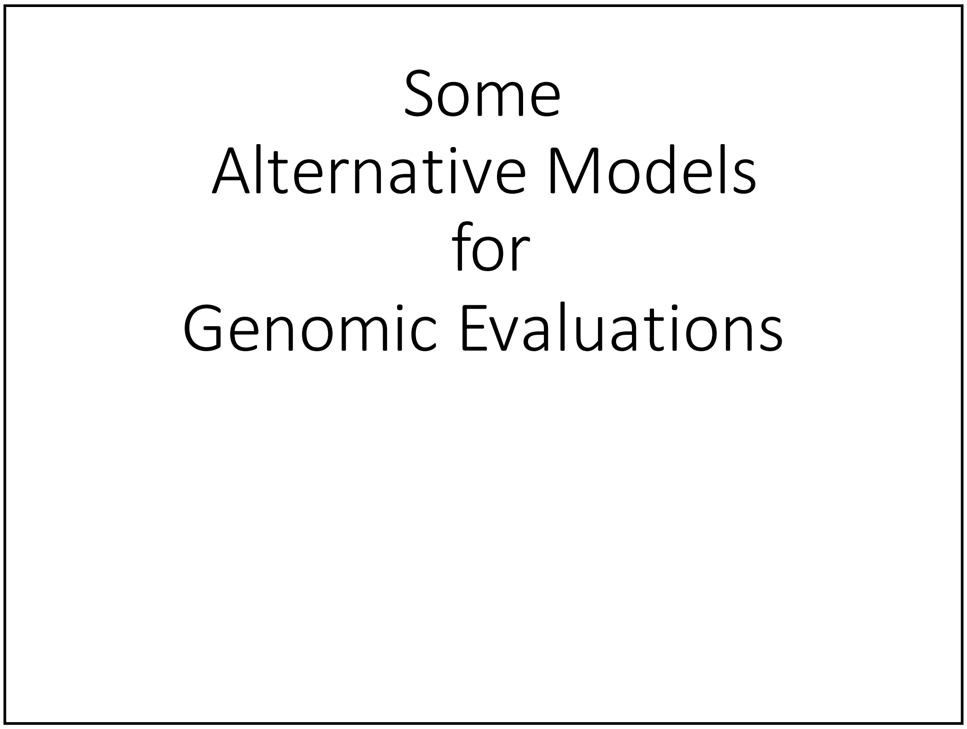
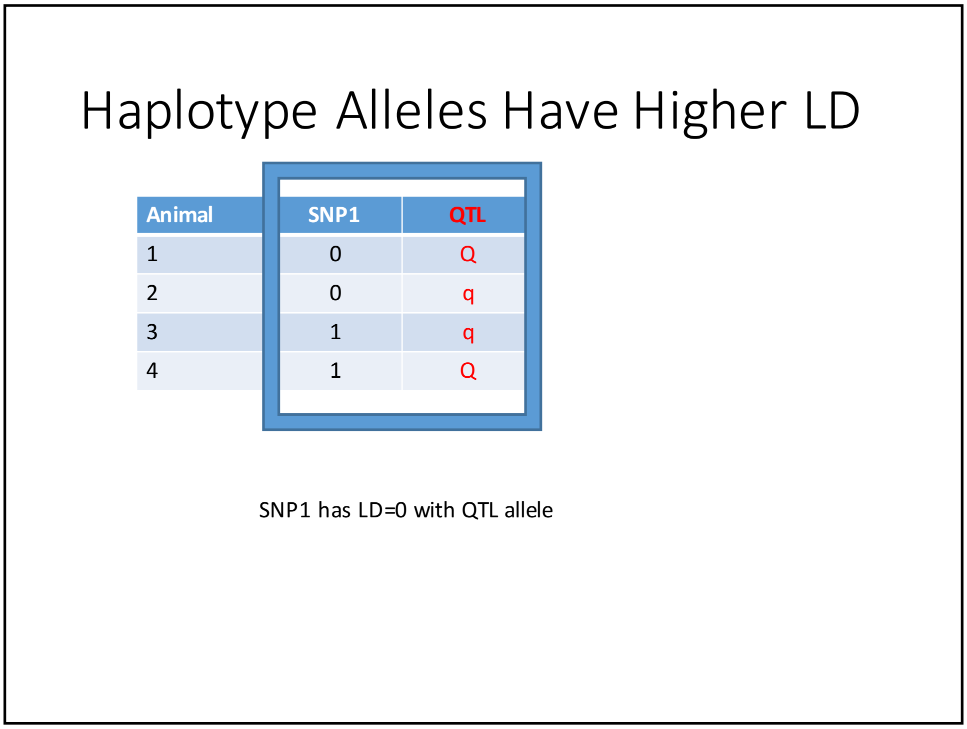


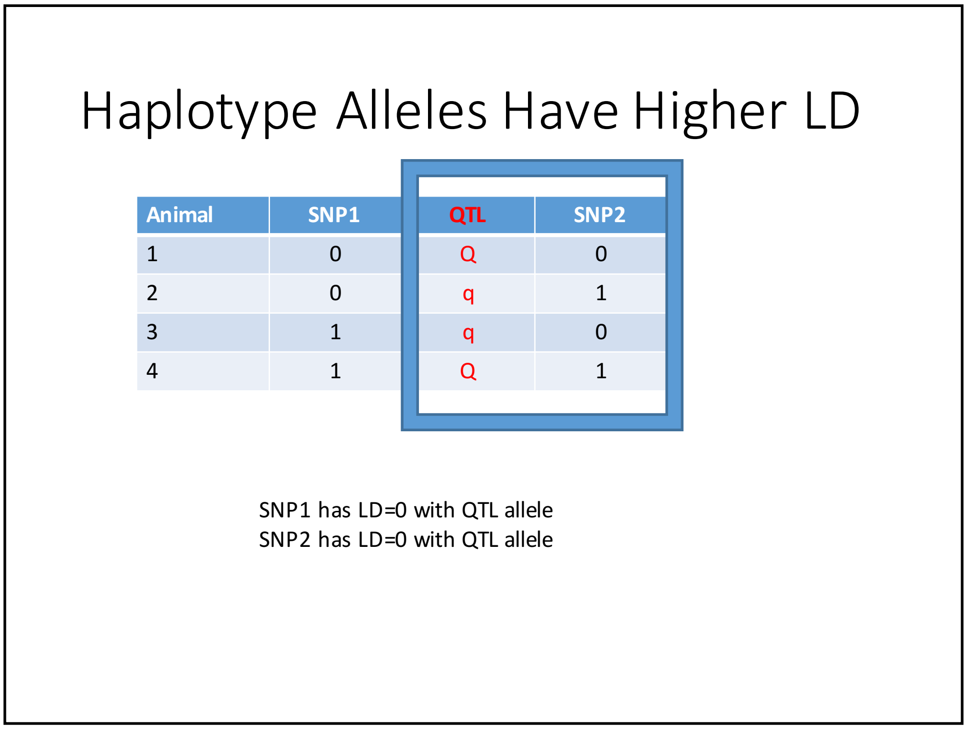
Some
Alternative Models
for
Genomic Evaluations



Haplotype Alleles Have Higher LD

Animal	SNP1	QTL
1	0	Q
2	0	q
3	1	q
4	1	Q

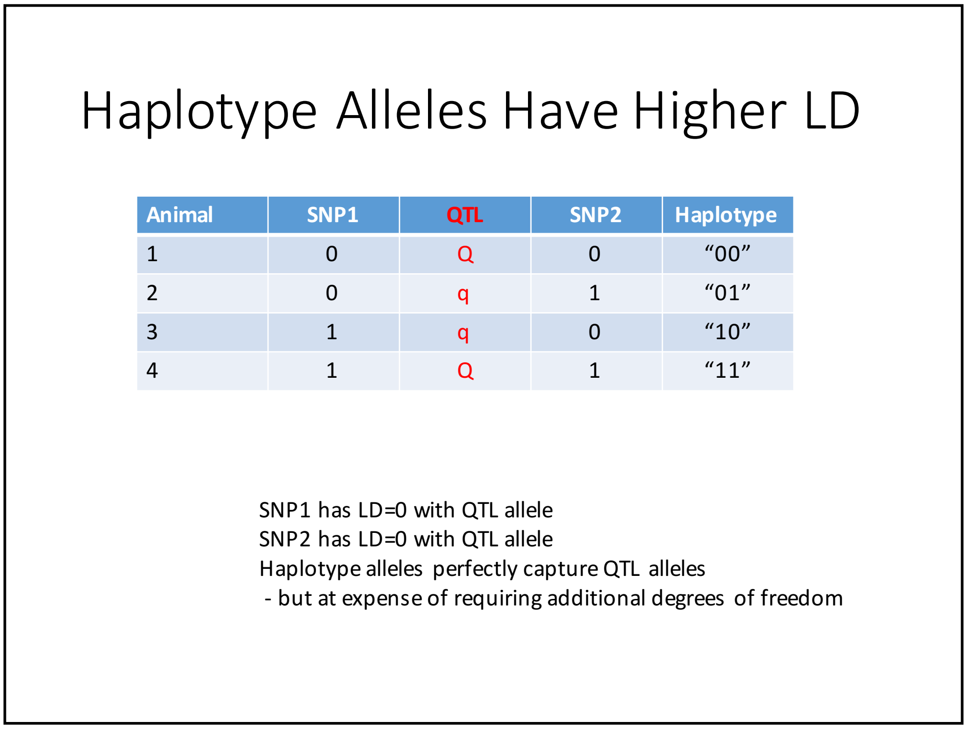
SNP1 has LD=0 with QTL allele



Haplotype Alleles Have Higher LD

Animal	SNP1	QTL	SNP2
1	0	Q	0
2	0	q	1
3	1	q	0
4	1	Q	1

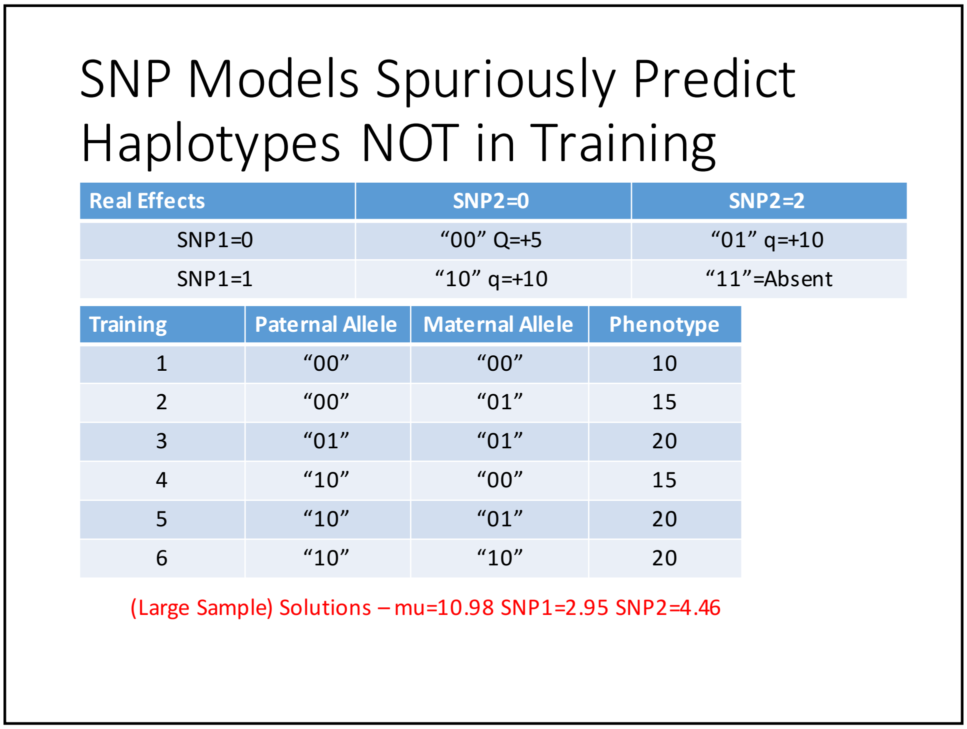
SNP1 has LD=0 with QTL allele
SNP2 has LD=0 with QTL allele



Haplotype Alleles Have Higher LD

Animal	SNP1	QTL	SNP2	Haplotype
1	0	Q	0	“00”
2	0	q	1	“01”
3	1	q	0	“10”
4	1	Q	1	“11”

SNP1 has LD=0 with QTL allele
SNP2 has LD=0 with QTL allele
Haplotype alleles perfectly capture QTL alleles
- but at expense of requiring additional degrees of freedom

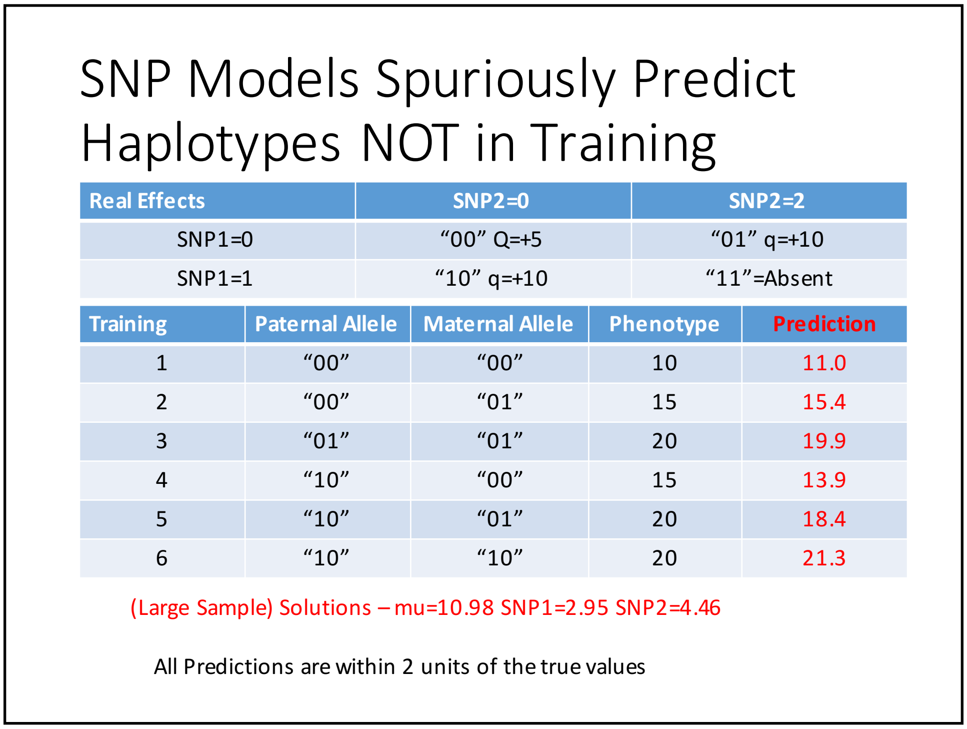


SNP Models Spuriously Predict Haplotypes NOT in Training

Real Effects	SNP2=0	SNP2=2
SNP1=0	“00” Q=+5	“01” q=+10
SNP1=1	“10” q=+10	“11”=Absent

Training	Paternal Allele	Maternal Allele	Phenotype
1	“00”	“00”	10
2	“00”	“01”	15
3	“01”	“01”	20
4	“10”	“00”	15
5	“10”	“01”	20
6	“10”	“10”	20

(Large Sample) Solutions – mu=10.98 SNP1=2.95 SNP2=4.46



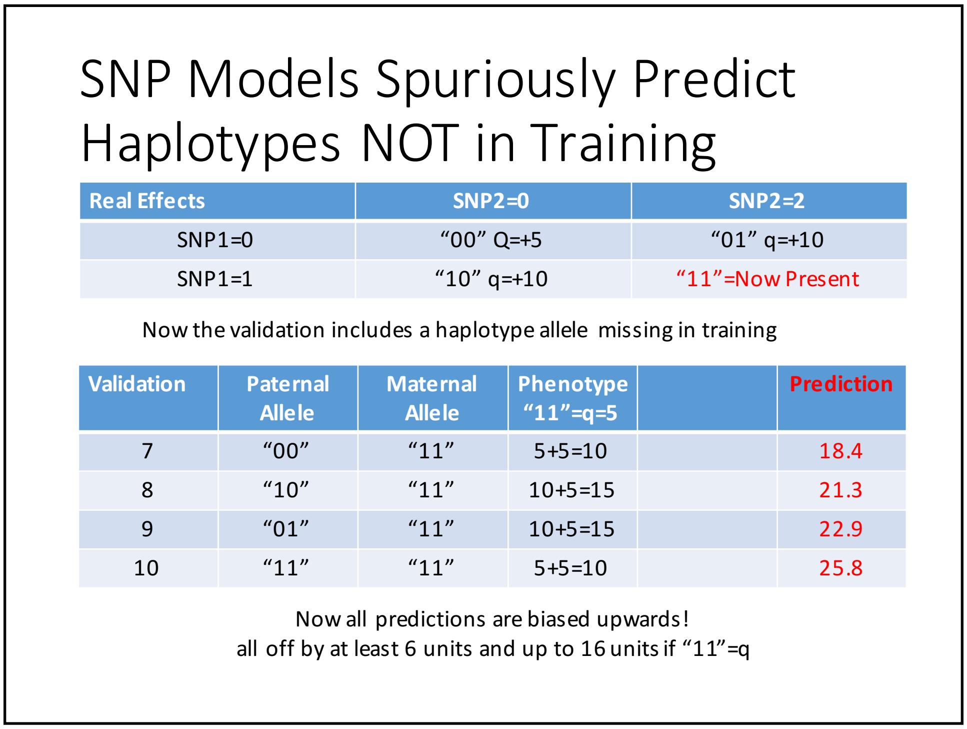
SNP Models Spuriously Predict Haplotypes NOT in Training

Real Effects	SNP2=0	SNP2=2
SNP1=0	“00” Q=+5	“01” q=+10
SNP1=1	“10” q=+10	“11”=Absent

Training	Paternal Allele	Maternal Allele	Phenotype	Prediction
1	“00”	“00”	10	11.0
2	“00”	“01”	15	15.4
3	“01”	“01”	20	19.9
4	“10”	“00”	15	13.9
5	“10”	“01”	20	18.4
6	“10”	“10”	20	21.3

(Large Sample) Solutions – mu=10.98 SNP1=2.95 SNP2=4.46

All Predictions are within 2 units of the true values



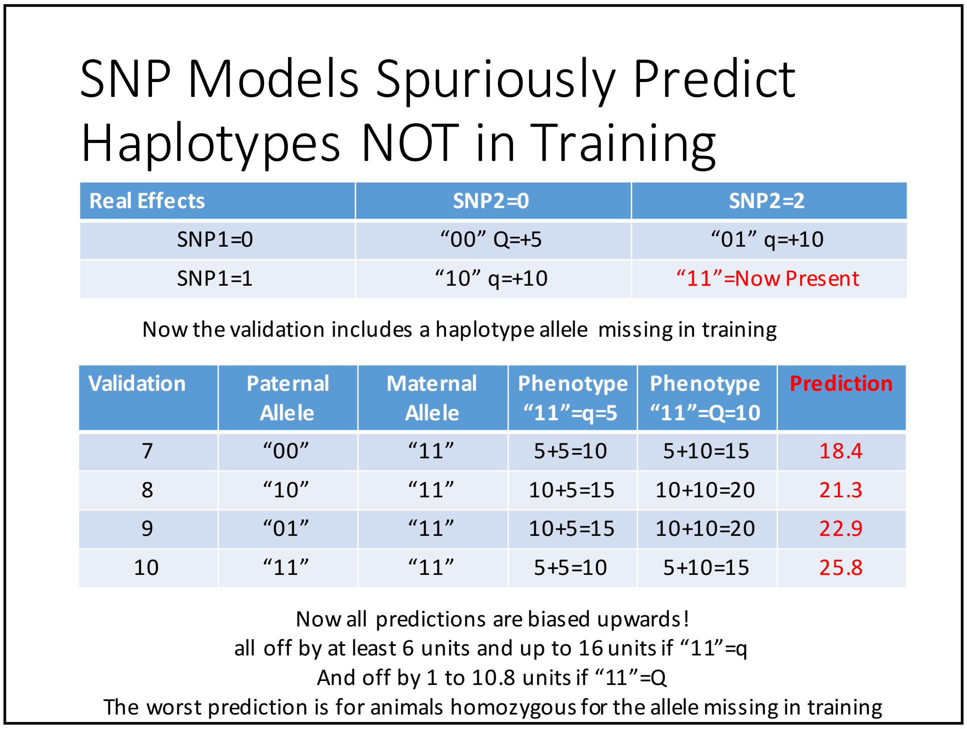
SNP Models Spuriously Predict Haplotypes NOT in Training

Real Effects	SNP2=0	SNP2=2
SNP1=0	“00” Q=+5	“01” q=+10
SNP1=1	“10” q=+10	“11”=Now Present

Now the validation includes a haplotype allele missing in training

Validation	Paternal Allele	Maternal Allele	Phenotype “11”=q=5		Prediction
7	“00”	“11”	5+5=10		18.4
8	“10”	“11”	10+5=15		21.3
9	“01”	“11”	10+5=15		22.9
10	“11”	“11”	5+5=10		25.8

Now all predictions are biased upwards!
all off by at least 6 units and up to 16 units if “11”=q



SNP Models Spuriously Predict Haplotypes NOT in Training

Real Effects	SNP2=0	SNP2=2
SNP1=0	“00” Q=+5	“01” q=+10
SNP1=1	“10” q=+10	“11”=Now Present

Now the validation includes a haplotype allele missing in training

Validation	Paternal Allele	Maternal Allele	Phenotype “11”=q=5	Phenotype “11”=Q=10	Prediction
7	“00”	“11”	5+5=10	5+10=15	18.4
8	“10”	“11”	10+5=15	10+10=20	21.3
9	“01”	“11”	10+5=15	10+10=20	22.9
10	“11”	“11”	5+5=10	5+10=15	25.8

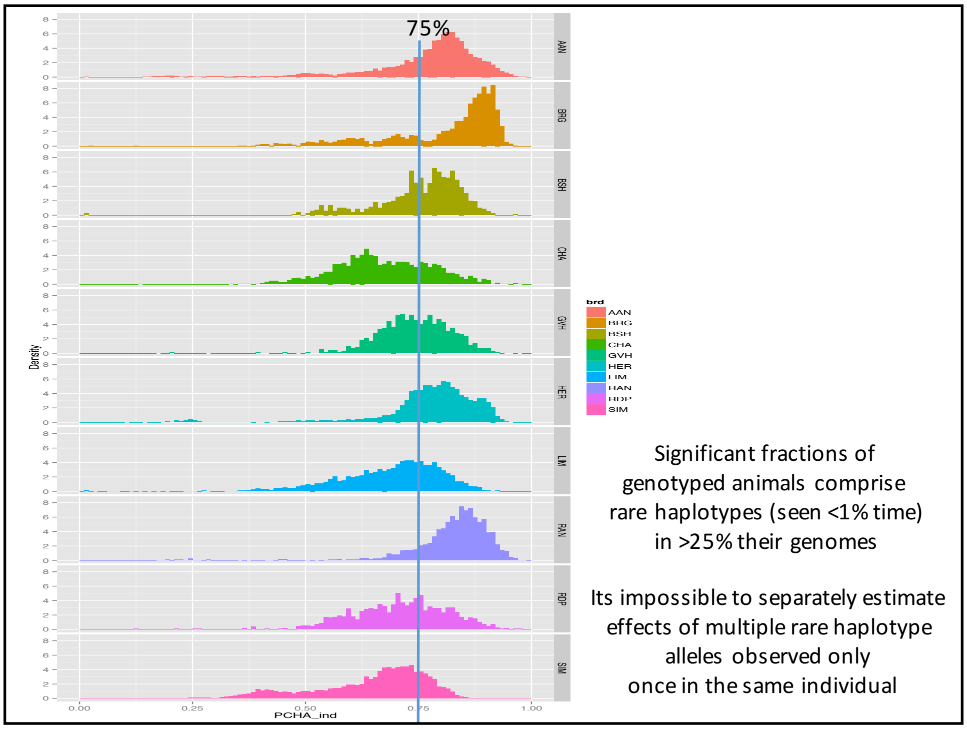
Now all predictions are biased upwards!
all off by at least 6 units and up to 16 units if “11”=q
And off by 1 to 10.8 units if “11”=Q
The worst prediction is for animals homozygous for the allele missing in training

Conclusions - There are good reasons to fit:

- Common haplotype alleles
  - They can capture more LD with QTL than SNP sets that do not include the causal mutations
  - The model means that alleles seen for the first time in validation will be recognized as unknown and predicted to be “average” rather than possibly extreme
- And SNP alleles (as well), particularly if these include possible causal mutations as they could fit the causal mutation with 1 degree of freedom
  - assuming a mixture model

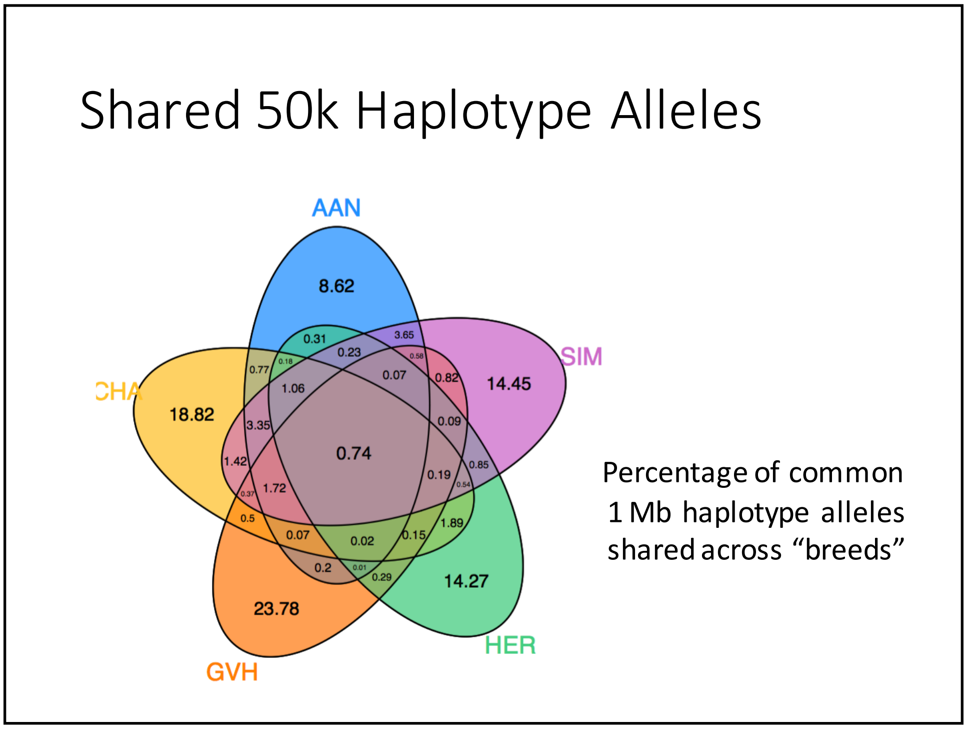
Haplotypes Detected from 50K

- Average 1Mb window contains about 20 SNP
- There are >1 million haplotypes possible per window
- Many of the observed haplotypes are only seen once
  - Result of a genotyping error
  - Maternally inherited alleles tracing to maternal founders
- Common haplotypes (seen at least 1% time) average from 15-20 variants per 1 Mb window
- Its not possible to reliably estimate rare haplotypes (seen <1% time)
- But most animals carry numerous rare haplotypes



Significant fractions of genotyped animals comprise rare haplotypes (seen <1% time) in >25% their genomes

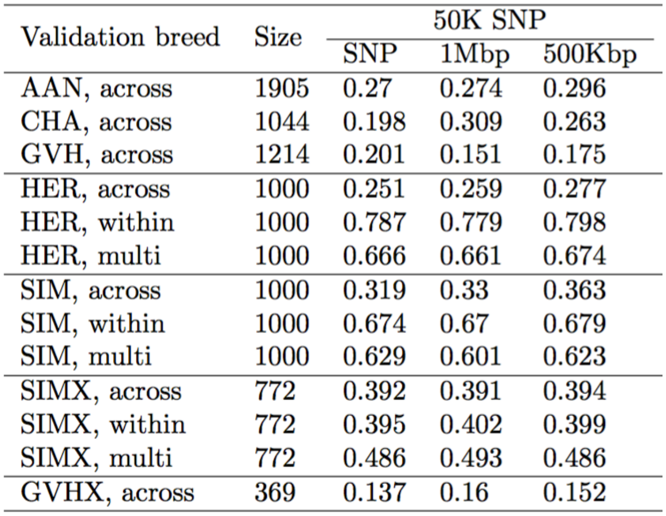
Its impossible to separately estimate effects of multiple rare haplotype alleles observed only once in the same individual



Shared 50k Haplotype Alleles

Percentage of common 1 Mb haplotype alleles shared across "breeds"

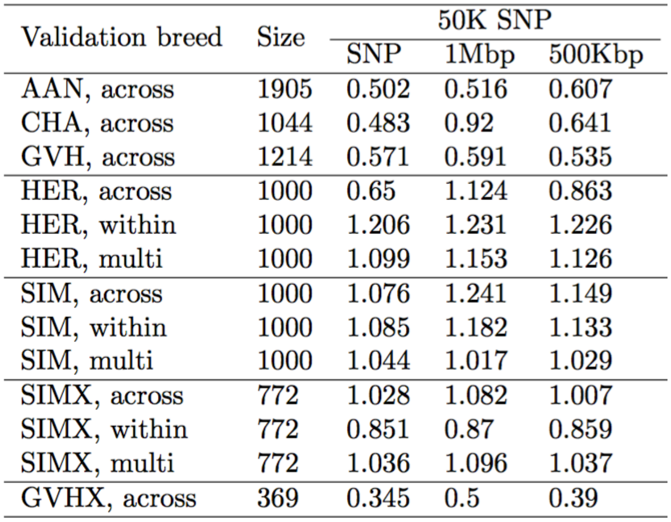
Birthweight Predictions



Validation breed	Size	50K SNP		
		SNP	1Mbp	500Kbp
AAN, across	1905	0.27	0.274	0.296
CHA, across	1044	0.198	0.309	0.263
GVH, across	1214	0.201	0.151	0.175
HER, across	1000	0.251	0.259	0.277
HER, within	1000	0.787	0.779	0.798
HER, multi	1000	0.666	0.661	0.674
SIM, across	1000	0.319	0.33	0.363
SIM, within	1000	0.674	0.67	0.679
SIM, multi	1000	0.629	0.601	0.623
SIMX, across	772	0.392	0.391	0.394
SIMX, within	772	0.395	0.402	0.399
SIMX, multi	772	0.486	0.493	0.486
GVHX, across	369	0.137	0.16	0.152

Correlations between genomic prediction and DEPD

Birthweight predictions



Validation breed	Size	50K SNP		
		SNP	1Mbp	500Kbp
AAN, across	1905	0.502	0.516	0.607
CHA, across	1044	0.483	0.92	0.641
GVH, across	1214	0.571	0.591	0.535
HER, across	1000	0.65	1.124	0.863
HER, within	1000	1.206	1.231	1.226
HER, multi	1000	1.099	1.153	1.126
SIM, across	1000	1.076	1.241	1.149
SIM, within	1000	1.085	1.182	1.133
SIM, multi	1000	1.044	1.017	1.029
SIMX, across	772	1.028	1.082	1.007
SIMX, within	772	0.851	0.87	0.859
SIMX, multi	772	1.036	1.096	1.037
GVHX, across	369	0.345	0.5	0.39

Regressions of DEPD on genomic prediction

Haplotype Models

- Did not markedly improve the accuracy of prediction across breeds
- Did not markedly reduce the bias even in prediction across breeds
- Similar results were obtained using imputed 700K markers and resultant haplotypes, and for different window sizes
  - But narrower (than 1 Mb) windows were usually better

"Hybrid" Mixed Model Equations

$$\begin{bmatrix} X'X & X'ZM & X_n'Z_n \\ M'Z'X & M'Z'ZM+\phi & M_n'Z_n'Z_n \\ Z_n'X_n & Z_n'Z_nM_n & Z_n'Z_n+A^{nn}\lambda \end{bmatrix}\begin{bmatrix} b \\ \alpha \\ \varepsilon_n \end{bmatrix}=\begin{bmatrix} X'y \\ M'Z'y \\ Z_n'y_n \end{bmatrix}$$

$$where\ X=\begin{bmatrix} X_n \\ X_g \end{bmatrix}, Z=\begin{bmatrix} Z_n & 0 \\ 0 & Z_g \end{bmatrix}, M=\begin{bmatrix} M_n \\ M_g \end{bmatrix}=\begin{bmatrix} A_{ng}A_{gg}^{-1}M_g \\ M_g \end{bmatrix}, y=\begin{bmatrix} y_n \\ y_g \end{bmatrix}$$

$$with\ EBV\ given\ by$$

$$\widehat{u_g}=M_g\widehat{\alpha}$$

$$\widehat{u_n}=M_n\widehat{\alpha}+\widehat{\varepsilon_n}$$

NB Single-Step GBLUP
is a special case of the above
(but in this equivalent model no inversion is needed)

$$M_n=A_{ng}A_{gg}^{-1}M_g$$

An extension to the single-step hybrid model

- with additional polygenic effect to capture variation not captured by markers
  - Allows models comparable to SS-GBLUP where

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = \begin{bmatrix} X_n \\ X_g \end{bmatrix} b + \begin{bmatrix} Z_n & 0 \\ 0 & Z_g \end{bmatrix} \begin{bmatrix} A_{ng} A_{gg}^{-1} M_g \\ M_g \end{bmatrix} \alpha_d + \begin{bmatrix} Z_n & 0 \\ 0 & Z_g \end{bmatrix} a_d + \begin{bmatrix} Z_n \\ 0 \end{bmatrix} \varepsilon_{nd} + \begin{bmatrix} e_n \\ e_g \end{bmatrix}$$

$$y \quad = \quad X \quad b + \quad Z \qquad M \quad \alpha_d + \quad Z \quad a_d \; + \; Z_{\_n} \varepsilon_{nd} + e$$

$$\lambda_\alpha = \frac{\sigma_e^2}{\sigma_\alpha^2}$$

$$\sigma_\alpha^2 = \frac{c\sigma_g^2}{2\overline{p\,q}k(1-\pi)}$$

$$\lambda_\alpha = \frac{2\overline{p\,q}k(1-\pi)\sigma_e^2}{c\,\sigma_g^2} \textit{ where } c = \textit{proportion of genetic variance accounted for by markers}$$

$$\lambda_\varepsilon = \frac{\sigma_e^2}{c\,\sigma_g^2}$$

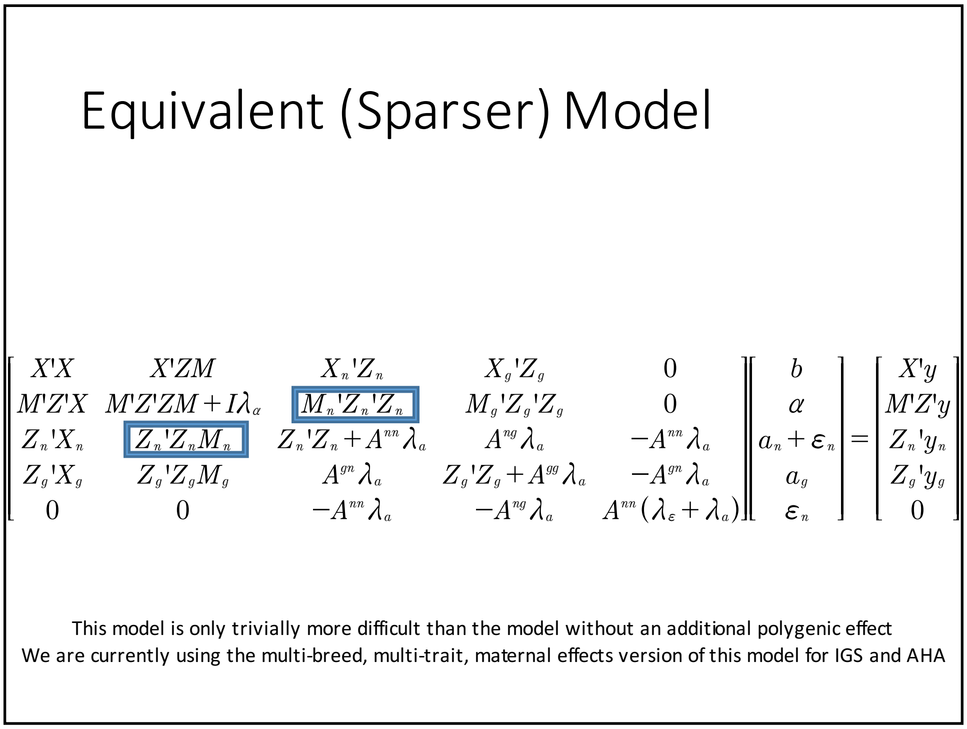
$$\lambda_a = \frac{\sigma_e^2}{(1-c)\sigma_g^2}$$

$$y = Xb + ZM\alpha_d + Za_d + Z_{\_n}\varepsilon_{nd} + e$$

$$\begin{bmatrix} X'X & X'ZM & X'Z & X'Z_{\_n} \\ M'Z'X & M'Z'ZM + I\lambda_\alpha & M'Z'Z & M'Z'Z_{\_n} \\ Z'X & Z'ZM & Z'Z + A^{-1}\lambda_a & Z'Z_{\_n} \\ Z_{\_n}'X & Z_{\_n}'ZM & Z_{\_n}'Z & Z_{\_n}'Z_{\_n} + A^{nn}\lambda_\varepsilon \end{bmatrix} \begin{bmatrix} b \\ \alpha_d \\ a_d \\ \varepsilon_{nd} \end{bmatrix} = \begin{bmatrix} X'y \\ M'Z'y \\ Z'y \\ Z_{\_n}'y \end{bmatrix}$$

Partitioning the "a" part of these equations into "g" and "n" results in the following

$$\begin{bmatrix} X'X & X'ZM & X_n'Z_n & X_g'Z_g & X_n'Z_n \\ M'Z'X & M'Z'ZM + I\lambda_\alpha & \boxed{M_n'Z_n'Z_n} & M_g'Z_g'Z_g & \boxed{M_n'Z_n'Z_n} \\ Z_n'X_n & \boxed{Z_n'Z_nM_n} & Z_n'Z_n + A^{nn}\lambda_a & A^{ng}\lambda_a & Z_n'Z_n \\ Z_g'X_g & Z_g'Z_gM_g & A^{gn}\lambda_a & Z_g'Z_g + A^{gg}\lambda_a & 0 \\ Z_n'X_n & \boxed{Z_n'Z_nM_n} & Z_n'Z_n & 0 & Z_n'Z_n + A^{nn}\lambda_\varepsilon \end{bmatrix} \begin{bmatrix} b \\ \alpha \\ a_n \\ a_g \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} X'y \\ M'Z'y \\ Z_n'y_n \\ Z_g'y_g \\ Z_n'y_n \end{bmatrix}$$



Equivalent (Sparser) Model

$$\begin{bmatrix} X'X & X'ZM & X_n'Z_n & X_g'Z_g & 0 \\ M'Z'X & M'Z'ZM + I\lambda_\alpha & \boxed{M_n'Z_n'Z_n} & M_g'Z_g'Z_g & 0 \\ Z_n'X_n & \boxed{Z_n'Z_nM_n} & Z_n'Z_n + A^{nn}\lambda_a & A^{ng}\lambda_a & -A^{nn}\lambda_a \\ Z_g'X_g & Z_g'Z_gM_g & A^{gn}\lambda_a & Z_g'Z_g + A^{gg}\lambda_a & -A^{gn}\lambda_a \\ 0 & 0 & -A^{nn}\lambda_a & -A^{ng}\lambda_a & A^{nn}(\lambda_\varepsilon + \lambda_a) \end{bmatrix} \begin{bmatrix} b \\ \alpha \\ a_n + \varepsilon_n \\ a_g \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} X'y \\ M'Z'y \\ Z_n'y_n \\ Z_g'y_g \\ 0 \end{bmatrix}$$

This model is only trivially more difficult than the model without an additional polygenic effect

We are currently using the multi-breed, multi-trait, maternal effects version of this model for IGS and AHA