


# Applied Statistical Methods In Animal Science

Peter von Rohr

22.02.2021

# Administration

- ▶ Course: 2 hours of lecture (2 V)
- ▶ Plan: 2 V  $\rightarrow$  1 U + 1 V (i.e., 1 hour of lecture intersperced with time to do exercises)
- ▶ Exercises: Work on problems in R  Exercise platform, w02
- ▶ Material: course notes, slides, solution to exercises
- ▶ Exam: written, date: 31.05.2021, 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

	Week	Date	Topic
Applied Statistics	1	22.02	Introduction to Applied Statistical Methods
	2	01.03	Linear Fixed Effect Models
	3	08.03	GBLUP - Marker-Effects Models
	4	15.03	GBLUP - Breeding Value Models
	5	22.03	Lasso
	6	29.03	Bayesian Approaches
	7	05.04	<b>Easter Monday</b>
	8	12.04	Introduction to <u>Genetic Evaluation</u> of Livestock
	9	19.04	Model Selection
	10	26.04	Variance Components
	11	03.05	Genetic Groups and Longitudinal Data
	12	10.05	Genomic Selection
	13	17.05	Questions, Test Exam
	14	24.05	<b>Pfingstmontag</b>
	15	31.05	Exams →
			for both courses

# Information


- ▶ Website: <https://charlotte-ngs.github.io/gelasmss2021/>
- ▶ Topics for master thesis: [https://charlotte-ngs.github.io/gelasmss2021/misc/MasterThesisTopics\\_SS2021.html](https://charlotte-ngs.github.io/gelasmss2021/misc/MasterThesisTopics_SS2021.html)
- ▶ Exam: 31.05.2021 08:15 – 09:00

# This Course

Bachelor Statistics: Multiple Linear Regression (MLR)

Applied Statistics: Aim: Further develop the concepts started in MLR

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

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

Methods 2, 3 and 4 are solving problems found with method 1

Assumption: Population of livestock animals. From animals of this population, we have a dataset of observations, and genomic information

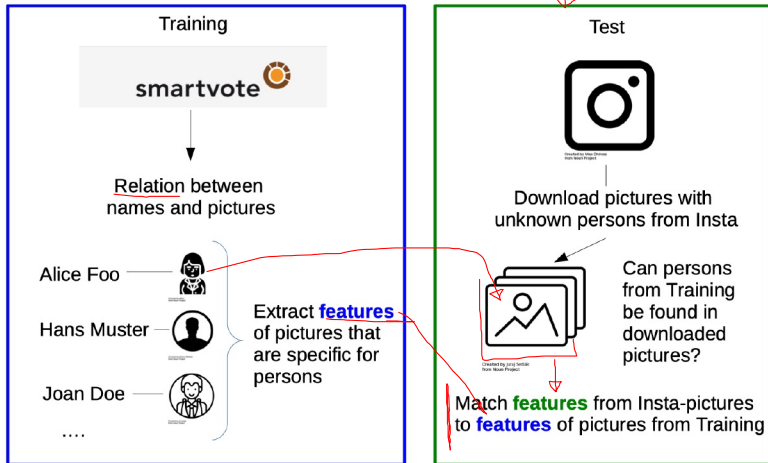
# Significance

Ex: Corona Pandemic:

Governments: Develop measures and rules of behavior based on the number of infections, R-value which the reproduction number

- ▶ 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 campaigns in the US
  - ▶ Health care
  - ▶ Face recognition
  - ▶ Agriculture: Smart Farming
  - ▶ Animal Science

## Face-Recognition





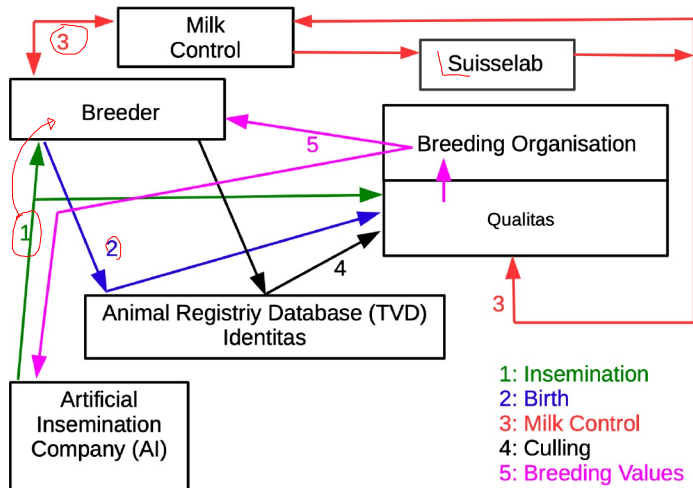
# Traditional Animal Breeding

traditional or "pre-genomic" era (before 2006)  
==> breeding values are predicted only based  
on phenotypic information and pedigree data

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

# Data Logistics

After birth of a calf, lactation starts



1: Insemination

2: Birth

3: Milk Control

4: Culling

5: Breeding Values

3 times a year:

April, August, Dec

# 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:

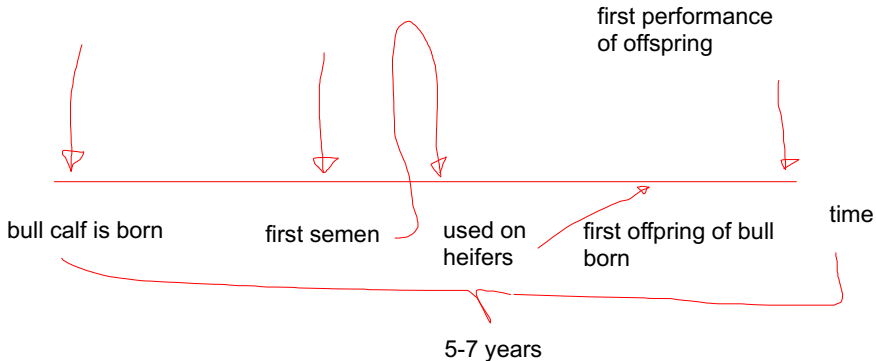
"> Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829"

- ▶ Popularisation: Use genomic selection to save about 90% of the total costs of cattle breeding program

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

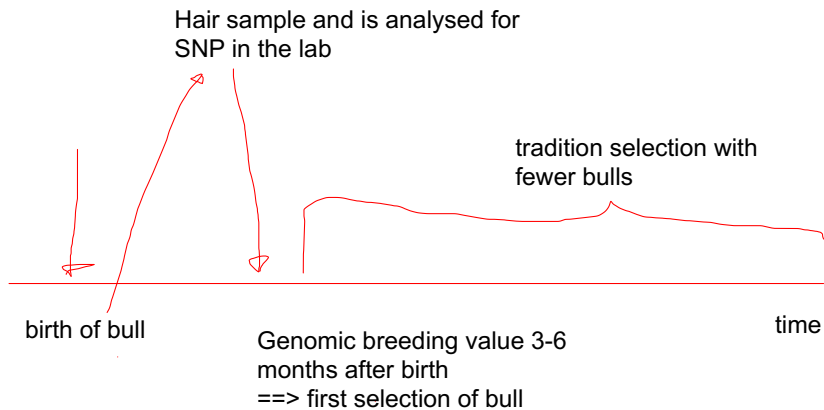
Traditional breeding programs in dairy cattle:

- selection of bulls is based on evaluation of daughter performance
- most important traits can only be observed in cows



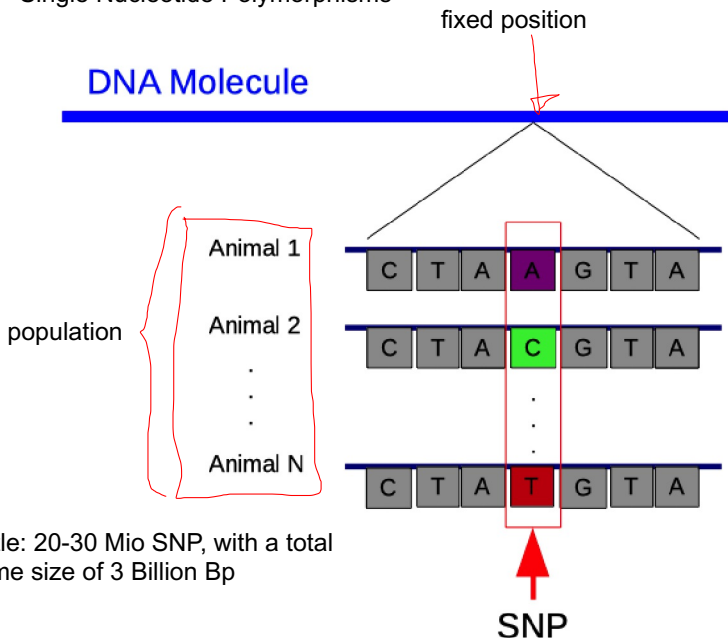
Progeny tests of bulls: start with 300-400 bulls in test, kept 15-20

## Breeding program with Genomic Selection



Cost saving: Reduction of time until the first selection decision from 7 years to 6 months.

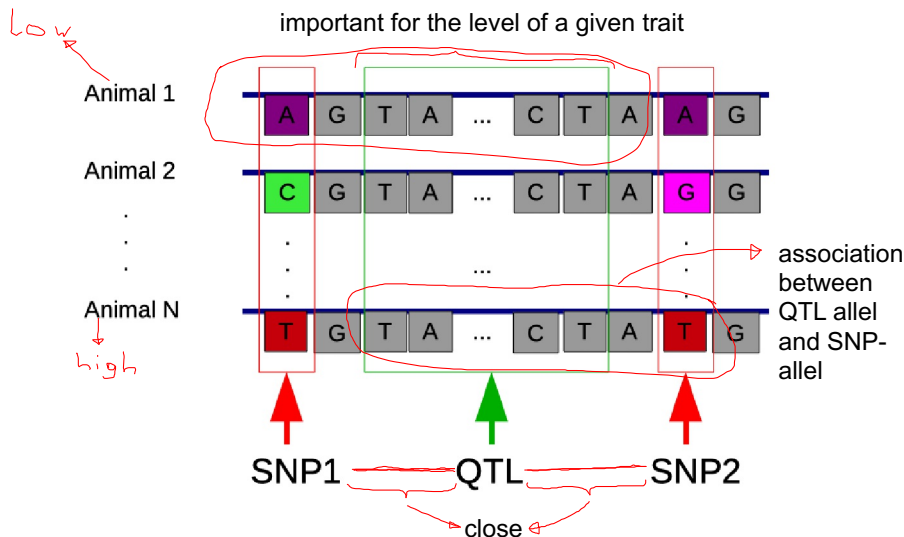
# SNP Single Nucleotide Polymorphisms



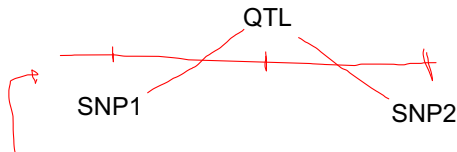
in cattle: 20-30 Mio SNP, with a total genome size of 3 Billion Bp

# QTL

## Quantitative Trait Locus, with unknown positions



# Linkage



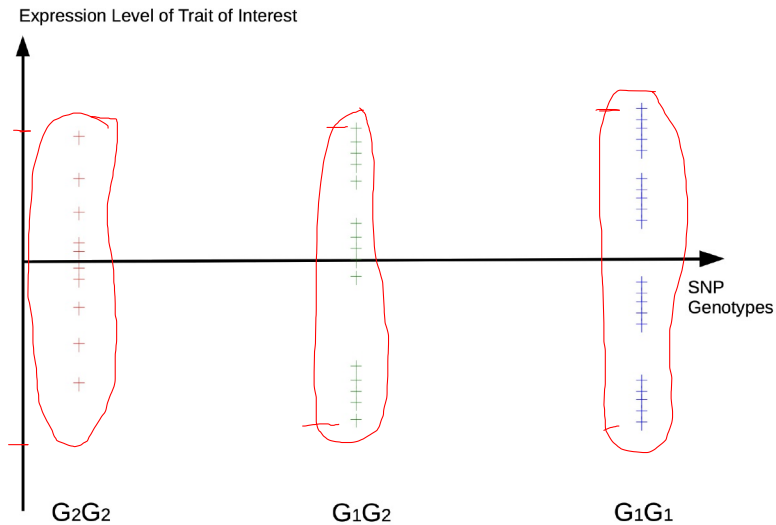
- ▶ Flanking SNPs and QTL not independent passed on from parents to progeny
- ▶ Favorable QTL-allele linked with a given SNP-allele
- ▶ QTL is unknown, but use SNPs close to QTL as information for selection



# Monogenic Model

- ▶ Assume quantitative trait is influenced by one locus only
- ▶ Locus is bi-allelic  $\rightarrow$  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  $d = 0$

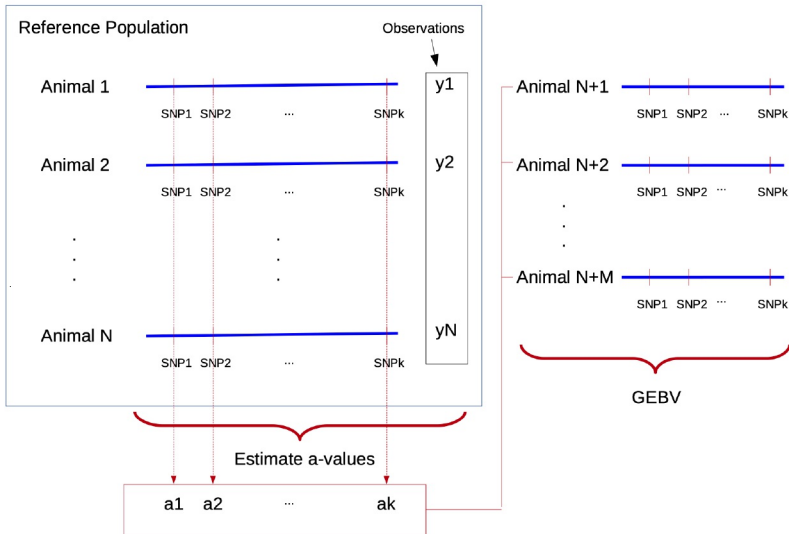
# Genomic Breeding Value

- ▶ 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



# 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

