# The role of covariance between genome segments in genetic variation and selection

**Henner Simianer & Eduardo Pimentel**
Department of Animal Sciences
Animal Breeding and Genetics Group
Georg-August-University Göttingen, Germany

1

Genomic information allows a dissection of the total genomic variability into physical sub-units, e.g. chromosomes
– see presentation of Eduardo Pimentel (this session)

- ⇨ Do we find covariances between chromosomes?
- ⇨ How do they contribute to the total genetic variance?
- ⇨ Which role do they play in selection and inheritance to the next generation?

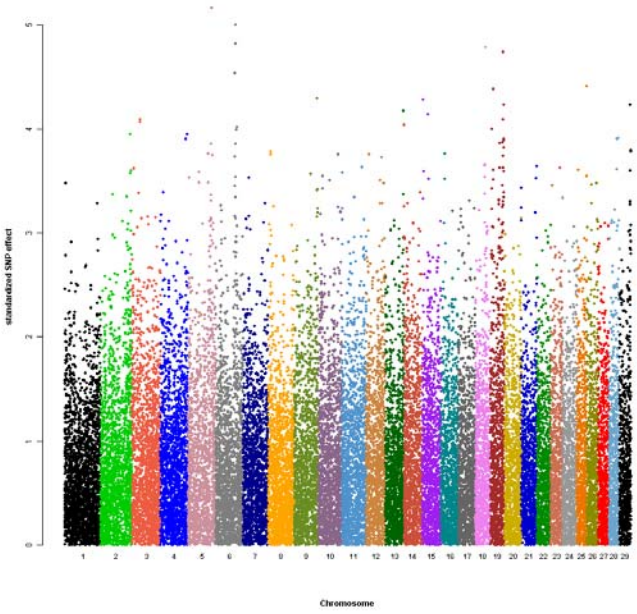
2

**Data:**

- SNP genotypes (Illumina BovineSNP50 BeadChip) of 2307 progeny tested Holstein Friesian bulls (>97% from birth years 1998 – 2002)
- After filtering w.r.t.
  - ⇨ call rate > 95%
  - ⇨ 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. reconstruction of missing genotypes
- Phenotypes = EBVs for SCS ($h^2$ = .163, $r_{AI}$ = .88)
- SNP effects estimated via random regression BLUP

3




Standardised estimated SNP effects

Somatic cell score

4

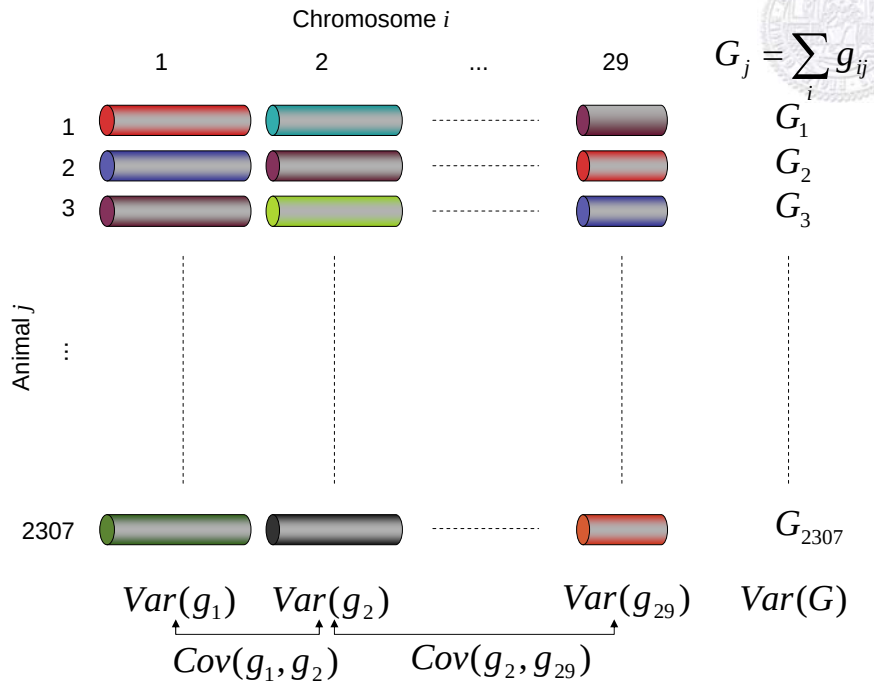
**Assumptions**

- Estimated SNP effects reflect true genetic effects in the neighbourhood of the SNP, but
- single SNP effects are estimated with low accuracy and high error covariance between adjacent SNPs
- accumulation of SNP effects over a physical region leads to more stable estimates reflecting the accumulated genetic variability in the region

Region = whole genome → gEBV ( $G$ )

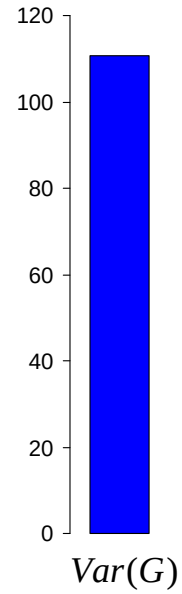
Region = autosome → 29 cEBVs ($g_i$; $\sum_{i=1}^{29} g_i = G$ )

5



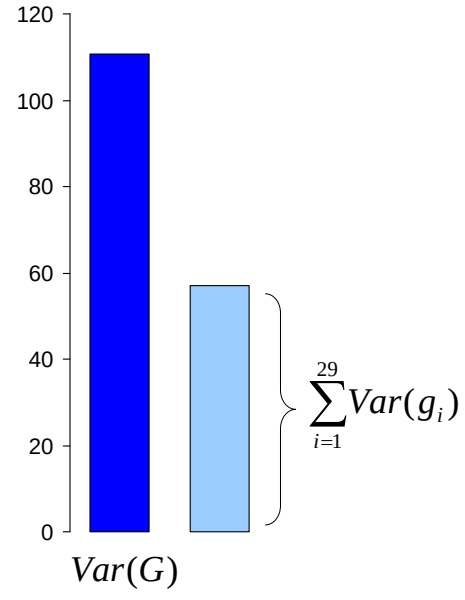

6

Variance of gEBV for SCS



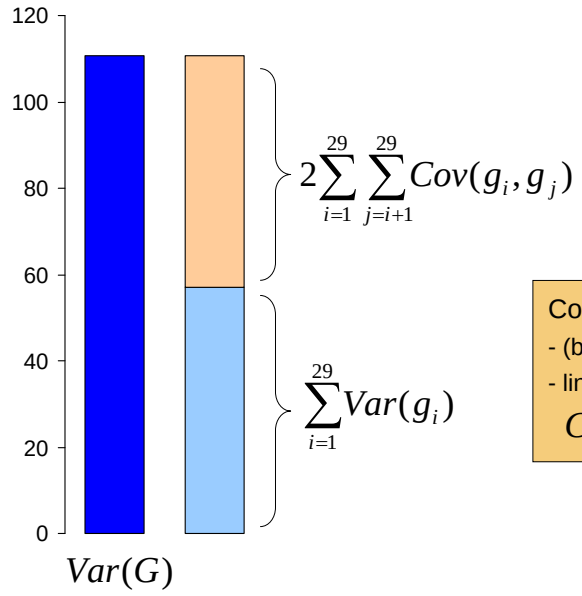

7

Variance of gEBV for SCS




8

Variance of gEBV for SCS




Covariance due to
- (biological) epistasis (?)
- linkage disequilibrium

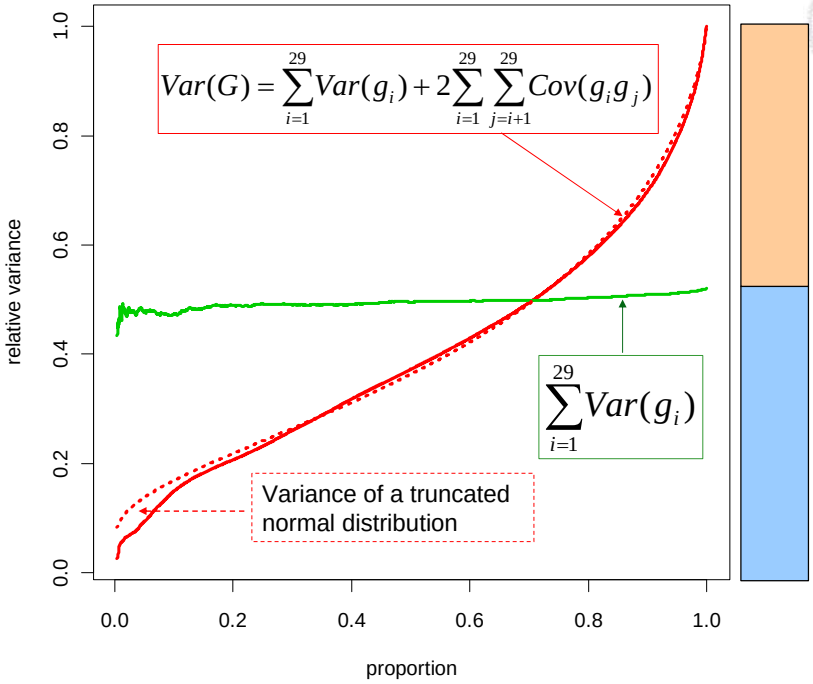
$$Cov(l_i l_j) = 2D_{ij}\alpha_i\alpha_j$$

9

**Effect of selection on the genomic variance**

- ⇨ Sort the animals by the gEBV for one trait
- ⇨ Select the p% best ones
- ⇨ Determine the variance of the gEBV and the cEBVs in the selected fraction

10

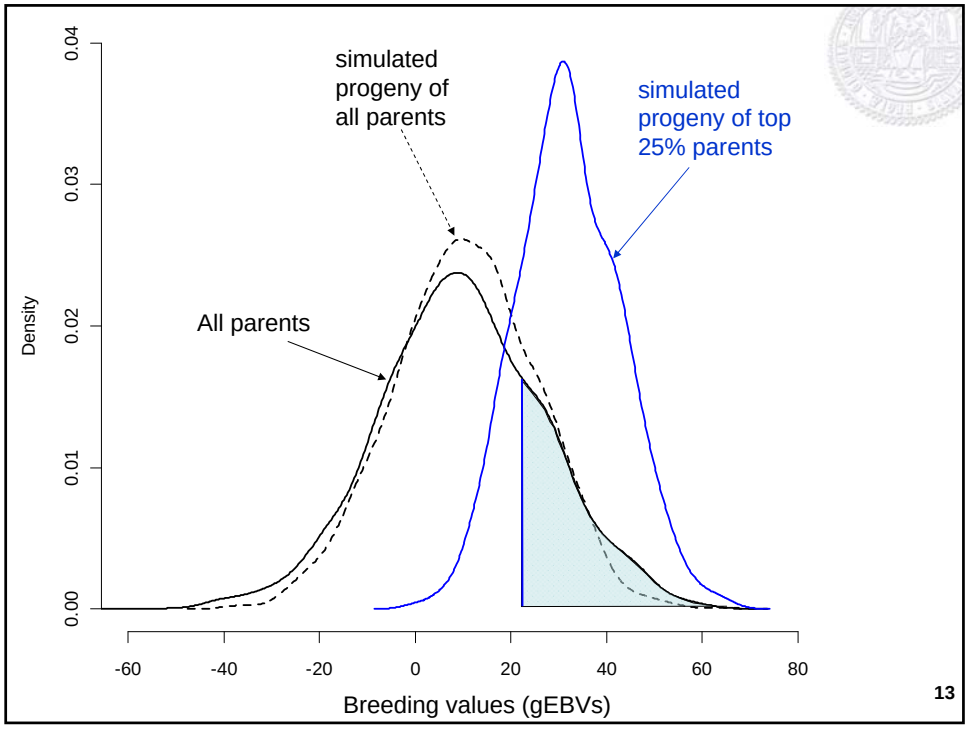


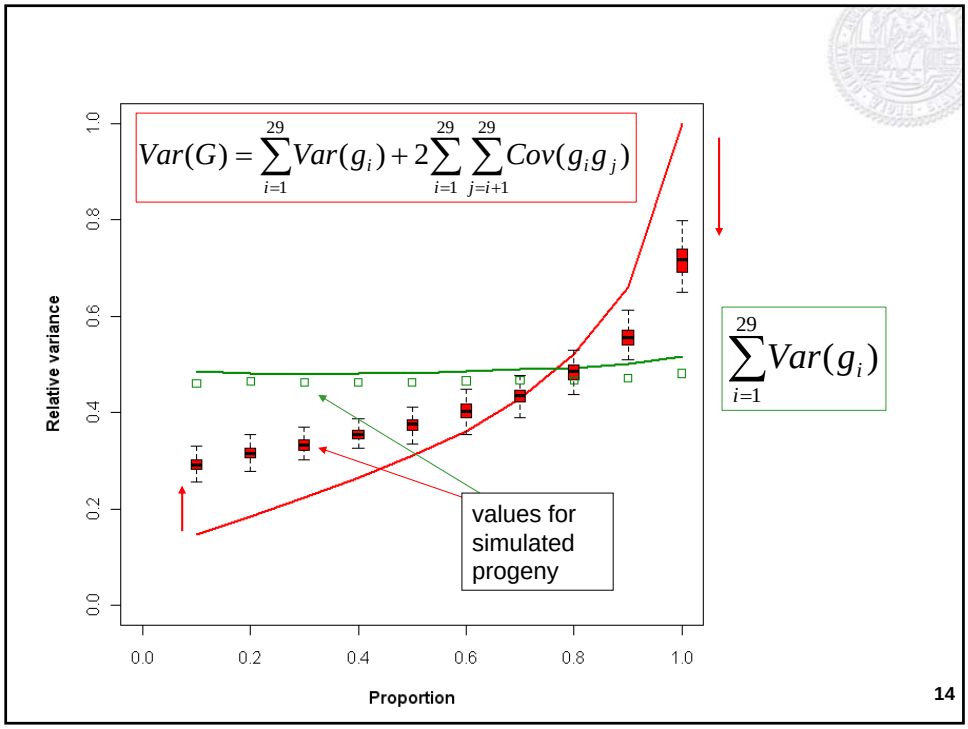

11

## Simulated progeny after selection

- ⇨ choose the top p % of the bulls according to gEBV as potential parents
- ⇨ select at random one ‚sire' and one ‚dam'
- ⇨ produce an offspring (random sampling of chromosomes, recombinations following a Poisson process)
- ⇨ calculate variances and covariances in the simulated offspring
- ⇨ repeat this 1000 times

12


simulated
progeny of
all parents
simulated
progeny of top
25% parents
All parents
Density
0.00
0.01
0.02
0.03
0.04
-60
-40
-20
0
20
40
60
80
Breeding values (gEBVs)
13

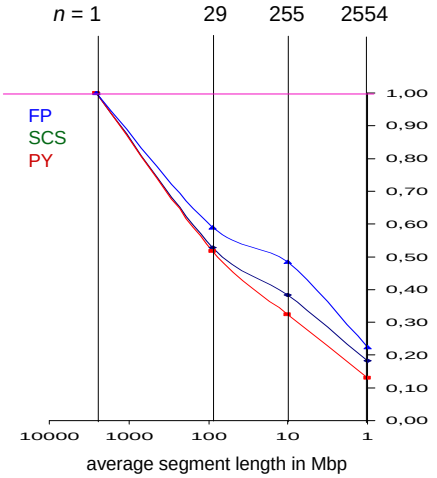

$Var(G) = \sum_{i=1}^{29} Var(g_i) + 2\sum_{i=1}^{29}\sum_{j=i+1}^{29} Cov(g_i g_j)$
$\sum_{i=1}^{29} Var(g_i)$
values for
simulated
progeny
Relative variance
0.0
0.2
0.4
0.6
0.8
1.0
Proportion
14

## Conclusions

- In the analysed sample, covariance between chromosomes is positive and explains ~ 50 per cent of the total genomic variance (extremes overrepresented?)
- Truncation selection affects the composition of the genomic variance:
  - total genomic variance behaves as expected (truncated normal)
  - variances of chromosomal effects remain constant
  - the entire variance reduction is due to the reduction and change of sign of the covariances
- One simulated generation of recombination and random mating halves the covariances
- Results are largely as expected under the infinitesimal model and the Bulmer theory of genetic variance under selection

15

- Is the covariance (part of) the explanation for the ‚dark matter' phenomenon?




Proportion of total genomic variance explained by $\sum_{i=1}^{n} Var(g_i)$

Maher, B. (2008) Nature, 456

## Some thoughts regarding practical consequences

- Can we generate more usable genetic variance by removing/ reducing the negative covariances?
- How would a genomic selection and mating scheme be designed to achieve this goal?
- What are the consequences for doing (re-) calibrations in (genomically) highly selected samples?
- What does it mean for QTL mapping if 50% of the genetic variance is actually due to non-mappable covariances?

17

This research was funded by the Federal Ministry of Education and Research (BMBF) within the project *"FUGATO-plus GenoTrack"* which is financially supported by the Federal Ministry of Education and Research (BMBF), the Förderverein Biotechnologieforschung e.V. (FBF) and Lohmann Tierzucht GmbH, Cuxhaven

Data used were kindly provided by
- Christian Albrechts Universität Kiel (Dr. J. Tetens, Prof. G. Thaller,)
- Helmholtz-Zentrum München (P. Lichtner, Prof. T. Meitinger)
- vit w.V., Verden (F. Reinhardt)

which is gratefully acknowledged

Thank you!

18