

Effect of different genomic relationship matrices on accuracy and scale

Ignacy Misztal, Ching-Yi Chen, Ignacio Aguilar
University of Georgia

Zulma Vitezica, Andres Legarra, INRA Toulouse
Bill Muir, Purdue University

Single-step genomic evaluation - combines genomic and pedigree relationships

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} [\mathbf{G} - \mathbf{A}_{22}] \begin{bmatrix} \mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

A – pedigree-based relationship matrix

G – genomic relationship matrix

1- ungenotyped animals

2- genotyped animals

Legarra et al. (2009)

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Aguilar et al. (2010)

Single-step implementation at UGA

- Module genomic in BLUPF90 package (Aguilar et al. 2011)
- OPTION SNP_File xxx in RENUMF90
- Up to 50k genotypes
- Lots of options with new options added continuously
- One option: filtering for minor allele frequencies (MAF)

Background

- Distribution of allele frequencies $\sim \text{UN}(0, 0.5)$
- SNP with low MAF:
 - Less informative
 - Possibly some causative SNP
- Elimination of SNP with low MAF can:
 - Decrease computations
 - Decrease/Increase accuracy of evaluation
- Properties of G assuming different allele frequencies :
 - If current, nice properties
 - If equal, highest accuracy in dairy

Goals

- Examine effect of increasing MAF for G constructed assuming equal (GE) or current (GC) allele frequencies on:
 - Properties of G and G^{-1}
 - Accuracy of evaluation
 - Biases of evaluation

Data – chicken (Cobb Vantress)

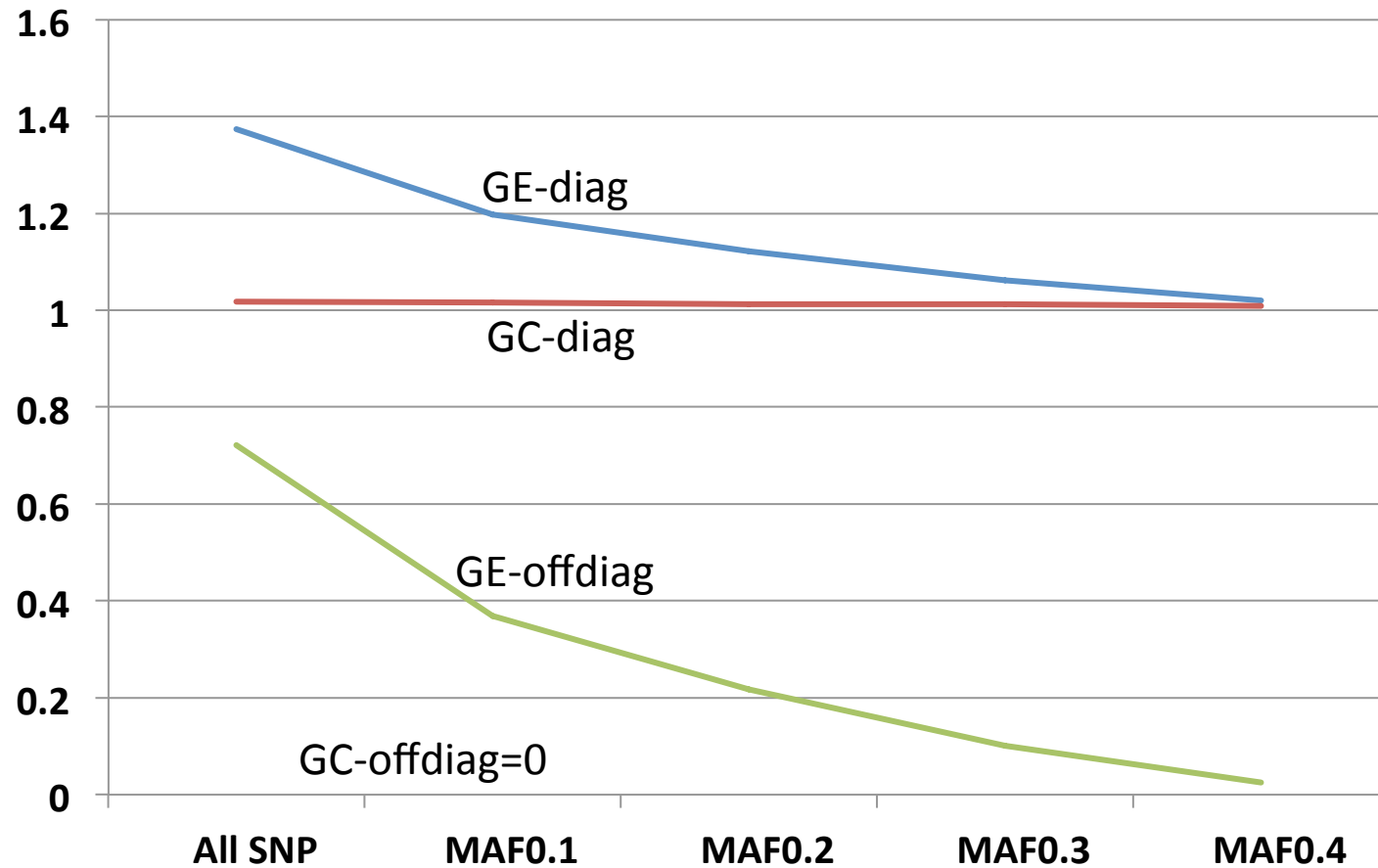
Item	No. of records	Mean	SD
Genotyped			
BW, 100g	4,113	25.79	3.20
BM, cm ²	3,923	43.98	5.75
All			
BW, 100g	287,614	26.25	4.91
BM, cm ²	69,057	45.52	7.34

BW= body weight at 6 weeks; BM= ultrasound area of breast meat.

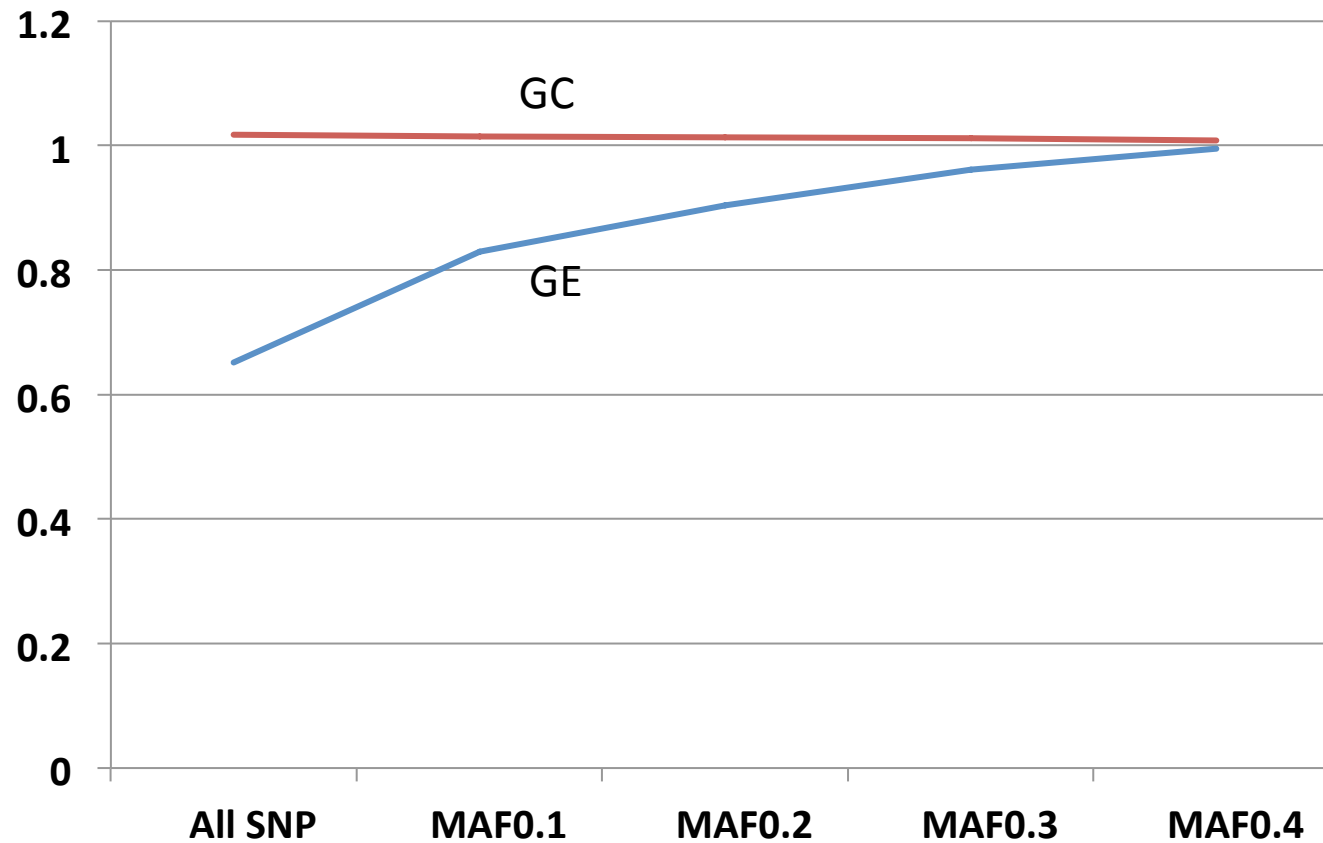
Number of SNP and MAF

Item	No. SNP
All SNP	57,636
MAF0.1	41,749
MAF0.2	32,885
MAF0.3	22,619
MAF0.4	11,253

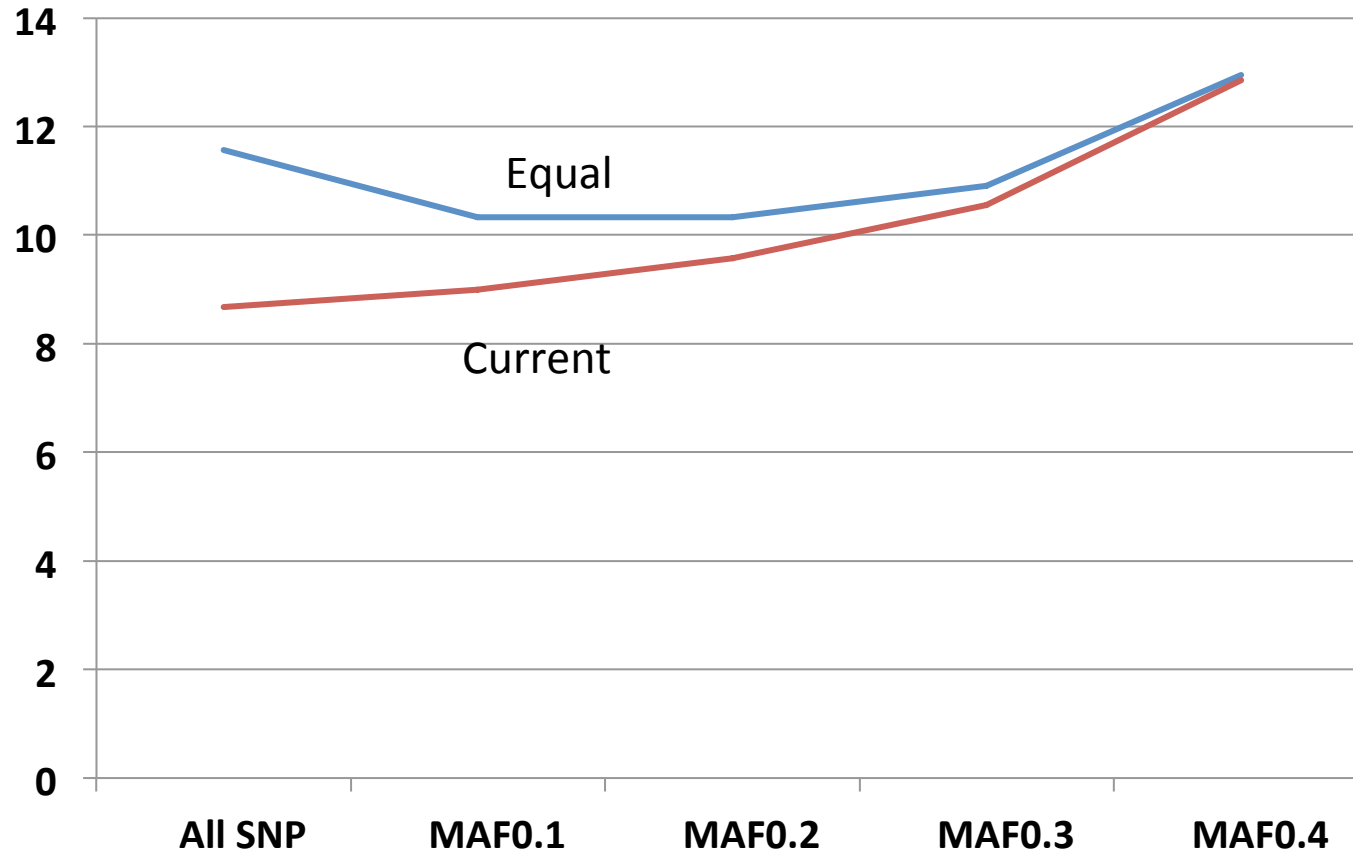
Mean of diagonal and off-diagonal elements of G



Difference between average diagonal and off-diagonal of G



Mean of diagonal of G^{-1}



Inv(GE): OffDiag= -0.000

Inv(GC): OffDiag= -0.000

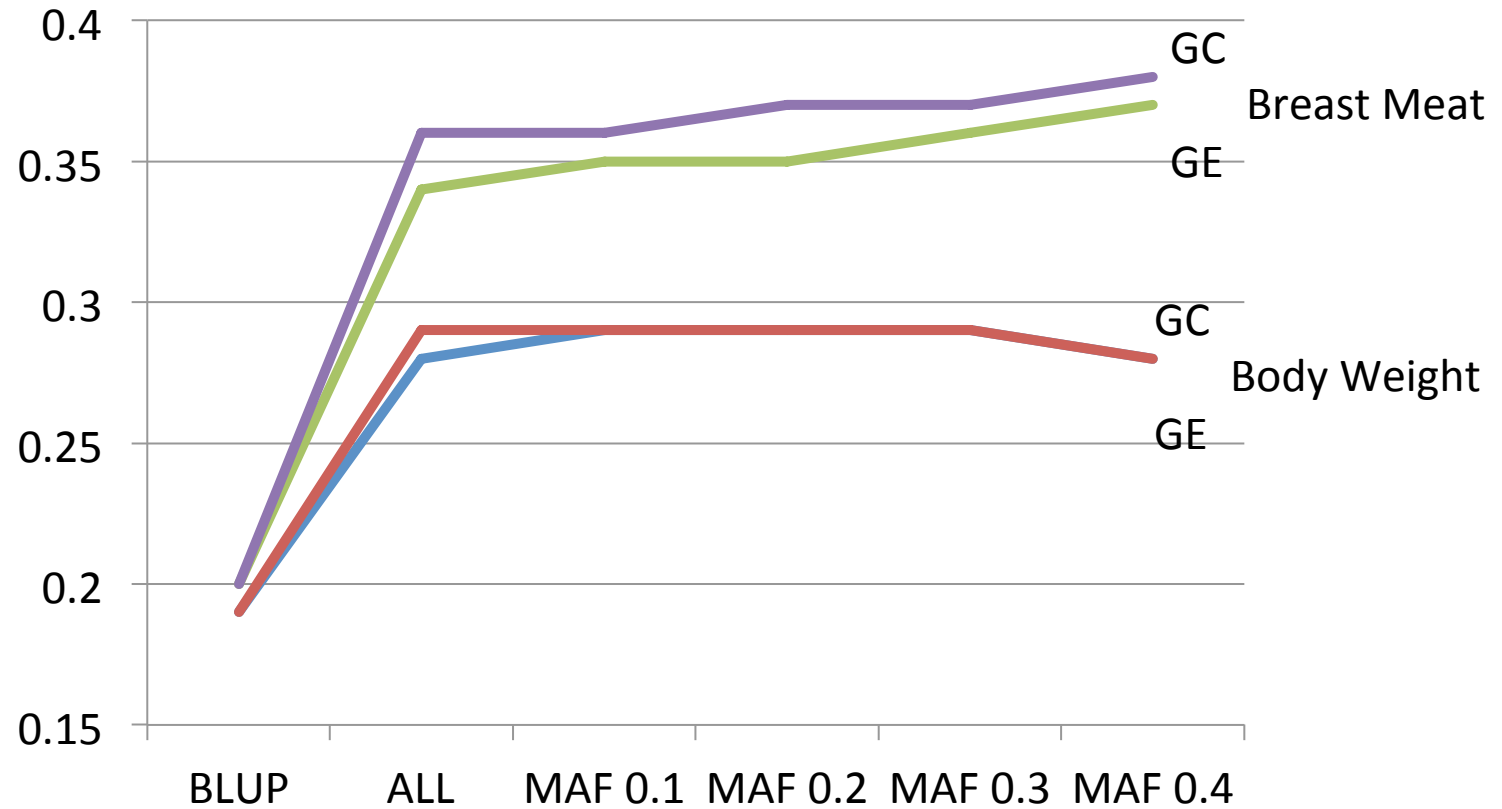
Diagonal element of properly scaled G

$$g^{ii} \approx 1 + n_{\text{par}}/2 + n_{\text{prog}}/2 + n_{\text{gen}}/2 + f(1/\#\text{SNP})$$

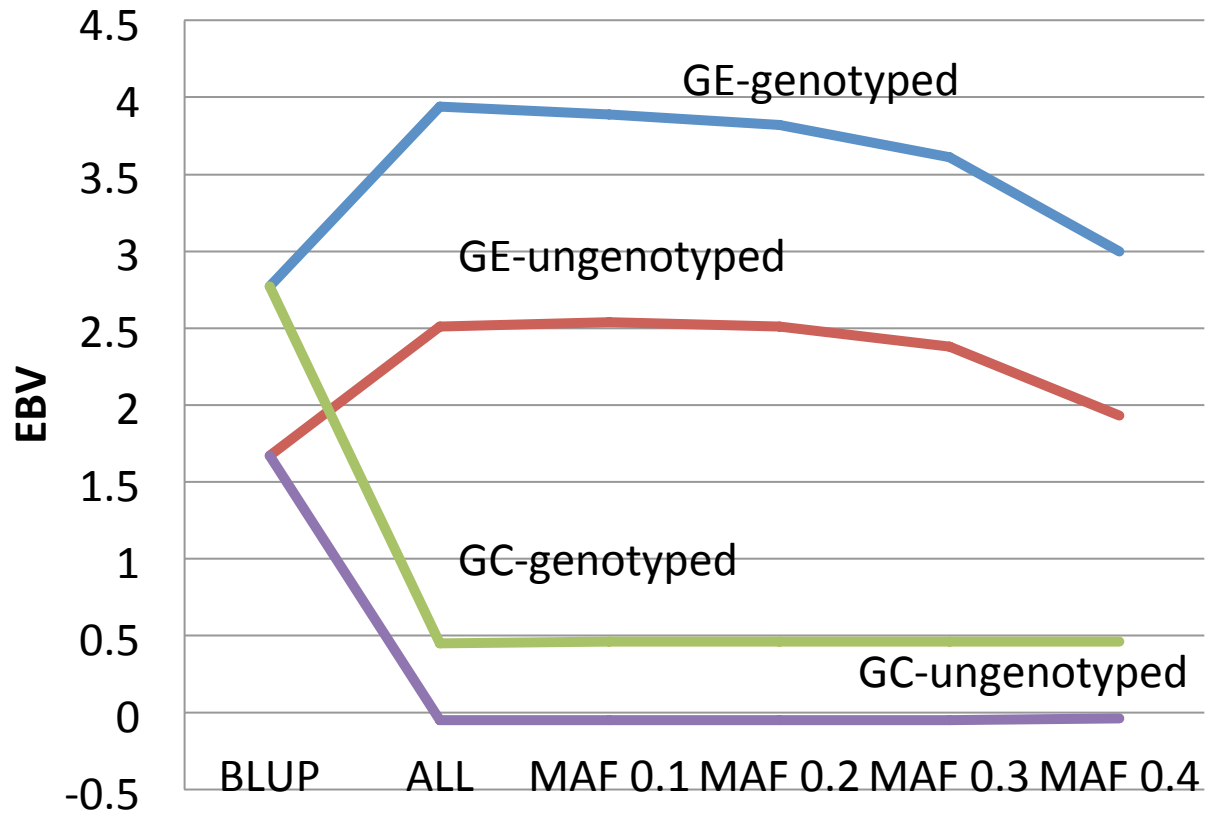
number of parents
number of progenies
“genomic” progenies
noise

- ⊠ Inflation of computed accuracies
- ⊠ Check on imputation quality

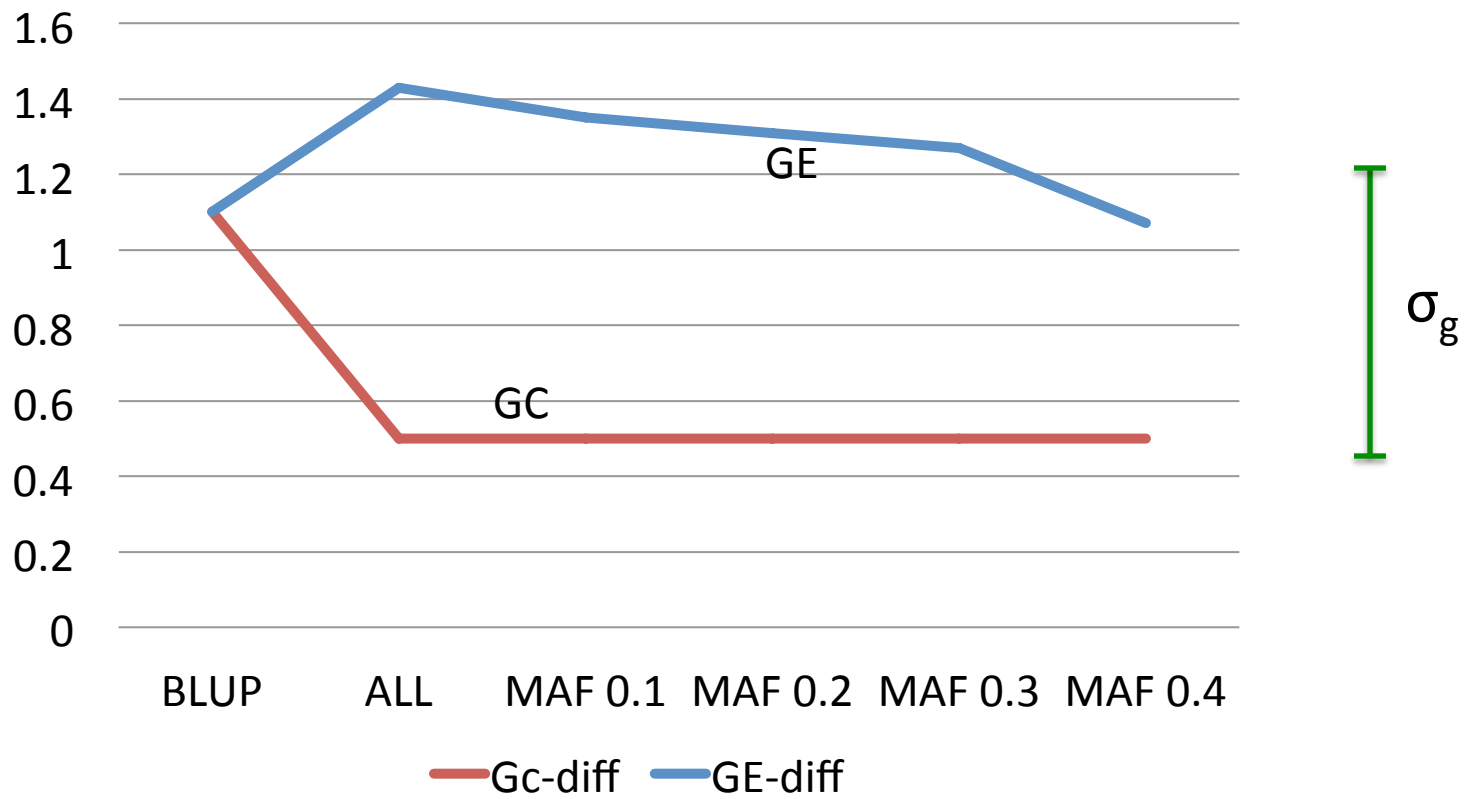
Predictivity for the last generation



EBV for genotyped and all animals - BW



Differences between EBV of genotyped and all animals - BW



Why biases with G?

$$\text{Var}(\mathbf{u}_g) = \mathbf{G}$$

G - true variance

$$\mathbf{G}^* = \mathbf{G} - \alpha$$

G* - calculated variance

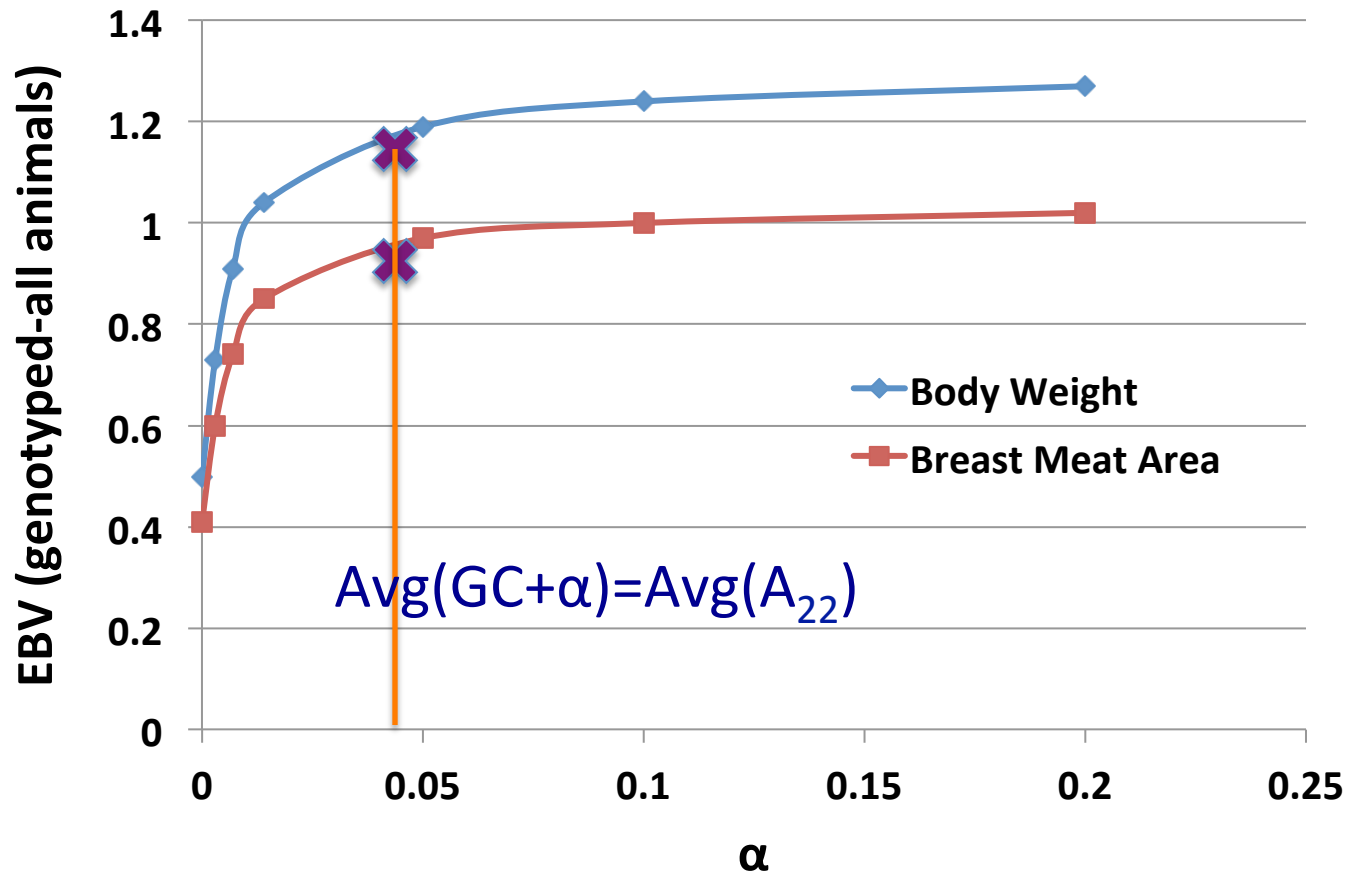
$$\mathbf{u}_g^* = \mathbf{u}_g + \mu \quad \text{var}(\mu) = \alpha$$

Constant is added to EBV
of all genotyped animals

How to eliminate bias?

- $G = G^* + \alpha$
- BLUP assumed unbiased
- Find α : $\text{Avg}(\text{EBV}_{\text{genotyped}} - \text{EBV}_{\text{all}})$ as in BLUP

Difference of EBV between genotyped and all animals as a function of α for GC



X BLUP

Vitezica et al., 2011 – equivalence with Fst
Fst in Young et al., 2010

Conclusions

- Reduction of SNP with MAF down to 0.4:
 - Minimally affect the accuracy
 - Increases average diagonals of G^{-1}
 - Inflates computed accuracies
- Different G cause upward or downward biases
- Biases removed when averages of G and A_{22} the same
 - More complication when multiple lines/breeds

Partial release of programs in few weeks

Acknowledgements

Cobb-Vantress

NIFA grants

Holstein Association

PIC