Genome wide evaluation using dominance

Robin Wellmann, Jörn Bennewitz

Department of Animal Husbandry and Animal Breeding

University of Hohenheim

May 12, 2010

Outline

- Why include dominance?
- Breeding values and dominance values
- Simulation
- Prediction of genomic breeding values
  - BLUP without dominance
  - BLUP with dominance
  - Stepwise procedure
- Comparison of different methods
- Conclusions

Why include dominance?

The inclusion of dominance

- could **increase the accuracy** of predicted breeding values,
- could be used to find mating pairs with good combining ability by **recovering inbreeding depression** and **utilizing possible overdominance**.

Breeding values and dominance values

Breeding values and dominance values

According to Falconer (1996), the breeding value of individual i is

$$BV_i = \sum_{n \in \mathcal{Q}} (a_n + d_n(q_n - p_n))(v_{ni} + m_{ni}),$$

and the dominance deviation is

$$DV_i = \sum_{n \in \mathcal{Q}} -2d_n(v_{ni} - p_n)(m_{ni} - p_n),$$

where

$v_{ni} \in \{0, 1\}$	paternal allele of individual i at QTL n,
$m_{ni} \in \{0, 1\}$	maternal allele of individual i at QTL n,
a_n	additive effect of QTL n,
d_n	dominance effect of QTL n
p_n	frequency of allele 1 at QTL n,
q_n	frequency of allele 0 at QTL n.

Breeding values and dominance values

Breeding value and dominance deviation are estimated as

$$EBV_i = \sum_{n \in \mathcal{M}} (\hat{a}_n + \hat{d}_n(q_n - p_n))(v_{ni} + m_{ni}),$$

and

$$EDV_i = \sum_{n \in \mathcal{M}} -2\hat{d}_n(v_{ni} - p_n)(m_{ni} - p_n),$$

where $\hat{a}_n$ and $\hat{d}_n$ are predicted marker effects.

Breeding values and dominance values

Breeding value and dominance deviation are estimated as

$$EBV_i = \sum_{n \in \mathcal{M}} (\hat{a}_n + \hat{d}_n(q_n - p_n))(v_{ni} + m_{ni}),$$

and

$$EDV_i = \sum_{n \in \mathcal{M}} -2\hat{d}_n(v_{ni} - p_n)(m_{ni} - p_n),$$

where $\hat{a}_n$ and $\hat{d}_n$ are predicted marker effects.

We compared different methods to predict marker effects by simulation.

Simulation

Simulation

We simulated a population

- that has the same LD pattern as the target population (see Villa-Angulo et al., 2009),
- where each trait has a different distribution of additive effects and dominance degrees,
- that has a smaller genome than the target population in order to reduce computation time.

Characteristics of the QTL effects:

- The distribution of the additive effects A_n was a mixture of a double exponential distribution and a normal distribution, i.e.

$$A_n \sim 0.95 \cdot \mathcal{L}(0, \sigma^2_{\mathcal{L}}) + 0.05 \cdot \mathcal{N}(0, (5\sigma_{\mathcal{L}})^2),$$

  where $\sigma_{\mathcal{L}}$ was chosen such that $\mathrm{Var}(A_n) = \sigma^2_A$.
- Normally distributed dominance degrees $G_n = \frac{D_n}{|A_n|}$ have mean μ_G and variance σ^2_G.
- Additive effects and dominance degrees are independent.
- No epistasis.

Simulation

Characteristics of the simulated population:

- Fisher-Wright diploid population,
- independent crossovers,
- 1 chromosome which equals 1 Morgan,
- 1666 markers per Morgan,
- 120 QTL on average per Morgan,
- no selection,
- N_e decreased from 1000 to 100 within 400 generations,
- marker effects were predicted from 1000 individuals.

Methods

to predict marker effects

BLUP with and without dominance

$$
\begin{aligned}
Y &= \mu 1 + Z_A A + E, && \text{(without dominance)} \\
Y &= \mu 1 + \beta F + Z_A A + Z_D(D - \mu_D) + E, && \text{(with dominance)}
\end{aligned}
$$

where

Y	vector with phenotypic values,
μ	overall mean,
F	vector with estimated inbreeding coefficients,
A	vector with additive effects of markers,
D	vector with dominance effects of markers,
Z_A	gene content matrix with entries 0,1 and 2,
Z_D	indicator matrix for heterozygosity with entries 0 and 1,
E	error

BLUP with and without dominance

$$
\begin{aligned}
Y &= \mu 1 + Z_A A + E, && \text{(without dominance)}\\
Y &= \mu 1 + \beta F + Z_A A + Z_D(D - \mu_D) + E, && \text{(with dominance)}
\end{aligned}
$$

where

$$
\begin{aligned}
E(A) &= E(E) = 0,\\
E(D) &= \mu_D,\\
\text{Var}(A) &= \sigma^2_A I,\\
\text{Var}(D) &= \sigma^2_D I,\\
\text{Var}(E) &= \sigma^2_E I,
\end{aligned}
$$

A and D are independent,
Random effects are normally distributed,
Variances captured by markers equal V_A and V_D.

Stepwise procedure

Steps:

1) A and $D - \mu_D$ were predicted with BLUP, using the model

$$Y = \mu 1 + \beta F + Z_A A + Z_D(D - \mu_D) + E,$$

but the prediction of $D - \mu_D$ was discarded.

2) Observations were corrected for predicted additive effects and inbreeding depression as $\tilde{Y} = Y - \hat{\mu}1 - \hat{\beta}F - Z_A\hat{A}$.

3) The centered dominance effects were predicted again for the corrected observations by assuming large variances for QTL where the predicted additive effect was large, using the model

$$\tilde{Y} = Z_D(D - \mu_D) + E.$$

Stepwise procedure

Steps:

4) The expectations of the dominance effects were estimated by dividing estimated inbreeding depression between QTL, putting more weight on QTL with large predicted additive effects.
5) The dominance effects were obtained by adding the estimated expectations and the predicted centered dominance effects.

Stepwise procedure

Steps:

4) The expectations of the dominance effects were estimated by dividing estimated inbreeding depression between QTL, putting more weight on QTL with large predicted additive effects.
5) The dominance effects were obtained by adding the estimated expectations and the predicted centered dominance effects.

$\Rightarrow$ This method utilizes that additive effects and dominance effects are dependent.

Comparison of methods: nonlinear regression

Comparison of methods: nonlinear regression

Observation:

smaller genome $\rightarrow$ smaller number of QTL

$\rightarrow$ larger variation of variance components

Problem:

- How to account for the variation of variance components?

Comparison of methods: nonlinear regression

Observation:

smaller genome $\rightarrow$ smaller number of QTL

$\rightarrow$ larger variation of variance components

Problem:

- How to account for the variation of variance components?
  - A multiple regression was carried out with the variance components as explanatory variables and the accuracy as the dependent variable.

Comparison of methods: nonlinear regression

As summarized by Meuwissen (2009) we have approximately

$$r_{BV} = \sqrt{\frac{Nh^2}{Nh^2 + Q_e}},$$

where

r_{BV}	accuracy of predicted breeding values, i.e. correlation between true and predicted breeding values
h^2	narrow sense heritability,
N	number of training records,
Q_e	effecive number of QTL loci.

Comparison of methods: nonlinear regression

This can be simplified to

$$r_{BV} = \sqrt{\frac{ah^2}{1+ah^2}},$$

where

$$a = \frac{N}{Q_e}.$$

Comparison of methods: nonlinear regression

Solving for ah^2 gives

$$\frac{r_{BV}^2}{1 - r_{BV}^2} = ah^2.$$

Comparison of methods: nonlinear regression

Therefore, we assumed the linear model

$$\frac{r_{BV,j}^2}{1 - r_{BV,j}^2} = a_1 h_j^2 + a_2 d_j^2 + a_3 |\mathcal{I}_j| + e_j,$$

where

$r_{BV,j}$	accuracy of predicted breeding value for trait j
h_j^2	narrow sense heritability,
d_j^2	ratio of dominance variance to phenotypic variance,
$\mathcal{I}_j$	Inbreeding depression (Decline of the trait value when inbreeding coefficient increases from 0% to 100%).
e_j	error.

Comparison of methods: nonlinear regression

Therefore, we assumed the linear model

$$\frac{r_{BV,j}^2}{1 - r_{BV,j}^2} = a_1 h_j^2 + a_2 d_j^2 + a_3 |\mathcal{I}_j| + e_j,$$

where

$r_{BV,j}$	accuracy of predicted breeding value for trait j
h_j^2	narrow sense heritability,
d_j^2	ratio of dominance variance to phenotypic variance,
$\mathcal{I}_j$	Inbreeding depression (Decline of the trait value when inbreeding coefficient increases from 0% to 100%).
e_j	error.

$\rightarrow$ Model is also used to fit accuracy of dominance values $r_{DV,j}$.

Results and conclusions

Results

Regression coef. for accuracy of breeding value

	h_j^2	d_j^2	$\lvert\mathcal{I}_j\rvert$
	a_1	a_2	a_3
BLUP without dominance	9.0	3.2	-0.3
BLUP with dominance	9.0	4.4	-0.3
Stepwise procedure	9.2	5.4	-0.3

Regression coef. for accuracy of dominance deviation

	h_j^2	d_j^2	$\lvert\mathcal{I}_j\rvert$
	a_1	a_2	a_3
BLUP with dominance	0.2	2.4	1.0
Stepwise procedure	0.8	6.2	0.9

Average values: $d^2 = 0.035, \mathcal{I} = 0.43, h^2 = 0.25.$

Conclusions

- Small genomes with few QTL cause substantial variation of variance components between replicates. A nonlinear regression approach can utilize the variation of variance components.
- BLUP is not optimal for the prediction of genomic breeding values because it can not account for the non-normal joint distribution of additive and dominance effects.

References

- Falconer, D. S., Mackay, T. F. C. (1996). Introduction to quantitative genetics. London, UK: Longman
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* **41**:35
- Villa-Angulo, R., Matukumalli, L. K., Gill et al. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics* **10**:19.

Acknowledgement

The study was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG.

Thank you for your attention!