Genomic Selection

Peter von Rohr

2023-12-08

Introduction

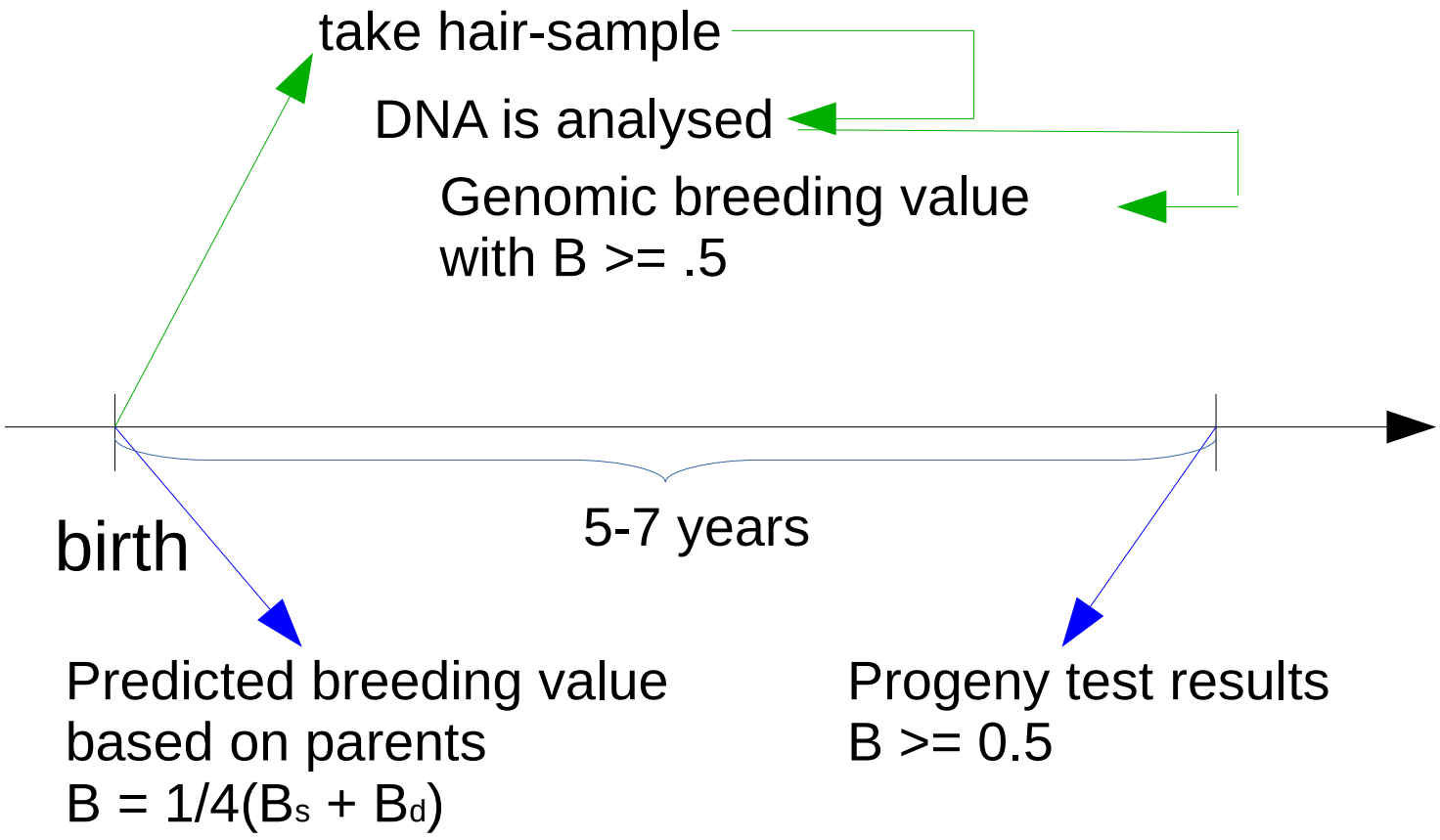
- Proposed in 2001
- Widely adopted in 2007/2008
- Costs of breeding program reduced due to shorter generation intervals
- In cattle: young sire selection versus selection based on sire proofs
- In pigs: early selection among full sibbs
- Inbreeding must be considered

Terminology

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

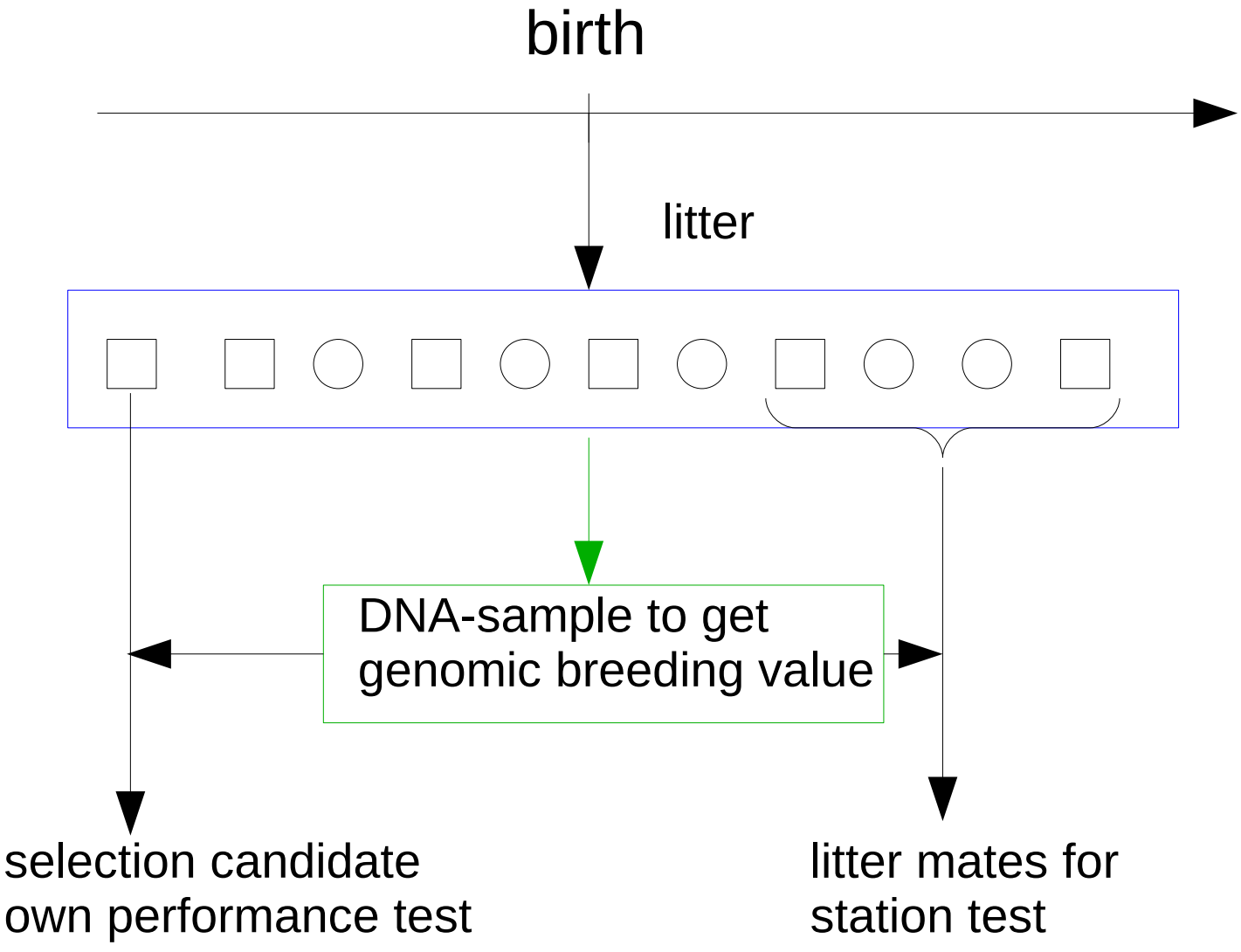
Benefits in Cattle

With genomic information



Without genomic information

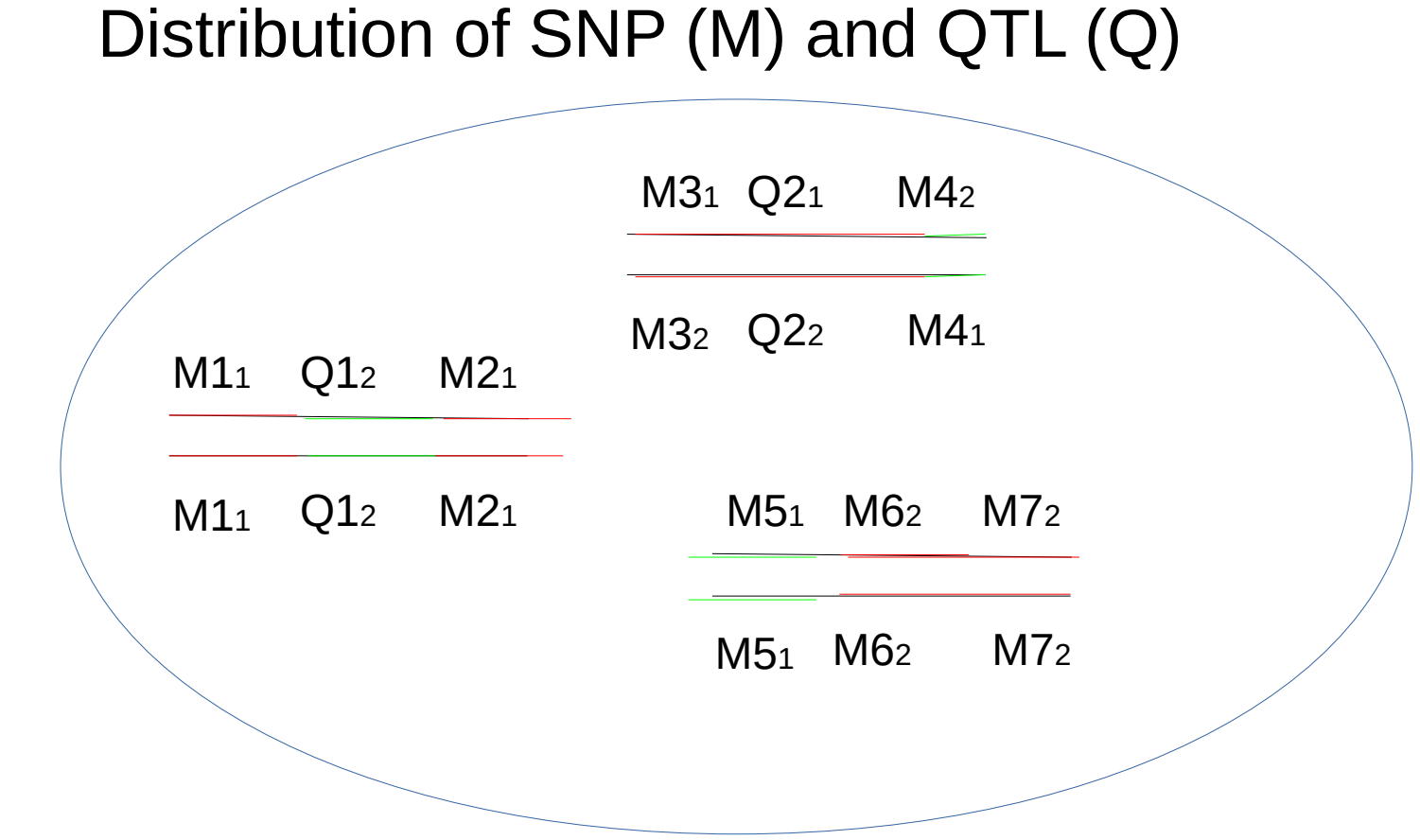
Benefits in Pigs



Genetic Model

- 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

Two types of models are used

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

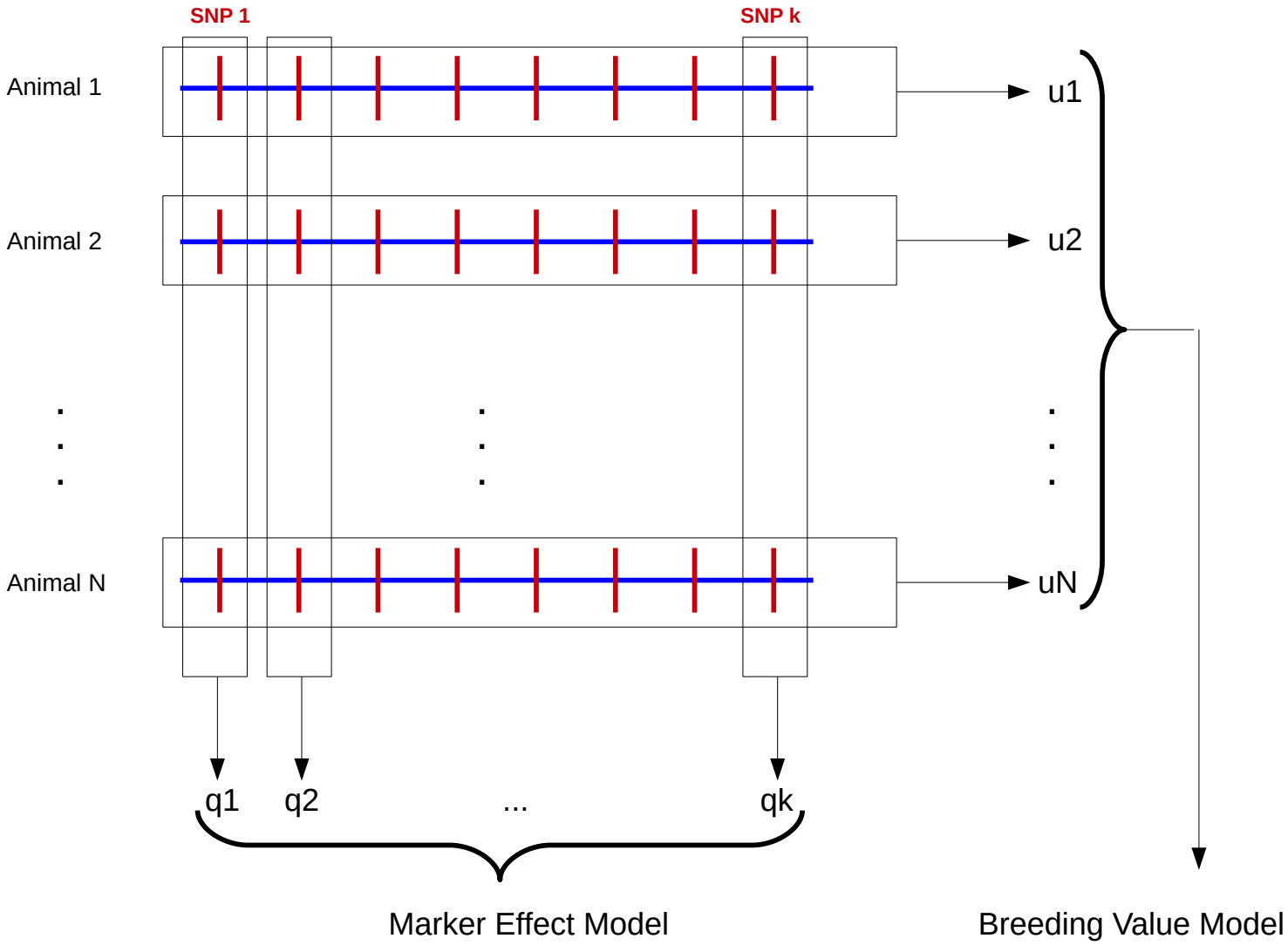
MEM

- 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)

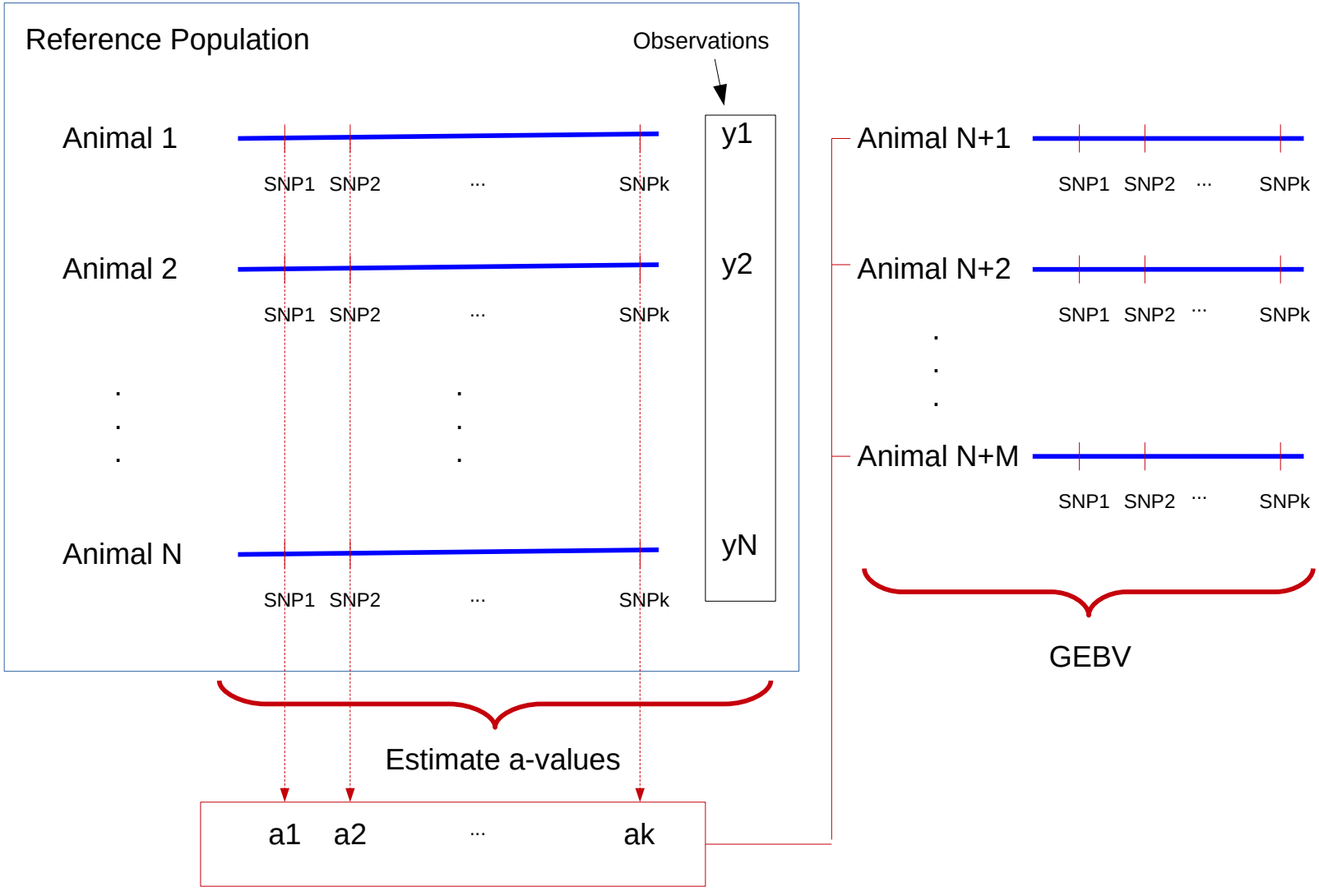
MEM versus BVM



Logistic Procedures

- 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



Single Step GBLUP

- Use a mixed linear effect model
- Genomic breeding values g are random effects

$$y = Xb + Zg + e$$

with

- $E(e) = 0$, $var(e) = I * \sigma_e^2$
- $E(g) = 0$, $var(g) = G * \sigma_g^2$
- Genomic relationship matrix G

Solution Via Mixed Model Equations

- All animals have genotypes and observations

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda * G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

with $\lambda = \sigma_e^2 / \sigma_g^2$.

Animals Without Observations

- Young animals do not have observations
- Partition $\hat{g}$ into
  - $\hat{g}_1$ animals with observations and
  - $\hat{g}_2$ animals without observations
- Resulting Mixed Model Equations are (assume $\lambda = 1$)

$$\begin{bmatrix} X^T X & X^T Z & 0 \\ Z^T X & Z^T Z + G^{(11)} & G^{(12)} \\ 0 & G^{(21)} & G^{(22)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \\ 0 \end{bmatrix}$$

Predicted Genomic Breeding Values

- Last line of Mixed model equations

$$G^{(21)} \cdot \hat{g}_1 + G^{(22)} \cdot \hat{g}_2 = 0$$

Solutions

- Solving for $\hat{g}_2$

$$\hat{g}_2 = -\left(G^{(22)}\right)^{-1} \cdot G^{(21)} \cdot \hat{g}_1$$

Genomic Relationship Matrix

- Breeding value model uses genomic breeding values g as random effects
- Variance-covariance matrix of g are proposed to be proportional to matrix G

$$var(g) = G * \sigma_g^2$$

where G is called **genomic relationship matrix** (GRM)

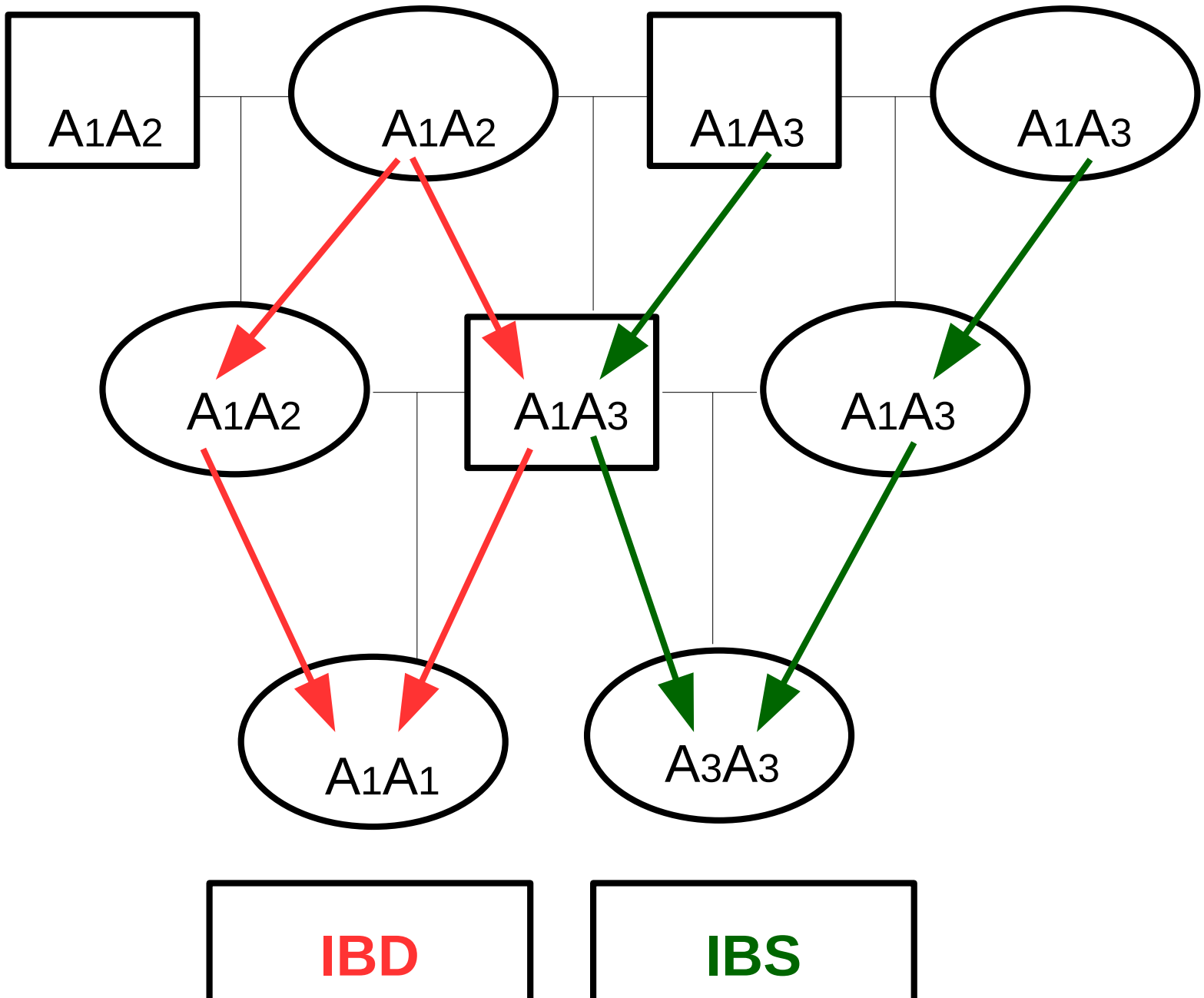
Properties of G

- genomic breeding values g are linear combinations of q
- g as deviations, that means $E(g) = 0$
- $var(g)$ as product between G and σ_g^2 where G is the genomic relationship matrix
- G should be similar to A

Change of Identity Concept

- A is based on identity by descent
- G is based on identity by state (including ibd), assuming that the same allele has the same effect
- IBS can only be observed with SNP-genotype data

Identity



Linear Combination

- SNP marker effects (a values) from marker effect model are in vector q
- Genomic breeding values from breeding value model are determined by

$$g = U \cdot q$$

- Matrix U is determined by desired properties of g

Deviation

- Genomic breeding values are defined as deviation from a certain basis

$\rightarrow E(g) = 0$

- How to determine matrix U such that $E(g) = 0$

Equivalence Between Models

Decomposition of phenotypic observation y_i with

- Marker effect model

$$y_i = w_i^T \cdot q + e_i$$

- Breeding value model

$$y_i = g_i + e_i$$

- g_i and $w_i^T \cdot q$ represent the same genetic effects and should be equivalent in terms of variability

Expected Values

- Required: $E(g_i) = 0$
- But: $E({w_i}^T \cdot q) = q^T \cdot E(w_i)$
- Take q as constant SNP effects
- Assume w_i to be the random variable with:

$$w_i = \begin{cases} 1 & \text{with probability} \quad p^2 \\ 0 & \text{with probability} \quad 2p(1-p) \\ -1 & \text{with probability} \quad (1-p)^2 \end{cases}$$

$\rightarrow E(w_i)$: For a single locus

$$E(w_i) = 1*p^2+0*2p(1-p)+(-1)(1-p)^2 = p^2-1+2p-p^2 = 2p-1 \neq 0$$

Specification of g

- Set

$$g_i = (w_i^T - s_i^T) \cdot q$$

with $s_i = E(w_i) = 2p - 1$

- Resulting in

$$g = U \cdot q = (W - S) \cdot q$$

with matrix S having columns j with all elements equal to $2p_j - 1$ where p_j is the allele frequency of the SNP allele associated with the positive effect.

Genetic Variance

- Requirement: $var(g) = G * \sigma_g^2$
- Result from Gianola et al. (2009):

$$\sigma_g^2 = \sigma_q^2 * \sum_{j=1}^{k}(1 - 2p_j(1 - p_j))$$

- From earlier: $g = U \cdot q$

$$var(g) = var(U \cdot q) = U \cdot var(q) \cdot U^T = UU^T\sigma_q^2$$

- Combining

$$var(g) = UU^T\sigma_q^2 = G * \sigma_q^2 * \sum_{j=1}^{k}(1 - 2p_j(1 - p_j))$$

Genomic Relationship Matrix

$$G = \frac{UU^T}{\sum_{j=1}^{k}(1 - 2p_j(1 - p_j))}$$

How To Compute G

- Read matrix W
- For each column j of W compute frequency p_j
- Compute matrix S and $\sum_{j=1}^{k}(1-2p_j(1-p_j))$ from p_j
- Compute U from W and S
- Compute G