Genome partitioning of genetic variance for production traits in Holstein cattle

Eduardo Pimentel, M. Erbe, H. Simianer and S. König

Department of Animal Sciences

Animal Breeding and Genetics Group

Georg-August-University Göttingen, Germany




Introduction

- SNP chips with very many markers
- Genomic (realized) relationships
- Genome-wise or chromosome-wise

Objective

- use dense marker information to partition genetic variance of important dairy traits across chromosomes

Material

- 2294 Holstein bulls (call rate > 0.97)

- 39557 SNPs (Illumina BovineSNP50 BeadChip)
  - call rate > 0.95
  - MAF > 0.05
  - known position
  - no missing (imputed with fastPHASE)

Material

- EBVs for milk (kg), fat (%), protein (%), SCS

Variable	Mean	Std. dev.	Minimum	Maximum
EBV_milk	730.3836	611.4743	-1362.00	2892.00
EBV_fat	-0.0974	0.2939	-1.01	1.05
EBV_protein	-0.0280	0.1187	-0.47	0.50
EBV_SCS	100.3749	12.0940	58.00	139.00
Accuracy_mkg	94.2367	2.1118	88.00	99.00
Accuracy_scs	87.9743	4.2974	76.00	99.00

Methods

- Marker-based kinship coefficients
  - Eding & Meuwissen (2001)

$$S_{xy,l} = 1/4[I_{11} + I_{12} + I_{21} + I_{22}]$$

where I_{ij} is an indicator variable which is 1 when allele i on locus l in the first individual and allele j on the same locus in the second individual are identical, otherwise it is 0.

S_{xy} is an unbiased estimator of kinship when founder alleles are unique

$$E(S_{ij}) = f_{ij} + (1 - f_{ij})s$$

P(IBD) P(AIS)

$$\hat{f}_{ij} = \frac{S_{ij} - s}{1 - s}$$

Methods

- Marker-based kinship coefficients
  - Eding & Meuwissen (2001)

$$\hat{f}_{ij} = \frac{S_{ij} - s}{1 - s}$$

$s = S_{ff} = \sum q_k^2$, where S_{ff} is the similarity in the founder population and q_k is the frequency of the kth allele in the founder population

  - Gengler et. al (2007)

Methods

- Variance component estimation
  - Method 1 (REML)

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon} \qquad \begin{cases} \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma^2_{\boldsymbol{\varepsilon}}) \\ \mathbf{g} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}\sigma^2_{\mathbf{g}}) \end{cases}$$

- full model: whole genome
- reduced model: all chromosomes but one
- fitted with ASReml®
- reduction in proportion of variance due to $\mathbf{g}$

Methods

- Variance component estimation
  - Method 2 (regression)

Heredity **84** (2000) 427–436

Received 20 July 1999, accepted 15 November 1999

Estimating variance components in natural populations using inferred relationships

STUART C. THOMAS*, JOSEPHINE M. PEMBERTON & WILLIAM G. HILL

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, U.K.

$$Z_i = \left[(y_i - \bar{y})(y_i' - \bar{y})\right]/\hat{\sigma}^2 \qquad Z_i = 2r_i h^2 + e_i$$

- chromosomal kinship

Methods

- Variance component estimation
  - Method 3 (WGA)
    - Step 1:

$$y_i = \mu + \sum_{j=1}^{p} x_{ij} b_j + e_i$$

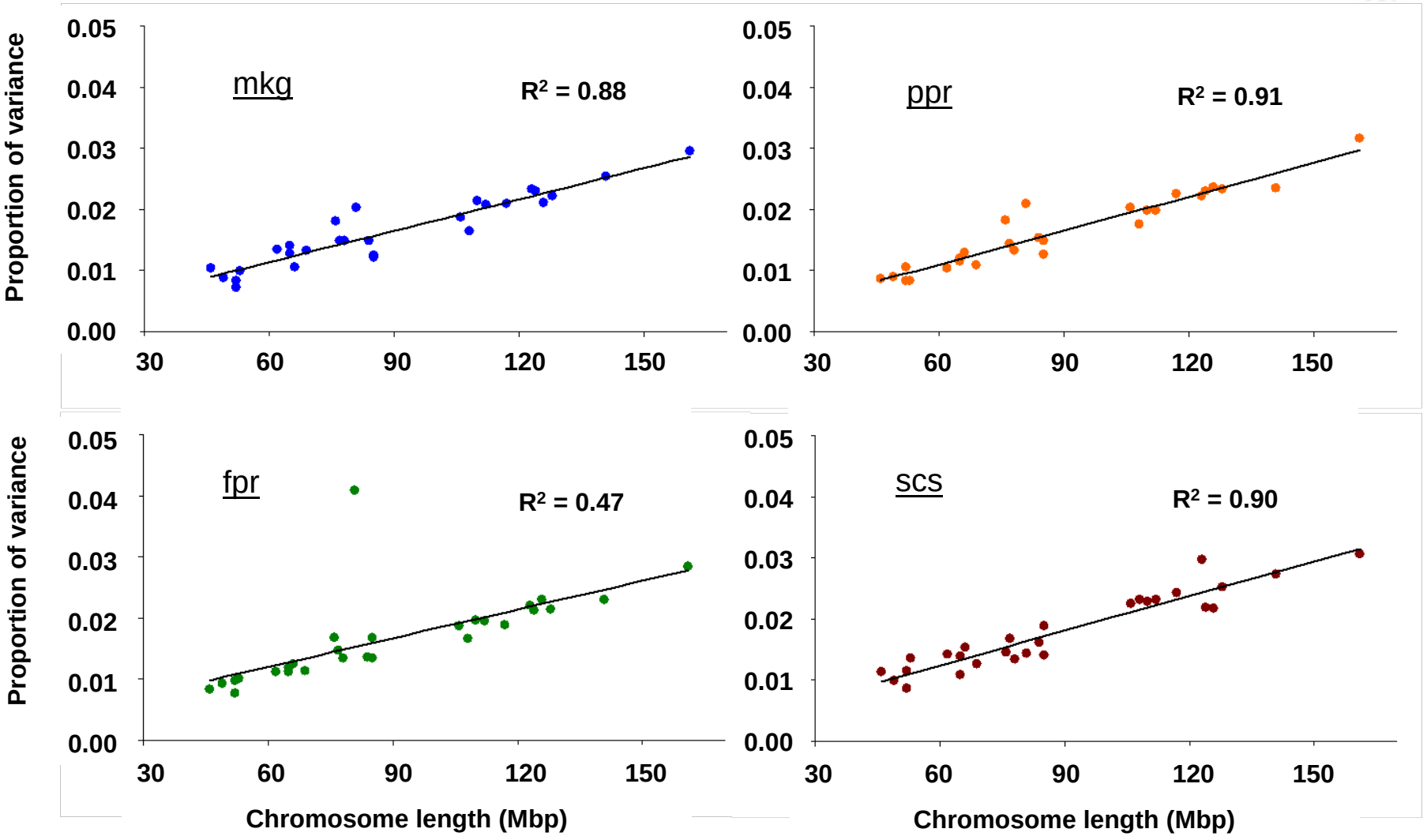
$$\begin{bmatrix} \hat{\mu} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \iota^t W \iota & \iota^t W X \\ X^t W \iota & X^t W X + I\lambda \end{bmatrix}^{-1} \begin{bmatrix} \iota^t W y \\ X^t W y \end{bmatrix}$$

$$w_{ii} = \hat{r}_{EBV_i, TBV_i}$$

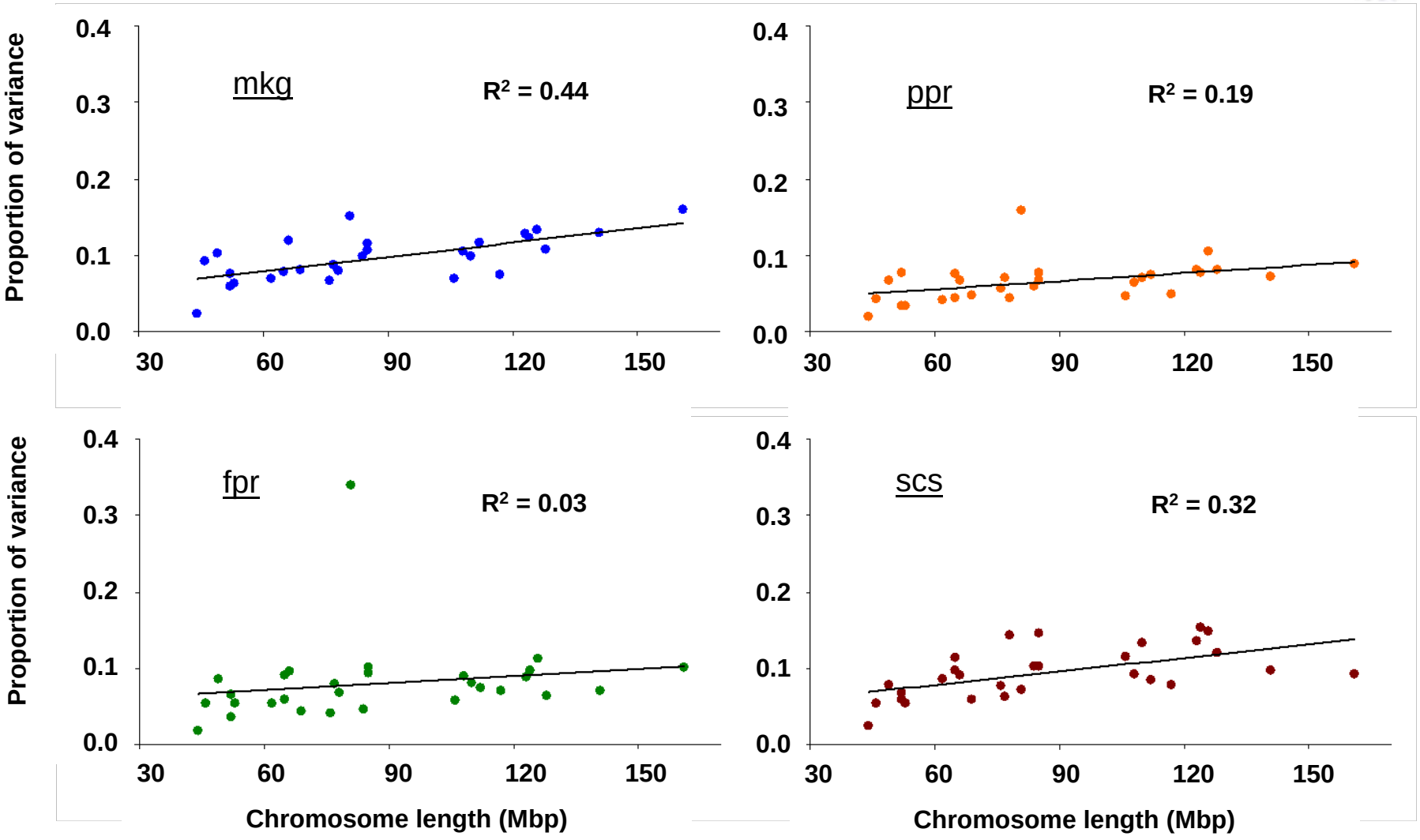
$$\lambda = \frac{\sigma_e^2}{\sigma_{SNP}^2} \qquad \sigma_{SNP}^2 = \frac{\sigma_a^2}{p}$$

    - Step 2: $\hat{\sigma}_{SNP}^2 = 2p(1-p)\hat{b}_j^2$

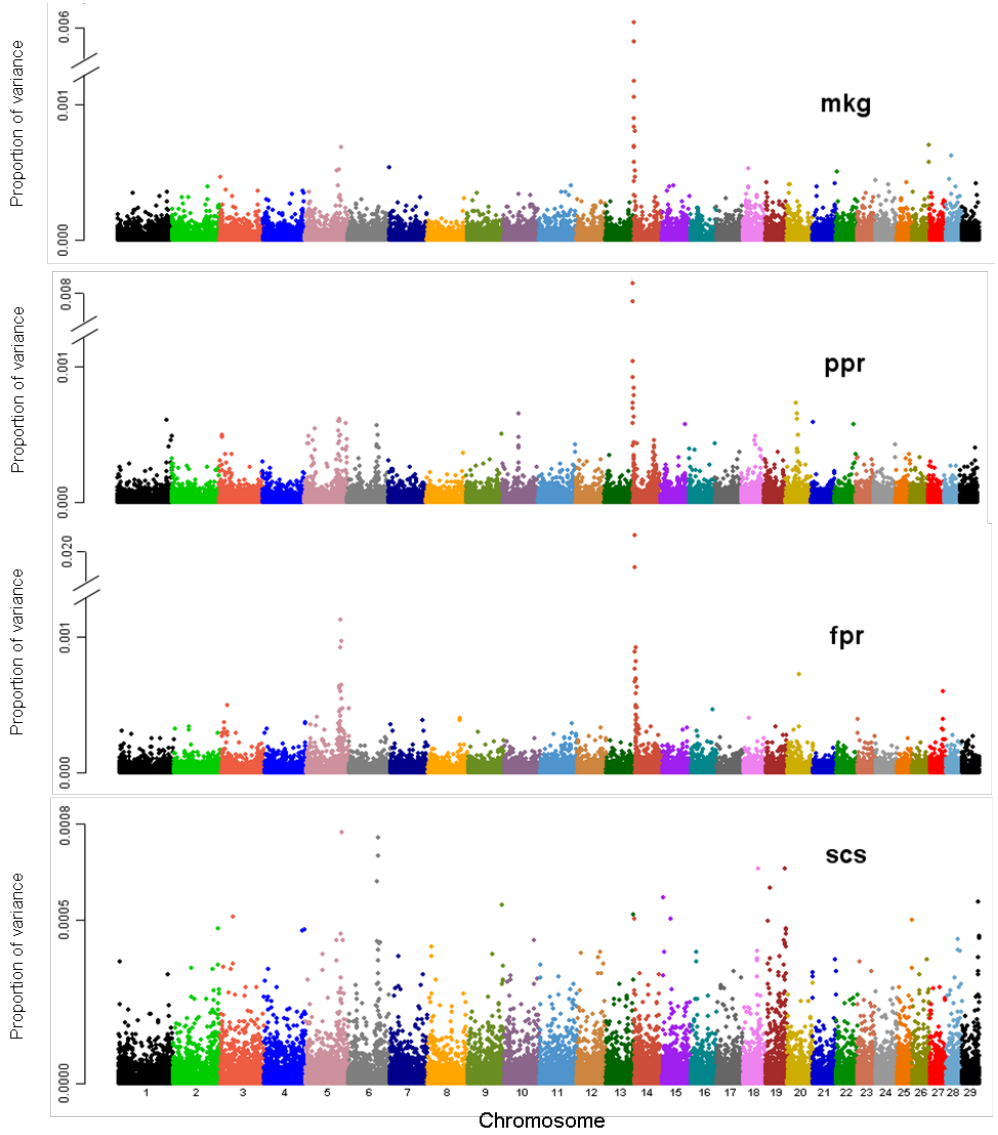
Results – method 1



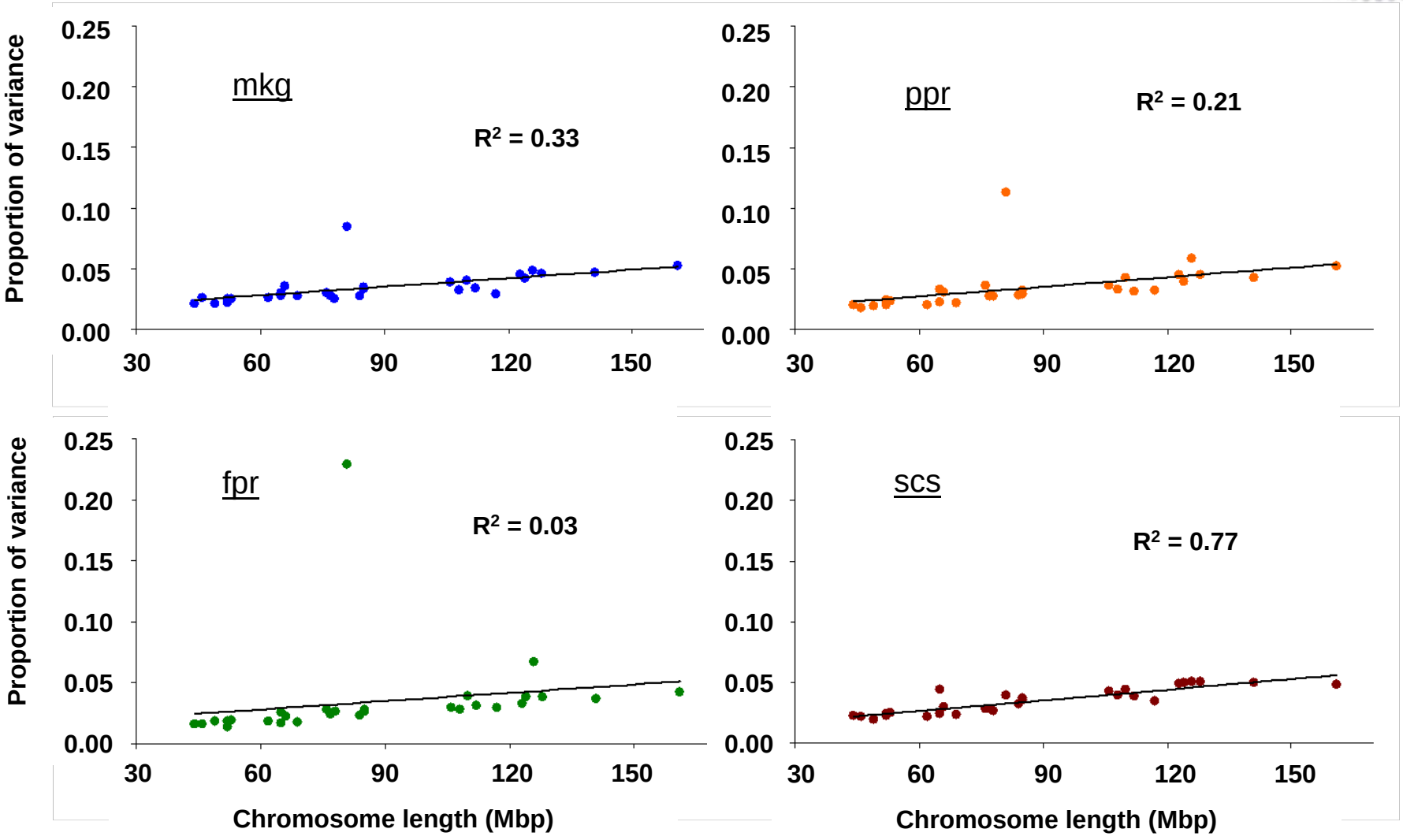
Results – method 2



Results – method 3



Results – method 3



Discussion

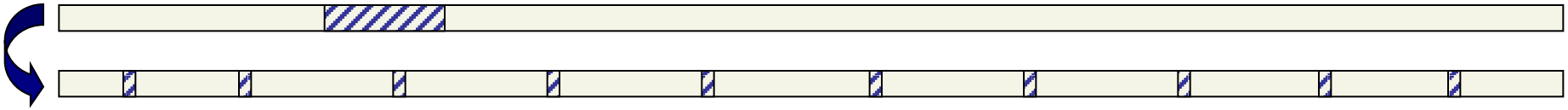
R^2 fitting a regression without Chrom. 14:

	Milk (kg)	Fat (%)	Protein (%)	SCS
Method 1	0.91	0.95	0.94	0.90
Method 2	0.52	0.27	0.40	0.32
Method 3	0.82	0.64	0.77	0.79

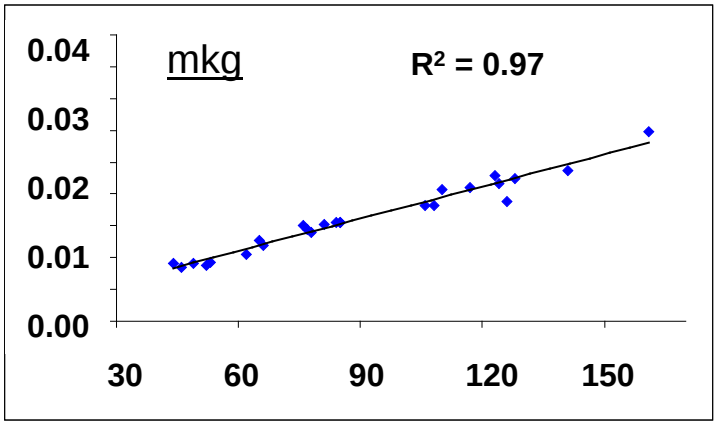
Visscher et al. 2007 (height in humans):

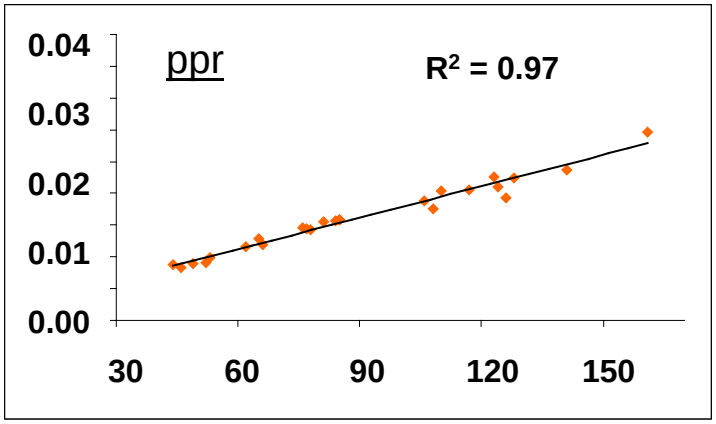
"In general, the longer the chromosome, the more variation it explains."

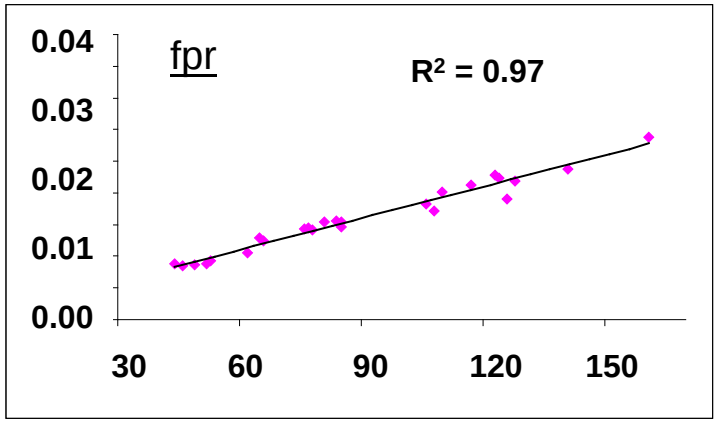
Results – ‘control’



“segment“ (# of markers) of the size of a chromosome, but “pulverized“ randomly across the genome







Conclusions

- Results from methods 1 to 3:
  - larger chromosomes explain more variance
  - exception for genes with large effects
- Results from “control”:
  - more markers explain more variance
  - suggests the use of denser panels
- Results from method 3:
  - suggests variances for use in genomic selection

Acknowledgments

This study is part of the project FUGATO-plus GenoTrack and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven.