

Genauigkeit genomischer Zuchtwerte: Cosegregation oder Linkage Disequilibrium?

Christian Stricker und Rohan L. Fernando

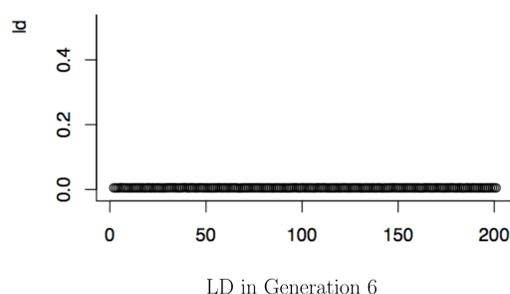
agn Genetics GmbH, Davos, Schweiz und Department of Animal Science, Iowa State University, Ames, USA

Zusammenfassung

Für eine Basispopulation von 500 Tieren werden Beobachtungen an 2000 SNPs mit einer Allelfrequenz von 0.5 simuliert. 100 der 2000 SNPs werden als QTL mit einem Substitutionseffekt $a \sim N(0, 1)$ ausgewählt. Nach 5 Generationen Zufallspaarung wird in Generation 6 kein Linkage Disequilibrium erreicht. Werden phänotypische Werte mit einer Heritabilität von 0.25 generiert und zufällig 1000 Tiere als Kandidaten ausgewählt, dann ergibt sich eine Korrelation (Genauigkeit) zwischen geschätzten und wahren Zuchtwerten von 0.67. Zur Schätzung der Zuchtwerte wurde ein Bayesian random regression Modell verwendet. Als Trainingspopulation dienten 1000 Tiere, die anderen 1000 wurden als Validierungspopulation verwendet. Die Genauigkeit der Zuchtwerte basiert auf jenen Tieren, die in der Trainingspopulation eine genomische Verwandtschaft von >0 aufwiesen. Sie wird andererseits aber negativ beeinflusst von Tieren, die untereinander keine genomische Verwandtschaft aufweisen. Wird eine Halbgeschwisterpopulation generiert, indem 10 Väter an 100 Mütter angepaart werden, d.h. 1000 Nachkommen generiert, dann steigt die Genauigkeit der Zuchtwertschätzung in 1000 zufällig ausgewählten Kandidaten auf 0.75 an. Dies ist auf die erhöhte genomische Verwandtschaft zurückzuführen. Wird anschliessend 500 weitere Generationen lang zufällig gepaart, dann entsteht Linkage Disequilibrium und erhöht die Genauigkeit in 1000 zufällig ausgewählten Kandidaten auf 0.92. Diese Beispiele zeigen exemplarisch auf, dass Bayesian random regression Models sowohl Cosegregation als auch Linkage Disequilibrium als Informationsquelle nutzen.

Simulation Linkage Disequilibrium Zufallspaarung

- 10 cM Intervall eines Chromosoms
- 2000 Loci
- keine Mutationen
- 5 Generationen Zufallspaarung
- Populationsgrösse 500

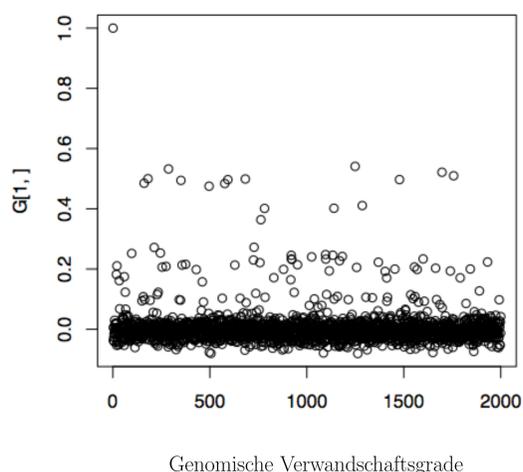


Simulation Phänotypen & Genotypen

- 100 Loci zufällig ausgewählt
- Effekt $N(0, 1)$
- $\sigma_g^2 = 66.3$
- Restvarianz $N(0, 3 * \sigma_g)$
- $\Rightarrow h^2 = 0.25$

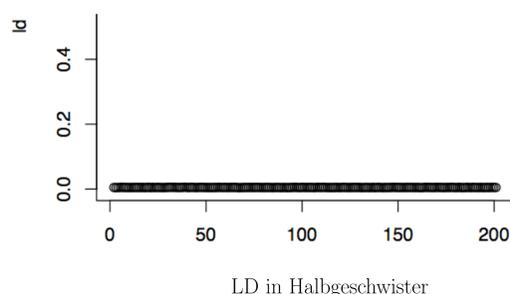
Genomic Prediction

- BayesC $\pi = 0$
- chain length 2000
- 1000 zufällige Tiere als Validierung
- $\text{Cor}(\hat{a}_{val}, a_{val}) = 0.64$



Simulation Linkage Disequilibrium Halbgeschwister

- gleiches genetisches Modell
- 10 Väter
- 100 Mütter/Väter
- 2010 Tiere total (10/1000/1000)

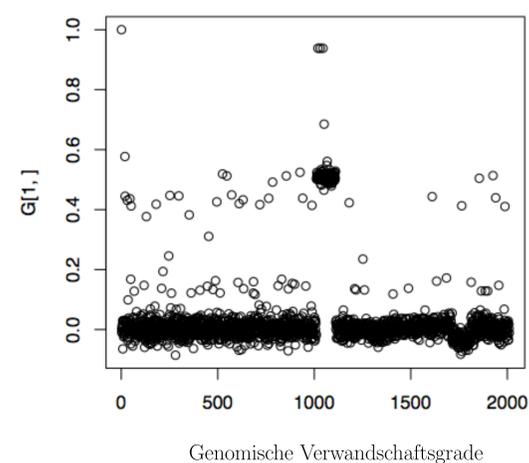


Simulation Phänotypen & Genotypen

- gleich wie bei Zufallspaarung, d.h.
 - 100 Loci zufällig ausgewählt
 - Effekt $N(0, 1)$
 - $\sigma_g^2 = 79.31$
 - Restvarianz $N(0, 3 * \sigma_g)$
 - $\Rightarrow h^2 = 0.24$

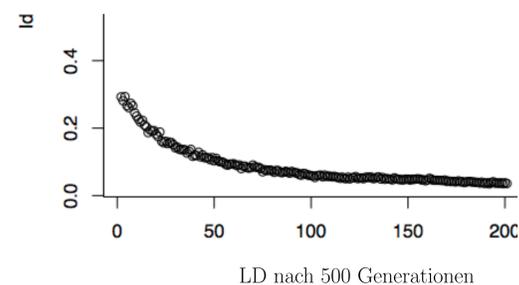
Genomic Prediction

- wie vorher, d.h.
 - BayesC $\pi = 0$
 - chain length 2000
 - 1000 zuf. Tiere als Validierung
- $\text{Cor}(\hat{a}_{val}, a_{val}) = 0.75$



+500 Generationen Zufallspaarung

- gleiches genetisches Modell
- 506 Generationen
- Populationsgrösse 500

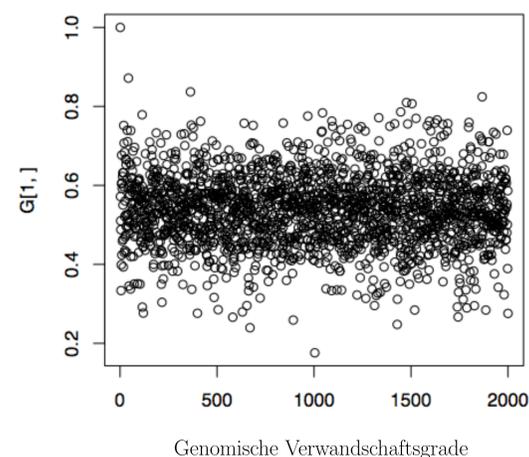


Simulation Phänotypen & Genotypen

- gleich wie bei Zufallspaarung $h^2 = 0.25$

Genomic Prediction

- wie vorher, d.h.
 - BayesC $\pi = 0$
 - chain length 2000
 - 1000 zuf. Tiere als Validierung
- $\text{Cor}(\hat{a}_{val}, a_{val}) = 0.92$



Schlussfolgerungen

- Genomische Zuchtwertschätzung funktioniert auch ohne LD
- Genomische ZWS besser als klassische ZWS, weil Abweichungen von additiv genetischer Verwandtschaftsmatrix zur Schätzung verwendet werden.
- Marker haben Cosegregations- und LD-Information
- Cosegregations- und LD-Information kann auch bei $> 50k$ SNP-Arrays auf wenige segregierende Haplotypen reduziert werden \Rightarrow Haplotyping statt Imputing!