

Genomic selection with weighted GBLUP and APY single step

Raphael Mrode (ILRI)

*Third ACGG Program Management Team Meeting
Abuja, Nigeria, 28-30 November 2016*



Rationale for the topic

- Various ecotypes of chicken are being tested with expected differences in performance in different agro-ecological zones
 - Need to understand the genetic basis of such differences in performance
 - This may involve GWAS or other candidate gene approach
 - Marker Assisted selection or other approaches may be employed to use such information in selective breeding
 - This is usually faced with a lot of practical problem
 - An alternative approach could be the use of weighted GBLUP which is readily easily to implement

Rationale for the topic

- Data will be collected at various levels of the project
 - Same ecotypes in several countries
 - In stations within countries with opportunity for more detailed phenotyping
 - At farm level within countries
 - Only a proportion of these will be genotyped
 - Single Step offers us the framework to account for all various data structure within and across countries using phenotypes, pedigree and genotypes in a single analysis
 - APY offers us the framework to reduce the dimensionality of the analysis

Weighted GBLUP

- Genomic selection with GBLUP involves constructing a genomic selection relationship \mathbf{G} among animals
- \mathbf{G} is plugged into the normal MME equations to get estimates of direct genomic breeding values

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

Weighted GBLUP

- The major assumptions of G and GBLUP is that all markers contribute equally to the genetic variance of a trait.
- This assumption is violated for traits that are controlled by a small number of quantitative trait loci (QTL) or individual QTL with large effects.
- Larger gains can be expected for traits with few QTL if we modify GBLUP to account for this
- Weighted GBLUP (wGBLUP) involves weighting SNPs or regions of the genome based on prior information on SNP effects

Possible weights in GBLUP

- Weights for \mathbf{G} - based on SNP effects
 - Single –marker regression effects ($P < 0.05$)
 - Use Log10 P values from GWAS
 - Estimates from Bayesian approaches
- Tiezzi and Maltecca (2015 –GSE 47:27) used several approaches but will focus on results from Single –marker regression effects
- SNPs were weighted by the expected variance of SNPs to compute $\mathbf{G}\mathbf{w}$:

$$w_i = 2P_i (1 - P_i)S_i \text{ with } S_i = 1 \text{ if SNP is significant or zero}$$

Accuracies for Milk Yield and Fat percent

Milk yield --- very polygenic traits

Fat percent – QTL – DGAT in chromosome 14

	Milk yield	Fat percent
A matrix	0.404	0.350
G matrix	0.560	0.548
Gw Matrix	0.560	0.580

Impact of WGLUP depends on heritability of trait and where few QTL accounted for up to 25% the genetic variance

Genomic prediction with Single –Step

- ssBLUP similar to GBLUP but with **G** replaced with matrix **H**
- **H** is formed by incorporating both pedigree-derived and genomic-derived relationships in to single matrix

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

Advantages and disadvantages

- Advantages 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

Advantages and disadvantages

- Can be used to analyze ACGG data pooled across stations, farms and countries with genotypic information incorporated
- Disadvantage
 - Obtaining the inverse of H with large population. A direct inverse has a cubic cost and can be computed efficiently for perhaps up to 150,000 individuals.
- APY constituent one of the ways to overcome this obstacle

Algorithm for Proven and young animals

- It involves
 - Partition of genotyped animals to core (c) and non-core animals (n)
 - Core animals could only be sires; sires and cows
 - Better properties when animals treated as core are well related to animals treated as non-core
 - Computing \mathbf{G}^{-1} needed for \mathbf{H}^{-1} that only involves the inverse of core animals

Algorithm for Proven and young animals

The inverse needed for all genotyped animals is:

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_{nn}^{-1} \begin{bmatrix} -\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$

Note that it involves only the inverse of G for core animals

Correlations of GEBVs between ssGBLUP and APY-ssGBLUP

Definition of sub-set	Animals in sub-set	Correlations	Rounds top convergence
All	100,000	1.00	567
Sires	23174	0.994	432
Sires + cows	50389	0.995	428
Cows	27215	0.992	797
Sires > 5 progeny	16434	0.992	415

Fragomeni et al 2015 (JDS : 98:4090)

Theory of APY

- Theory for the APY assumes that the optimal size of the subset (maximizing accuracy of genomic predictions) is due to a limited dimensionality of the \mathbf{G}
- With large numbers of SNP markers and genotyped animals, the dimensionality of genomic information defined by the eigenvalue of \mathbf{G} which is approximately a linear function of effective population size.
- Consequently, an inverse of the \mathbf{G} based on limited recursion can be computed inexpensively for a large number of individuals.

Conclusion

- Weighted GBLUP offers a quick route to put more selective pressure on identified genomic regions with more effect on traits of interest
- APY-ssGBLUP could be a framework to implement genomic prediction for ACGG data across stations, farms and countries

Acknowledgements

This work is financed by the Bill & Melinda Gates Foundation

It contributes to the CGIAR Research Program on Livestock and Fish

better lives through livestock

ilri.org

ILRI thanks all donors and organizations who globally supported its work through their contributions to the **CGIAR system**

Patron: Professor Peter C Doherty AC, FAA, FRS

Animal scientist, Nobel Prize Laureate for Physiology or Medicine–1996

Box 30709, Nairobi 00100 Kenya
Phone +254 20 422 3000
Fax +254 20 422 3001
Email ilri-kenya@cgiar.org

ilri.org
better lives through livestock
ILRI is a CGIAR research centre

Box 5689, Addis Ababa, Ethiopia
Phone +251 11 617 2000
Fax +251 11 667 6923
Email ilri-ethiopia@cgiar.org

ILRI has offices in East Africa • South Asia • Southeast and East Asia • Southern Africa • West Africa



This presentation is licensed for use under the Creative Commons Attribution 4.0 International Licence.