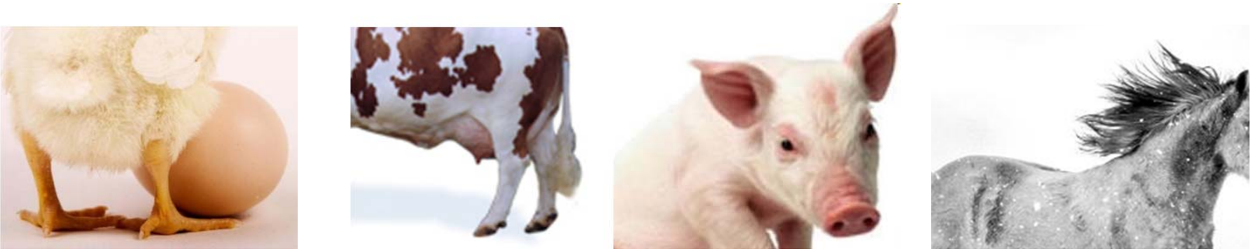
Estimation of the effective number of genes underlying quantitative traits based on chromosomal partitioning of the genomic variance

H. Simianer[a], M. Erbe[a], S. König[b], E. Pimentel[ab]

[a] Department of Animal Sciences
Animal Breeding and Genetics Group
Georg-August-University Göttingen, Germany

[b] Department of Animal Breeding
University of Kassel, Witzenhausen, Germany



Old question: how many genes are involved in the inheritance of a quantitative trait?

Results are highly variable:

- Chamberlain *et al.* (2007): ~ 30 QTL underlying milk yield in Holsteins
- Daetwyler *et al.* (2010) applied to data of Luan *et al.* (2009): ~ 750 genes underlying milk yield in Norwegian Red Cattle
- Reed *et al.* (2008): ~ 6000 genes underlying growth in mice

When does a gene count as ‚involved'?

Data

- SNP genotypes (Illumina BovineSNP50 BeadChip) of 2294 progeny-tested bulls (born 1981 – 2003)

- Pedigree with 21‘646 animals, back to 1906.

- After filtering w.r.t.
  - ⇨ call rate > 97%
  - ⇨ MAF > 0.05
  - ⇨ known autosomal position

  39’557 SNPs on 29 autosomes (2562 on BTA1 ⇲ 742 on BTA28)

- Haplotype reconstruction with fastPHASE (Scheet & Stephens, 2001) incl. Imputation of missing genotypes

- Phenotypes = EBVs for milk yield (Mkg), fat percentage (F%), protein percentage (P%) and somatic cell score (SCS)

Data processing

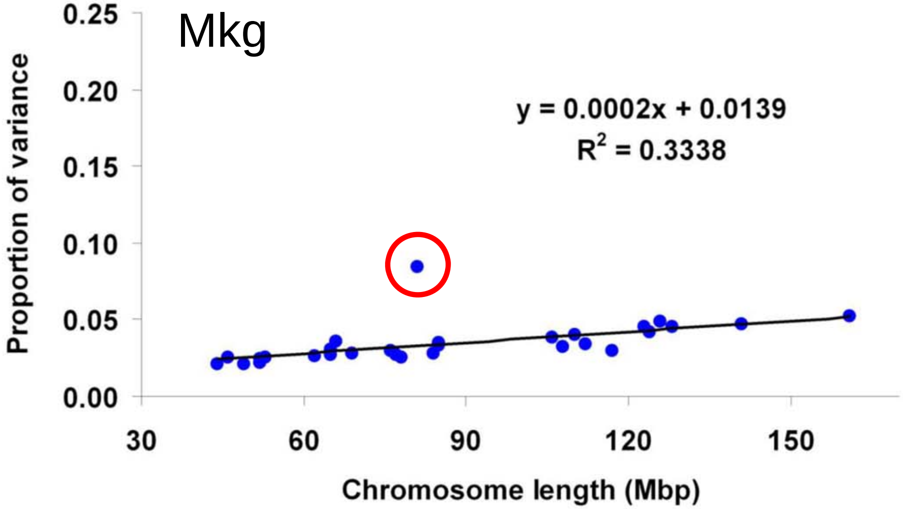
- Genomic SNP effects estimated using random regression BLUP (Meuwissen et al. 2001), variance components estimated from the data using REML → $\hat{\alpha}_i$
- Allele frequency at SNP i in the founder generation estimated by the approach of Gengler et al. (2007) → $\hat{p}_i$
- Variance for chromosome *j* with n_j SNPs calculated as

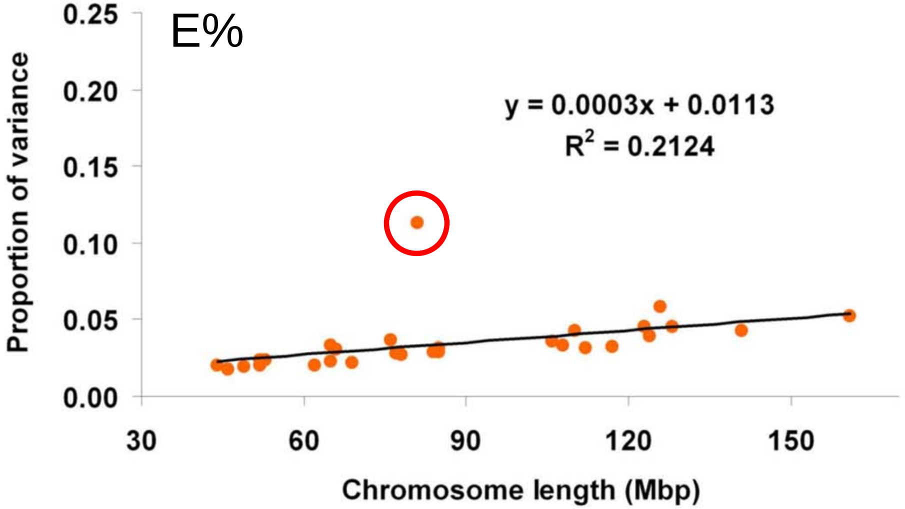
$$V_j = \sum_{i=1}^{n_j} 2\hat{p}_i(1-\hat{p}_i)\hat{\alpha}_i^2$$

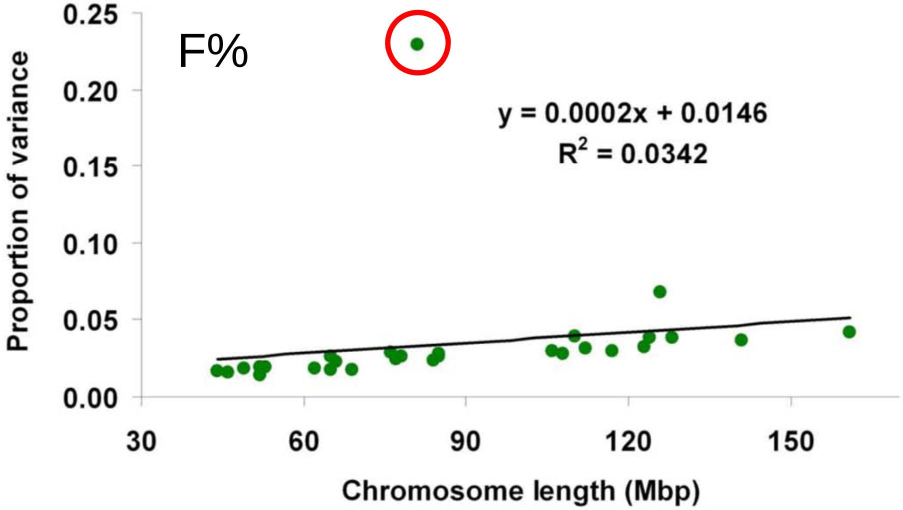
- Proportion of variance of chromosome *j* calculated as

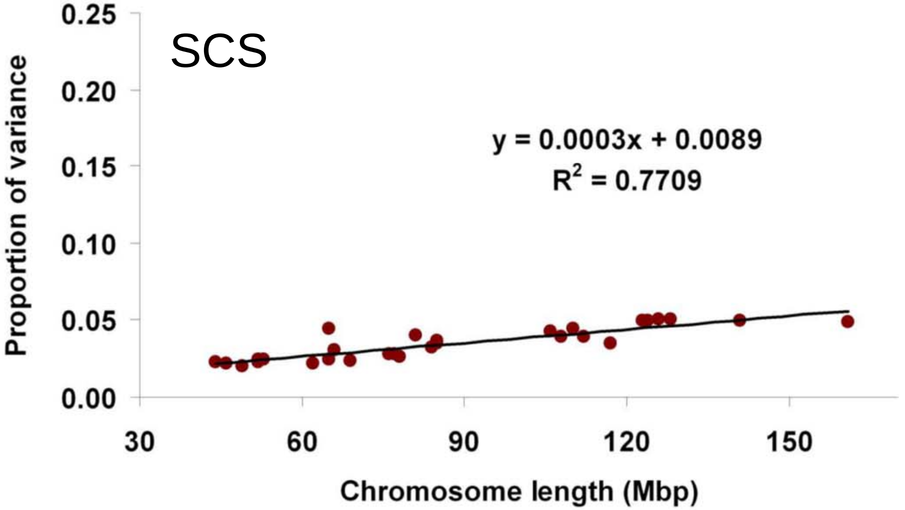
$$Q_j = V_j / \textstyle\sum_{k=1}^{29} V_k$$

Regression of the proportion of variance on the physical chromosome length

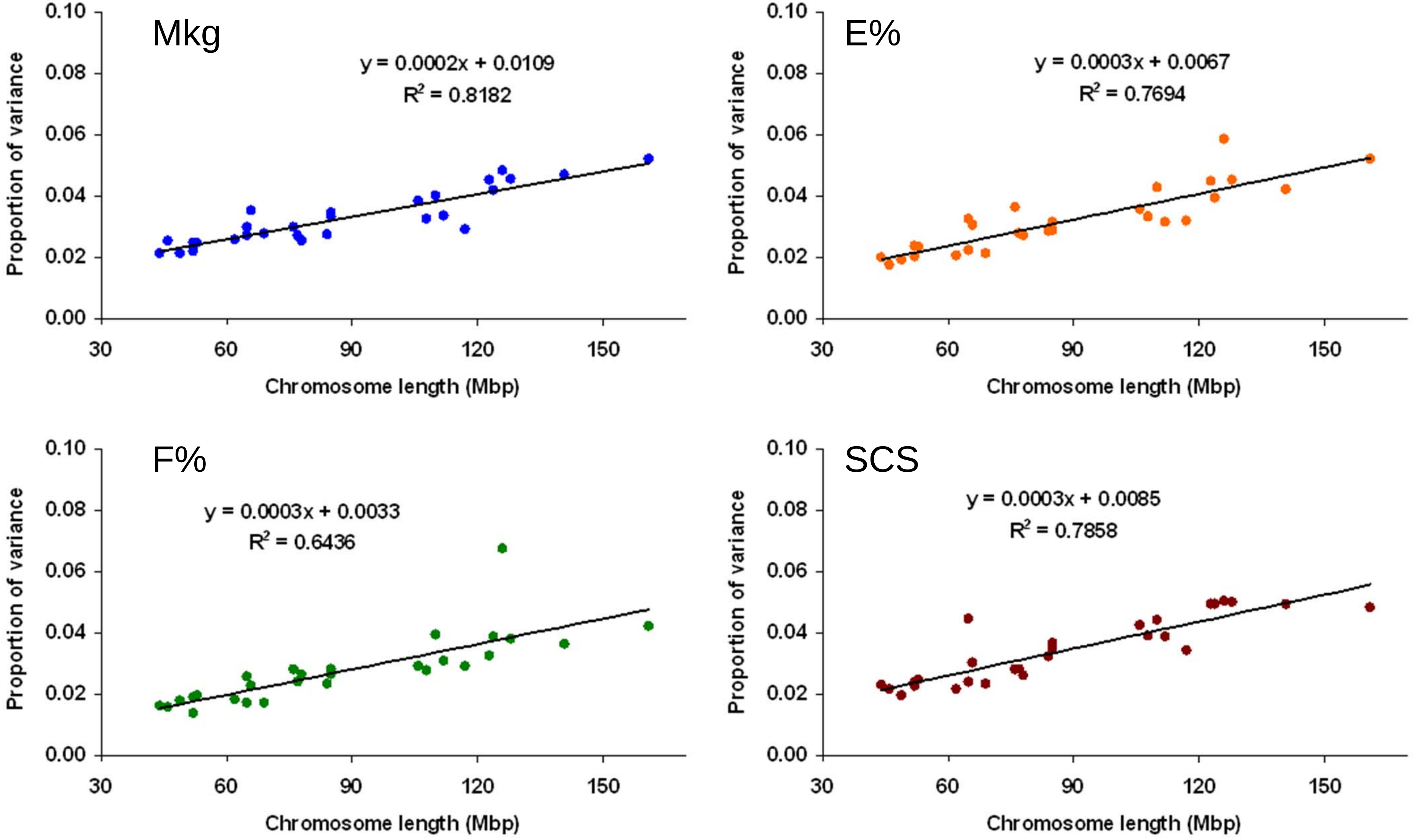








Regression of the proportion of variance on the physical chromosome length – BTA 14 excluded

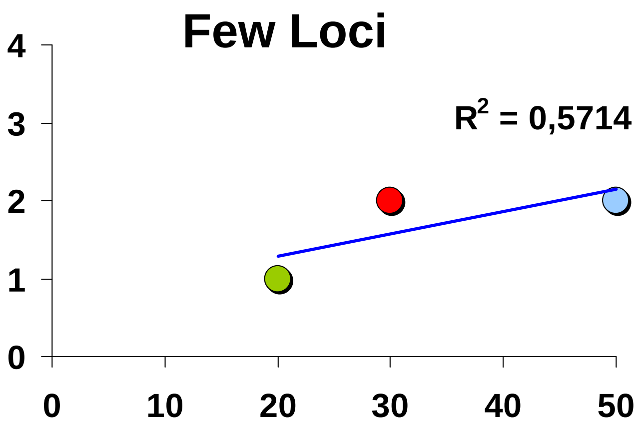


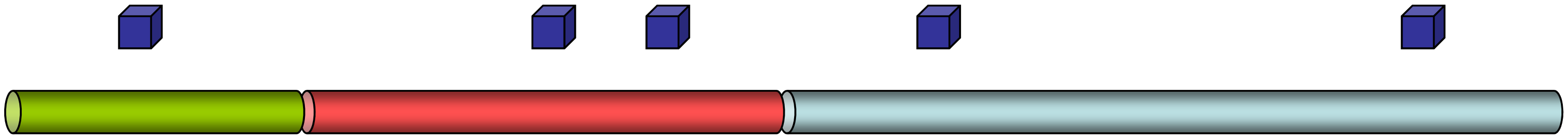
The concept of the effective number of genes

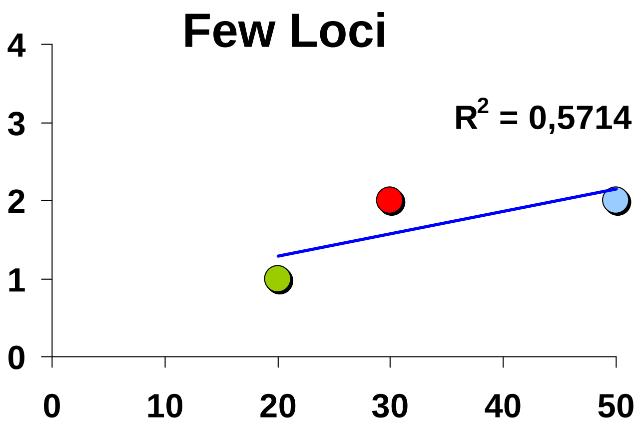
Effective number of genes *N_g* is the number of hypothetical genes which

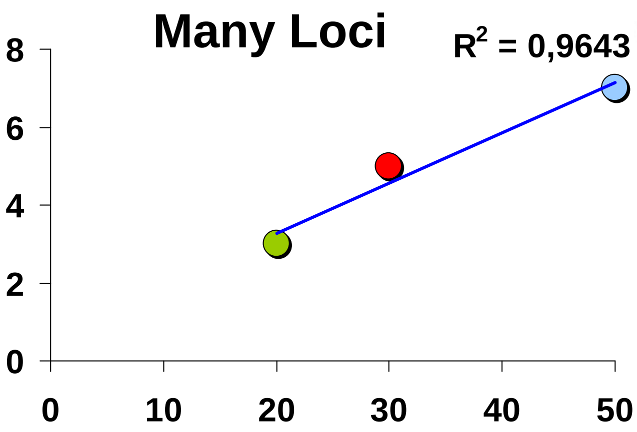
- have the same genetic variance
- are in linkage equilibrium
- are randomly (Poisson) distributed across the genome

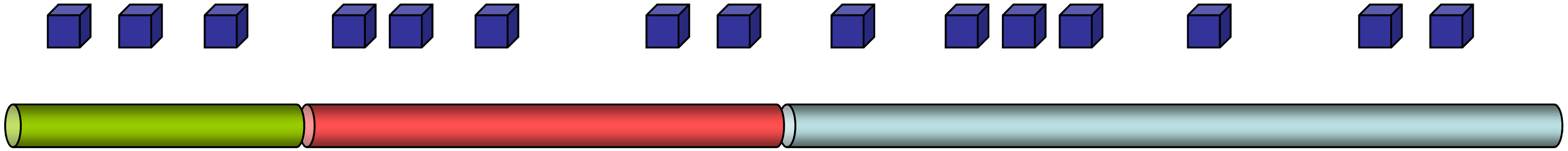
that leads to the same coefficient of determination R^2 in variance partitioning as obtained in the empirical analysis


Few Loci
R² = 0,5714
4
3
2
1
0
0
10
20
30
40
50



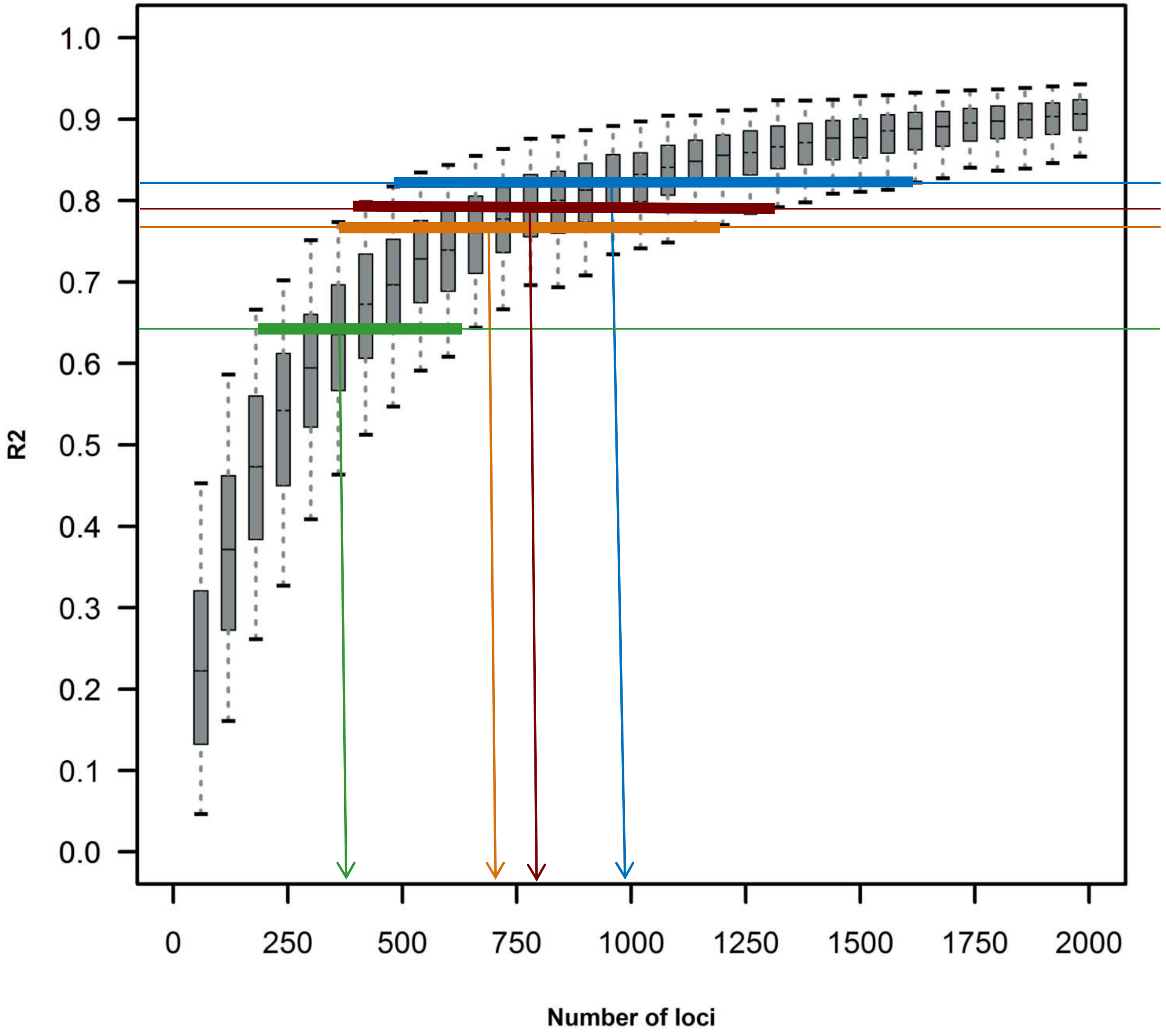

Few Loci
$R^2 = 0,5714$
4
3
2
1
0
0
10
20
30
40
50


Many Loci
$R^2 = 0,9643$
8
6
4
2
0
0
10
20
30
40
50



Simulation results

1000 replicates per number of loci, without BTA 14



Mkg: N_g = 949 [462;1593]

SCS: N_g = 779 [366;1337]

P%: N_g = 710 [337;1213]

F%: N_g = 378 [157;696]

Final remarks

- The concept provides a sensible and consistent definition of the number of genes underlying quantitative traits
- The approach is conservative, i.e. N_g is a lower bound estimate.
- Unequal variance of genes and non-random distribution of genes increase the true number of genes relative to N_g
- The resulting values are in the expected range, substantial variation between traits is observed (F%: 378 → Mkg: 949)
- The link of N_g with the accuracy of genomic prediction and the relative performance of different estimation procedures is subject to further studies

If you want to read the whole story ...

frontiers in
GENETICS

ORIGINAL RESEARCH ARTICLE
published: 02 May 2011
doi: 10.3389/fgene.2011.00019

Genome partitioning of genetic variation for milk production and composition traits in Holstein cattle

Eduardo da Cruz Gouveia Pimentel[1]*, Malena Erbe[1], Sven König[2] and Henner Simianer[1]

[1] Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University, Göttingen, Germany
[2] Department of Animal Breeding, University of Kassel, Witzenhausen, Germany

This research was funded by the Federal Ministry of Education and Research (BMBF) within the AgroClustEr *„Synbreed – Synergistic plant and animal breeding" (FKZ 0315526)*, in collaboration with *"FUGATO-plus GenoTrack" (FKZ 0315134C)*

Synbreed FUGATO Bundesministerium für Bildung und Forschung FBF LOHMANN TIERZUCHT KWS

Thank you!