27.08.2012 EAAP – Closing the phenomic gap

Milk metabolites as new traits and their role for genetic evalution of traditional milk traits

Nina Melzer, Dörte Wittenburg, Dirk Repsilber

LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

General information

genetic variation

phenotypic variation

SNPs (single nucleotide polymorphisms)

milk traits:
fat content
protein content

Genomic selection

50K

classical approach

Genomic selection

50K

classical approach

50K

reduced classical approach

Genomic selection

50K

classical approach

3-6K

reduced classical approach

Our approach

50K

classical approach

3-6K

reduced classical approach

50K

Metabolite profiles

metabolite approach

Our approach

50K

classical approach

3-6K

reduced classical approach

50K

Metabolite profiles

metabolite approach

Our approach

50K

classical approach

3-6K

reduced classical approach

Metabolite profiles

metabolite approach

Our approach

50K

classical approach

3-6K

reduced classical approach

Metabolite profiles

metabolite approach

Experimental data

1,305 Holstein cows (1. lactation):

- **genotypes**
  - Illumina 50k SNP-chip
  - in total: 40,317 SNPs

- **milk metabolite profiles – hydrophilic phase of milk**
  - amino acids
  - sugar
  - carbon acids
  - in total: 190 metabolites (~ 2,000 milk metabolites are expected in the literature)

- **milk traits**
  - fat content
  - protein content
  - pH value

Statistical Model

$$y_{ijkl} = ah_i * stp_j + gld_k + b_1 * ltp + b_2 * ltp^2 + se_l + \varepsilon_{ijkl}$$

with

farm (ah)	i = 1,...,18,
test day (stp)	j = 1,...,39,
GC-MS batch (gld)	k = 1,...,47,
day of lactation (ltp)	ltp ϵ {21,...,120},
sire effect (se)	l = 1,...,214

Workflow - Methods

Metabolome

metabolite concentrations

	1	2	3	4	5	6	7	8
cow 1	0.1	0.2	0.1	0.02	1.2	5.2	0.7	0.03
cow 2	0.1	0.4	0.2	0.01	1.0	4.8	0.5	0.12
cow 3	0.1	0.4	0.1	0.06	1.4	5.3	0.6	0.06
cow 4	0.1	0.3	0.4	0.01	1.1	4.9	0.4	0.09

step 1.

Milk trait

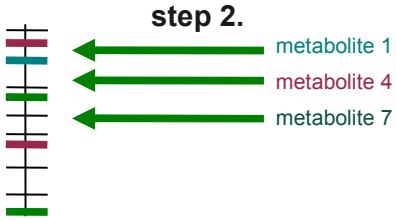
Step 1:
Determining important SNPs for an investigated milk trait.

- random forest (Breiman 2001)
- partial least squares (Wold 1966)

Workflow - Methods

Genome - 50K
Illumina SNP chip

metabolite approach



Metabolome

metabolite concentrations

	1	2	3	4	5	6	7	8
cow 1	0.1	0.2	0.1	0.02	1.2	5.2	0.7	0.03
cow 2	0.1	0.4	0.2	0.01	1.0	4.8	0.5	0.12
cow 3	0.1	0.4	0.1	0.06	1.4	5.3	0.6	0.06
cow 4	0.1	0.3	0.4	0.01	1.1	4.9	0.4	0.09

step 1.

Milk trait

Step 1:
Determining important SNPs for an investigated milk trait.
- random forest (Breiman 2001)
- partial least squares (Wold 1966)

Step 2:
Using a stochastic variable selection (SVS, Ishwaran & Rao 2005) method to determine important SNPs.

Workflow - Methods

Genome - 50K
Illumina SNP chip

Metabolome

metabolite concentrations

	1	2	3	4	5	6	7	8
cow 1	0.1	0.2	0.1	0.02	1.2	5.2	0.7	0.03
cow 2	0.1	0.4	0.2	0.01	1.0	4.8	0.5	0.12
cow 3	0.1	0.4	0.1	0.06	1.4	5.3	0.6	0.06
cow 4	0.1	0.3	0.4	0.01	1.1	4.9	0.4	0.09

metabolite approach

step 2.

metabolite 1
metabolite 4
metabolite 7

step 1.

reduced classical approach

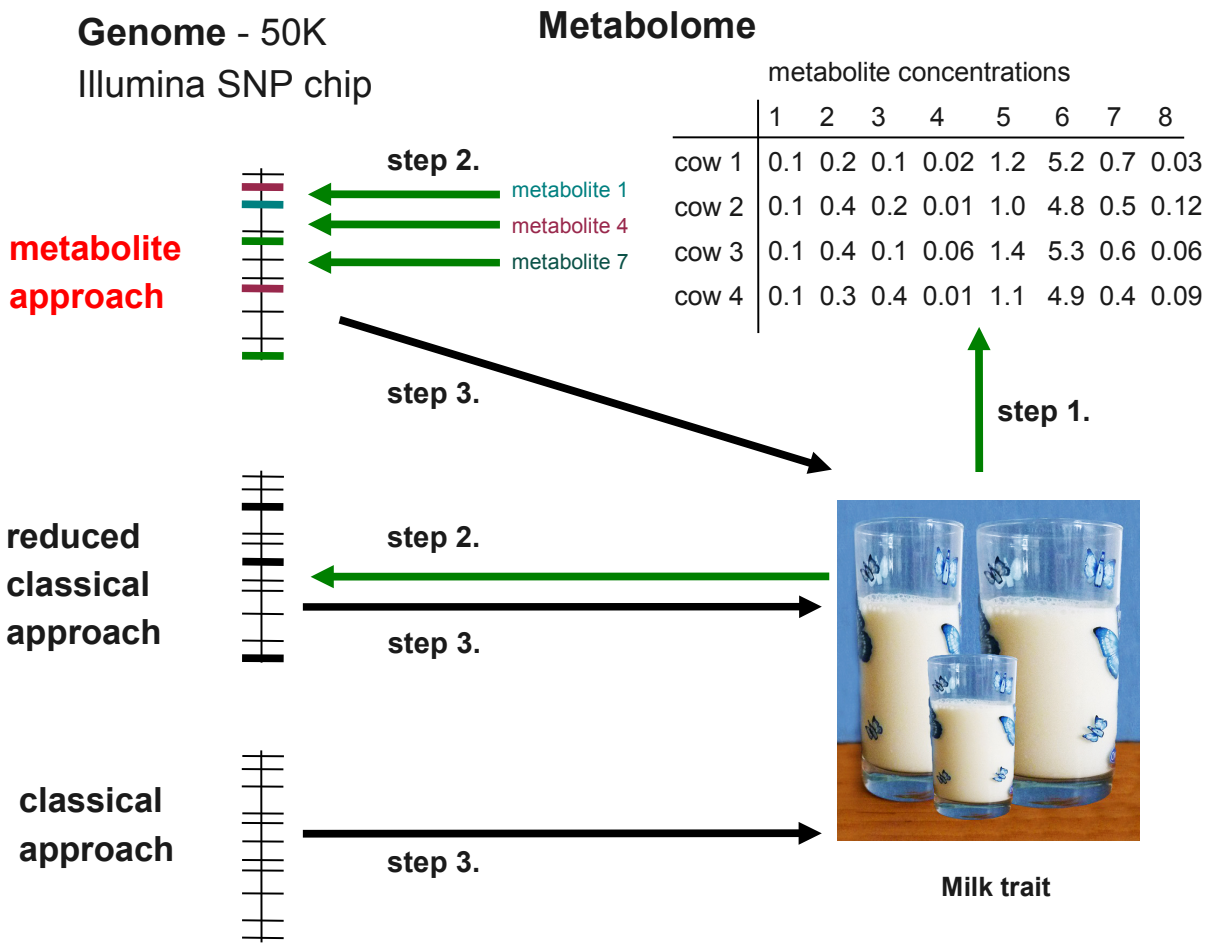
step 2.

Milk trait

Step 1:
Determining important SNPs for an investigated milk trait.
- random forest (Breiman 2001)
- partial least squares (Wold 1966)

Step 2:
Using a stochastic variable selection (SVS, Ishwaran & Rao 2005) method to determine important SNPs.

Workflow - Methods



Step 1:
Determining important SNPs for an investigated milk trait.
- random forest (Breiman 2001)
- partial least squares (Wold 1966)

Step 2:
Using a stochastic variable selection (SVS, Ishwaran & Rao 2005) method to determine important SNPs.

Step 3:
Predicting milk trait from SNP-subset using SVS.

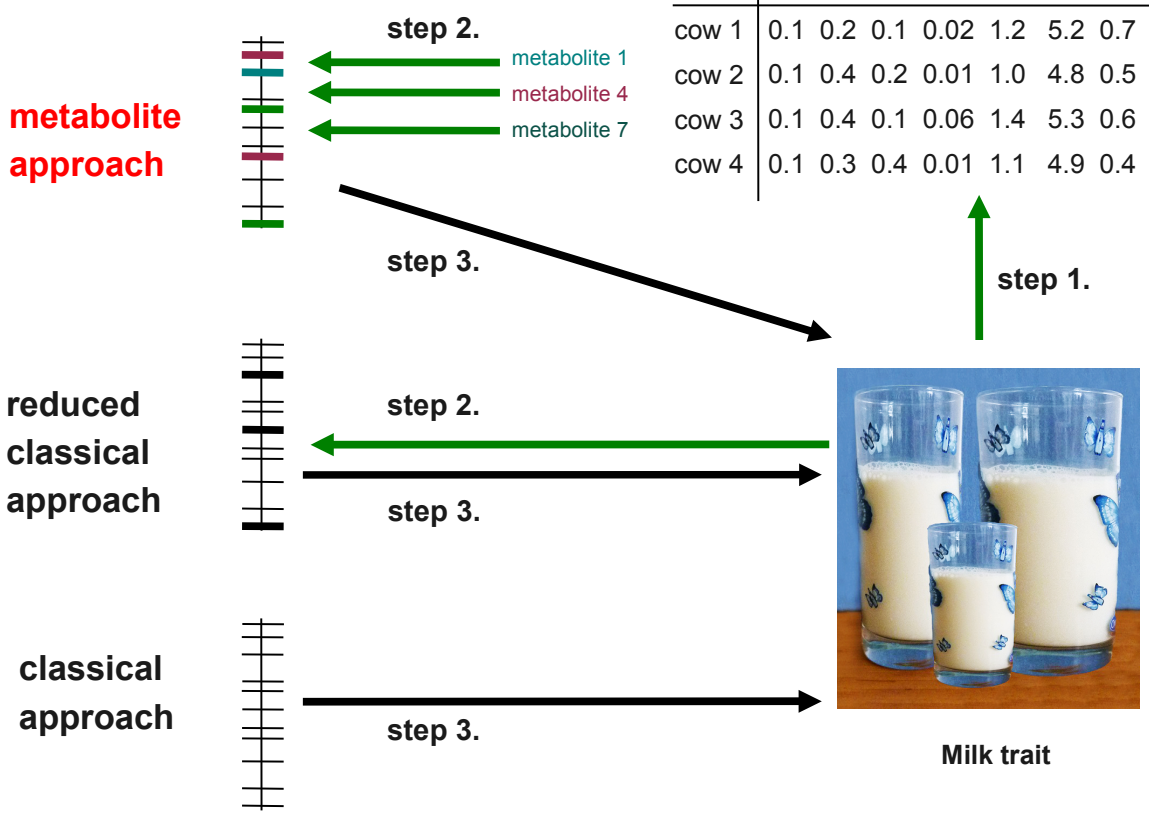
Workflow - Methods

Genome - 50K
Illumina SNP chip

Metabolome

metabolite concentrations

	1	2	3	4	5	6	7	8
cow 1	0.1	0.2	0.1	0.02	1.2	5.2	0.7	0.03
cow 2	0.1	0.4	0.2	0.01	1.0	4.8	0.5	0.12
cow 3	0.1	0.4	0.1	0.06	1.4	5.3	0.6	0.06
cow 4	0.1	0.3	0.4	0.01	1.1	4.9	0.4	0.09



Step 1:
Determining important SNPs for an investigated milk trait.
- random forest (Breiman 2001)
- partial least squares (Wold 1966)

Step 2:
Using a stochastic variable selection (SVS, Ishwaran & Rao 2005) method to determine important SNPs.

Step 3:
Predicting milk trait from SNP-subset using SVS.

10-fold cross-validation

Results – Step 1 (metabolite ← milk trait)

Average prediction precision (correlation between estimated and observed milk trait values):

	Random forest	Partial least squares
Protein	0.63	0.64
Fat	0.35	0.34
pH value	0.37	0.34

Results – Step 1 (metabolite ← milk trait)

Average prediction precision (correlation between estimated and observed milk trait values):

	Random forest	Partial least squares
Protein	0.63	0.64
Fat	0.35	0.34
pH value	0.37	0.34

Average number of detected important metabolites (examples):

Protein	16	(Spermidine, Asparagine, Aspartic acid)
Fat	11	(1,3- Dihydroxyacetone, Arabitol)
pH value	10	(Glycine)

More information can be found in Melzer et al. 2012 (J. Dairy Sci.)

Results – Step 2 (important SNPs ← trait)

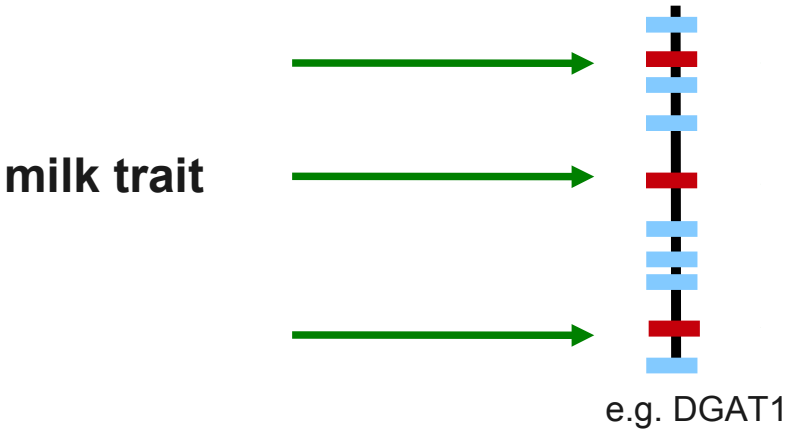
Average number of detected SNPs:

Approach	Fat (%)	Protein (%)	pH value
Reduced classical approach	26	83	82
Metabolite approach	129	302	114
Single important metabolites	[4;42]	[5;55]	[4;54]

Results – Step 2 (important SNPs ← trait)

Average number of detected SNPs:

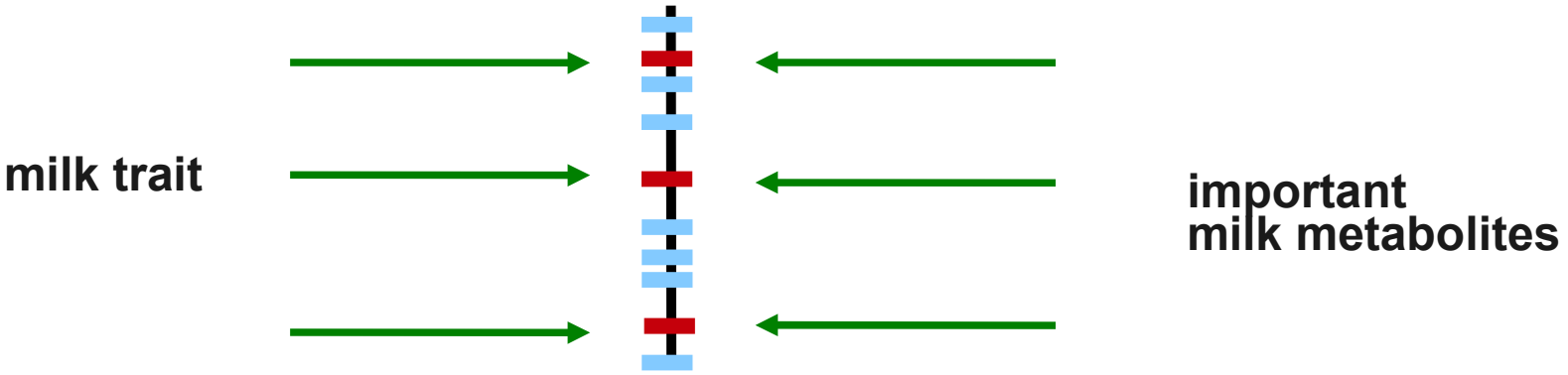
Approach	Fat (%)	Protein (%)	pH value
Reduced classical approach	26	83	82
Metabolite approach	129	302	114
Single important metabolites	[4;42]	[5;55]	[4;54]



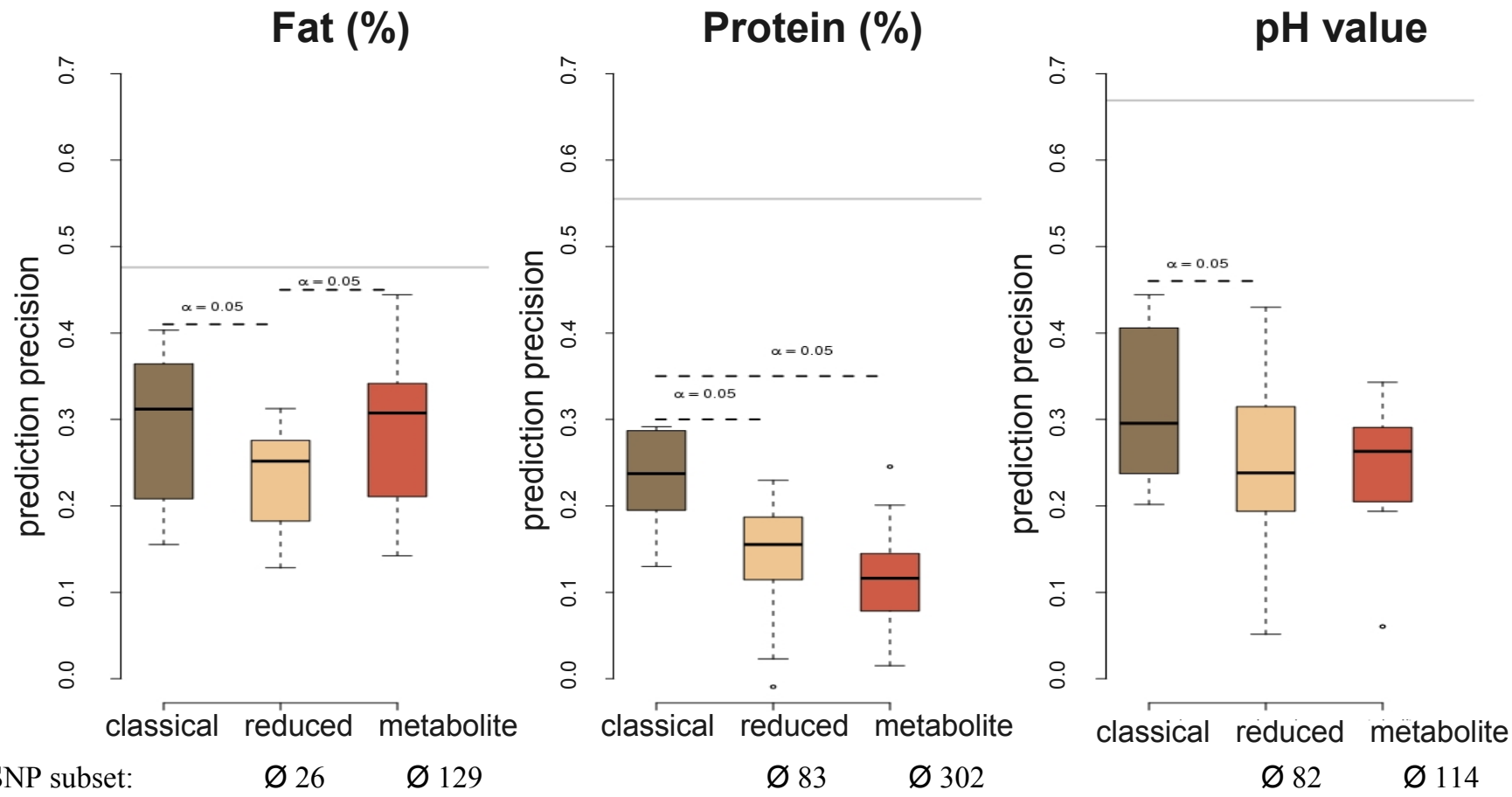
Results – Step 2 (important SNPs ← trait)

Average number of detected SNPs:

Approach	Fat (%)	Protein (%)	pH value
Reduced classical approach	26	83	82
Metabolite approach	129	302	114
Single important metabolites	[4;42]	[5;55]	[4;54]



Results – Step 3 (SNPs ⟶ milk trait)



SNP subset: Ø 26 Ø 129 Ø 83 Ø 302 Ø 82 Ø 114

Summary

Step 1:

- both regression methods revealed similar important metabolites
- biological role of some important metabolites with respect to a specific milk trait was investigated

Summary

Step 1:
- both regression methods revealed similar important metabolites
- biological role of some important metabolites with respect to a specific milk trait was investigated

Step 2:
- mostly fewer important SNPs were detected for single important metabolites
- important SNPs with high genetic effects for milk traits also showed an impact on at least one of the important metabolites

→ considering the heritability of metabolites (Poster Session 12 No. 23: "Milk metabolites and their genetic variability", Wittenburg et al., 2012)

Summary

Step 1:
- both regression methods revealed similar important metabolites
- biological role of some important metabolites with respect to a specific milk trait was investigated

Step 2:
- mostly fewer important SNPs were detected for single important metabolites
- important SNPs with high genetic effects for milk traits also showed an impact on at least one of the important metabolites

→ considering the heritability of metabolites (Poster Session 12 No. 23: "Milk metabolites and their genetic variability", Wittenburg et al., 2012)

Step 3:
- observed prediction precisions were mostly higher for the metabolite approach compared to the reduced classical approach

Conclusion

- considering an intermediate level in the classical genotype-phenotype map enables various investigation opportunities

- for two investigated milk traits, fat and pH value, the genotype-phenotype map is mirrored on the metabolome level

Our metabolite approach seems to be promising and in addition provides functional information.

Thank you for you attention!

Thanks a lot for assistance us:

- Dr. F. Reinhardt, E. Pasmann (VIT Verden)
- Dr. U. Kesting, Dr. S. Jakubowski, Dr. S. Hartwig, S. Wolf (LKV Güstrow)
- Prof. Dr. Meitinger, Dr. P. Lichtner (HelmholzZentrum München)
- Prof. L. Willmitzer, Ä. Eckardt (Max-Planck-Institut)
- Dr. H. Hammon, C. Reiko (FBN, WG nutrition physiology)
- PD Dr. J. Vanselow, Dr. R. Fürbaß, M. Nimz and M. Anders, M. Spitschak (FBN, WG Gene Expression)
- PD Dr. C. Kühn (FBN, WG QTL Regions)
- A. Rief, R. Grahl (FBN)

This study was part of the FUGATO plus project Bovine Integrative Bioinformatics for Genomic Selection (BovIBI) with financial support of the German Federal Ministry of Education and Research (BMBF).

References

[1] Breiman, L. 2001. Random forests. Machine Learning 45(1):5–32.

[2] Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J.-J., Kvasz, A., Mni, M., Simon, P., Frere, J.-M., Coppieters, W. & Georges, M. (2004), Genetic and functional confirmation of the causality of the dgat1 k232a quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. U. S. A. 101(8), 2398–2403.

[3] Ishwaran, H. & Rao, J. S. (2005), Spike and slab variable selection: Frequentist and bayesian strategies. Ann. Stat. 33, 730.

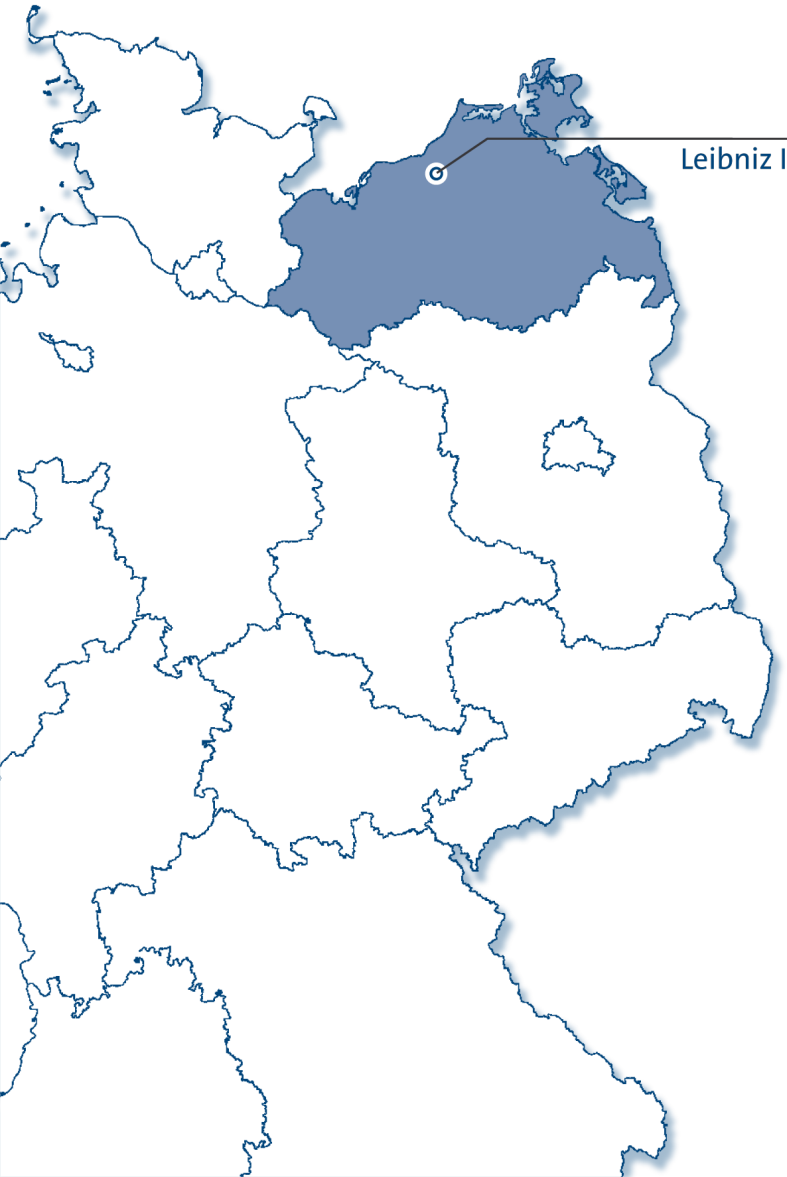
[4] Meuwissen T.H., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome- wide dense marker maps. Genetics.157(4):1819–29

[5] Wittenburg, D. & Reinsch, N. (2011), Application of spike and slab variable selection for the genome-wide estimation of genetic effects and their complexity. In Book of Abstracts of the 62nd Annual Meeting of the EAAP, Wageningen Academic Publishers, 62nd Annual Meeting of the EAAP in Stavanger, Norway, p. 116, ISSN 1382-6077.

[6] Wold, H. 1966. Estimation of Principal Components and Related Models by Iterative Least squares. Academic Press, New York, pp. 391–420.

LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

HOW TO FIND US



Leibniz Institute for Farm Animal Biology FBN

Wilhelm-Stahl-Allee 2
18196 Dummerstorf

Contact

Nina Melzer

Phone: +49 38208 68 934
Fax: +49 38208 68 902
E-Mail: melzer@fbn-dummerstorf.de
Internet: www.fbn-dummerstorf.de