

Using structural equation models to increase of accuracy of QTL estimation in cattle

Jakimowicz M.¹, Suchocki T.^{1,2}, Żarnecki A.², Skarwecka M.², Szyda J.^{1,2}, Jagusiak W.³

¹ Wrocław University of Environmental and Life Sciences, Department of Genetics, THETA Biostatistic Group

²National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland.

³ Department of Genetics, Animal Breeding and Ethology, University of Agriculture in Krakow, al. Mickiewicza 24/28, 30-059 Krakow, Poland



Outline

- ➊ Aim of study
- ➋ Introduction
- ➌ Data
- ➍ Methods
- ➎ Results
- ➏ Conclusion & further steps

Aim of study



The main aim of this study was to validate whether including information about relationships between traits can improve QTL estimation accuracy.

Introduction

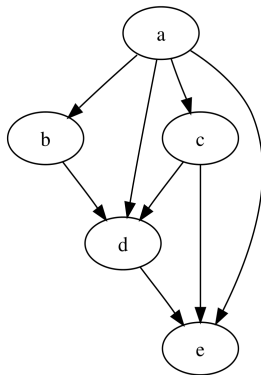


Introduction: QTL estimation in cattle

- 1 Due to economical reasons QTL estimation in cattle isn't a new thing
- 2 The current methods are not the most optimal, therefore the methodology is constantly being improved.
- 3 One idea for improving the results is to incorporate information about relationships between features into currently used models
- 4 Structural equation modelling (SEM) can include information about the relationship between traits to the model.

Introduction: Bayesian Network (BN)

- one of the methods that allows to determine the relationship between variables
- probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG)
- two stages: structure learning, parameter learning





Introduction: Conception rate

Conception rate (CR) is one of the traits included in the fertility group. It is very important group of traits due to impact on animal welfare and longevity, and for the economical reasons. CR is defined as success or failure to conceive after each artificial insemination (AI) or natural mating in cows.

Data



Data: Genotypes

- 5 121 Polish Holstein-Friesian cows.
- genotyping: Illumina BovineSNP50 BeadChip Version 3
- filtration: minor allele frequency of at least 0.01, the technical quality of the chip, with a minimum call rate of 98%
- included 46 216 SNPs



Data: Phenotypes

- conception rate measured in first three lactations - conception rate 1, conception rate 2, conception rate 3 (CR1, CR2, CR3)
- cows were calved between 1996 and 2022
- on average, there were 4 records for each cow

Methods



Genetic parameters

To estimate variance components following model was used

$$y = X\beta + Z\alpha + e, \quad (1)$$

- y - trait values
- X - design matrix for β
- β - vector of fixed effects comprising a general mean
- α - vector of random additive polygenic effects
- Z - design matrix for α
- e - residuals



Bayesian network

- residuals from mixed model \rightarrow posterior probability
- to structure learning, hill climbing algorithm was used
 - iterative algorithm that starts with an arbitrary solution to a problem, then attempts to find a better solution by making an incremental change to the solution
- to estimate parameters, Maximum Likelihood algorithm was used



Structural Equation Modeling

$$y = \Lambda y + Ws + X\beta + Z\alpha + \epsilon \quad (2)$$

- y - vector of phenotypes
- Λ - matrix of regression coefficient obtained from learned structure of Bayesian network
- W - matrix of genotypes codes of SNP makers
- s - effect of SNP



Differences in effects between models

In order to check how big are the differences in the estimators of the snp effects between the two models, the following statistic was calculated

$$diff_i = |a_i - b_i| \quad (3)$$

- $diff_i$ - difference for i th snp
- a_i - estimator of effect form gwas model for i th snp
- b_i - estimator of effect form gwas-bn model for i th snp

Results



Results: Summary statistics for CR

variable	n	mean	sd	se
CR1	5121	74.49	31.26	0.44
CR2	5121	34.48	40.66	0.57
CR3	5121	14.52	31.03	0.43

Table: Summary statistic for conception rate in each lactation



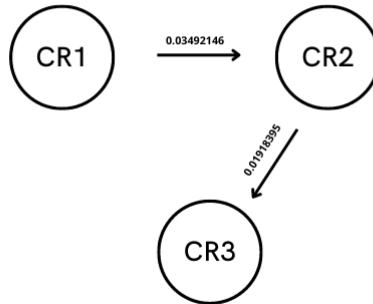
Results: genetic parameters

	genetic	residual	heritability
CR1	74.43	733.25	0.09
CR2	127.29	888.96	0.13
CR3	134.47	871.97	0.13

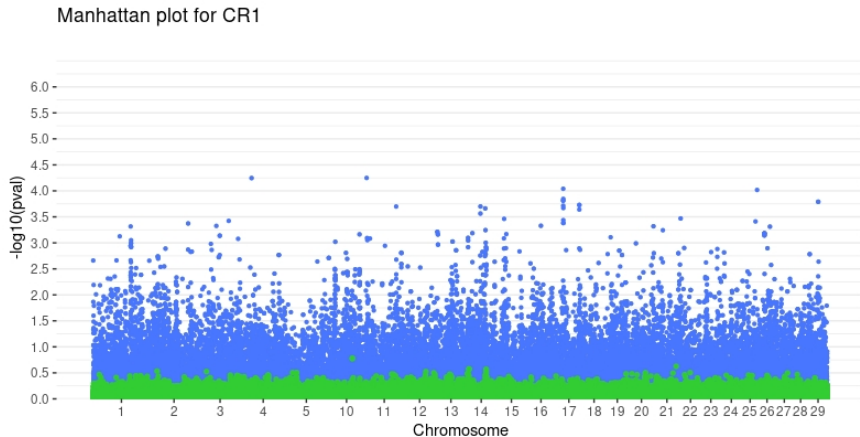
Table: Estimators of genetic parameters for conception rate

Results: Bayesian network

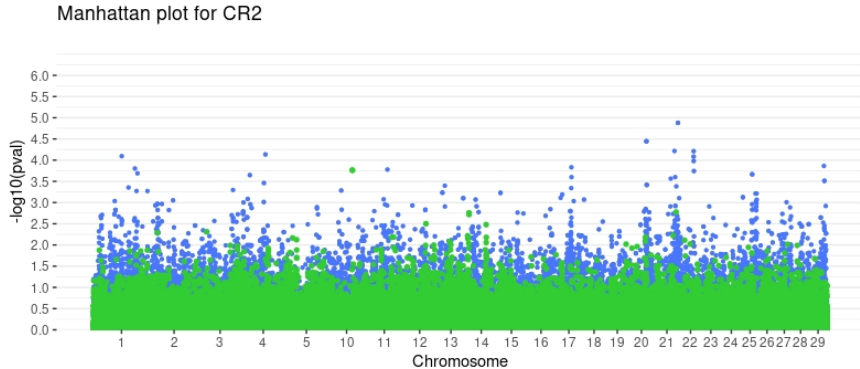
Conception rate for 1st, 2nd and 3rd
lactation



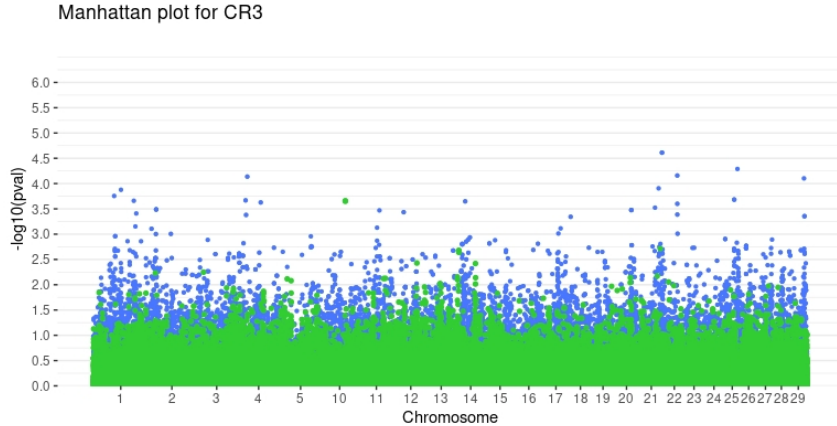
Results: gwas vs. gwas-bn for CR1



Results: gwas vs. gwas-bn for CR2



Results: gwas vs. gwas-bn for CR3





Results: Differences in effects

	CR1	CR2	CR3
mean	0.0024	0.0015	0.0036
sd	0.0021	0.0013	0.0032



Conclusions

- be careful when adding effects to your model and do so with great understanding
- the use of bn reduces the number of false positives



What's next?

- 1 re-examine the data
- 2 validate the models



Acknowledgements

Calculations have been carried out using resources provided by Wrocław Centre for Networking and Supercomputing (<http://wcss.pl>), grant No. 509.

THANK YOU FOR YOUR ATTENTION

Biostatistics Group
Department of Genetics
Wrocław University of Environmental
and Life Sciences
<http://theta.edu.pl>

