

The covariance between genotypic effects in half-sib families

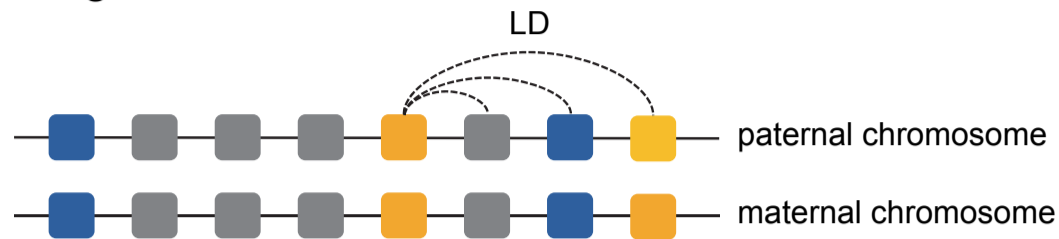
FBN / Wilhelm-Stahl-Allee 2 / 18196 Dummerstorf / www.fbn-dummerstorf.de

Dörte Wittenburg, Friedrich Teuscher, Jan Klosa, Norbert Reinsch



Background

Linkage and linkage disequilibrium (LD) cause dependencies between genomic markers.



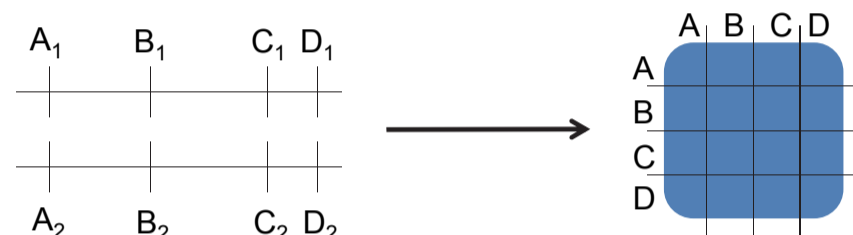
Typical livestock population: half-sib families (e.g. dairy cattle)

Conclusions

- For **small sample sizes** ($n=100$), a preciser estimation of parameters (genetic effects of single loci and variance components) and breeding values was obtained in simulations.
- By including additional information about the population structure and genome, a more reliable identification of genome segments with impact on health or performance is possible.

Objectives

The theoretical covariance between SNPs is derived and included in statistical methods for the genome-based phenotype prediction.



Outlook

- Complex validation study incl. parallelisation, application to real data and interpretation; biological functions of identified genome segments
- Derivation of theoretical covariance matrices also for non-additive effects, multiple families and different family structures (mixture of full and half sibs)

Methods

THEORY OF COVARIANCE

Expected value conditional on sire diplotype S

$$E(X_{t,i} X_{t,j} | S) = \sum_{\mathcal{D}} P(\mathcal{D} | S) E(X_{t,i} X_{t,j} | S \cap \mathcal{D})$$

Covariance of single matings (Bonk et al., 2015, submitted)

Weighted covariance of all possible matings

$$K_{i,j} := \text{cov}(X_{t,i}, X_{t,j} | S) = E(X_{t,i} X_{t,j} | S) - E(X_{t,i} | S) E(X_{t,j} | S)$$

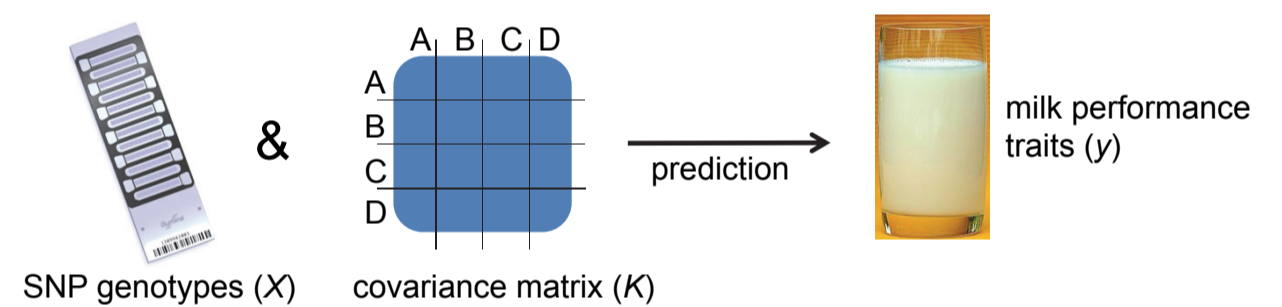
(loci i, j , animal t , sire S , dam \mathcal{D})

Requirements

- Known linkage phase and recombination rate of sire
- Known haplotype frequencies of dam population

⇒ Inference from progeny genotypes

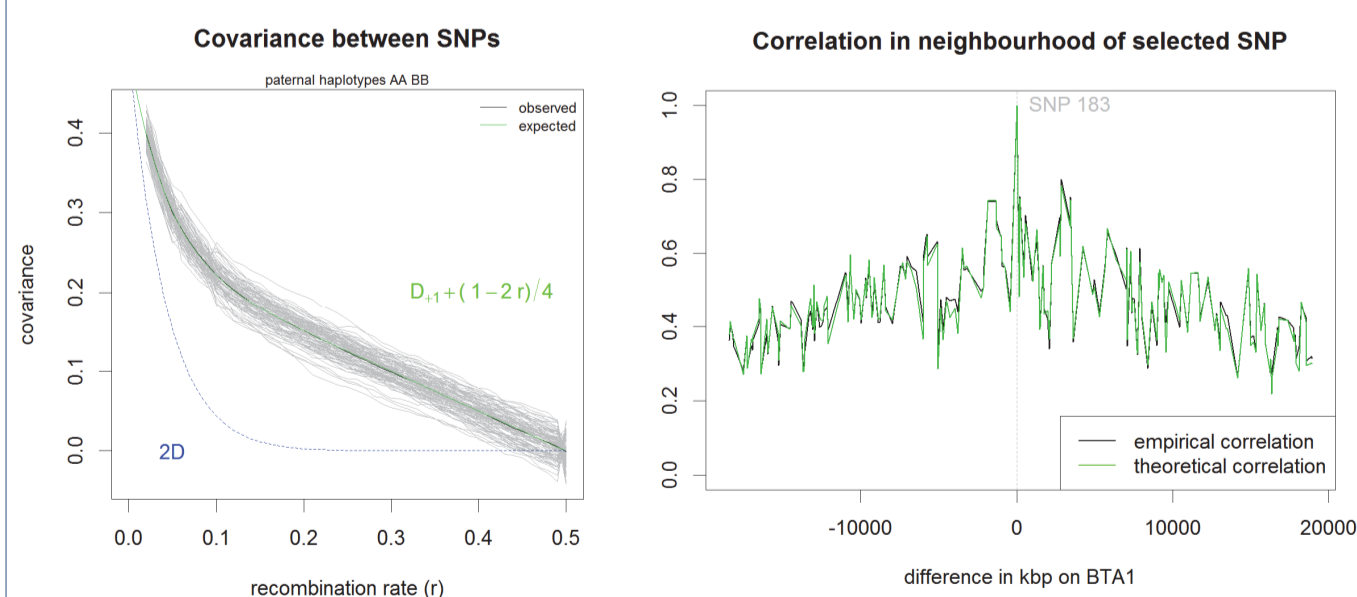
INCLUSION OF COVARIANCE MATRIX



Bayesian shrinkage approach with covariance matrix

- Correlated prior, similar to Wang et al. (1994, GSE)
- Model: $y = X m + e$ with $\text{cov}(e) = I \sigma_e^2$
- Prior: $m | K, \sigma^2 \sim N(0, K \sigma^2)$
- Posterior: $m | \text{else} \sim N(\Sigma X' y, \Sigma \sigma_e^2)$ with $\Sigma = (X' X + K^{-1} \sigma_e^2 / \sigma^2)^{-1}$
- Adaptive priors based on decomposition of K^{-1}

Results



Example: Covariance between SNPs within half-sib family in simulated ($n=1000$, left) and real data ($n=106$, right).

Prior	MSE σ_e^2	MSE σ_a^2	Corr(TBV, EBV)
Uncorrelated P1	0.151	0.082	0.754
Correlated P2	0.016	0.021	0.839
Adaptive P3	0.088	0.091	0.646
Adaptive P4	0.038	0.075	0.679

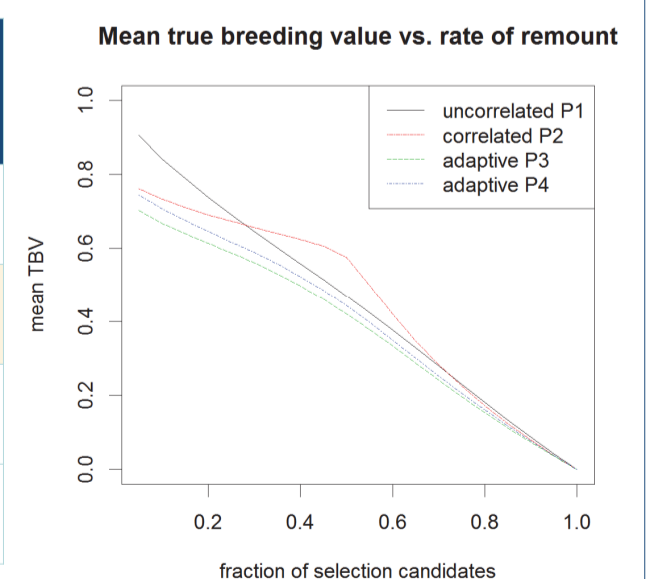


Table: Mean squared error of estimated variance components and accuracy of breeding value prediction; simulated data with 5 QTLs, 100 half sibs and 100 repetitions.

