Centre for
Tropical Livestock
Genetics and Health

Genomic Based Predictions

Professor Raphael Mrode

Genetic and Genomic Approaches for Livestock Improvement

ILRI Campus, Nairobi, Kenya.

February 16 – 21 2026

Genetic Markers

Centre for Tropical Livestock Genetics and Health

- The genome may be defined as is an organism's complete set of DNA, including all of its genes and it contains all the information needed for an individual to develop and function.

- Genetic marker may be defined as a DNA sequence with a known location on a chromosome, that can be observed or detected and whose inheritance can be monitored.
- It may be comprised of long DNA sequence like minisatellites or a short sequence (microsatellites) or those associated a single base-pair change called single nucleotide polymorphism (SNP)

Limitations of Microsatellites

Centre for Tropical Livestock Genetics and Health

- Initially microsatellites were used as genetics in the 1980s and 1990s.
- Microsatellites are set of short repeated DNA sequences at a particular locus on a chromosome, which vary in number in different individuals and so can be used as markers

- Most significant genetic marker can be 10 cM or more from the QTL, therefore QTL are not mapped precisely.

- The association between marker and QTL may not persist through the population.

- The phase between marker and QTL may have to estimated for each family

Single Nucleotide Polymorphism (SNP).

Centre for Tropical Livestock Genetics and Health

- SNP is a DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome differs between paired chromosomes in an individual.
- For example, two sequenced DNA fragments from an individual, AAGC**C**TA to AAGC**T**TA, contain a difference in a single nucleotide.

- In this case we say that there are two *alleles*: C and T. Almost all common SNPs have only two alleles.

Genotyping chips

- Simultaneous genotyping of many SNP
- From few dozens up to several million SNP
- Two main technology providers, Illumina and Affymetrix
- Illumina products in cattle
  - 3000 (➔7000➔ 10000➔ 20000=« LD »)
  - **54 000=« 50k »**
  - 777 000=« HD »

FIGURE 1: BOVINESNP50 BEADCHIP

The BovineSNP50 BeadChip features more than 54,000 evenly-spaced SNPs across the entire bovine genome.

Genomic Selection (GS)

Centre for Tropical Livestock Genetics and Health

- GS - the use of genomic breeding values (GEBV) for the selection of animals.

- Genomic selection requires that markers (SNPs) are in linkage disequilibrium (LD) with the QTLs across the whole population

- Thus, the use of SNPs as markers enables all QTL in the genome to be traced through the tracing of chromosome segments defined by adjacent SNPs.

Steps in Genomic Selection (GS)

Centre for Tropical Livestock Genetics and Health

- Genotype animals with phenotypes (sires with daughters records for sex limited traits)

Genotyped and phenotyped animals ⟷ **Reference population**

- Estimate SNP solutions (SNP Key) in the reference population
- Validate in another data set but records excluded to determine accuracy of SNP key

Genotyped & phenotyped but phenotypes excluded ⟷ **Validation candidates**

- Genotype animals at birth or young age (no phenotypes) and use SNP key to prediction their GEBV and do selection

Genotyped but no phenotypes ⟷ **Selection candidates**

ROSLIN SRUC ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE 50 2024 CGIAR

Main advantages of Genomics

Centre for Tropical Livestock Genetics and Health

- Young bulls can be genotyped early in life and breeding values computed
- Can be used to select young bulls to be progeny tested, thereby reducing cost
- Higher accuracy of about 20-40% for young bulls above parent average and high accuracies for cow evaluations
- Reduction in generation interval
- Strategic genotyping of connected herds to handle difficult to measure traits : fertility, heat tolerance, methane reduction, diseases traits

Genomic Selection : efficiency

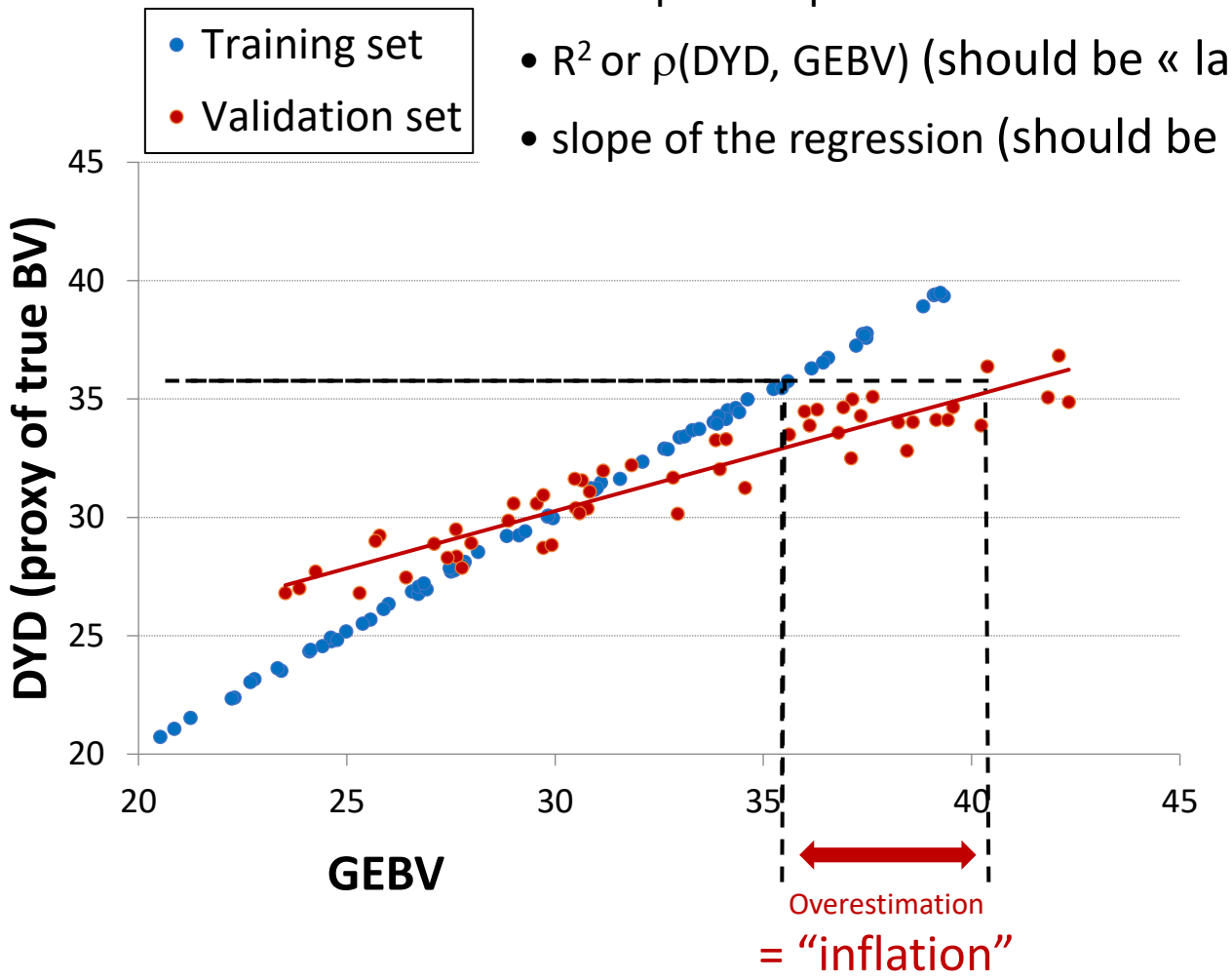
Centre for Tropical Livestock Genetics and Health

- Two main factors :
  - Accuracy of SNP effect estimation
    - size of reference population
    - heritability of the trait
    - statistical methodology used
  - Linkage Disequilibrium (LD) between markers and QTL
    - marker density
    - effective size of the population => number of « independent » segments
    - Relationship between candidates and reference population

Validation test

Two important parameters :

- R^2 or ρ(DYD, GEBV) (should be « large enough »)
- slope of the regression (should be close to 1)



General linear model

The general linear model underlying genomic evaluation is of the form

$\mathbf{y} = \mathbf{Xb} + \sum_{i}^{m} \mathbf{M}\mathbf{g}_i + \mathbf{e}$

where

m is the number of SNPs ; **y** is the data vector,

b the vector for mean or fixed effects

$\mathbf{g}_i$ the genetic effect of the i^{th} SNP genotype and **e** is the error.

The matrix **M** is of the dimension *n* (number of animals) and *m*, and $\mathbf{M}_i$ relates the i^{th} SNPs to data

It is assumed that all the additive genetic variance is explained by all the markers effects such that the estimate of animal's total genetic merit or breeding value (**a**) is: $\mathbf{a} = \sum_{i}^{m} \mathbf{M}_i \mathbf{g}_i$.

Data types used for genomic evaluation

Centre for Tropical Livestock Genetics and Health

- **y** = YD (Yield deviation) = Individual record corrected for all fixed and non genetic random effects

- y = DYD (Daughter yield deviation) = twice the average of all YDs of the daughter of a bull corrected for ½ genetic merit of their dams. Bulls usually have different number of daughters and so, we need to compute a weight equivalent to the numbers of daughters and the information on those daughters for each bull. This weight is called EDC (Equivalent Daughter Contribution)
- y = de-regressed proofs -- obtained simply by dividing the EBV of bull by the bull's reliability (corrected for parents' reliability) or by solving the MME to get the right-hand side
- EBVs --- NO

Coding and scaling genotypes

Centre for Tropical Livestock Genetics and Health

- The genotypes of animals (elements of **M**) are commonly coded as 2 and 0 for the two homozygotes (AA and BB) and 1 for the heterozygote (AB).

- Or if alleles are expressed in terms of nucleotides, and reference allele at a locus is G and the alternative allele is C, then code 0 = GG , 1 = GC and 2 = CC.

- The diagonal elements of **MM'** then indicate the individual relationship with itself (inbreeding) and the off-diagonal indicate the number of alleles shared by relatives

Scaling of genotypes

Centre for Tropical Livestock Genetics and Health

- SNPs → 2 alleles A/B but only one effect defined in M, therefore we are estimating the substitution effect m_i
- Commonly elements of **M** are scaled
  - to set the mean values of alleles effects to zero
  - account for differences in allele frequencies of the various SNPS

  - Let the frequency of the second or alternative allele at locus j be p_j
  - Elements of M can be scaled by subtracting $2p_j$.
  - If the element of column j of a matrix **P** equals $2p_j$, then matrix **Z**, which contained the scaled elements of **M** is : **Z = M - P**.
- Furthermore, the elements of Z be normalised by dividing the column for marker j by its standard deviation assumed to be
- $$\sqrt{2p_j(1-p_j)}$$

Mixed linear model for computing SNP effect

Centre for Tropical Livestock Genetics and Health

- The most common random model used assumes
  - the effect of the SNP are normally distributed,
  - all SNP are from a common normal distribution (eg. the same genetic variance for all SNPs).

- There are two equivalent models with these assumptions
- (1) SNP-BLUP - a model to estimate individual SNP effects simultaneously.
  - Then the DGV for animals are calculated from SNP solutions as DGV = $\mathbf{Z}\hat{\mathbf{g}}$, where $\hat{\mathbf{g}}$ are the estimates of SNP effects.

- (2) GBLUP - estimates DGV directly,
- Very similar to BLUP but uses **G** is the genomic relationship matrix, which is the realised proportion of the genome that animals share in common estimated from the SNP instead of the **A** matrix

SNP BLUP model

- In matrix form, model is
- **Y** = **Xb** + **Zg** + **e**
- **Y** = vector of observations: these can be de-regressed EBVs, phenotypes corrected for all fixed effects
- where **g** = vector of additive genetic effects corresponding to allele substitution effects for each SNP and **Z** = scaled matrix of genotypes
- MME are below with $\alpha = \sigma^2_e/\sigma^2_g$

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{I}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{pmatrix}$$

SNP-BLUP

- (1) **y** = YD (Yield deviation) = Individual record corrected for all fixed and non genetic random effects
  - If cows have several records, YD may be the average of all records
  - In such cases cows may have different number of records and a weight is needed for each cow.
  - An example of such a weight could be the inverse of the standard error for the mean of the records.

- (2) **y** = de-regressed breeding values of bulls, then
  - Each observation may be associated with differing reliabilities.

  - Thus a weighted analysis may be required to account for these differences in bull reliabilities.

  - Weight (wt_i) = effective daughter contribution or $wt_i = (1/ rel_{dtr}) - 1$, where rel_{dtr} is the bull's reliability from daughters with parent information excluded

SNP-BLUP

- The MME then are

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{I}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{pmatrix}$$

- where **R** = **D** and **D** is a diagonal matrix with diagonal element $i = wt_i$.

- In practise, the value of σ^2_g may not been known and σ^2_g could be obtained
- either as $\sigma^2_g = \sigma^2_a /m$, with m = the number of markers
- or as $\sigma^2_g = \sigma^2_a /2\Sigma p_j(1 - p_j)$
- and $\alpha = 2\Sigma p_j(1 - p_j) *[\sigma^2_e/\sigma^2_a]$

Example 1- Estimate SNP Effects using SNP BLUP

Centre for Tropical Livestock Genetics and Health

					FAT	SNP									
Animal	Sire	Dam	Mean	EDC	DYD	Genotype									
13	0	0	1	558	9.0	2	0	1	1	0	0	0	2	1	2
14	0	0	1	722	13.4	1	0	0	0	0	2	0	2	1	0
15	13	4	1	300	12.7	1	1	2	1	1	0	0	2	1	2
16	15	2	1	73	15.4	0	0	2	1	0	1	0	2	2	1
17	15	5	1	52	5.9	0	1	1	2	0	0	0	2	1	2
18	14	6	1	87	7.7	1	1	0	1	0	2	0	2	2	1
19	14	9	1	64	10.2	0	0	1	1	0	2	0	2	2	0
20	14	9	1	103	4.8	0	1	1	0	0	1	0	2	2	0
21	1	3	1	13	7.6	2	0	0	0	0	1	2	2	1	2
22	14	8	1	125	8.8	0	0	0	1	1	2	0	2	0	0
23	14	11	1	93	9.8	0	1	1	0	0	1	0	2	2	1
24	14	10	1	66	9.2	1	0	0	0	1	1	0	2	0	0
25	14	7	1	75	11.5	0	0	0	1	1	2	0	2	1	0
26	14	12	1	33	13.3	1	0	1	1	0	2	0	1	0	0

Example 1- Estimate SNP Effects using SNP BLUP

Centre for Tropical Livestock Genetics and Health

- The observations are the daughter yield deviations for fat yield
- The only fixed effect in the model is the mean
- It is assumed the genetic variance for fat yield is 35.241kg2 and residual variance of 245kg2
- Aim : To predict SNP effects for the 10 SNPs.
- Animals 13 to 20 as assumed as the reference population and 21 to 26 as validation candidates.
- The incidence matrix $\mathbf{X} = \mathbf{I}_q$, with q = 8, the number of animals in the reference population

ROSLIN SRUC ILRI 50 2024 CGIAR

Computing the matrices we need

Centre for Tropical Livestock Genetics and Health

- The incidence matrix $\mathbf{X} = \mathbf{I}_q$, with $q = 8$, the number of animals in the reference population
- $\mathbf{X'} = [\,1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\,]$

- The computation of Z requires calculating the allele frequency for each SNP.

Computing allele frequencies

- The allele frequency for the i^{th} SNP was computed as

$$\frac{\sum_{j}^{n} m_{ij}}{2 * n}$$

with n = 14, the number of animals with genotypes and m_{ij} are elements of **M**.

- Allele frequencies 0.321, 0.179, 0.357, 0.357, 0.143, 0.607, 0.071, 0.964, 0.571 and 0.393 respective.
- Using those frequencies $2\Sigma p_j(1 - p_j) = 3.5383$. Thus $\alpha = 3.5383*(245/35.242) = 24.598$

Z matrix

- **Z= M – P** and is

$$\mathbf{Z} = \begin{pmatrix} 1.357 & -0.357 & 0.286 & 0.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & -0.286 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & 0.643 & 1.286 & 0.286 & 0.714 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & 1.286 & 0.286 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & 0.643 & 0.286 & 1.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & 0.643 & -0.714 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & -0.786 \end{pmatrix}$$

- We have computed **X** and **Z.**
- Remaining matrices **X'Z** and **Z'X** and **Z'Z** are computed by multiplication. Then add **I**α to **Z'Z** then MME are formed.
- When solved we these solutions:

SNP solutions

Centre for Tropical Livestock Genetics and Health

- Mean effect
- 9.944
- SNP effects (ĝ)
- 1 0.087
- 2 -0.311
- 3 0.262
- 4 -0.080
- 5 0.110
- 6 0.139
- 7 0.000
- 8 0.000
- 9 -0.061
- 10 -0.016

Direct genomic breeding values (DGV) for animals

Centre for Tropical Livestock Genetics and Health

- The DGV for the reference animals (animals 13- 20) is then computed as $\mathbf{Z}\hat{\mathbf{g}}$.

- For the validation animals (animals 21 -26) , DGV = $\mathbf{Z}_2\hat{\mathbf{g}}$

  where $\mathbf{Z}_2$ contains the centralised genotypes for the validation candidates

Direct genomic breeding values (DGV) for validation candidates

Centre for Tropical Livestock Genetics and Health

$$\begin{bmatrix} \hat{a}_{21} \\ \hat{a}_{22} \\ \hat{a}_{23} \\ \hat{a}_{24} \\ \hat{a}_{25} \\ \hat{a}_{26} \end{bmatrix} = \begin{pmatrix} 1.357 & -0.357 & -0.714 & -0.714 & -0.286 & -0.214 & 1.857 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & -0714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & 0.714 & -0.214 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & -0.357 & -0.714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & -0.929 & -1.143 & -0.786 \end{pmatrix} \begin{pmatrix} 0.087 \\ -0.311 \\ 0.262 \\ -0.080 \\ 0.110 \\ 0.139 \\ 0.000 \\ 0.000 \\ -0.061 \\ -0.016 \end{pmatrix}$$

$$= \begin{pmatrix} 0.027 \\ 0.114 \\ -0.240 \\ 0.143 \\ 0.054 \\ 0.354 \end{pmatrix}$$

Validation- Illustration

- Warning: Data set is small and this is just for illustration
- Validation animals data:

Anim	Dyd	DGV
21	7.6	0.027
22	8.8	0.114
23	9.8	-0.240
24	9.2	0.143
25	11.5	0.054
26	13.3	0.354

Accuracy = correlation (r) = 0.49 ; Regression (bias) = 5.2
(underestimation = deflation)

GBLUP

Centre for Tropical Livestock Genetics and Health

- Equivalent model to SNP-BLUP
- BLUP MME but with $\mathbf{A}^{-1}$) replaced by $\mathbf{G}^{-1}$
- The DGV is computed directly as the sum of the SNP effects(**a = Zg**)
- Model is
- **y** = **Xb** + **Wa** + **e**
- where **a** = vector of DGVs and **W** is the design matrix linking records to animals
- Matrix X is as defined before and W is an identity matrix (a diagonal matrix with all diagonal elements = 1)

GBLUP

- Given that **a = Zg**
- Then var(**a**) = **ZZ'**σ^2_g**.**

- Note that σ^2_g = $\dfrac{\boldsymbol{\sigma}^2_a}{2\sum \mathbf{p_j(1-p_j)}}$

- then the matrix **ZZ'** can be scaled such that
-
- **G** = $\dfrac{\mathbf{ZZ'}}{\mathbf{2\sum p_j(1-p_j)}}$
-
- and var(**a**) = **G**σ^2_a .
- Division by 2Σpi(1–pi) makes **G** analogous to **A.**

G from the 10 SNPs

Centre for Tropical Livestock Genetics and Health

13 1.472
14 -0.446 0.746
15 0.988 -0.930 1.634
16 0.059 -0.446 0.422 0.907
17 0.685 -0.950 1.048 0.402 1.593
18 -0.163 0.180 -0.365 -0.163 -0.102 0.746
19 -0.708 0.201 -0.627 0.423 -0.365 0.201 0.786
20 -0.547 0.079 -0.183 0.301 -0.203 0.079 0.382 0.826
21 0.887 0.100 0.120 -0.526 -0.183 0.100 -0.728 -0.567 2.280
22 -0.789 0.402 -0.708 -0.506 -0.446 -0.163 0.140 -0.264 -0.526 1.190
23 -0.203 -0.143 0.160 0.362 0.140 0.140 0.160 0.604 -0.224 -0.486 0.665
24 -0.143 0.483 -0.345 -0.708 -0.648 -0.365 -0.345 -0.183 0.120 0.705 -0.405 1.068
25 -0.829 0.362 -0.748 -0.264 -0.486 0.079 0.382 -0.022 -0.567 0.867 -0.244 0.382 0.826
26 -0.264 0.362 -0.466 -0.264 -0.486 -0.203 0.100 -0.304 -0.284 0.584 -0.526 0.382 0.261 1.109

Some properties of G

Centre for Tropical Livestock Genetics and Health

- In H-W, Linkage equilibrium
  - Average of Diag(**G**) = 1
  - Average off-diagonal(**G**) =0
  - Average genetic value of genotyped individuals =0
  - This corresponds to the definition of base population
- With average inbreeding F
  - Average of Diag(**G**) = 1+F

Some properties of G

Centre for Tropical Livestock Genetics and Health

- Matrix **G** is positive semi-definite but can be singular if
  - Two individuals can have identical genotypes
  - If limited numbers of loci are considered
  - If the total number of alleles is less than the number of individuals genotyped. The rank of **ZZ′** cannot exceed the columns in **Z** if **Z** has fewer columns than rows.
- An improved, non-singular matrix as

  $\mathbf{G}_{new} = w\mathbf{G} + (1-w)\mathbf{A}$, if numbers of markers are limited

  with w varying from 0.95 to 0.99

  or

  $\mathbf{G}_{new} = \mathbf{G} + D$, where D is a diagonal matrix with values of about 0.01

Some properties of G

Centre for Tropical Livestock Genetics and Health

- The relationship matrix **A** based on pedigree is an
- average relationship which assumes infinite loci.
- Real relationships are a bit different due to finite
- genome size
- Therefore **A** is the expectation of realized relationships
- SNPs more informative than **A.**
- Two half-sibs might have a correlation of 0.3 or 0.2

MME for GBLUP

- MME are

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}W \\ W'R^{-1}X & W'R^{-1}W + G^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ W'R^{-1}y \end{pmatrix}$$

- where α now equals σ^2_e/σ^2_a.
- Advantages:
  - Existing software for genetic evaluation can be used by replacing **A** with **G**
  - systems of equations are of the size of animals which tend to be fewer than the number of SNP.
  - In pedigreed populations **G** discriminates among sibs, and other relatives, capture information on Mendelian sampling.
  - method is attractive for populations without good pedigree as **G** will capture this information among the genotyped individuals

Solutions for the example data

Centre for Tropical Livestock Genetics and Health

- Reference Animals
- 13 0.069
- 14 0.116
- 15 0.049
- 16 0.260
- 17 -0.500
- 18 -0.359
- 19 0.146
- 20 -0.231
- Selection or validation candidates
- 21 0.028
- 22 0.115
- 23 -0.240
- 24 0.143
- 25 0.054
- 26 0.353

Computing SNP solutions from GBLUP

Centre for Tropical Livestock Genetics and Health

- $\mathbf{g} = \frac{1}{2 \sum pj(1-pj)} \mathbf{Z'G^{-1}\hat{a}}$
- For the example data for GBLUP
- $\frac{1}{2 \sum pj(1-pj)} = \frac{1}{3.5383} = 0.2826$

- From the SNP model
- We 10 in the reference and

$$\mathbf{Z} = \begin{pmatrix} 1.357 & -0.357 & 0.286 & 0.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & -0.286 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & 0.643 & 1.286 & 0.286 & 0.714 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & 1.286 & 0.286 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & 0.643 & 0.286 & 1.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & 0.643 & -0.714 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & -0.786 \end{pmatrix}$$

- And for the validation animals , we had
-

$$[\mathbf{Z}_2] = \begin{pmatrix} 1.357 & -0.357 & -0.714 & -0.714 & -0.286 & -0.214 & 1.857 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & -0714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & 0.714 & -0.214 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & -0.357 & -0.714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & -0.929 & -1.143 & -0.786 \end{pmatrix}$$

- Let $\mathbf{Z}_3 = \begin{pmatrix} \mathbf{Z} \\ \mathbf{Z}_2 \end{pmatrix}$
- $\mathbf{g} = \frac{1}{3.5383}\mathbf{Z}_3'\mathbf{G}^{-1}\hat{\mathbf{a}}$
- $\hat{\mathbf{g}} = (0.087, -0.311, 0.262, -0.080, 0.110, 0.139, 0.000, 0.001, -0.061, -0.016)$

Single Step Method (ssGBLUP)

Centre for Tropical Livestock Genetics and Health

GBLUP computes genomic breeding values only for genotyped animals.

How can non-genotyped animals benefit from genomic information

Let $\mathbf{g_2}$ be the genetic (genomic) values of genotyped animals and $\mathbf{g_1}$ the genetic values of non genotyped animals

An estimate of $\mathbf{g_1}$ based on genomic information is obtained by regression of $\mathbf{g_1}$ on $\mathbf{g_2}$ and added to information from BLUP through the usual MME

ssGBLUP

- We define variance of vector of g_1 (non-genotyped) and g_2 (genotyped)

H = Variance of $\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$

$$\mathbf{H}=\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}=\begin{bmatrix} \overbrace{\mathbf{A}_{11}+\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\left(\mathbf{G}-\mathbf{A}_{22}\right)\mathbf{A}_{22}^{-1}\mathbf{A}_{21}}^{\textit{non genotyped}} & \overbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}}^{\textit{genotyped}} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix}$$

ssGBLUP

- Surprisely, $\mathbf{H}^{-1}$ has simple form:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

- Model is just as before but uses all data (genotyped and ungenotyped):

- MME are the usual but with $\mathbf{A}^{-1}$ replaced with $\mathbf{H}^{-1}$

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1'y} \\ \mathbf{Z'y} \end{bmatrix}$$

- Advantages of ssGBLUP include
  - Easy to implement with existing genetic evaluation software
  - Account for bias to some degree due to selective genotyping
  - Propagate genomic information across the whole population
  - Useful in combining information when recording and genotyping is only on sub-sets of the population

- Disadvantage of ssGBLUP
- May need to fine tune **G** to make it compatible with $\mathbf{A}_{22}$, in the form $\mathbf{G}^* = a + b\mathbf{G}$
  - where *a* can be understood as an "overall" relationship and *b* as a change in scale (or genetic variance).
  - VanRaden(2008) suggested a regression approach
  - another approach (see Legarra et al 2014) is setting $a = 1$ and $b = \overline{(\mathbf{A_{22}} - \mathbf{G})}$, where the bar implies average across values of **G** and **A**

  (Legarra et al 2014. - Single Step, a general approach for genomic selection. Livestock Science 166 (2014) 54–65)
-

Centre for Tropical Livestock Genetics and Health

Examples of Genomic Selection in east Africa

Example of genomic selection in LMIC country : Illustration with using Tanzania data

Centre for Tropical Livestock Genetics and Health

- The African Dairy Genetic Project at ILRI funded by
- BMGF and working in several African countries

- Tanzania data to illustrate of quick wins form genomics

- Limited pedigree: 88%, 11.4% and 0.6% with no,
- one and both parents identified

- Genomic Best Linear Unbiased Prediction (GBLUP) on about 2000 animals using HD SNP chip

- Use genomic relationship among animals (**G**) computed using SNP data

Accuracies from our prediction models for Tanzania

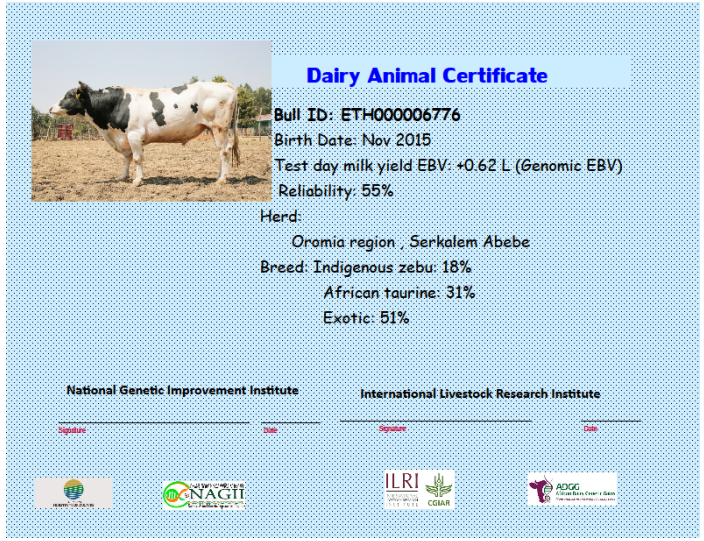
Milk yield, h2= 0.12 , body weight, h2 = 0.22 and rg = 0.34

Trait	Method	Correlation	Regression
Milk yield	FRM-GBLUP	0.57	1.1
	FRM-ssGBLUP	0.59	1.0
	RRM-GBLUP	0.55	1.0
	RRM-ssGBLUP	0.53	0.92
Body weight	FRM-GBLUP	0.83	1.0
	FRM-ssGBLUP	0.77	1.1

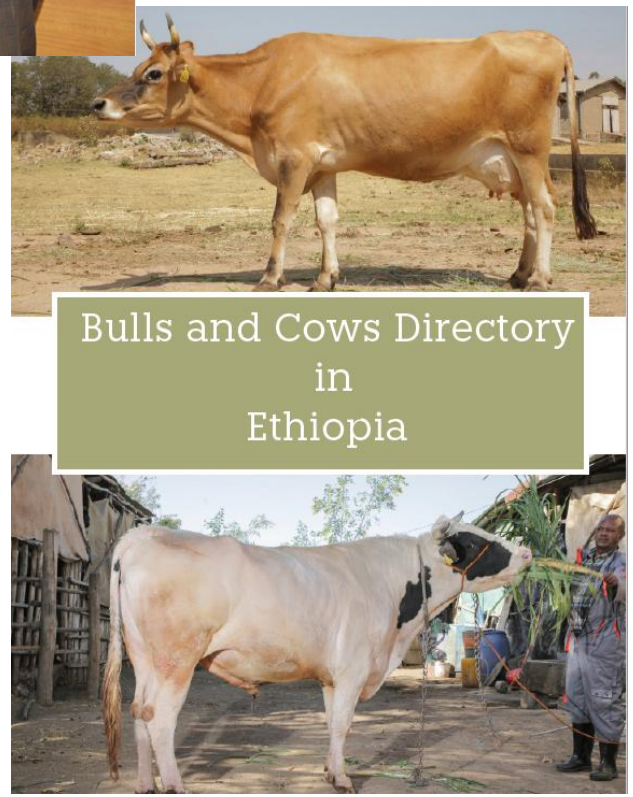
- FRM = fixed regression model
- RRM - random regression model

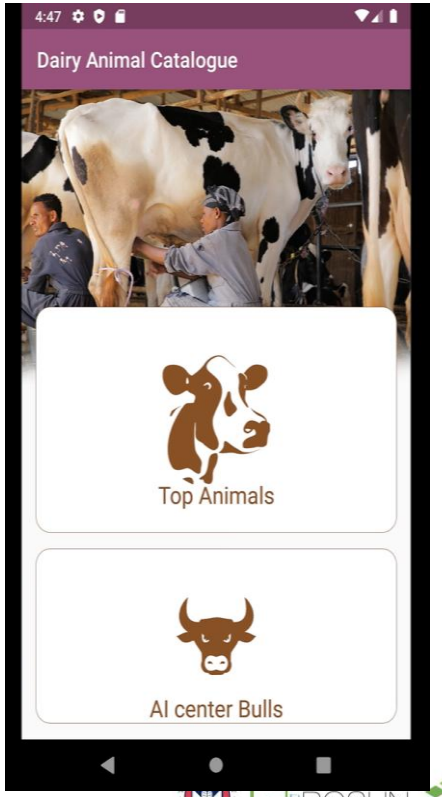
ADGG Achievements so far …

- Ethiopia First dairy animal parade held on Tuesday March 30, 2020, Fikiru Regessa, State Minister of Agriculture (extreme left), Selam Meseret ADGG Ethiopia National Coordinator (middle), and Asrat Tera, Director General of National Animal Genetics Institute (NAGII) Ethiopia.



- ILRI director general, Jimmy Smith (left) and Minister for Livestock and Fisheries Luhaga Mpina (2nd to left) present the award for best bull at a special bull and cow show at the Nane Nane exhibition center in Dodoma, Tanzania, June 2019. Photo ILR





Centre for Tropical Livestock Genetics and Health

CTLGH Funders

BILL & MELINDA GATES foundation

UKaid from the British people

UKRI Biotechnology and Biological Sciences Research Council

The University of Edinburgh · Roslin · SRUC · ILRI International Livestock Research Institute · 50 2024 · CGIAR