Genomic Relationship Matrix

Peter von Rohr

09.03.2020

Background

- 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)

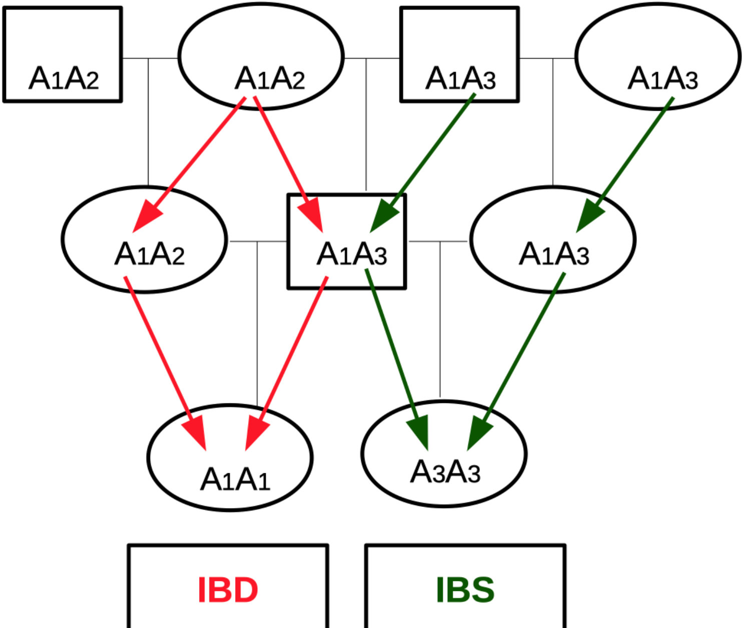
Properties of G

- 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:

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

$\rightarrow E(w_i)$: 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