



LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

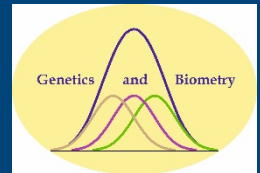
Estimation of recombination rates in a dairy cattle population

A. Hampel, F. Teuscher, D. Wittenburg

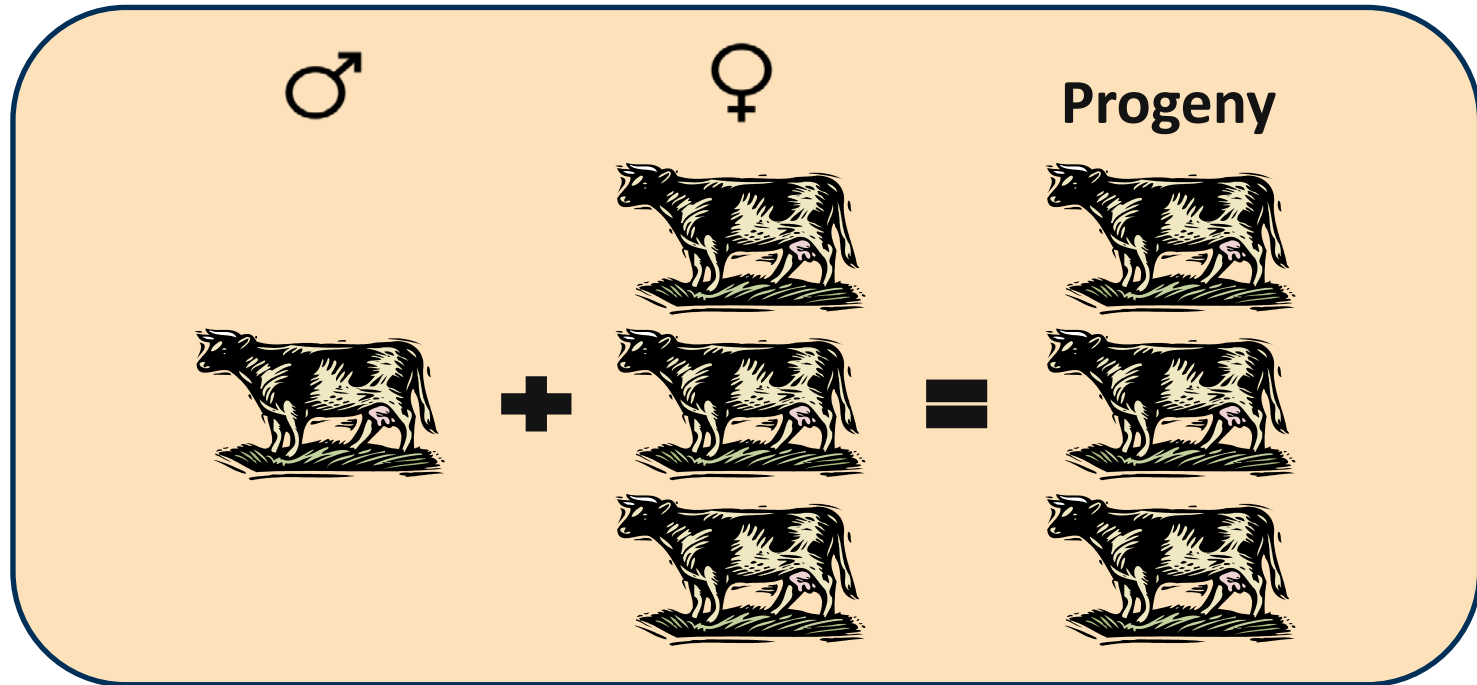
01.09.2016



LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY



Background: Half-sib family



Background: Population parameters

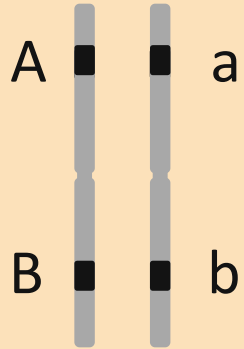
Population structure has influence on population parameters

Population LD
Population recombination rate

Maternal LD

Paternal recombination rate

Background: Linkage disequilibrium



➤ $A; a$... two alleles at a locus (allele frequency f_A, f_a)

➤ $B; b$... two alleles at a locus (allele frequency f_B, f_b)

➤ Frequencies of combinations in a population:

$$f_{AB}, f_{aB}, f_{Ab}, f_{ab}$$

Background: Linkage disequilibrium

Loci are in linkage equilibrium:

$$f_{AB}f_{ab} = f_{Ab}f_{aB}$$

D ... disequilibrium coefficient

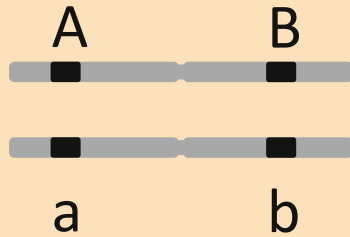
Loci are in linkage disequilibrium:

$$f_{AB}f_{ab} \neq f_{Ab}f_{aB}$$

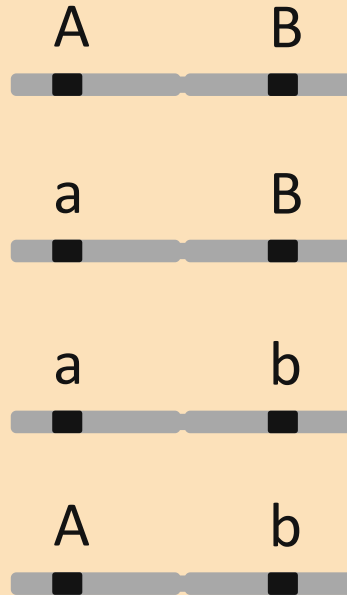
$$D = f_{AB}f_{ab} - f_{Ab}f_{aB}$$

Background: Recombination rate

Paternal diplotype



Daughter generation



Probability

$$\frac{1}{2}(1 - \theta)$$
$$\frac{1}{2}\theta$$
$$\frac{1}{2}(1 - \theta)$$
$$\frac{1}{2}\theta$$

θ ... Recombination rate

Objective: Estimation of LD and recombination rates

New Method

- Minimization approach with less computation time

*EM Method**

- Maximization approach with high computation time

- Both methods were applied to an empirical dataset
- Verification of the accuracy was performed in simulated half-sib families



Approaches: *EM Method*

Parameters

- Maternal haplotype frequencies
- Genotype counts from offspring
- Paternal recombination rate

Solved by applying the EM algorithm

Approaches: *New Method*

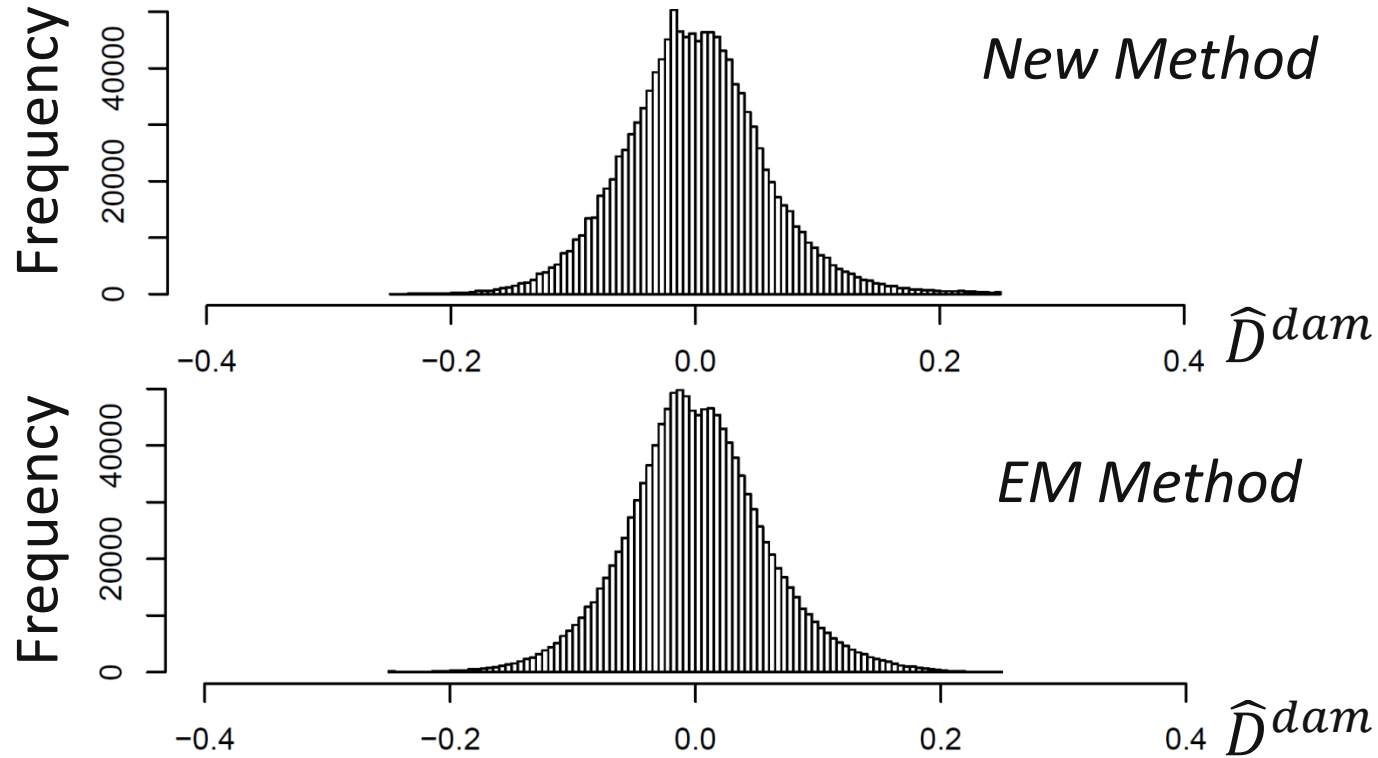
Parameters

- Empirical covariance between genotype codes for **additive**/**dominant** effects at two SNPs
e.g. **AA** → 1, **Aa** → 0, **aa** → -1 / e.g. **AA** → -1, **Aa** → 1, **aa** → -1
- Allele frequency in the maternal population
- LD of dam; LD of sire $\rightarrow \theta = \frac{1-4D^{sire}}{2}$

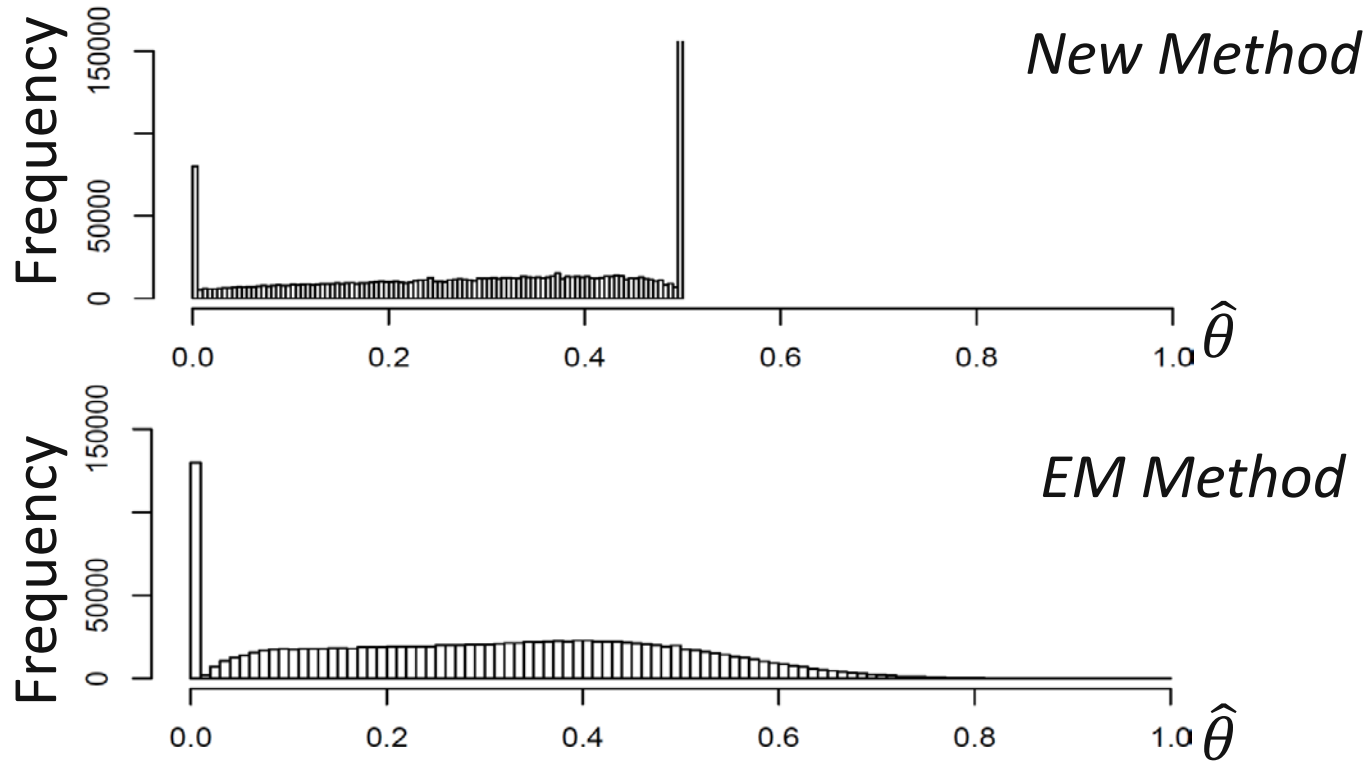
Empirical data set

- Comprised 1295 half-sibs of a dairy cattle population (Fugato-plus “BovIBI” data)
 - 40317 SNP-genotypes (29 autosomes)
- Estimation on BTA1
 - Maternal linkage disequilibrium
 - Paternal recombination rate

Results: LD on chromosome 1



Results: θ on chromosome 1



Simulation: Parameters

LD	Population size	Recombination rate
$D^{sire} = 0.15$	$N \in \{30, 100, 1000\}$	$\theta \in \{0.01, 0.05, 0.10,$
$D^{dam} = 0.05$		$0.20, 0.40, 0.50\}$

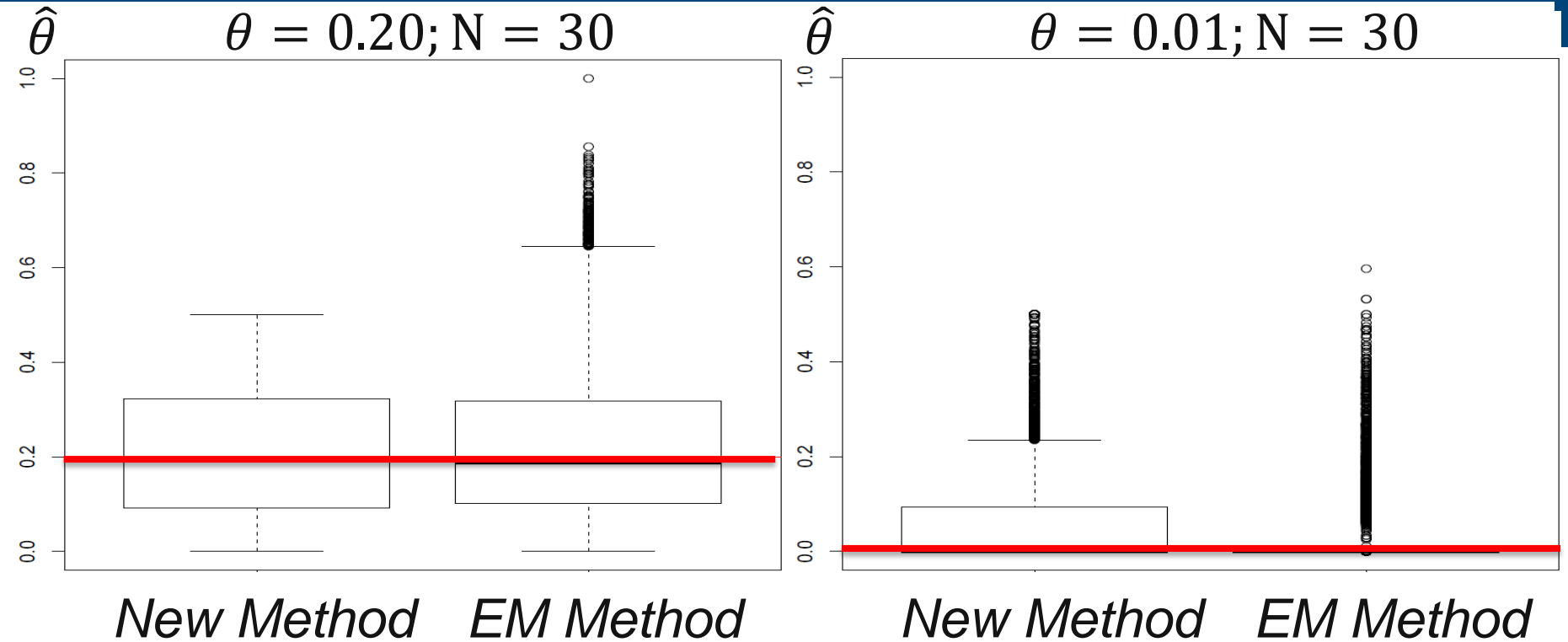
Simulation with own R programs, with 10 000 repetitions

Simulation: Criteria of Comparison

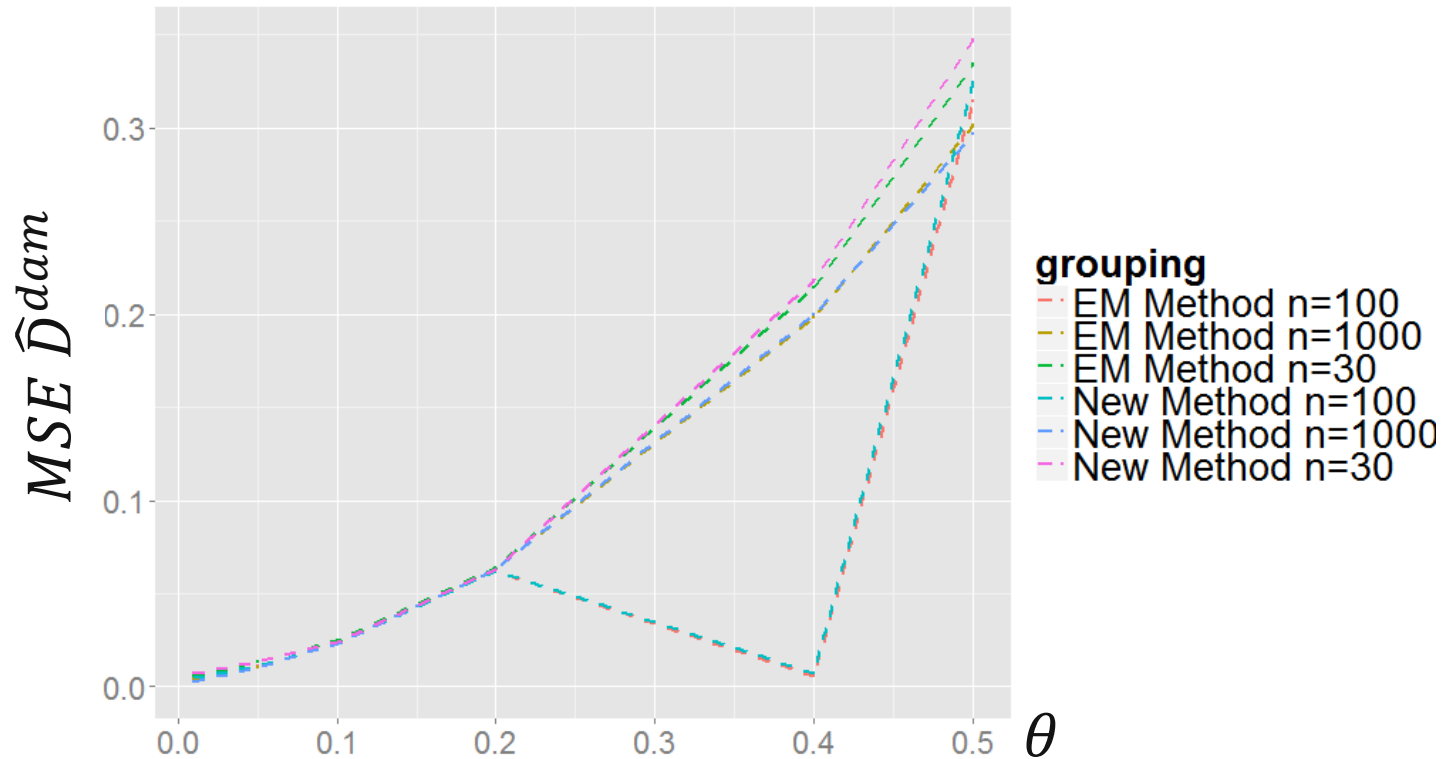
- Bias of paternal recombination rate
- Mean squared error (D^{dam})
- Computation time

Selection of results

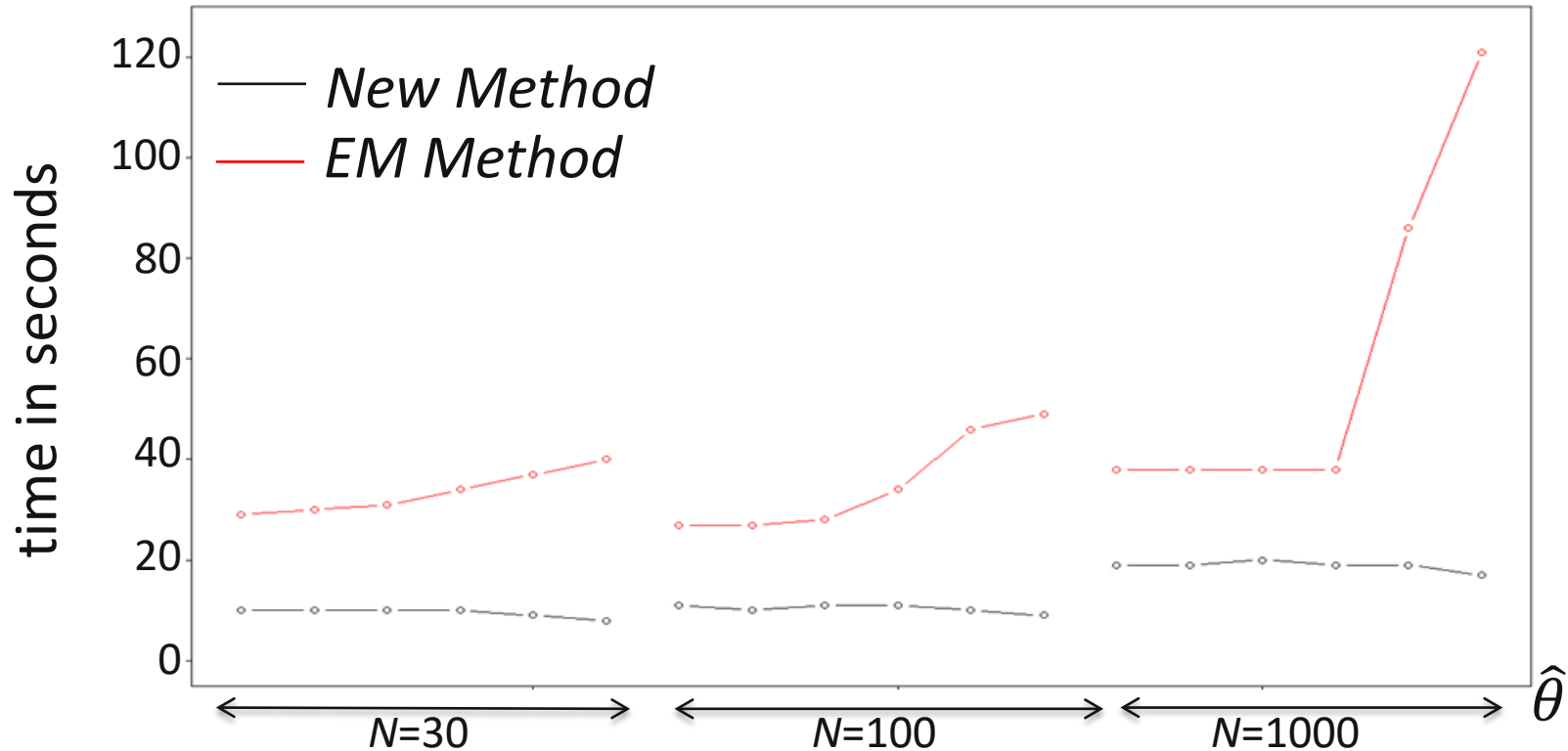
Results: Bias of θ



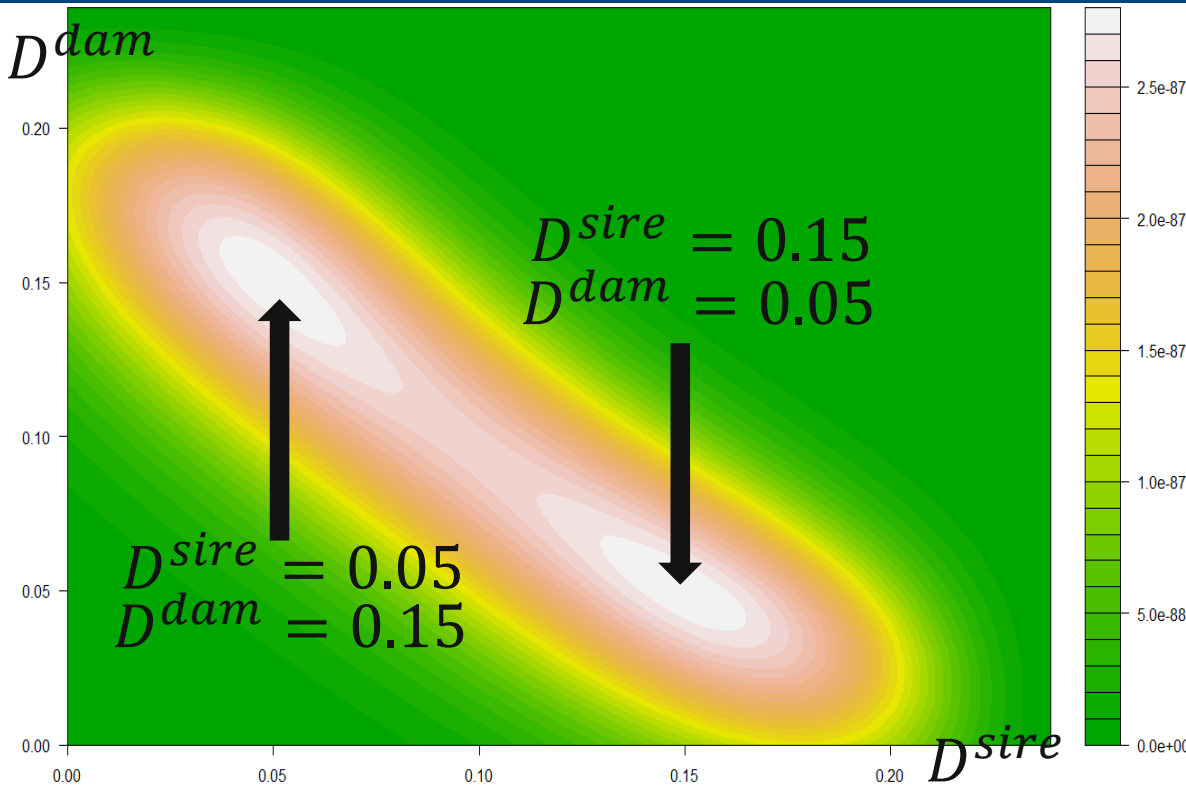
Results: MSE of D^{dam}



Results: Computation time



Results: Likelihood function



Two maxima of likelihood function

Example:

$$f_1^{dam} = f_2^{dam} = 0.5$$

Results: Complementary case

Simulation

$$f_{AB} = 0.30$$

$$f_{ab} = 0.30$$

$$f_{Ab} = 0.20$$

$$f_{aB} = 0.20$$

$$D^{sire} = 0.15$$

$$D^{dam} = 0.05$$

$$\theta = 0.20$$

Complementary case

$$f_{AB} = 0.40$$

$$f_{ab} = 0.40$$

$$f_{Ab} = 0.10$$

$$f_{aB} = 0.10$$

$$D^{sire} = 0.05$$

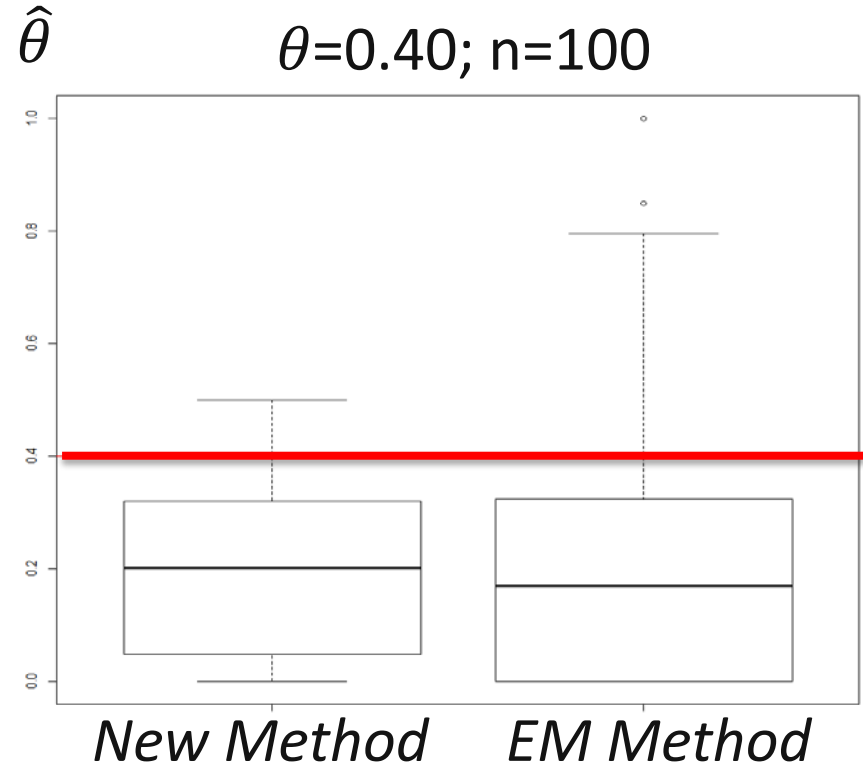
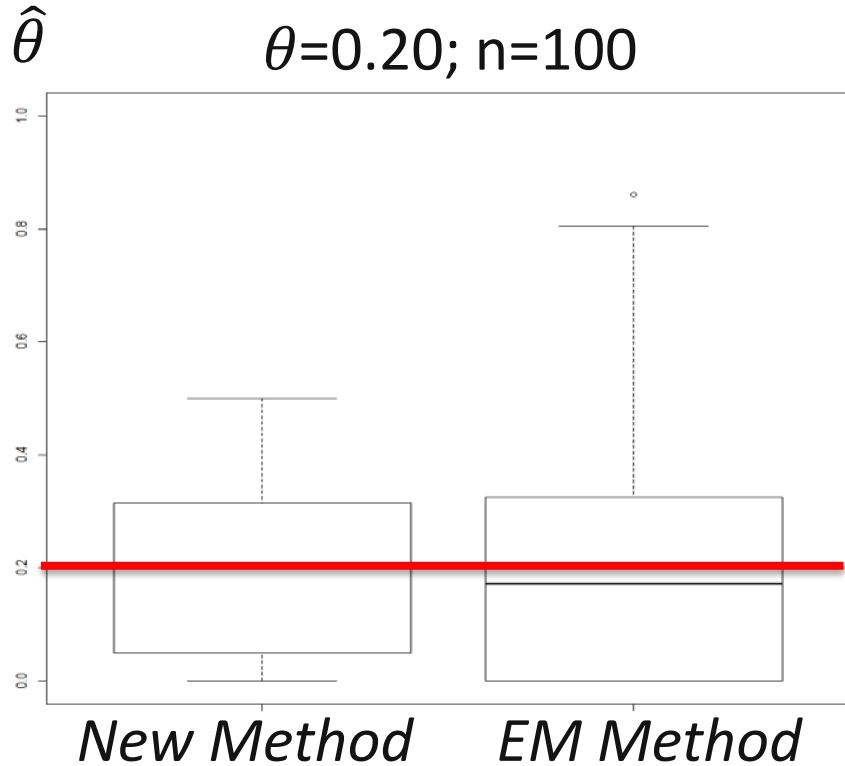
$$D^{dam} = 0.15$$

$$\theta = 0.40$$



recalculation

Results: Simulation of complementary case



Outlook and summary

New Method

- Based on a minimization approach
- Is three times computationally faster than the *EM Method*

Simulation

The *New Method* had more accurate estimates for higher recombination rates

Outlook and summary

Empirical data set

The comparison of results showed that both methods had a similar distribution of the maternal LD but the distribution of the recombination rate differed.

Outlook and summary

Maximization function

- Two possible solutions → two maxima
- Found in both New and EM Method
- Final solution depends on starting values

Outlook and summary

Next steps

- Criterion for distinction of both maxima (likelihood values)
- Estimation of good starting values
(combination of minimization and maximization approach)

Appendix: New Method

$$\widehat{cov}_{add} = D^{sire} + D^{dam}$$

$$\widehat{cov}_{dom} = 16D^{sire}D^{dam} + 4D^{sire}(1 - 2\hat{f}_1^{dam})(1 - 2\hat{f}_1^{dam})$$

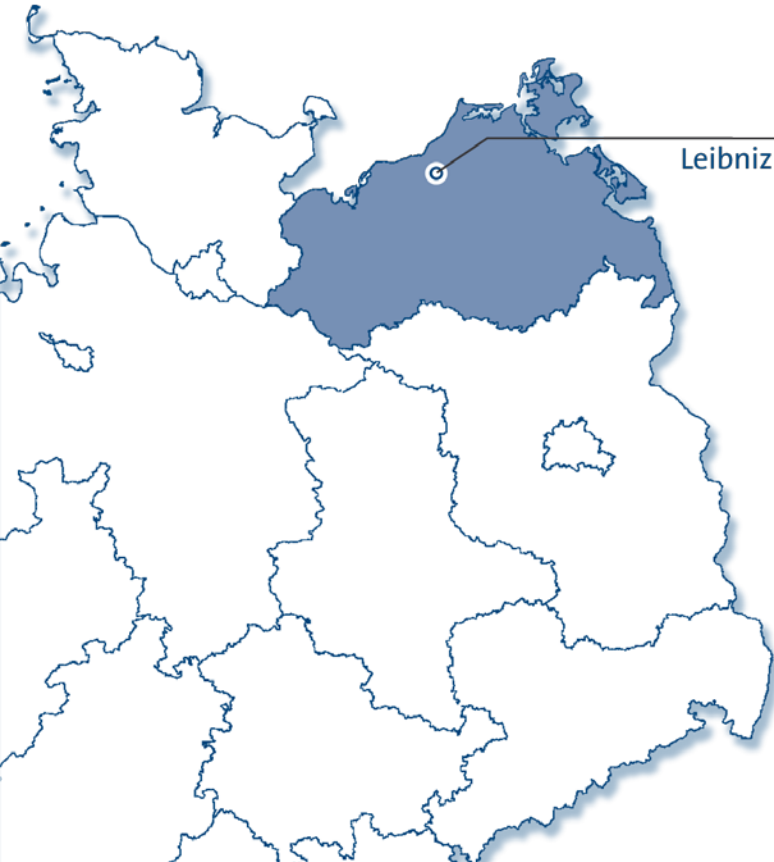
$$g_1 := D^{sire} + D^{dam}$$

$$g_2 := 16D^{sire}D^{dam} + 4D^{sire}(1 - 2\hat{f}_1^{dam})(1 - 2\hat{f}_1^{dam})$$

$$Q(D^{sire}, D^{dam}) = (\widehat{cov}_{add} - g_1)^2 + (\widehat{cov}_{dom} - g_2)^2$$

$$\arg \min_{D^{sire}, D^{dam}}(Q) = Q(D^{sire}, D^{dam})$$

Minimization
approach



Dummerstorf

Leibniz Institute for Farm Animal Biology FBN

Leibniz-Institut für Nutztierbiologie FBN

Wilhelm-Stahl-Allee 2
18196 Dummerstorf

contact

Hampel, Alexander

Telefon: +49 38208 68 908

E-Mail: hampel@fbn-dummerstorf.de

Internet: www.fbn-dummerstorf.de