So far: Not considered any genomic information for selecting parents
New: Use genomic information in the form of SNP-markers for our selection decision ==> Genomic Selection

Genomic Selection

Peter von Rohr

20 November 2020

Introduction

Starting in 2000 with the first draft of the complete sequence of the Human Genome, the area of genomics started to spread in all different areas of genetics

- Proposed in 2001
- Widely adopted in 2007/2008
- Costs of breeding program reduced due to shorter generation intervals
- In cattle: young sire selection versus selection based on sire proofs
- In pigs: early selection among full sibbs
- Inbreeding must be considered

Terminology

- **Genomic Selection**: use of genomic Information for selection decisions
- Genomic Information is used to predict **genomic breeding values**

Number of genotypes used in the prediction of GBV:

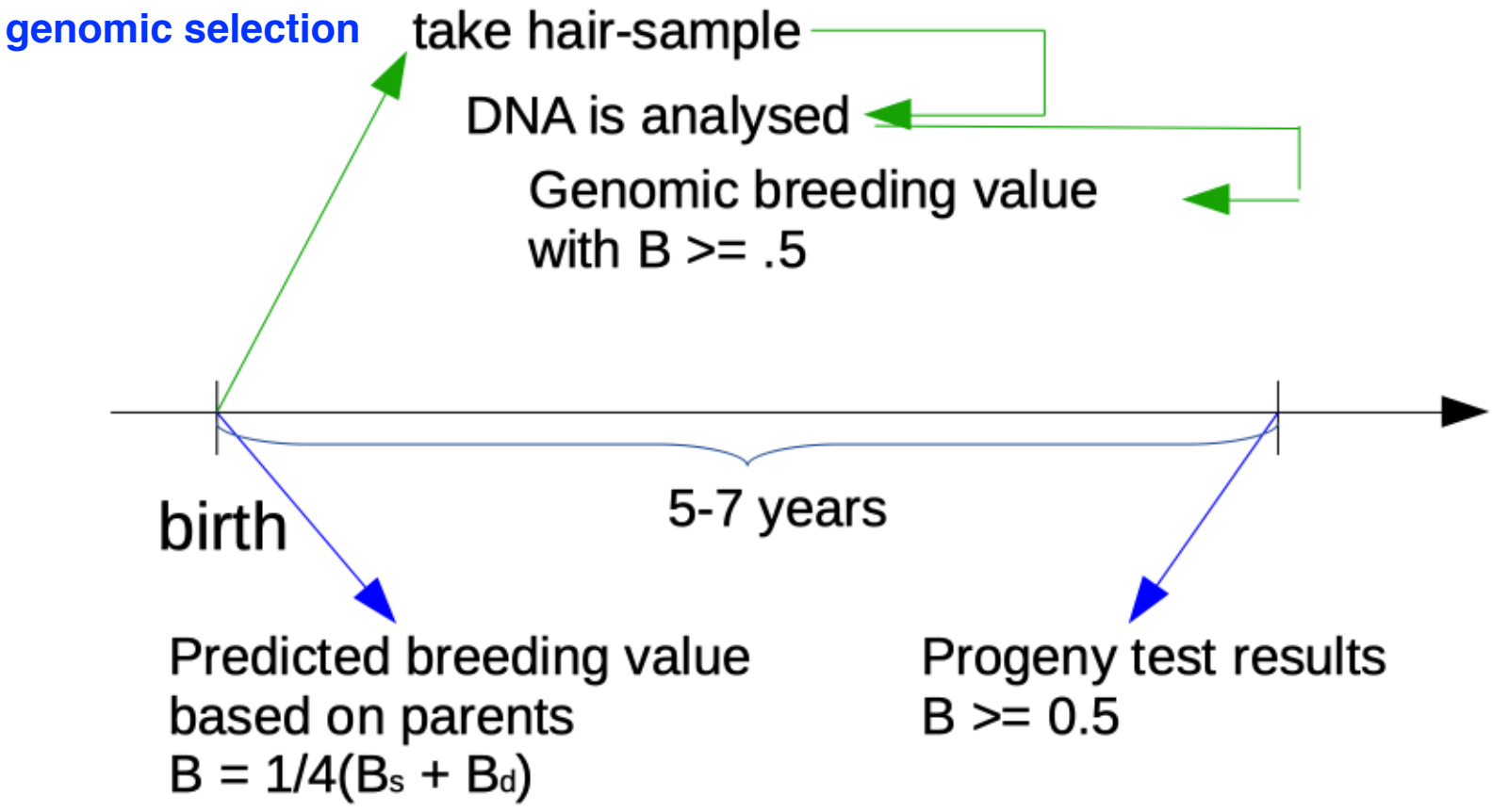
BrownSwiss: 80'000 genotypes
Holstein (red and black): 300'000 genotypes

Per genotype: 150'000 SNP markers

From the complete sequence of cattle, we can infer that there are about 20 Mio SNP

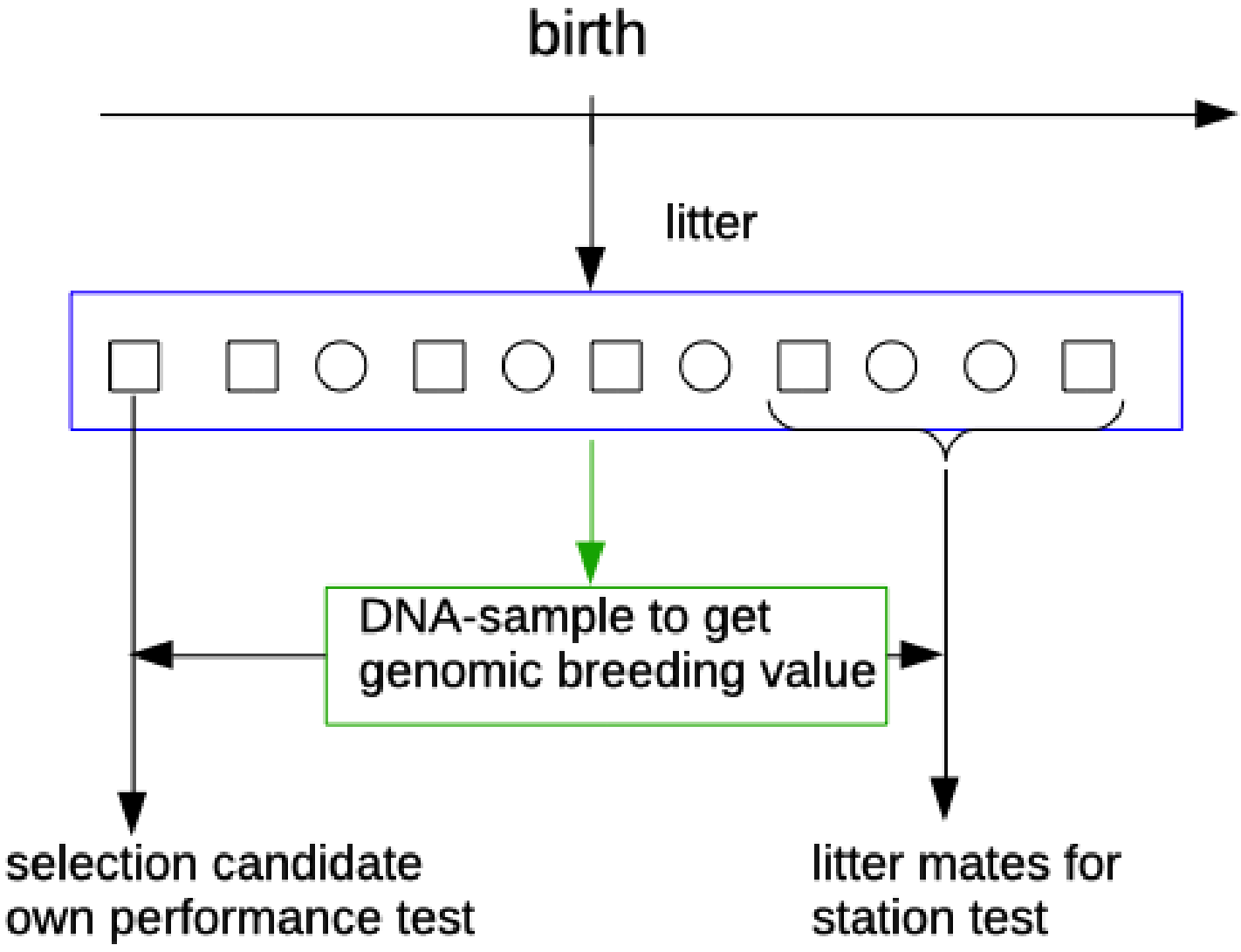
Benefits in Cattle

With genomic information



Without genomic information

Benefits in Pigs

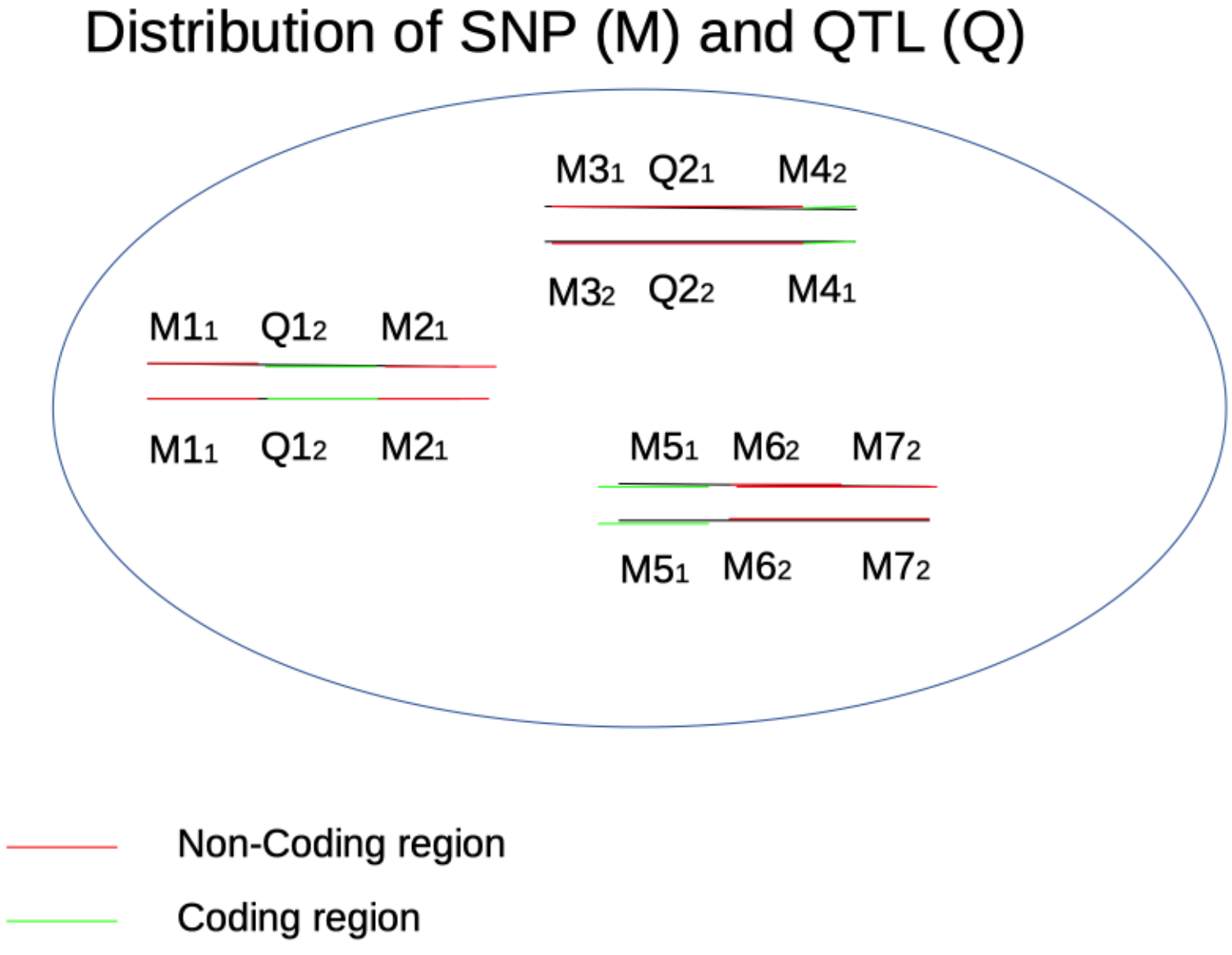


Genetic Model

Infinitesimal Model: Assume an infinite number loci that are affecting a quantitative trait
Polygenic Model: Assume a finite number of loci influencing a quantitative trait

- Recall: BLUP animal model is based on infinitesimal model
- Prediction of genomic breeding values is based on **polygenic model**
- In polygenic model: **Single Nucleotide Polymorphisms** (SNP) are used as markers
- Marker genotypes are expected to be associated with genotypes of **Quantitative Trait Loci** (QTL)

Polygenic Model



Statistical Models

*** In principle, the linear mixed effect models of the traditional BLUP animal model can be replaced by a fixed linear effect model (no additional random genetic breeding value in the model), but only fixed effects which correspond to the different SNP-genotypes**
*** Problem: Many SNP-positions lead to many different fixed Effects**

Two types of models are used

1. marker-effect models (MEM)
2. genomic-breeding-value based models (BVM)

MEM

From single locus model with three genotypes:
genotypic values of
*** G1G1: +a**
*** G1G2: +d**
*** G2G2: -a**

For MEM, we set d=0, for every SNP-position, the respective a-value must be estimated.

- marker effects (*a*-values) are fitted using
  - a simple linear model $\rightarrow$ marker effects are fixed
  - a linear mixed effects model $\rightarrow$ marker effects are random
- Problem of finding which markers are associated to QTL
- With high number of SNP compared to number of genotyped animals: very large systems of equations to solve

BVM

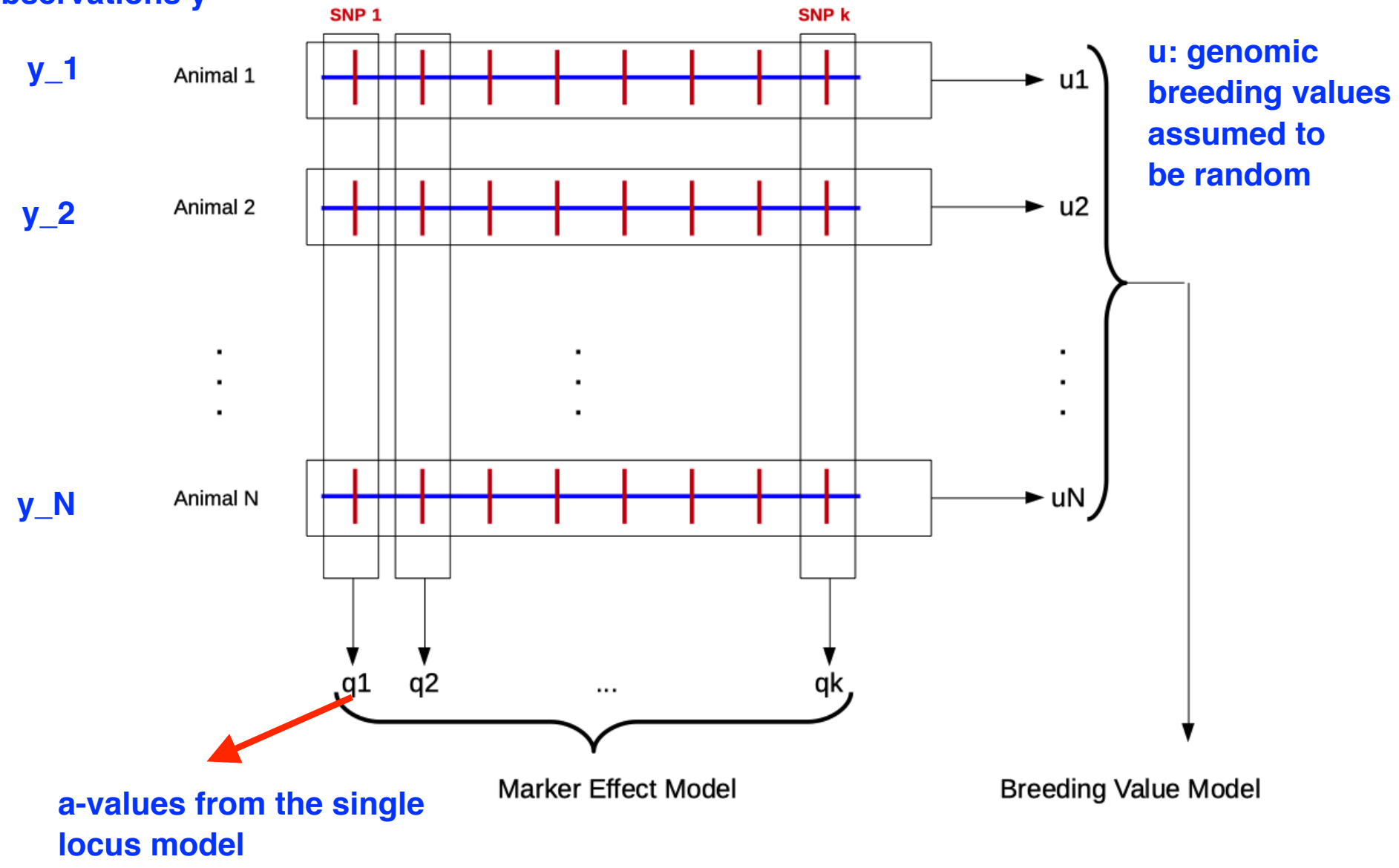
- genomic breeding values as random effects
- similar to animal model
- genomic relationship matrix (G) instead of numerator relationship matrix (A)

MEM versus BVM

Typical numbers of k: 150'000-800'000

Typical Numbers for N: 80'000 for Braunvieh, 250'000 for Holstein, Simmenthal

Observations y



Logistic Procedures

2-Step: reference population is assumed to have a perfect data set, that means every animal in the reference population has genomic information and also phenotypes
Step 1: Estimate marker effects (a-values) based on the data available in the reference population
Step 2: Estimate genomic breeding values for young animals that do not have phenotypic information yet, based on their genotypes and based on the results of step 1.

- Two Step:
  - use reference population to get marker effects using MEM
  - use marker effects to get to genomic breeding values
- Single Step
  - MEM or BVM in a single evaluation
  - difficulty how to combine animals with and without genotypes

Two Step Procedure