Applied Statistical Methods In Animal Science

Peter von Rohr

18.02.2019

Administration

- Course: 2 hours of lecture (2 V)
- Plan: 2 V $\rightarrow$ 1 U + 1V (1 hour of exercise and 1 hour of lecture)
- Exercises: Work on problems in R
- Material: course notes, slides, solution to exercises
- Exam: written, date: 27.05.2019, 08:15-09:00

Objectives

The students

- are familiar with the properties of **fixed linear effects models**
- are able to analyse simple data sets
- know why least squares cannot be used for genomic selection.
- know the statistical methods used in genomic selection, such as
  - BLUP-based approaches,
  - Bayesian procedures and
  - LASSO.
- are able to solve simple exercise problems using the statistical framework R.

Program

Woche	Datum	Thema
1	18.02	Introduction
2	25.02	Fixed Linear Effects Models
3	04.03	GBLUP - Marker-Effects Models
4	11.03	GBLUP - Breeding Value Models
5	18.03	Lasso
6	25.03	Bayesian Approaches

Information

- Website: https://charlotte-ngs.github.io/GELASMSS2019/
- Topics for master thesis
- Excursion: 13.05.2019, more information will follow
- Exam: 27.05.2019 08:15 – 09:00

This Course

- Use dataset that is used to predict genomic breeding values and introduce four methods

1. Fixed Linear Effects Models - Least Squares
2. GBLUP - genomic version of BLUP
3. LASSO - still fixed linear effects model, but modified parameter estimation
4. Bayesian approach to estimate unknown parameter

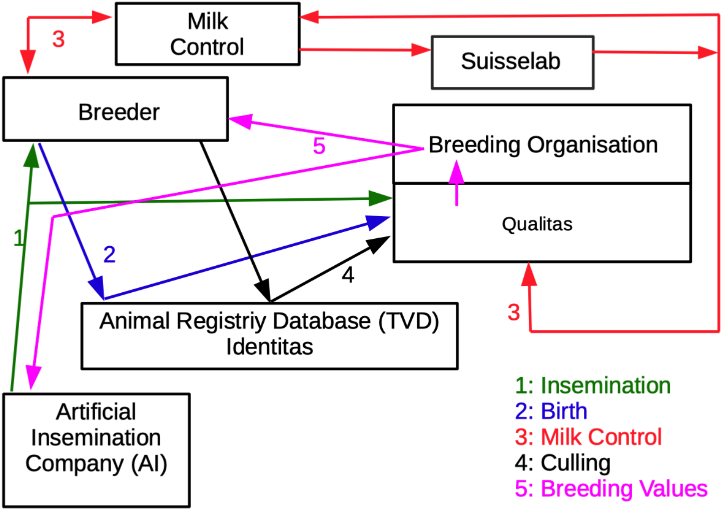
Significance

- Why is this important?
- Is this only relevant for animal breeding?
- What about the rest of animal science?
- General trend of collecting data has led to development of `Big Data`
- Examples
    - Presidential campains in the US
    - Health care
    - Agriculture: Smart Farming
    - Animal Science

Traditional Animal Breeding

- Before 2006
- Data collected for other purposes were used to predict breeding values
- Predicted breeding values as side-product

Data Logistics



1: Insemination
2: Birth
3: Milk Control
4: Culling
5: Breeding Values

Genomic Selection

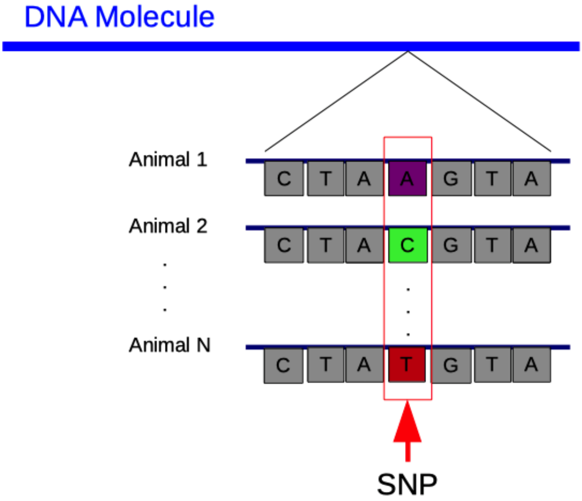
- Same goal as in traditional breeding: Find animals with best genetic potential as parents of next generation
- New: use additional source of information
- **Genomic** information
  - spread accross whole genome
  - single nucleotide polymorphisms (SNP)
- Introduction:

"> Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829"

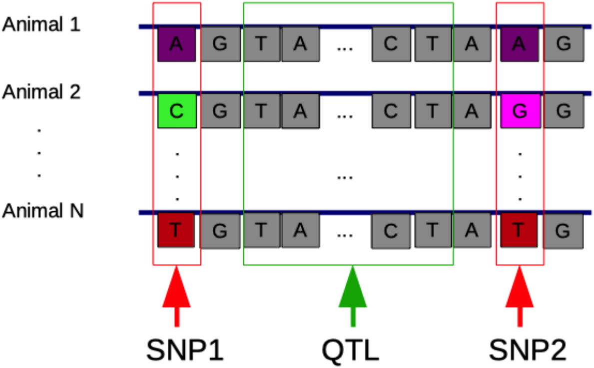
- Popularisation:

"> L. R. Schaeffer. Strategy for applying genome-wide selection in dairy cat- tle. Journal of Animal Breeding and Genetics, 123(4):218–223, 2006. ISSN 09312668. doi: 10.1111/j.1439-0388.2006.00595.x."

SNP


DNA Molecule
Animal 1
C T A A G T A
Animal 2
C T A C G T A
Animal N
C T A T G T A
SNP

QTL


Animal 1
Animal 2
Animal N
A G T A ... C T A A G
C G T A ... C T A G G
T G T A ... C T A T G
SNP1
QTL
SNP2

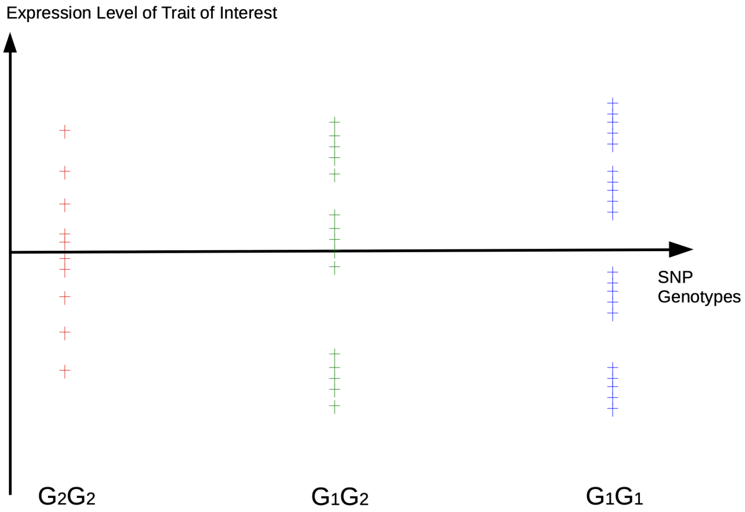
Linkage

- Flanking SNPs and QTL not independent passed on from parents to progeny
- Favorable QTL-allele linked with a given SNP-allele
- QTL is unknown, but use SNPs close to QTL as information for selection

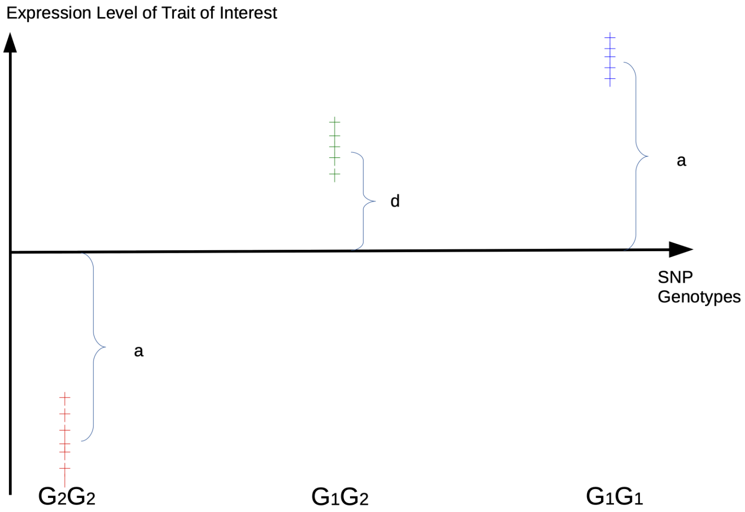
Monogenic Model

- Assume quantitative trait is influenced by one locus only
- Locus is bi-allelic → two alleles (G_1 and G_2) and three genotypes
- Look at Distribution of trait values for three different genotypes

Distribution No Effect



Distribution With Effect



Breeding Value

- Definition: Two times deviation from large number of offspring from population mean
- Assume: Hardy-Weinberg equilibrium
- Compute population mean as expected value of genotypic values
- Compute expected genotypic value of offspring for each of the three parental genotypes
- Assume purely additive loci, hence $d = 0$

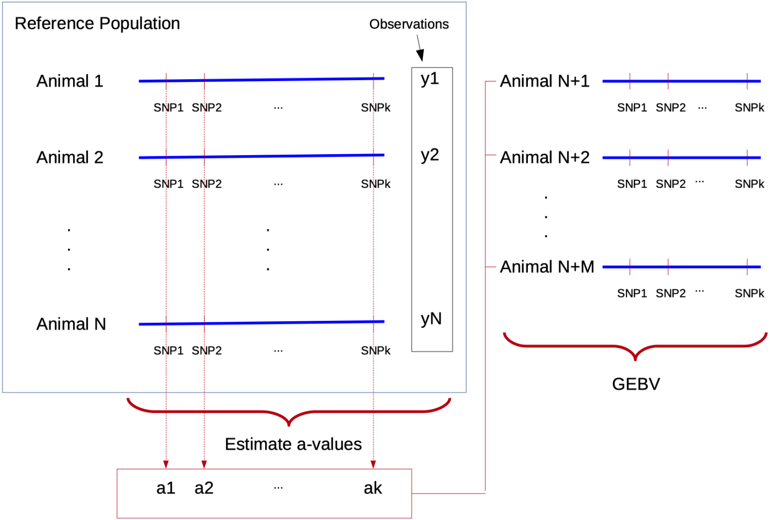
Genomic Breeding Value

- Take into account many loci
- Approximate unknown QTL with linked SNP
- Estimate *a*-effects from monogenic model
- Compute genomic breeding values for all loci based on *a* effects

Two Approaches

1. Two Step Procedure (used currently in Swiss Dairy Cattle)
2. Single Step

Two Step



Reference Population

Observations

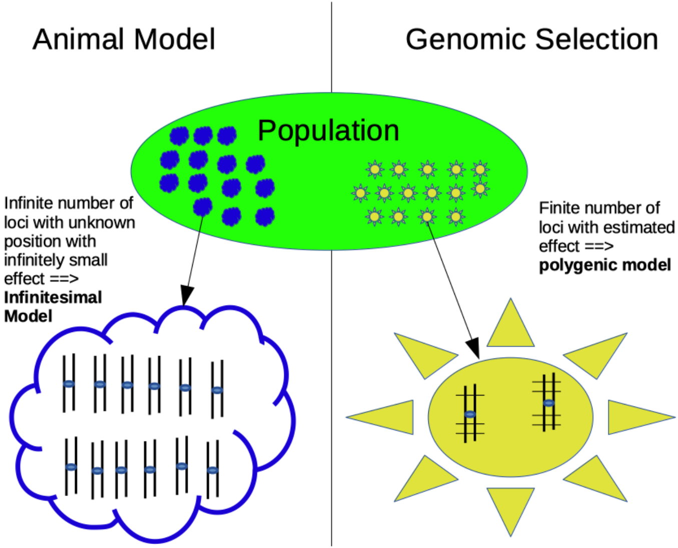
Estimate a-values

GEBV

Single Step

- Combine all information into one single BLUP-based analysis
- Problem: Determine covariance between animals with and without genomic information

Summary: Traditional versus Genomic Selection



Animal Model

Genomic Selection

Population

Infinite number of loci with unknown position with infinitely small effect ==> **Infinitesimal Model**

Finite number of loci with estimated effect ==> **polygenic model**