

Incorporation of LD information into the GBV model



Kacper Żukowski, Joanna Szyda

Wrocław University of Life Sciences
Institute of Animal Genetics

Table of contents

- Aim of study
- Data Set
- Methods
- Results
- Discussion

Aim of the Study

- incorporation of information on covariance between SNPs into GBV model
- differences between D , D' and r^2

Data Set (*animals*)

Simulated data set from XII QTL-MAS Workshop 2008.

- *6 chromosomes*
- *1000 SNPs per one chromosome*
- *50 male and female ancestors (-50 generation)*
- *50 generations*
- *4 generations*
 - *genotype + phenotype*
 - *15 males and 150 females*
 - *750 animals*
 - *(6 QTLs + polygenes)*
- *generation 0*
- *generations 1-3*
- $\bar{y} = 1.36$
- *generations (5-7)*
 - *genotype information*
 - *400 animals each*

Data Set (*SNP selection*)

SNP data sets:

- $MAF \geq 0.3$ → 3 328 SNPs
- every 5th SNP → 1 200 SNPs
- every 10th SNP → 600 SNPs
- every 20th SNP → 300 SNPs

LD estimation

<i>Haplotype</i>	<i>Frequency</i>	<i>Allele</i>	<i>Frequency</i>
$SNP1_1 SNP2_1$	x_{11}	$SNP1_1$	$\hat{p}_1 = x_{11} + x_{12}$
$SNP1_1 SNP2_2$	x_{12}	$SNP1_2$	$\hat{p}_2 = x_{21} + x_{22}$
$SNP1_2 SNP2_1$	x_{21}	$SNP2_1$	$\hat{q}_1 = x_{11} + x_{21}$
$SNP1_2 SNP2_2$	x_{22}	$SNP2_2$	$\hat{q}_2 = x_{12} + x_{22}$

- D (Lewontin and Kojima, 1960)

$$D = x_{11} - \hat{p}_1 \hat{q}_1$$

	SNP ₁₁	SNP ₁₂	Total
SNP ₂₁	$x_{11} = \hat{p}_1 \hat{q}_1 + D$	$x_{21} = \hat{p}_2 \hat{q}_1 - D$	$\hat{q}_1 = x_{11} + x_{21}$
SNP ₂₂	$x_{12} = \hat{p}_1 \hat{q}_2 - D$	$x_{22} = \hat{p}_2 \hat{q}_2 + D$	$\hat{q}_2 = x_{12} + x_{22}$
Total	$\hat{p}_1 = x_{11} + x_{12}$	$\hat{p}_2 = x_{21} + x_{22}$	1

LD estimation, Matrix norm

- r^2 (Hill and Robertson, 1968)

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$$

Matrix norm - Frobenius Norm

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{trace}(A^T A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}$$

SNP Data Set	Every 20	Every 10	Every 5	MAF ≥ 0.3
Norm between D and r^2	16.54	33.00	67.04	93.76

LD matrix



D

1,000	0,041	0,056	0,045	0,053	0,040
0,041	1,000	0,040	0,052	0,065	0,040
0,056	0,040	1,000	0,080	0,119	0,084
0,045	0,052	0,080	1,000	0,086	0,069
0,053	0,065	0,119	0,086	1,000	0,091
0,040	0,040	0,084	0,069	0,091	1,000

Model

$$y = \mu + Zg + e$$

$$\begin{bmatrix} -0.083496086 \\ 1.3162193298 \\ 5.4905762672 \\ 0.9890260696 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 2 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} SNP1 \\ SNP2 \\ SNP3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e1 \\ e2 \\ e3 \\ e4 \end{bmatrix}$$

y trait value
 μ general mean
 g vector of random SNP effect

$$g \sim N\left(0, LD \frac{\hat{\sigma}_a^2}{N_{SNP}}\right) \quad \hat{\sigma}_a^2 = 1.36$$

Z incidence matrix for g
 e residual

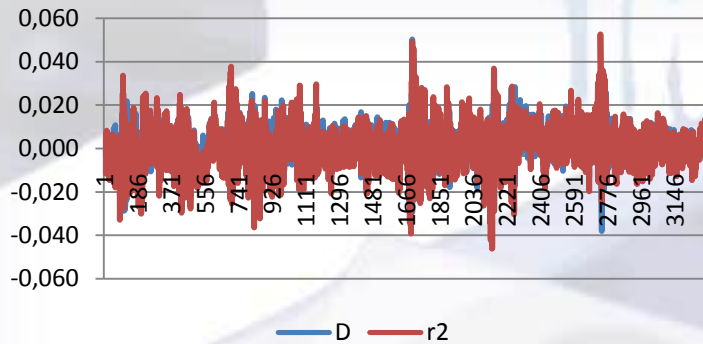
$$e \sim N(0, I\hat{\sigma}_e^2) \quad \hat{\sigma}_e^2 = 3.12$$

Technical aspects

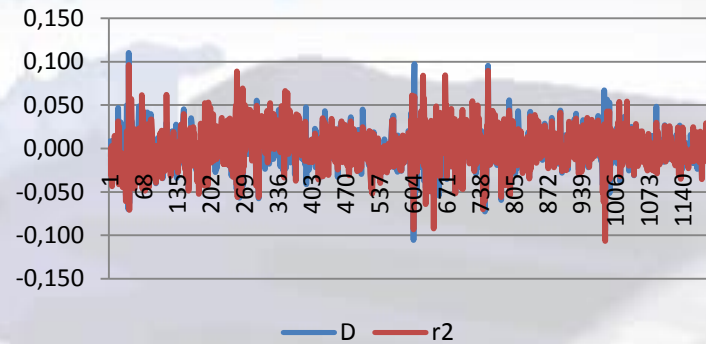
- Calculating LD
 - R package „SNPassoc”
 - http://davinci.crg.es/estivill_lab/snpassoc
- Calculating GBV - own functions in R

Results

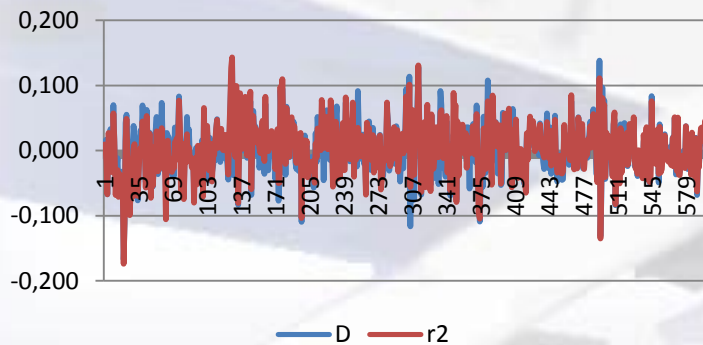
SNP additive effects for MAF ≥ 0.3



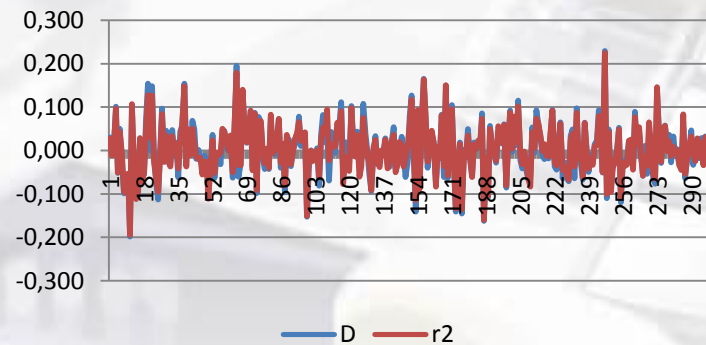
SNP additive effects for Every 5



SNP additive effects for Every 10

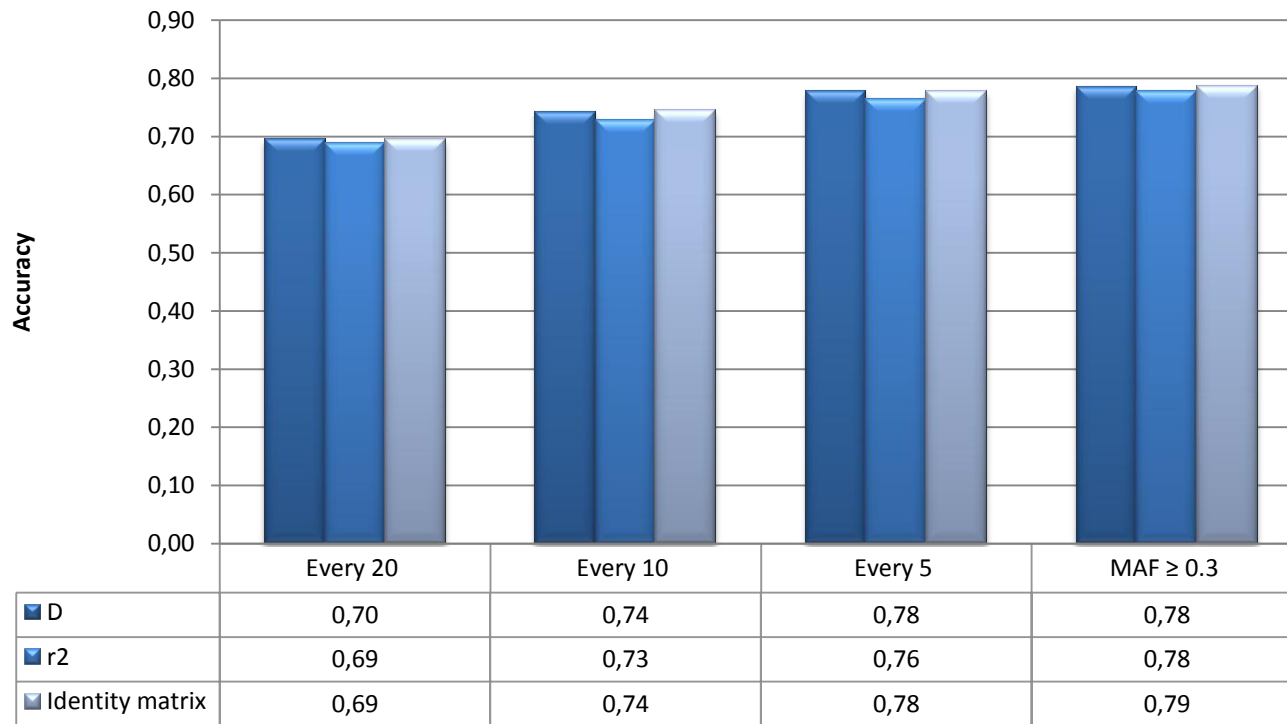


SNP additive effects for Every 20



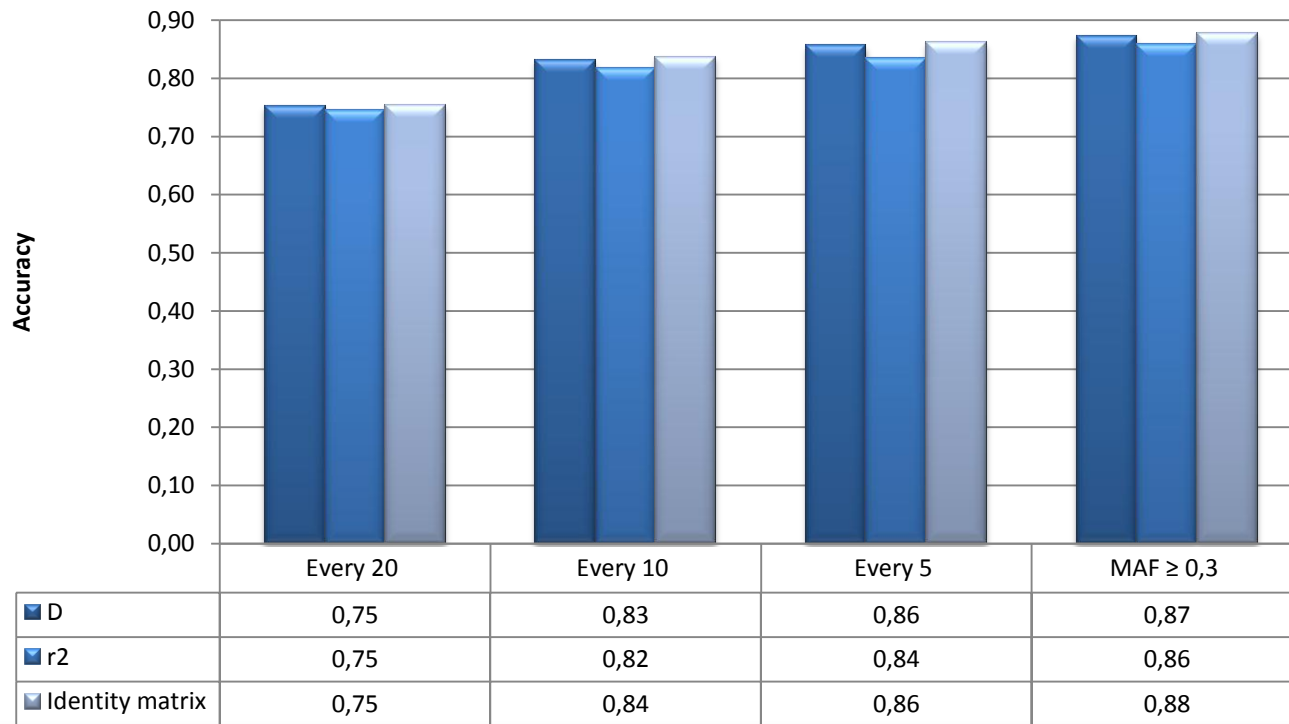
Results

Accuracy between TBV and GBV

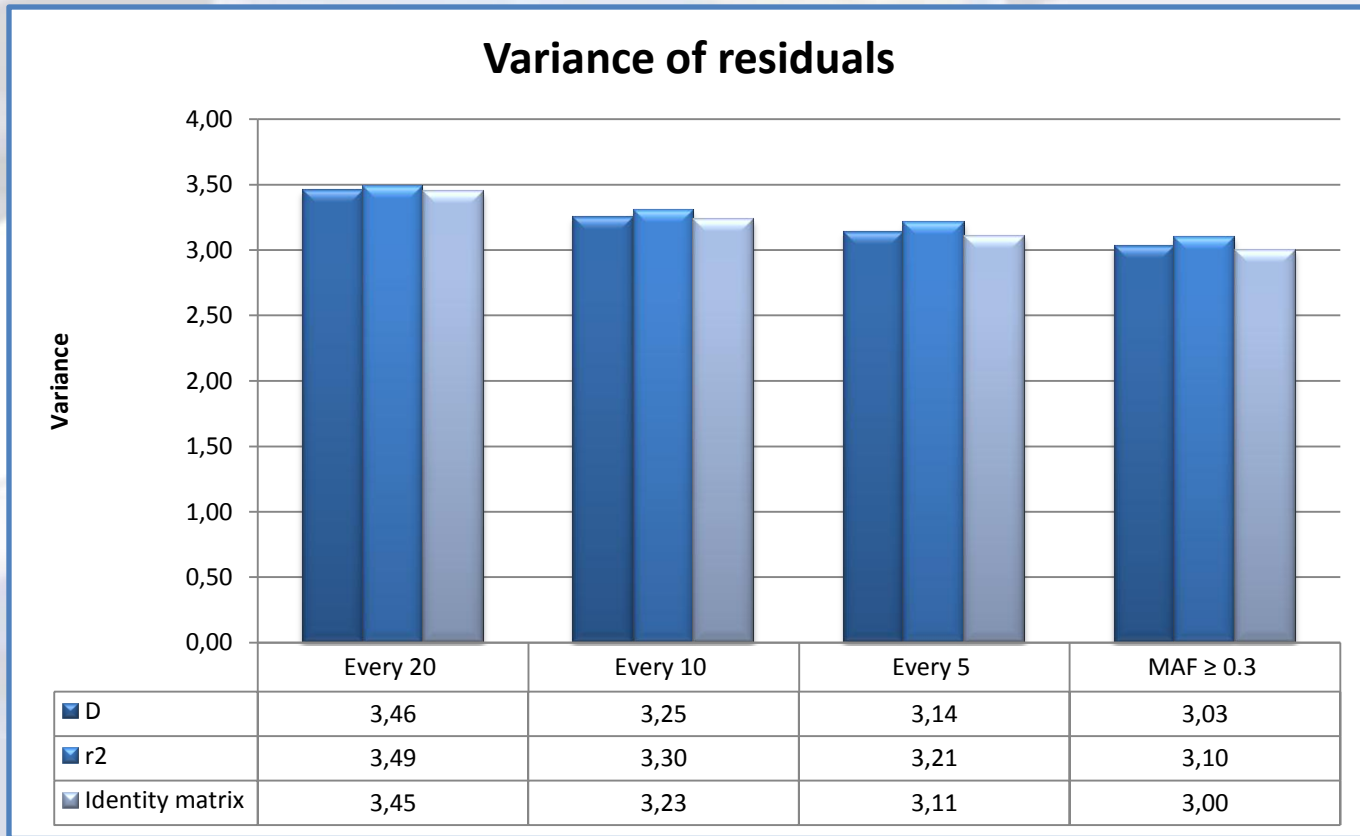


Results

Accuracy between TrueEBV and GBV



Results



Conclusions

- low correlations EBV-GBV
- SNP estimates differ across models
- LD as a covariance matrix don't contribute information to model
- D' isn't useful to this type of analysis

Thank You for Attention

Visit our web site
gen.up.wroc.pl/theta

Home Page About Us Contact Us Teaching

THETA
Statistical Genetics Group
Institute of Animal Genetics

New Lectures

- New lectures for BIOLOGY
- New lectures for ANIMAL HUSBANDRY

Plan

Nearest Seminars, Presentations, Conferences,

Behind Us

Genetic Seminars
21-10-2008; 1.00 pm
meeting place: seminar room

1. Kacper Żukowski; Quantitative Trait Locus-by-Environment Interaction for Milk Yield Traits on *Bos taurus* Autosome 6

Genetic Seminars
28-10-2008; 1.00 pm
meeting place: seminar room

1. Joanna Szyda, Kacper Żukowski; QTL-MAS workshop Salzburg. Ben Hayes's presentation - Linkage Disequilibrium to Genomic Selection (day 1).

Genetic Seminars
4-11-2008; 1.00 pm
meeting place: 7K

1. Katarzyna Rogowska; ??

SEMINAR
20,22-11-2008

Genetic Seminars
14-10-2008; 1.00 pm
meeting place: seminar room

1. Tomasz Suchocki; Modelling a QTL for production on BT46 in Chinese Holstein using a mixed model

Genetic Seminars
7-10-2008; 1.00 pm
meeting place: 7K

1. Tomasz Strzala; Charakterystyka genetyczna polskiej populacji zająca szaraka w oparciu o mitochondrialne DNA

Genetic Seminars
30-09-2008; 9.00 am

1. Tomasz Suchocki; Random Regression

Genetic Seminars
23-09-2008; 11.00 am

Seminars
Projects
Publications
Grants
Conferences
Collaboration

Benjamin Disraeli
There are three kinds of lies: lies, damned lies and statistics...

Joanna Szyda
Heliodor Wierzbicki
Irena Zwolińska-Bartczak
Adrian Drożdż
Anna Gontarek
Anna Macierzyńska
Tomasz Suchocki
Kacper Żukowski