Construction of a Genomic Relationship Matrix using Linked Markers

L Gomez-Raya, WM Rauw, L Silio,
MC Rodriguez, and LA Garcia-Cortés

gomez.luis@inia.es

Instituto Nacional de Investigaciones
Agroalimentarias, Madrid, Spain.

Introduction

- Genomic BLUP (GBLUP) is a linear mixed model incorporating a marker-based genomic relationship matrix (*G*-matrix) instead of a relationship matrix based on a known pedigree.

Introduction

Current construction of the *G*-matrix uses single SNPs ignoring that markers are linked, which may reduce genomic information contributed by relatives.

CCTTACTTA
GGAATGAAT

AGGTATTCA
TCCATAAGT

The goal of this study was to construct a *G*-matrix using linked markers

1) to develop formulae for centering and scaling the *G*-matrix using haplotype frequencies at two loci.

2) to compare results of the new *G* matrix with the standard *G* constructed with single loci in 435 sow of Iberian pigs.

METHODS

Assumptions

1) gene action is fully additive: *T/t* and *M/m*

2) each marker or pair contributes equally to the additive mean and variance of the trait.

Steps for constructing *G*-matrix with information for pairs of loci

1) estimation of linkage disequilibrium between all pairs of consecutive markers using the EM algorithm

2) computing all haplotype frequencies at each of two consecutive loci

Retain pairs with haplotype freq: $f_{TM}, f_{Tm}, f_{tM}, f_{tm} > 0.01$

3) develop formulae for centering and scaling the *G*-matrix

Scaling and centering using the biological parameterization, Su et al., (2012), Vitezica et al. (2013)

$$\boldsymbol{G} = \frac{(\boldsymbol{M}-\boldsymbol{P})(\boldsymbol{M}-\boldsymbol{P})'}{V_a}$$

M **is an haplotype-sharing or allele-sharing matrix with *na* rows corresponding to the number of animals**

P **is a matrix with elements containing the mean of each haplotype**

For single loci the columns of M are nl loci

For two loci the columns of M are 4 x nl loci

EACH LOCI CONTRIBUTE WITH FOUR HAPLOTYPES

Genotypic values for the four haplotypes (times a) for matrix *M*

	TM	**Tm**	**tM**	**tm**
TTMM	**2**	**0**	**0**	**0**
TTMm	**1**	**1**	**0**	**0**
TTmm	**0**	**2**	**0**	**0**
TtMM	**1**	**0**	**1**	**0**
TtMm	**λ**	**1-λ**	**1-λ**	**λ**
Ttmm	**0**	**1**	**0**	**1**
ttMM	**0**	**0**	**2**	**0**
ttMm	**0**	**0**	**1**	**1**
ttmm	**0**	**0**	**0**	**2**

CENTERING
Matrix *P*

Mean for the four haplotypes

Loci *T/t* and *M/m*

	TM	Tm	tM	tm
TTMM	2	0	0	0
TTMm	1	1	0	0
TTmm	0	2	0	0
TtMM	1	0	1	0
TtMm	λ	1-λ	1-λ	λ
Ttmm	0	1	0	1
ttMM	0	0	2	0
ttMm	0	0	1	1
ttmm	0	0	0	2

$$\lambda = \frac{f_{TM}f_{tm}}{f_{TM}f_{tm} + f_{Tm}f_{tM}}$$

$$\begin{aligned}\mu_{TM} &= 2(f_{TM}^2 + f_{TM}f_{Tm} + f_{TM}f_{tM} + f_{TM}f_{tm}) \\ &= 2f_{TM} \\ \mu_{Tm} &= 2f_{Tm} \\ \mu_{tM} &= 2f_{tM} \\ \mu_{tm} &= 2f_{tm}\end{aligned}$$

Scaling

Variance for the four haplotypes

Loci T/t and M/m

	TM	Tm	tM	tm
TTMM	2	0	0	0
TTMm	1	1	0	0
TTmm	0	2	0	0
TtMM	1	0	1	0
TtMm	λ	1-λ	1-λ	λ
Ttmm	0	1	0	1
ttMM	0	0	2	0
ttMm	0	0	1	1
ttmm	0	0	0	2

$$V_{TM} = f_{TM}(f_{TM} + f_{Tm} + f_{tM} + \lambda f_{tm})$$
$$V_{Tm} = f_{Tm}(f_{Tm} + f_{TM} + f_{tm} + (1-\lambda) f_{tM})$$
$$V_{tM} = f_{tM}(f_{tM} + f_{TM} + f_{tm} + (1-\lambda) f_{Tm})$$
$$V_{tm} = f_{tm}(f_{tm} + f_{Tm} + f_{tM} + \lambda f_{TM})$$

$$V_a = \frac{1}{4}\sum_{i=1}^{nl} V_{TM(i)} + V_{Tm(i)} + V_{tM(i)} + V_{tm(i)}$$

Material

435 sows of Iberian pigs.

Illumina 60K

Dependent variable
litter size

Independent variables
Station , parity
Permanent Environment
Animal

ASREML


JOSELITO
OFERTA
Jamón de bellota 5 Jotas
189,00 €
Jamón de Salamanca Joselito
199,00 €
Jamón bellota de Guijuelo
135,00 €
Jamón de cebo
75,00 €

RESULTS

Construction of *G* matrices

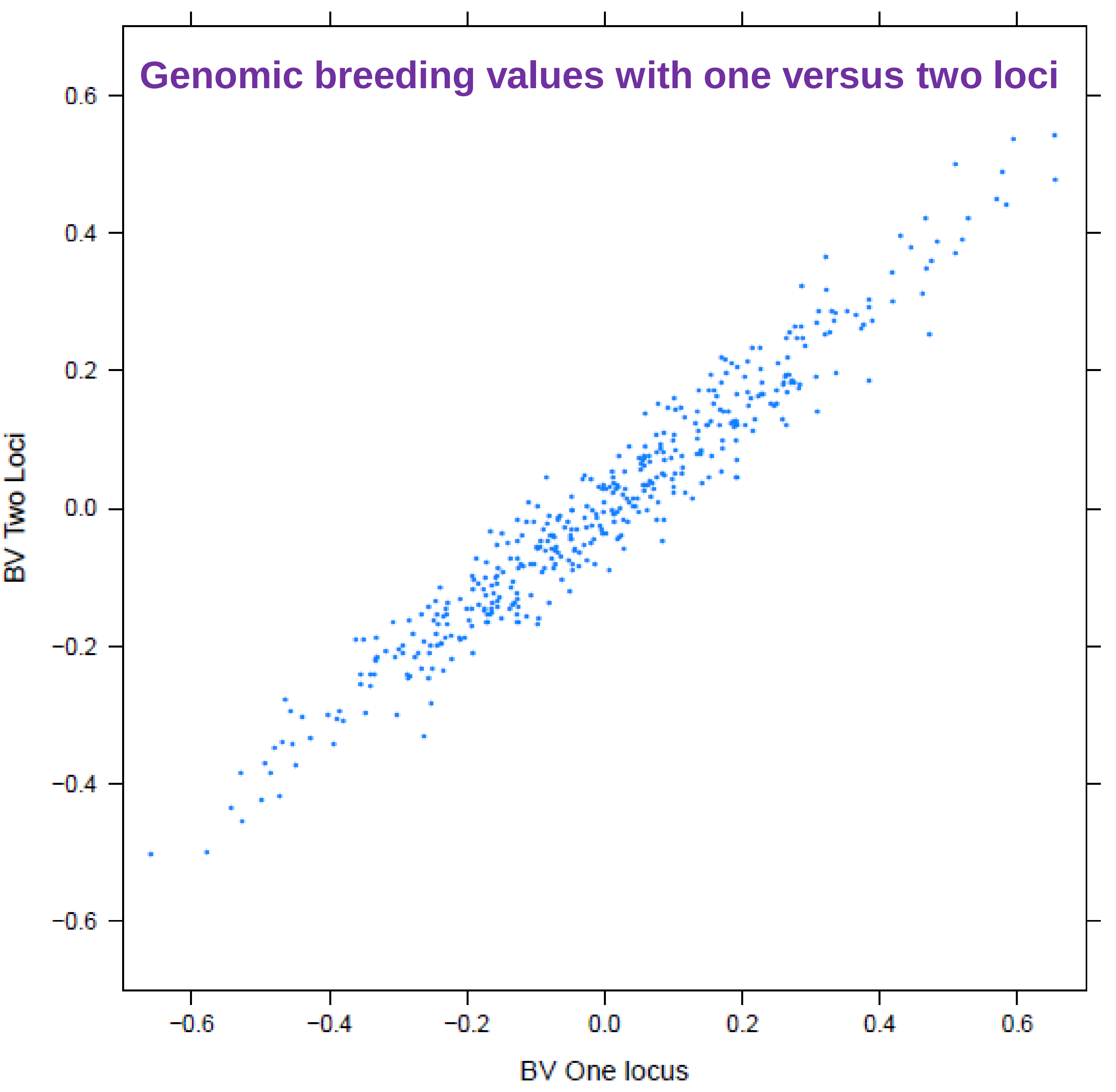
	One locus*	Two loci
# SNPs or pairs	26,359	5,194
Average diagonal elements	0.988	0.982
Average off-diagonal elements	-0.0023	-0.0007

One locus* as Van Raden and Vitezica et al.

Variance components with alternative *G* matrices

Component	One locus	Two loci
Animal	0.2044	0.2780
Permanent	0.6870	0.6171
Residual	4.1074	4.1074
h^2	0.041	0.056

Genomic breeding values with one versus two loci



Discussion and Conclusions

Components of variance are very similar whether either one or two loci are used to construct ***G***.

Breeding values are very highly correlated (0.97) when using one or two loci.

Discussion and Conclusions

Strong LD between adjacent pairs may reduce information from linked markers.

More markers are required to accommodate linkage in genomic evaluations.

For haplotypes with nl loci the columns of M are 2^{nl} x nl loci → Reduccion of the haplotypes.

Thanks!