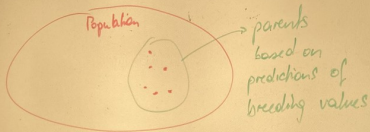


OHP Picture 1

Recap: Genomic Selection stands for the procedure where in addition to the previously available data (observations and pedigree) used for traditional prediction of breeding values, information on many loci across the whole genome is used to select candidates as parents for future generations.



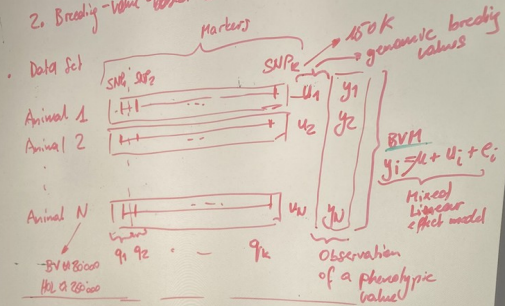
Predictions of breeding values

- | | | | |
|--------------------------------------|---|-------------------------------------|---------------|
| Genomic Prediction of breeding value | { | • Observations of phenotypic traits | } traditional |
| | | • Pedigrees | |
| | | • SNP - Marker information | |

OHP Picture 2

- Changes related to prediction of genomic breeding values
 - replace the infinitesimal model by the polygenic model

- Two statistical models that can be used
 1. Marker effect-model (MEM)
 2. Breeding-value-based model (BVM) \rightarrow equivalent



MEM: $y_i = \mu + q_1 + q_2 + \dots + q_k + e_i \rightarrow$ Fixed linear effect model

OHP Picture 3

With respect to parameter estimation, there are two possible approaches:

1. Two-step approach (Cattle)
(Pigs, cattle moving towards)
2. Single-step

Two-Step (mostly done with MEM)

Step 1: Estimation of marker-effects in a so called reference population

• Traditional breeding values with high reliability

• Genotyped

• Mostly bulls which are progeny tested

$\hat{a} \rightarrow$ de-regression \rightarrow $\begin{cases} \text{pedigree} \\ \text{phenotype} \end{cases}$

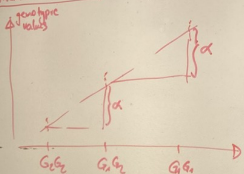
MEM: $y^* = \mu + q_1 + q_2 + \dots + q_k + E$ using fixed linear effect model.

OHP Picture 4

Data Set for MEM:

Animal	SNP1	SNP2	...	SNP _k	y
1	G ₁ G ₂	H ₁ H ₂	...	I ₁ I ₁	y ₁
2	G ₂ G ₂	H ₁ H ₁	...	I ₂ I ₂	
...	
N	G ₁ G ₁	H ₁ H ₂	...	I ₁ I ₁	

Quantitative Genetics



	U_{ij}
G ₁ G ₁	a
G ₁ G ₂	d
G ₂ G ₂	-a

In MEM, we assume $d=0$
 $\Rightarrow \alpha=a$

MEM for animal 1:

$$\left. \begin{aligned} y_1 &= \mu + 0 \cdot q_1 + 0 \cdot q_2 + \dots + 1 \cdot q_k + E_1 \\ y_2 &= \mu + (-1) \cdot q_1 + 1 \cdot q_2 + \dots + (-1) \cdot q_k + E_2 \end{aligned} \right\} q_j \text{ is the phenotypic value of SNP } j$$

OHP Picture 5

a With regard to estimating marker effects using fixed linear effect models, Least Squares cannot be used because $N \ll k$ where N is the number of animals in the reference population (5000) and k is the number of SNP-positions.

a. Solutions:

1. LASSO: reduce number of SNPs and estimate marker effects.

2. Bayesian procedure

3. Mixed linear effect model.

$$y = \underbrace{X\beta}_{\text{fixed}} + \underbrace{Wq + e}_{\text{random marker effects}}$$

$$\text{MRE: } \begin{bmatrix} X^T X & X^T W \\ W^T X & W^T W + I \cdot \lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{q} \end{bmatrix} = \begin{bmatrix} X^T y \\ W^T y \end{bmatrix}$$

$$\text{with } \lambda = \frac{\sigma_e^2}{\sigma_q^2}$$

$$\text{assume: } \text{var}(q) = I \cdot \sigma_q^2$$

OHP Picture 6

Result of Step 1: Estimates \hat{g} for marker effects

□ Step 2: From estimates of marker effects, genomic breeding values are computed by summing up appropriate marker effects for each genotyped animal.

Animal 1: Genotype:

$$G_1 G_2 \quad H_1 H_2 \quad \dots \quad I_1 I_1 \\ 0 \cdot \hat{q}_1 + 0 \cdot \hat{q}_2 + \dots + 1 \cdot \hat{q}_k = \hat{g}_1$$

Direct genomic breeding value for animal 1.

Animal 2:

$$G_2 G_2 \quad H_1 H_1 \quad \dots \quad I_2 I_2 \\ (-1) \hat{q}_1 + 1 \hat{q}_2 + \dots + (-1) \hat{q}_k = \hat{g}_2$$

Animal N: ...

$$\hat{g}_v$$

OHP Picture 7

Big advantage of genomic selection:

! Step 2 can be done with any animal with a genotype, not only for animals in the reference population.

⇒ If a calf is born, hair sample is used to extract DNA and to determine the genotype at the SNP-Markers

Animal $N+1$: Genotype
 G_1G_2 H_1H_2 --- I_1I_1

Using marker effect estimates, the direct genomic breeding value can be predicted.

Swiss Cattle:

- Marker effect estimation: 3 times per year (Step 1)
(April, August, December)
- Direct genomic breeding values: every 2 weeks (Step 2)

OHP Picture 8

Important Point to consider:

- High dependency on the data quality of the reference population.
- Genomic selection using the Two-Step procedure allows to make selection decisions very early in the live of an animal.
 - ⇒ the number of bull with progeny tests is reduced and the bulls with progeny test are pre-selected ⇒ reduces quality of marker effect estimates
- So far reference populations consisted only of male animals, but the large number of genotypes that are becoming available now are from female animals.
- Animals without genotype information, receive traditional prod. BV
- Some issues are addressed in Single Step Method using BMM

OHP Picture 9

□ Single Step using BVH

- Estimate/predict breeding values for animals in one evaluation.

- Linear Mixed Effect Models where the genomic breeding value is modelled as a random effect.

$$y = X\beta + Zg + e; \quad E(e) = 0, \text{Var}(e) = I \cdot \sigma_e^2$$

$$E(g) = 0, \text{var}(g) = G \cdot \sigma_g^2$$

where G is the genomic relationship matrix

- Solutions with MME
- Data set is the same as for the reference population in the Two-step approach.
 - ⇒ all animals have observations (y) and genotype information.
- Young animals, they do not have observations, but they have genotype information.

OHP Picture 10

Young Animals without Observations:

Animals	Genotypes z		k	Observations
	$G_1 G_1$	$H_1 H_2$	$I_1 I_1$	y_1
1	$G_1 G_1$	$H_1 H_2$	$I_1 I_1$	y_1
2	$G_1 G_2$	$H_1 H_1$	$I_1 I_1$	y_2
...				
N			$I_1 I_1$	
young animals	$G_1 G_2$	---	$I_1 I_1$	
	$G_1 G_1$		$I_1 I_2$	
	$G_1 G_2$		$I_1 I_2$	

Model:

$$y = X\beta + Zg + e$$

► vector g contains genomic breeding values for all animals. It can be partitioned into two parts

$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ where g_1 : animals with observations
 g_2 : young animals without observations

► variance-covariance matrix

$$\text{var}(g) = G \cdot \sigma_g^2 \rightarrow \text{var} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \cdot \sigma_g^2$$

OHP Picture 11

For MRE, G^{-1} is required which can be partitioned accordingly. Assuming

$$\text{var} \begin{bmatrix} \hat{\beta} \\ \hat{g}_e \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \cdot \hat{y}^2$$

$$\rightarrow G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \rightarrow G^{-1} = \begin{bmatrix} G^{(11)} & G^{(12)} \\ G^{(21)} & G^{(22)} \end{bmatrix}$$

the times per year

$$\rightarrow \text{MRE: } \begin{bmatrix} X^T X & X^T y & 0 \\ \Sigma^T X & \Sigma^T y + G^{(11)} \lambda & G^{(12)} \lambda \\ 0 & \lambda \cdot G^{(21)} & \lambda \cdot G^{(22)} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g}_e \\ \hat{g}_e \end{bmatrix} = \begin{bmatrix} X^T y \\ \Sigma^T y \\ 0 \end{bmatrix}$$

genomic breeding values for gay animals

every second week

$$\begin{aligned} 0 \cdot \hat{\beta} + \lambda \cdot G^{(11)} \hat{g}_e + \lambda \cdot G^{(12)} \hat{g}_e &= 0 \\ G^{(21)} \hat{g}_e + G^{(22)} \hat{g}_e &= 0 \\ G^{(22)} \hat{g}_e &= -G^{(21)} \hat{g}_e \\ \hat{g}_e &= -[G^{(22)}]^{-1} \cdot G^{(21)} \hat{g}_e \end{aligned}$$

OHP Picture 12

Genomic Relationship Matrix (G)

- Variance-Covariance matrix of genomic breeding values (g)

$$\text{var}(g) = G \cdot \sigma_g^2$$

- Properties of genomic breeding values (g)

- linear combinations of marker effects g

- deviations $\Rightarrow E(g) = 0$

- G should be similar to A

- \Rightarrow diagonal elements should be dominant

- \Rightarrow high off-diagonal elements only for animals with ^{close} relationship

- Genetic background

- * A is based on identity by descent

- * G is additionally also considering identity by state relationships

OHP Picture 13

- Genomic Breeding values g are a linear combination of marker effects q , i.e.,

$$g = U \cdot q$$

- Determine U :

- $E(g) = 0$

- Equivalence between MEM and BVH

$$y_i = \mu + w_i^T \cdot q + e_i \quad (\text{MEM})$$

$$y_i = \mu + g_i + e_i \quad (\text{BVH})$$

- Because $g_i = w_i^T \cdot q$

$$\text{with } E[g_i] = 0 \Rightarrow E[w_i^T \cdot q] = q \cdot E[w_i^T]$$

Markov effect

with random variable w_i encoding the genotype

$$w = \begin{cases} 1 \\ 0 \\ -1 \end{cases}$$

with probability p^2 ($\Leftrightarrow G_1 G_1$)

$$2pq$$

$$G_1 G_2$$

$$G_2 G_1$$

$$q^2$$

$$E[w_i] = 2p - 1 \neq 0$$

OHP Picture 14

a correction: Set

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

where $s_i = 2p - 1$

with p being the allele frequency of the positive allele

a Matrix-Vector:

$$g = U \cdot q = (W - S) q \quad \text{with that } E(g) = 0$$

- Requirement 2: $\text{var}(g) = G \cdot \bar{\sigma}_g^2$

$$\bar{\sigma}_g^2 = \underbrace{\bar{\sigma}_q^2}_{\text{marker-effect variance}} \cdot \sum_{j=1}^k (1 - 2p_j)(1 - p_j)$$

- Together with $g = U \cdot q$

$$\text{var}(g) = U \cdot \underbrace{\text{var}(q)}_{I \bar{\sigma}_q^2 \text{ (see MEM)}} \cdot U^T = U U^T \cdot \bar{\sigma}_q^2$$

$$\rightarrow \text{var}(g) = U U^T \cdot \bar{\sigma}_q^2 = G \cdot \bar{\sigma}_q^2 \cdot \sum_{j=1}^k (1 - 2p_j)(1 - p_j)$$

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