	274
5	2708	1563	11486	262



Need for Retraining

Wolc et al. (GSE, 2011)

Size of training and validation data				
Ge-ne-ra-tion	Training data: # geno-typed	# with own record	# progeny with genotyped parents Early	Vali-dation data size
<1	777	295	2443	322
2				295
3				357
4				274
5				262



How to build a Reference Population for a Closed Breeding Population?

Accuracy of GEBV is greater if

- More individuals are genotyped and phenotyped
- Heritability of phenotype is higher

  → genotype individuals that have high EBV accuracy

  + use deregressed EBV
- Selection candidates are more related to training data

  → include parental generation in training

- How important is it to achieve a 'target' accuracy of genomics right from the start?
  - You're going to need to retrain anyway?
  - Build-up training data during GS implementation

Strategy for Implementation of Genomic Selection within a breed/line (for Pigs & Poultry?)

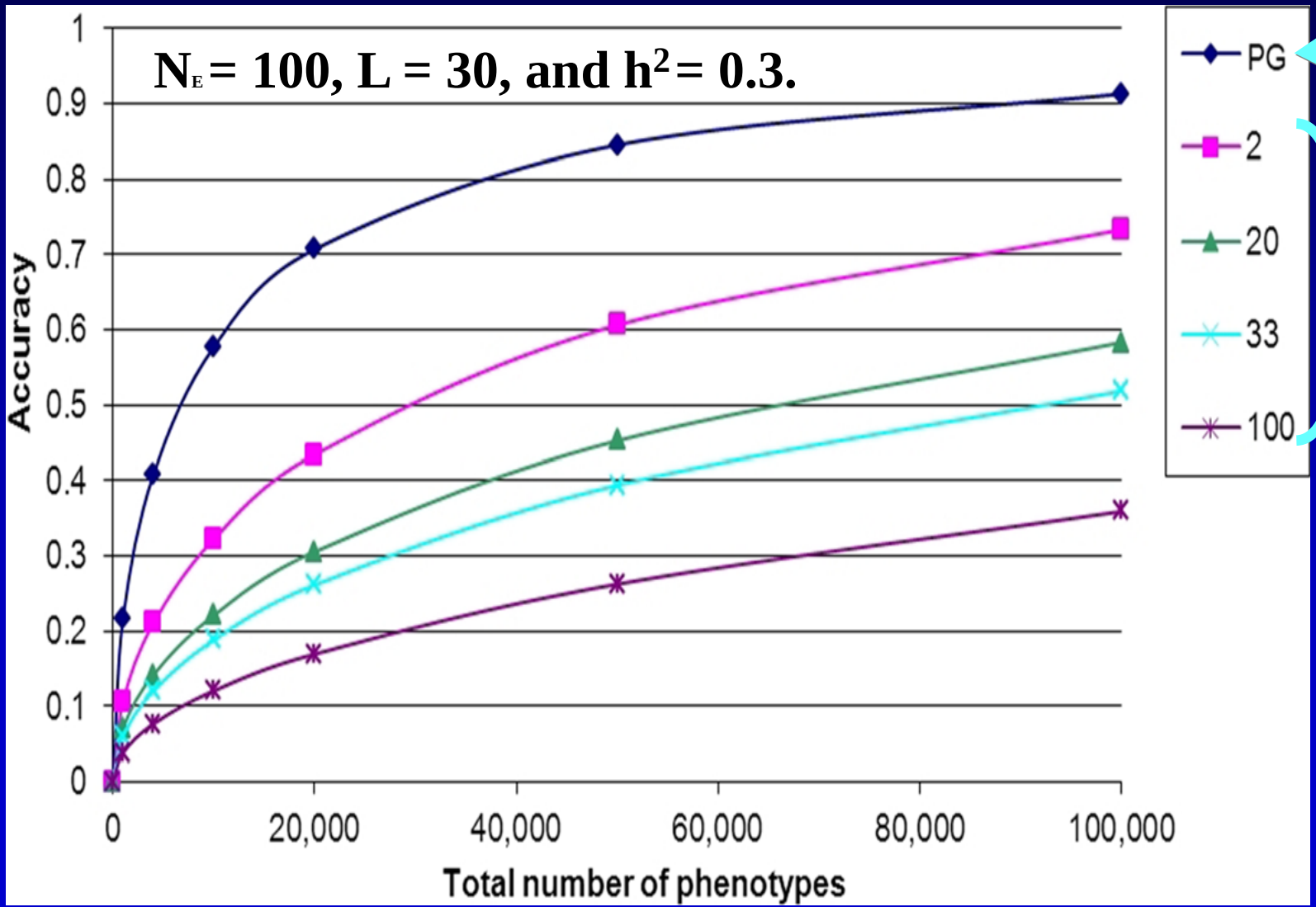
1. Genotype ≥3 generations of parents with HD panel
   - Use for initial training
2. Genotype selection candidates with ELD panel
   - Impute HD genotypes and compute G-EBV for selection
3. Re-genotype selected sires (and dams) with HD panel
4. Retrain with new data on LD/HD-genotyped animals

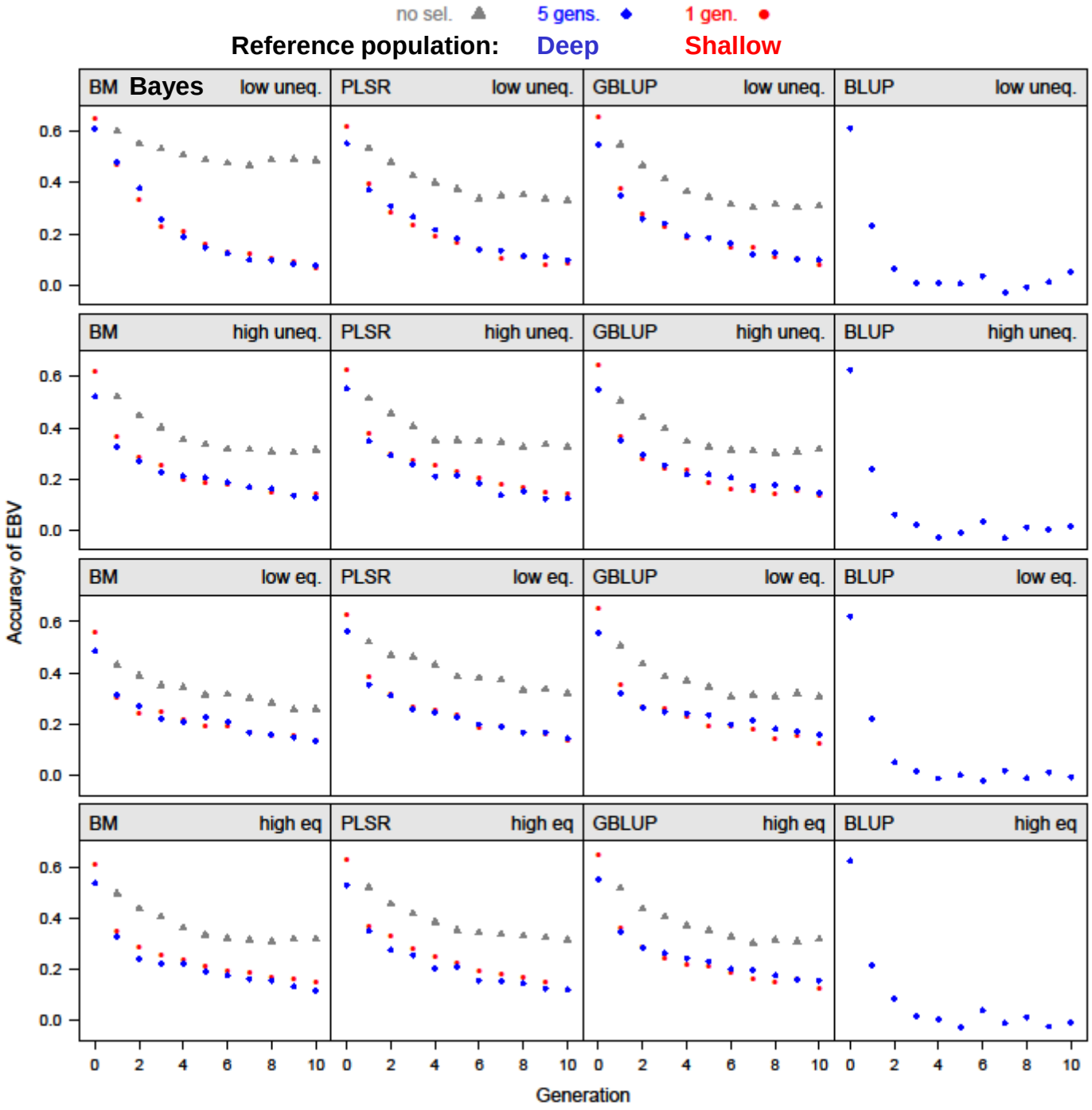
Reference Population for 'New Traits'

when # phenotypes is limited and genotyping is not

Genotype individuals with phenotype, rather than parents

Grevenhof, Bijma, van Arendonk GSE 2012





Bastiaansen et al GSE '12

Deep vs. Shallow Reference pop.

N=500 in reference

-In 1 generation

- or across 5 gens.

Accuracy of EBV averaged over 30 replicates

No Retraining

Low uneq = low # QTL (30) unequal variance

Low eq. = low # QTL (30) equal variance

High uneq = high # QTL (300) unequal variance

High eq. = high # QTL (300) equal variance

Shallow has advantage only in first generation

Which individuals should be entered into central test stations?

Potential bull dams? Konig and Swalve JDS 2009

Limited gain in accuracy EBV of bulldams with addition of own record

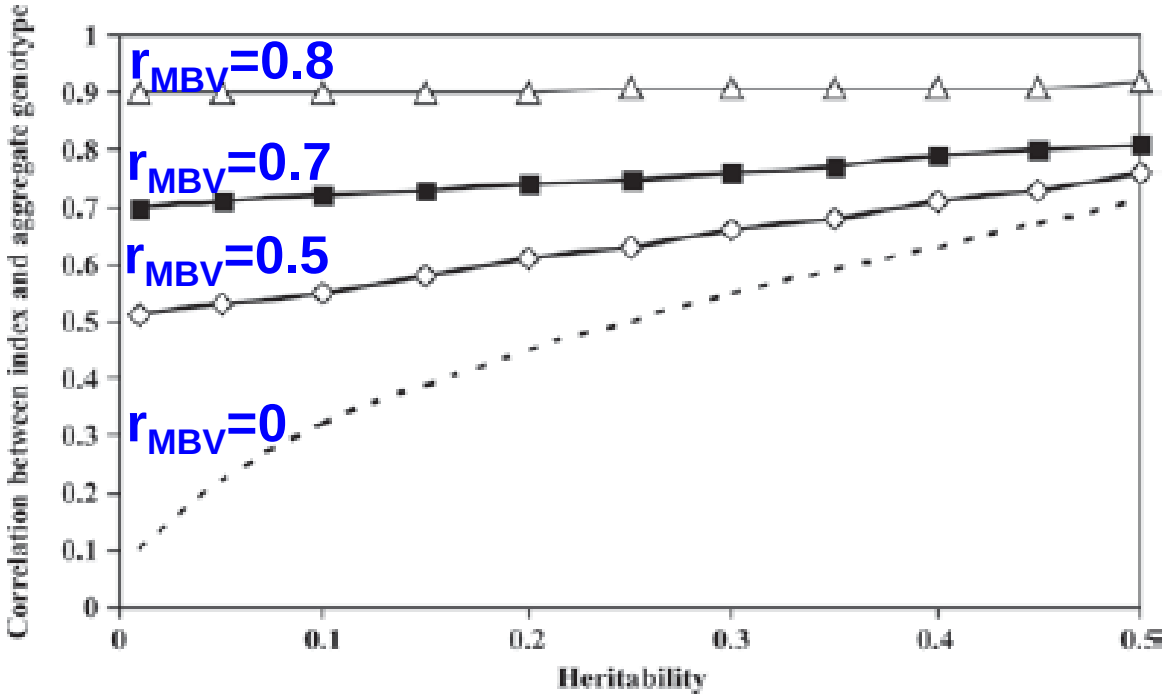


Figure 2. Correlation between index and aggregate genotype (r_{TI}) for scenario I by altering the heritability of the trait and the correlation between the true breeding value and genomic EBV (r_{mg}). Dotted line: index without genomic information; solid line with open diamonds: $r_{mg} = 0.5$; solid line with solid squares: $r_{mg} = 0.7$; solid line with open triangles: $r_{mg} = 0.9$.

Low density genotyping and Imputation

Jack Dekkers

Animal Breeding & Genetics
Department of Animal Science
Iowa State University

IOWA STATE UNIVERSITY

Animal Breeding & Genetics

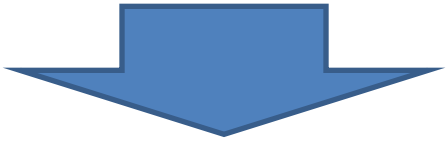
ANIMAL SCIENCE

Implementing GS in Pig/Poultry Programs

Problem

High cost of genotyping ←→value of an individual

Very large numbers of selection candidates



Impossible to implement genomic selection based on high density genotyping in cost efficient manner

Solution

Combination of strategic genotyping and imputation

Information used for imputation

- **LD across the population**
  - To impute from medium density (>10,000 SNP) to high density – up to sequence
- **Linkage within families**
  - To impute from very low density (<1000 SNP) to high density

Imputation using population-wide LD

	Haplotypes
HD-genotype	ACAAGGATTCCGAT
HD-genotype	GCTATCATGCCTAT
LD-genotype	--T---A-----T--

Imputation using population-wide LD

	Haplotypes
HD-genotype	ACAAGGATTCCGAT
HD-genotype	GCTATCATGCCTAT
LD-genotype	--T---A-----T--

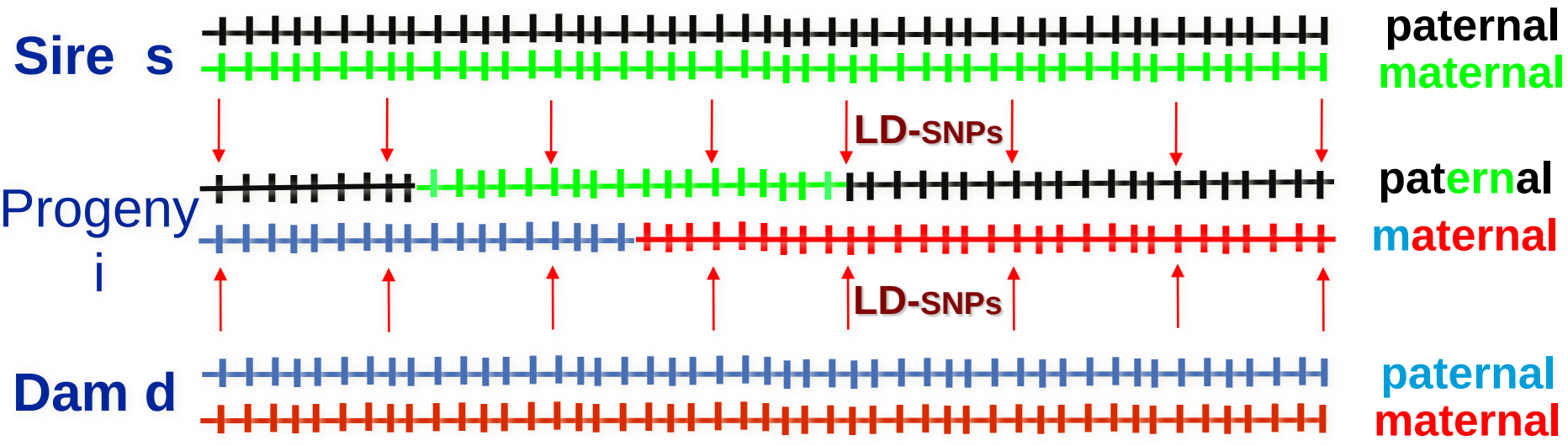
Imputation using population-wide LD

	Haplotypes
HD-genotype	ACAAGGATTCCGAT
HD-genotype	GCTATCATGCCTAT
LD-genotype	GCTATCATGCCTAT

Information used for imputation

- **LD across the population**
  - To impute from medium density (>10,000 SNP) to high density – up to sequence
- **Linkage within families** – as explained before
  - To impute from very low density (<1000 SNP) to high density

Imputation based on Linkage Information



HD-GS → $EBV_i = \sum_{SNP\ k} \left(g^m_{ik} + g^p_{ik} \right)$

Sum estimates of effects of maternal and paternal SNP alleles

LD-GS → $EBV_i = \sum_{SNP\ k} \left(p^{md}_{ik} g^m_{dk} + p^{pd}_{ik} g^p_{dk} + p^{ms}_{ik} g^m_{sk} + p^{ps}_{ik} g^p_{sk} \right)$

Probability that i received dam's maternal allele at SNP k

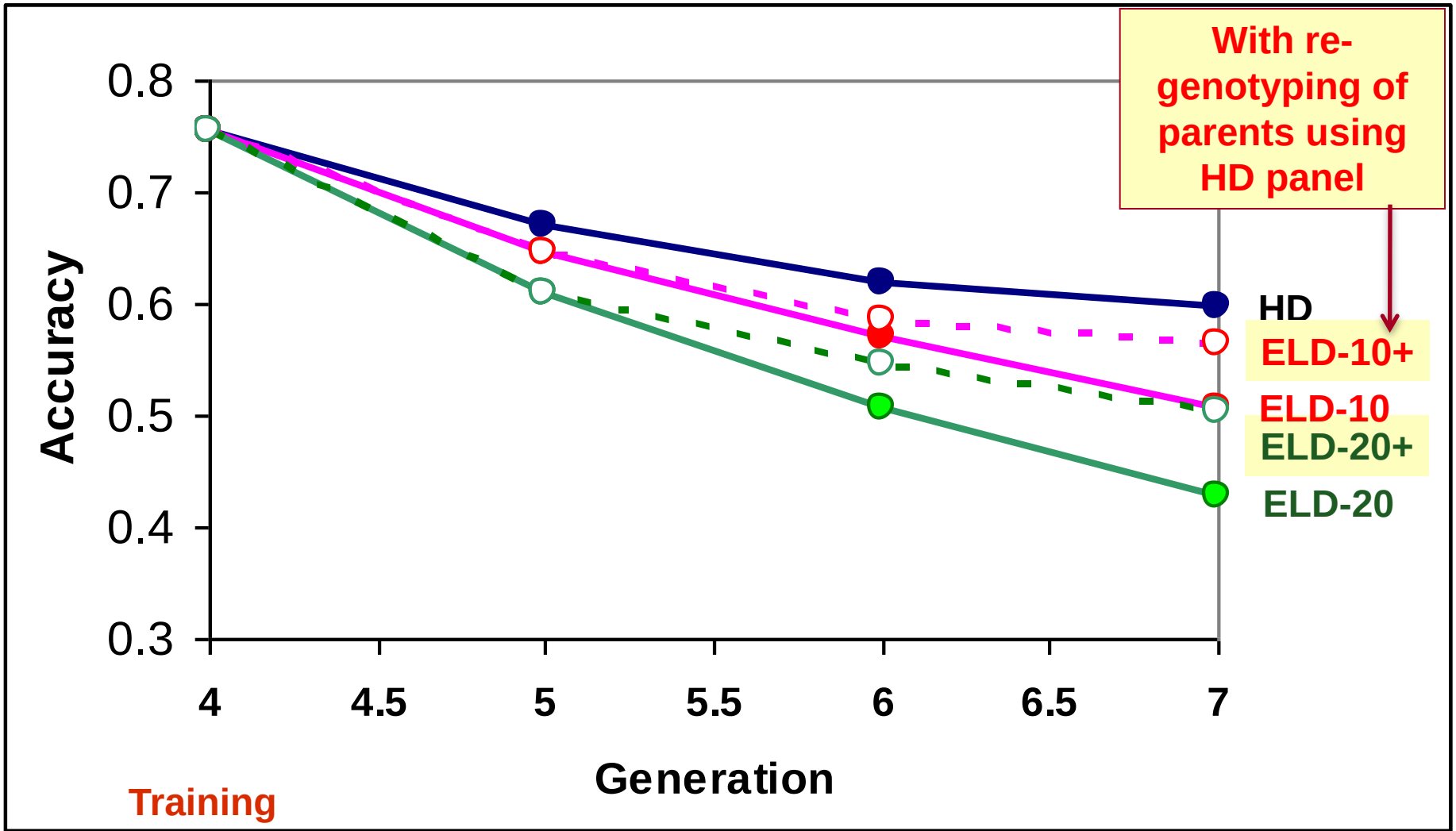
PDM = Probability of Descent of Marker allele

Requirements:

- **Ordered/phased HD SNP genotypes of parents**
- **Imputation of HD SNP genotypes on progeny**

Accuracy of G-EBV based on High- vs Low-Density SNP genotyping

Simulation (Habier et al. 2009 Genetics)



Genetic Excellence®

Imputation results in HyLine data

Neil O'Sullivan, Janet Fulton, Petek Settar and Jesus Arango

HY-LINE INTERNATIONAL

Anna Wolc, David Habier, John Hickey, Mehdi Sargolzaei, Dorian Garrick, Rohan Fernando, Nathan Bowerman, Chunkao Wang, Jack Dekkers,

IOWA STATE UNIVERSITY, USA
POZNAN UNIVERSITY OF LIFE SCIENCES, POLAND
UNIVERSITY OF NEW ENGLAND, AUSTRALIA
UNIVERSITY OF GUELPH, CANADA

Hy-Line®

Hy-Line data

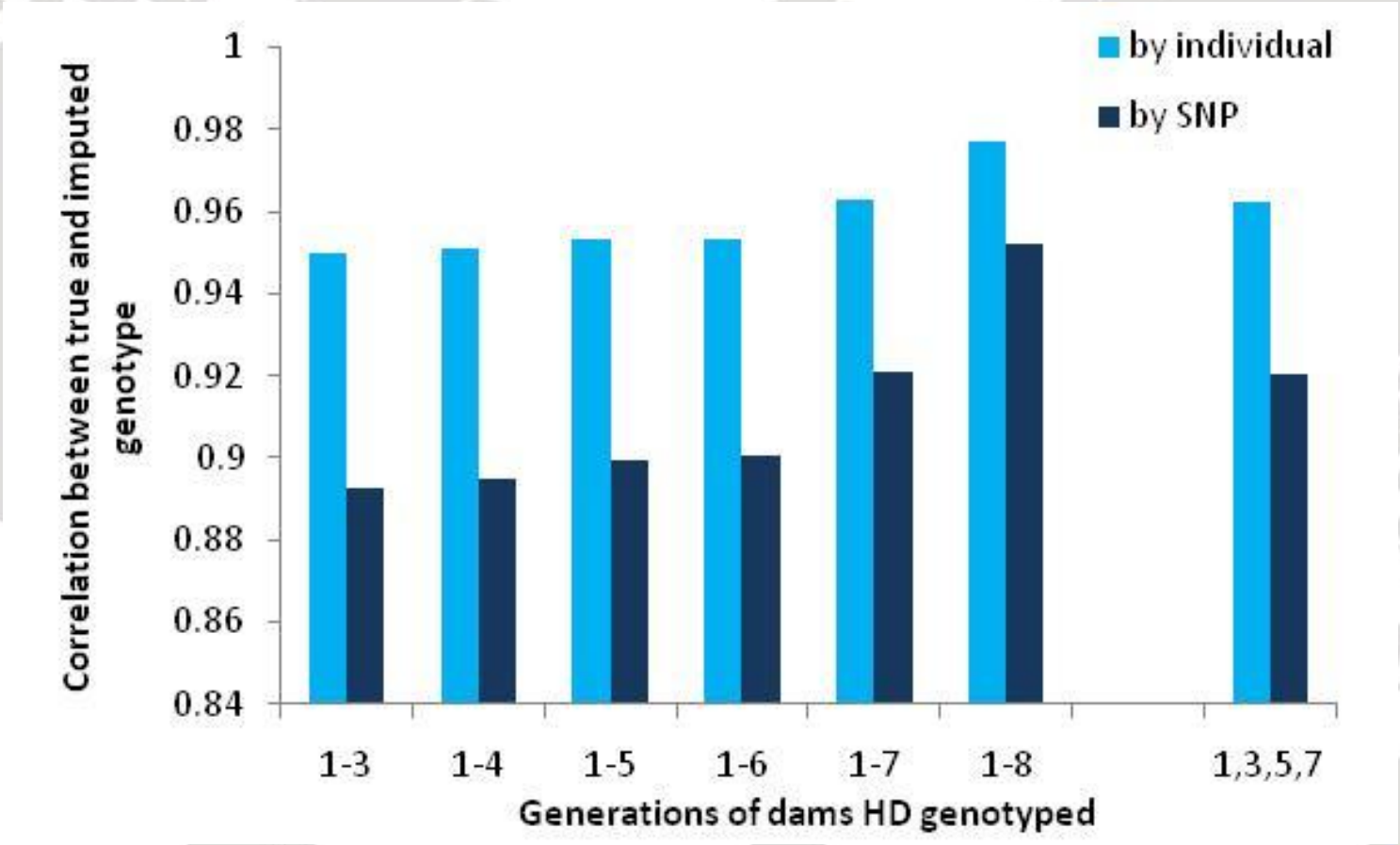
- 8 generations of HD sires and dams
- Selection candidates : 544 individuals from generation 9

- **High Density genotypes** - 4,893 segregating SNPs on chromosome 1

- **Low Density genotypes** – Simulated panel of 73 ~equally spaced SNPs (equivalent to ~400 SNP across the genome)

Hy-Line®

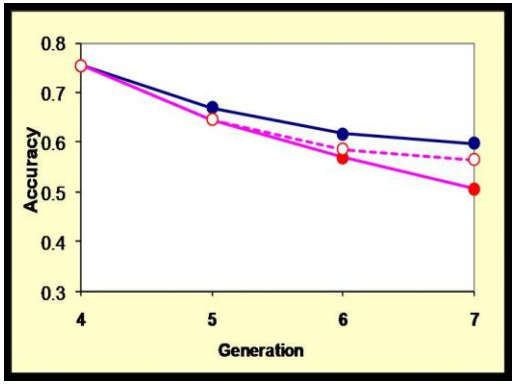
Imputation with multiple generations of Low Density genotyped females



Accuracy of imputed genotypes in generation 9 with 8 generations of sires and various generations of dams HD genotyped, and the remaining generations of dams Low Density genotyped.

Hy-Line®

Genomic Selection using Low-Density SNPs



Conclusions

GS can be implemented by genotyping selection candidates for <400 SNPs spread across the genome

- Loss in accuracy limited: < 5 % - if parents re-genotyped HD
  - sufficient to genotype only sires
- Cost effectiveness depends on cost of Low- vs. High-density genotyping
  $20 ←??→ $150
- Loss in accuracy ~ independent of # QTL and # traits
- LD-genotyped individuals can also be used for training
- Allows imputing to higher densities / sequence from founders