S44.3
wittenburg@fbn-dummerstorf.de

61th Annual EAAP Meeting in Heraklion

Including non-additive effects in Bayesian methods for the prediction of genetic values from genome-wide SNP data

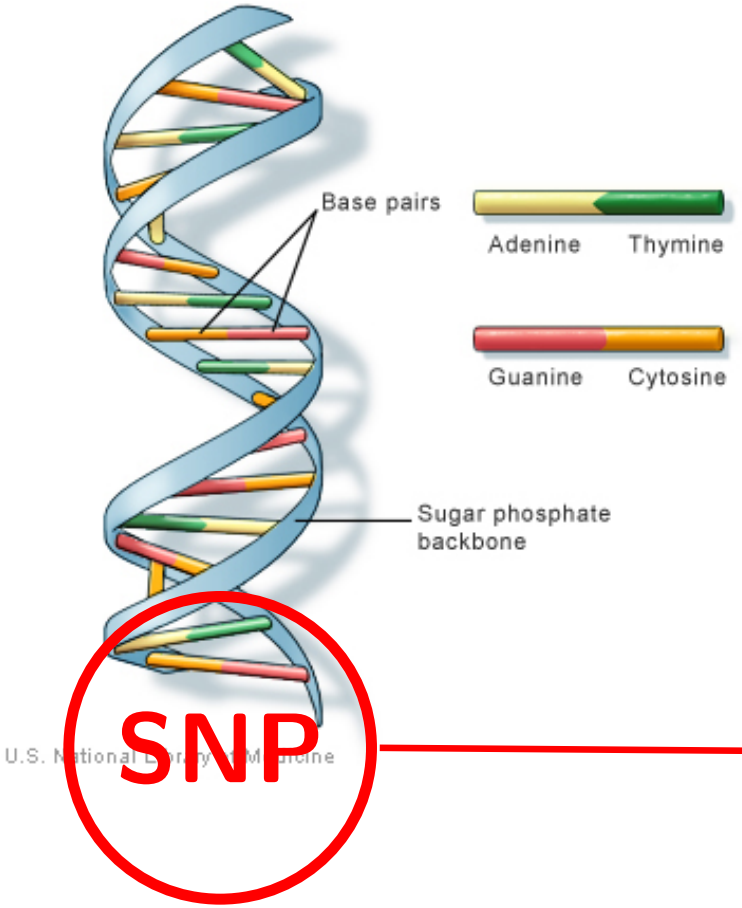
Dörte Wittenburg, Nina Melzer, Norbert Reinsch

August 26, 2010

LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

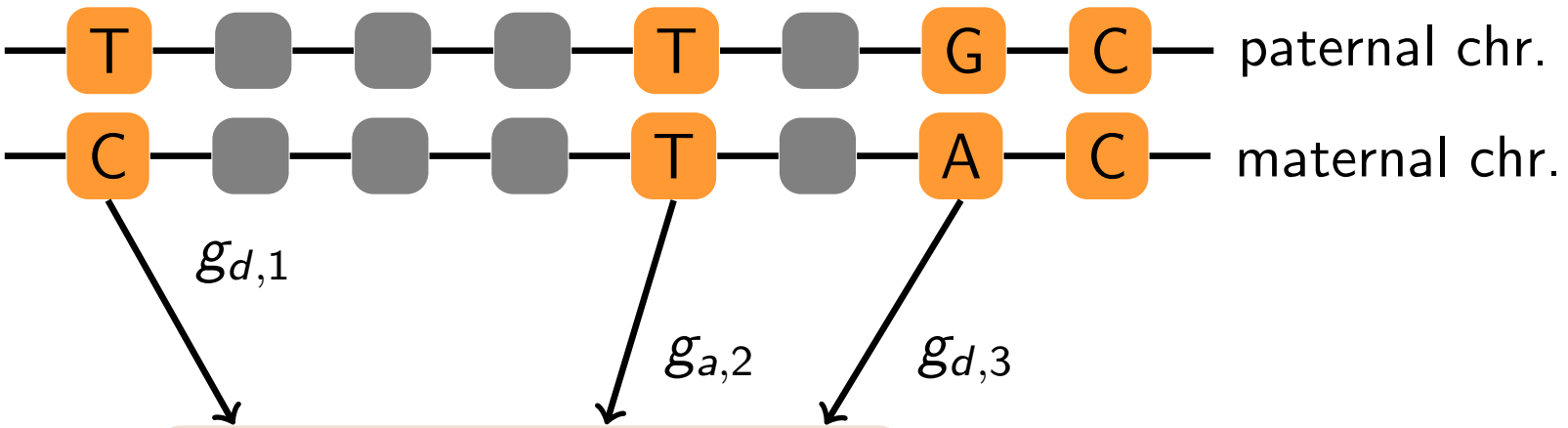
Content

Aim



Prediction of genetic values from genome-wide SNP marker

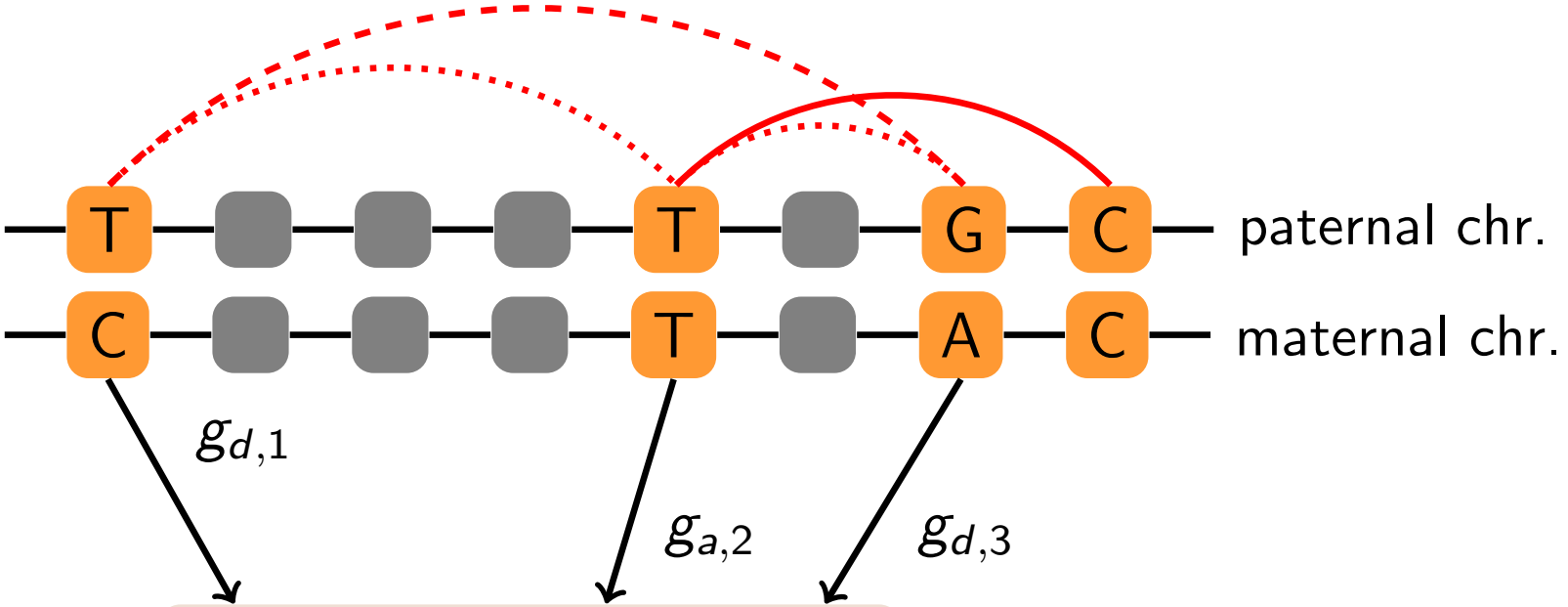
Linear model



$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{X}\mathbf{g}_a + \mathbf{D}\mathbf{g}_d + \mathbf{e}$$

Model II regression: $X_{i,j} \in \{-1, 0, 1\}$ and $D_{i,j} \in \{0, 1, 0\}$
$i = 1, \ldots, n$ (observations), $j = 1, \ldots, m$ (loci)

Linear model with epistatic effects



$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{X}\mathbf{g}_a + \mathbf{D}\mathbf{g}_d + \mathbf{e} \qquad \text{(M-dom)}$$

$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{X}\mathbf{g}_a + \mathbf{D}\mathbf{g}_d + \sum_{s \in \{aa,ad,da,dd\}} \mathbf{W}_s\mathbf{g}_s + \mathbf{e} \qquad \text{(M-epi)}$$

Reparameterisation of **X** and **D**

Requirements

1. at different loci:
   main genetic effects independently distributed
2. at one locus $j \in \{1, \ldots, m\}$:
   uncorrelated genotypic effects $\mathrm{Cov}(X_{i,j} g_{a,j}, D_{i,j} g_{d,j}) = 0$

Orthogonalisation method

- use genotype probabilities*
- additional standardisation for numerical stability

*[Alvarez-Castro & Carlborg, 2007]

Bayesian solver

Gold standard: MCMC method BayesB*

➜ high computing time (even for additive effects)

Now: approximative Bayesian approach fBayesB†

- iterative procedure
- developed under pure additivity
- extended to non-additive effects

*[Meuwissen et al., 2001]

†[Meuwissen et al., 2009]

Bayesian solver

Gold standard: MCMC method BayesB*
➜ high computing time (even for additive effects)

Now: approximative Bayesian approach fBayesB†

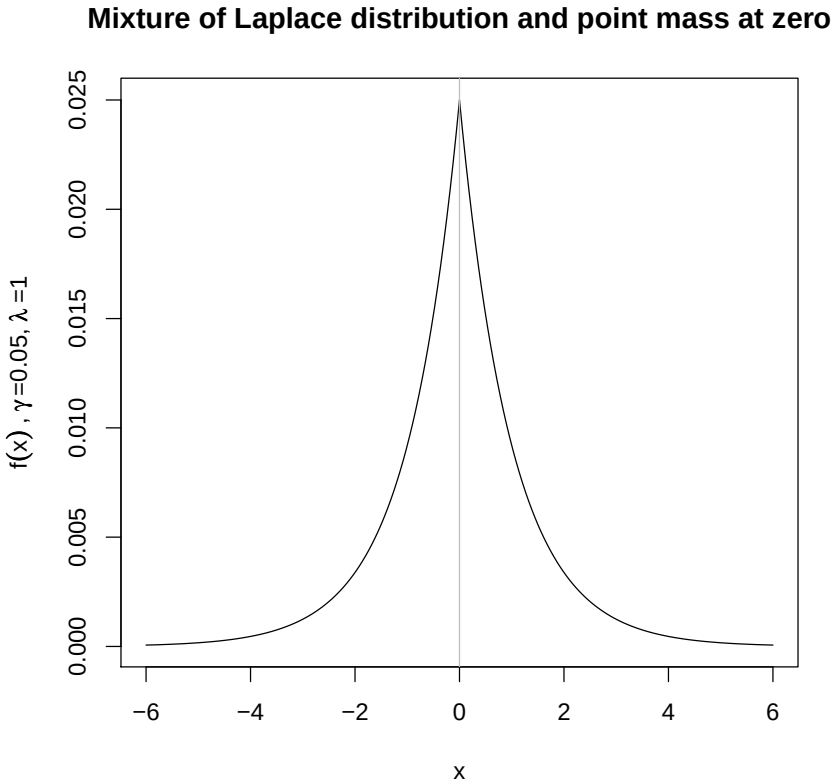
- iterative procedure
- developed under pure additivity
- extended to non-additive effects

*[Meuwissen et al., 2001]
†[Meuwissen et al., 2009]

Linear model in a Bayesian framework

Idea: one SNP effect at a time $\mathbf{Y}_{-j} = \mathbf{X}_{j,s} g_{j,s} + \mathbf{e}$

For each type of genetic effect $s \in \{a, d, aa, ad, da, dd\}$



Prior distribution:

$$\begin{aligned} g_{s,j} &\sim L^*(\gamma_s, \lambda_s) \\ \Pr(g_{s,j} = 0) &= 1 - \gamma_s \\ \mathrm{Var}(g_{s,j}) &= \gamma_s \frac{2}{\lambda_s^2} \end{aligned}$$

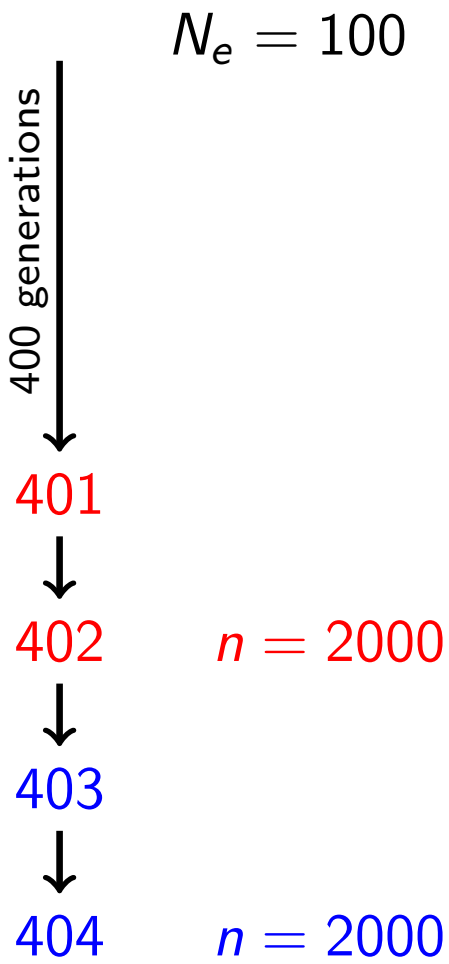
Posterior expectation:

$$\widehat{g}_{s,j} = E(g_{s,j} | \mathbf{Y} = \mathbf{y}_{-j})$$

Simulation study

- 52 273 SNP markers on 30 Morgan
- mutation-drift model
  mutation rate $2.5 \cdot 10^{-3}$ ($DH = 0.33$)*
- recombination events ($LD = 0.12$)*
- 23 loci with main effects
- optional: epistatic effects for 6 loci pairs for each type of interaction
- residual variance according to broad-sense heritability $H^2 = 0.5$

➜ every 10th marker ($m = 5\,227$)

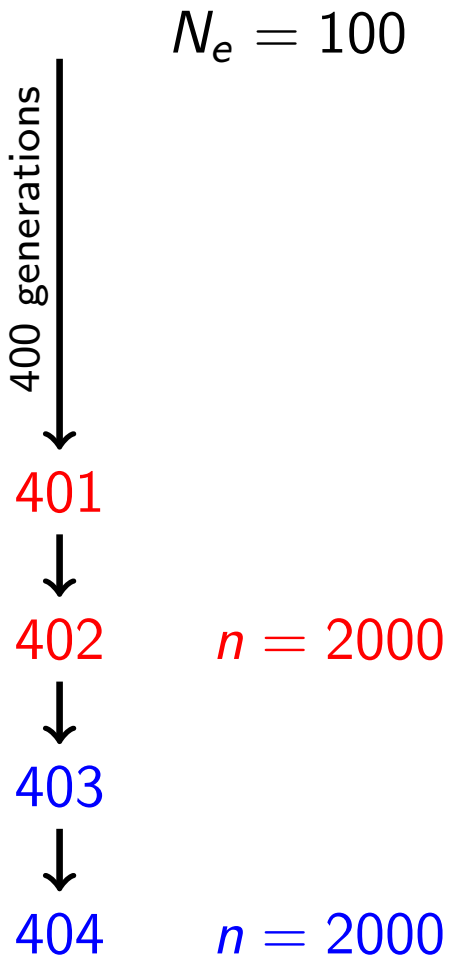


*DH degree of heterozygosity; LD linkage disequilibrium

Simulation study

- 52 273 SNP markers on 30 Morgan
- mutation-drift model
  mutation rate $2.5 \cdot 10^{-3}$ ($DH = 0.33$)*
- recombination events ($LD = 0.12$)*
- 23 loci with main effects
- optional: epistatic effects for 6 loci pairs for each type of interaction
- residual variance according to broad-sense heritability $H^2 = 0.5$

➜ every 10th marker ($m = 5\,227$)

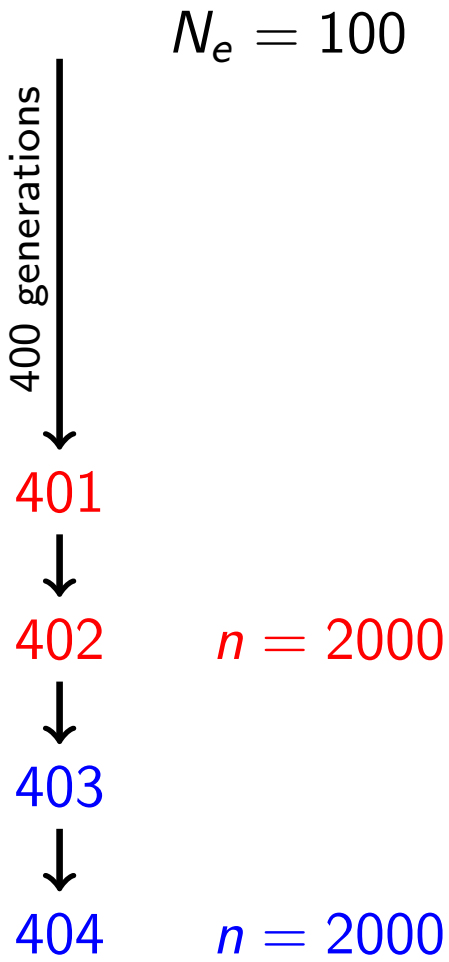


*DH degree of heterozygosity; LD linkage disequilibrium

Simulation study

- 52 273 SNP markers on 30 Morgan
- mutation-drift model
  mutation rate $2.5 \cdot 10^{-3}$ ($DH = 0.33$)*
- recombination events ($LD = 0.12$)*
- 23 loci with main effects
- optional: epistatic effects for 6 loci pairs for each type of interaction
- residual variance according to broad-sense heritability $H^2 = 0.5$

➜ every 10th marker ($m = 5\,227$)



*DH degree of heterozygosity; LD linkage disequilibrium

Average estimated variance components

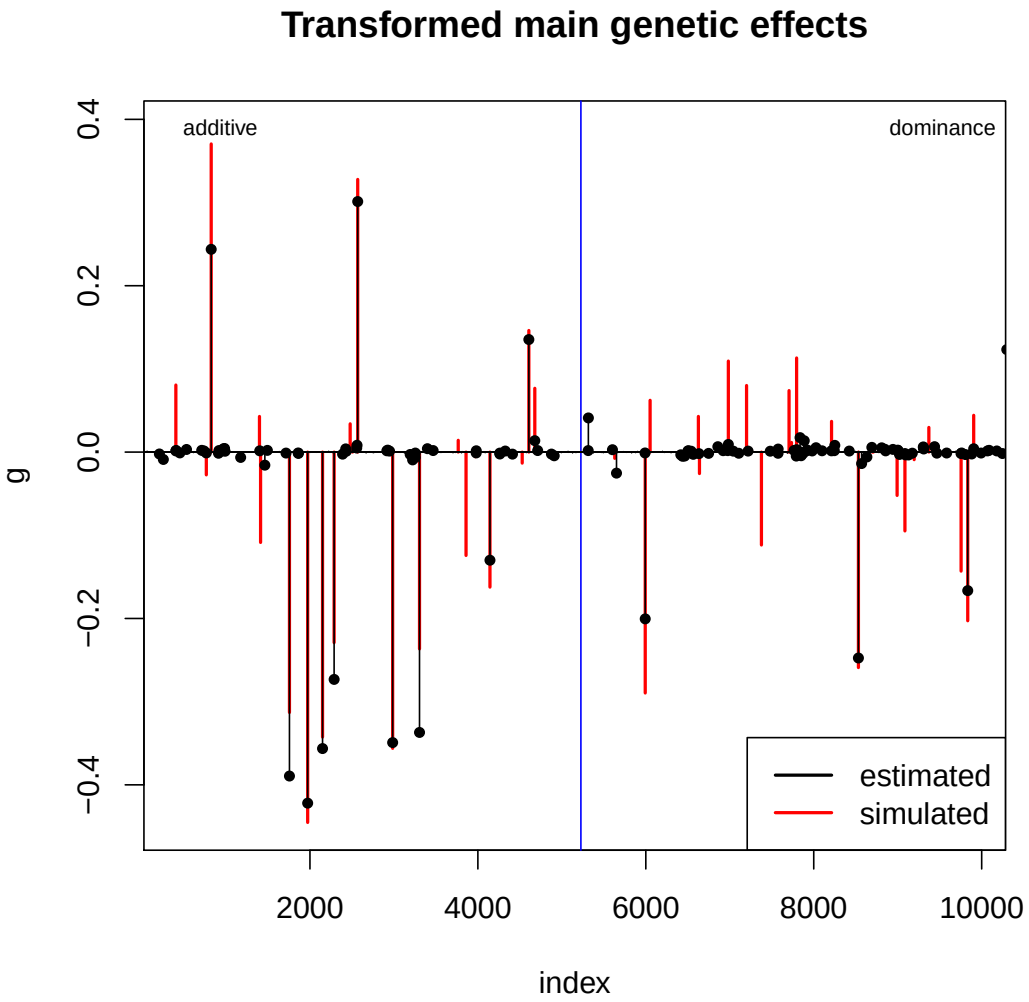
Simulation without epistasis

		σ^2_a	σ^2_d	σ^2_{aa}	σ^2_{ad}	σ^2_{da}	σ^2_{dd}
(M-dom)	BayesB	0.746	0.039	–	–	–	–
(M-dom)	fBayesB	0.742	0.035	–	–	–	–
(M-epi)	fBayesB	0.748	0.039	0.008	0.007	0.007	0.008
Simulated		**0.757**	**0.040**	–	–	–	–

Simulation with epistasis

		σ^2_a	σ^2_d	σ^2_{aa}	σ^2_{ad}	σ^2_{da}	σ^2_{dd}
(M-dom)	BayesB	1.324	0.176	–	–	–	–
(M-dom)	fBayesB	1.310	0.161	–	–	–	–
(M-epi)	fBayesB	1.338	0.193	0.299	0.138	0.065	0.057
Simulated		**1.409**	**0.217**	**0.346**	**0.133**	**0.089**	**0.020**

Estimated genetic effects



Simulation example with epistasis

- good estimates of size and position of big effects
- insufficient results for moderate to small effects

Accuracy of fBayesB

- average correlation between simulated and predicted genetic values*, $H^2 = 0.5$
  - $0.95 - 0.97$ simulation without epistasis
  - $0.74 - 0.85$ simulation with epistasis
- decrease in accuracy about 5 % when $H^2 = 0.3$
- accuracy of breeding value prediction for selection candidates (10 best) at high level* $0.93 - 0.98$

less computing time than BayesB: 1 sec vs. 4 h (M-dom)

. . . **but**

*BayesB +1 %

Accuracy of fBayesB

- average correlation between simulated and predicted genetic values*, $H^2 = 0.5$
  - $0.95 - 0.97$ simulation without epistasis
  - $0.74 - 0.85$ simulation with epistasis
- decrease in accuracy about 5 % when $H^2 = 0.3$

- accuracy of breeding value prediction for selection candidates (10 best) at high level* $0.93 - 0.98$

less computing time than BayesB: 1 sec vs. 4 h (M-dom)

. . . **but**

*BayesB +1 %

Accuracy of fBayesB

- average correlation between simulated and predicted genetic values*, $H^2 = 0.5$
  - $0.95 - 0.97$ simulation without epistasis
  - $0.74 - 0.85$ simulation with epistasis
- decrease in accuracy about 5 % when $H^2 = 0.3$

- accuracy of breeding value prediction for selection candidates (10 best) at high level* $0.93 - 0.98$

less computing time than BayesB: 1 sec vs. 4 h (M-dom)

. . . **but**

*BayesB +1 %

Discussion

1. Loss of accuracy for increasing # QTL or # markers
   - $\leq$ 0.81 for 23 QTL and 52 273 markers
   - $\leq$ 0.80 for 230 QTL and 5 227 markers
   - $\leq$ 0.61 for 230 QTL and 52 273 markers
2. fBayesB sensitive in choice of hyper-parameter γ_s (M-epi)
   - cross validation for initialisation?
   - emBayesB*

*[Stepherd et al., 2010]

Discussion

1. Loss of accuracy for increasing # QTL or # markers
   - $\leq$ 0.81 for 23 QTL and 52 273 markers
   - $\leq$ 0.80 for 230 QTL and 5 227 markers
   - $\leq$ 0.61 for 230 QTL and 52 273 markers
2. fBayesB sensitive in choice of hyper-parameter γ_s (M-epi)
   - cross validation for initialisation?
   - emBayesB*

*[Stepherd et al., 2010]

Conclusions:

- **fBayesB is convenient for genetic value prediction including non-additive effects**
- **sensitive in the choice of hyper-parameter**

This work is part of the FUGATO^{+} project **BovIBI** financed by the Federal Ministry of Education and Research (BMBF).

Thanks for your attention!

Conclusions:

- **fBayesB is convenient for genetic value prediction including non-additive effects**
- **sensitive in the choice of hyper-parameter**

This work is part of the FUGATO^{+} project **BovIBI** financed by the Federal Ministry of Education and Research (BMBF).

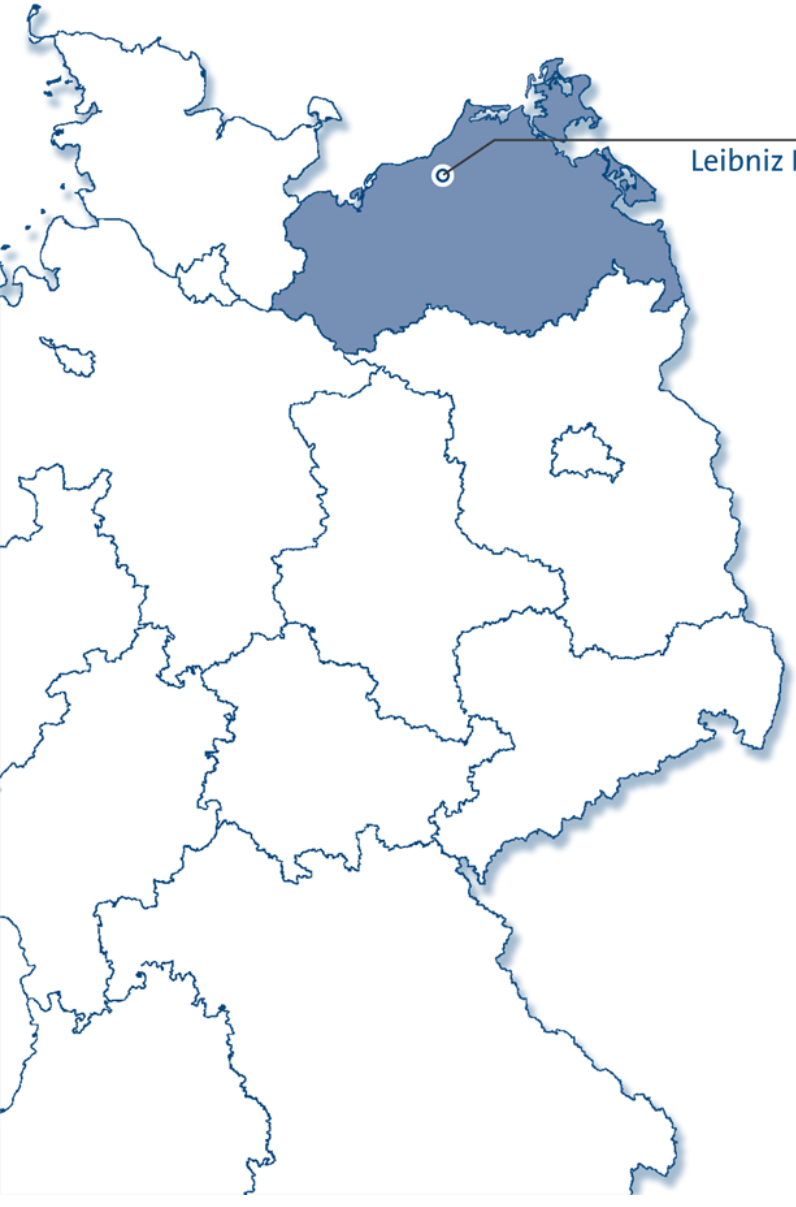
References


Álvarez-Castro, J. and Carlborg, Ö (2007)
A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis
Genetics 176: 1151–1167


Stepherd, R., Meuwissen, T. and Wooliams, J. (2010)
A fast EM algorithm for genomic selection
Proceedings of the 9th WCGALP in Leipzig, Germany, abstract ID 542


Meuwissen, T., Solberg, T., Shepherd, R. and Woolliams, J. (2009)
A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value
Genetics Selection Evolution 41: 2


Meuwissen, T., Hayes, B. and Goddard, M. (2001)
Prediction of total genetic value using genome-wide dense marker maps
Genetics 157: 1819–1829



Leibniz Institute for Farm Animal Biology FBN

Wilhelm-Stahl-Allee 2

18196 Dummerstorf, Germany

Contact

Dr. Dörte Wittenburg

Phone: +49 38208 68930

Fax: +49 38208 68902

Email: wittenburg@fbn-dummerstorf.de

Internet: www.fbn-dummerstorf.de