Feasibility of joint genomic evaluations for smallholder dairy data in Tanzania and Ethiopia

R. Mrode[1,2], J. Ojango[1], C. Ekine-Dzivenu[1], D. Komwihangilo[3], A. Tera[4] and M. Okeyo[1]

[1]*International Livestock Research Institute (ILRI), PO Box 30709, Nairobi, Kenya*
[2]*Scotland's Rural College, Easter Bush, Midlothian EH25 9RG, Scotland, UK*
[3]*Tanzania Livestock Research Institute, P. O. Box 834, Dodoma, Tanzania*
[4]*Livestock Development Institute, P.O. Box 22692, Addis Ababa, Ethiopia*

Abstract

One of the approaches to increasing the accuracy of genomic predictions when the reference population size is limited is by joint across country evaluations. Such an approach is straightforward only when one breed is involved. However, in small holder dairy systems in East Africa, the cows are mostly crossbreds and breeds and breed types used in the crossing and the breed composition of cows tend to be different among countries. The feasibility and benefits of joint across country evaluations have not been demonstrated for such systems. This paper therefore examines the feasibility of across joint country genomic evaluations for milk yield in Tanzania and Ethiopia. PCA analysis indicated that the first principal component accounted for about 50 percent of the variation between bulls and cows genotyped in both countries. The genetic correlation between both countries from a bivariate analysis was rather low at 0.12±0.10. Marginal increases in the accuracy for genomic prediction of about 2.5% were observed for Ethiopia from a joint analysis of all data from both countries or only including a sub-set of 500 related from Tanzania. The increase was only 1.7% for Tanzania when a sub-set of 500 related cows were included from Ethiopia. The current data structure indicates no advantage from joint analysis of data from both countries, but the study does highlight the need to deliberately exchange top ranking bulls between both countries.

Key words: across country evaluations, small holder dairy cattle, genomic prediction, accuracy

Introduction

One of the approaches to increasing the accuracy of genomic predictions when the reference population size is limited is by joint across country evaluations such as the Intergenomics or inclusion of foreign related bulls in national evaluations. (Li et. al., 2015) Such an approach is straightforward when only one breed is involved, and MACE proofs are available for bulls in the countries concerned. However, in small holder dairy systems in East Africa, the cows are crossbreds with several breeds and breed types involved in the cross breeding and the breed composition of cows tend to be different among countries. In addition, the degree of genetic connectedness among countries is uncertain as breeding

policies have been driven by government and the choice of bulls for breeding highly dependent on what the AI companies, most of which are currently promote sires that are evaluated outside Africa, offer to farmers. Moreover, in the data used here, there is complete lack of pedigree information in Tanzania and some of the foreign bulls used in Ethiopia are assigned new numbers, therefore genetic links can only be determined between the two countries using marker information. One of the objectives of the African dairy Genetics Gains program (ADGG; (https://africadgg.wordpress.com/category/adgg/)) involves exploring the feasibility of across country evaluations among the countries involved in the project. This paper therefore explores the feasibility of joint across country

genomic evaluations for milk yield in Tanzania and Ethiopia by examining the degree of connectedness between both countries and the impact of such joint evaluation on accuracies of genomic predictions.

Materials and Methods

Genotypes

Genotypic data for this study was generated from about 5600 cows and bulls genotyped in each country with the GeneSeek Genomic Profiler (GGP) Bovine 50K chip. About 47843 SNPs were returned from the laboratory and after the usual edits, 40581 SNPs were available for analysis. These were imputed to the Illumina HD chip using a reference population consisting of crossbreds from a previous Dairy Genetics East Africa project and several European Holstein-Friesian, Jersey, Guernsey, and Ayrshire purebred animals (Aliloo et al, 2018).

Phenotypic data

Phenotypic data consisted of 1916 cows from Tanzania with 19599 daily milk test date (MTD) yields from 456 herds. The corresponding data set from Ethiopia consisted of 1642 cows with 16530 MTD from 772 herds. The MTD were collected by data recording agents visiting each farm monthly from November 2016 until May, 2019. MTD from up to four parities were represented in the data but some of the initial cows recruited in the project lacked data in their early parities. These cows were crosses between indigenous African breed which are ancient admixtures of African Bos taurus and Bos indicus (N'dama and Nellore) and 5 exotic dairy breeds (Ayrshire, Friesian, Holstein, Guernsey, and Jersey). The percentage of indigenous and exotic genes for cows were estimated by an admixture analysis (Ojango et al, 2014). The exotic dairy percentage of each cow was computed as the total proportion of the estimated percentage contributions of each of the 5 exotic dairy breeds. Four breed classes of animals were then created based on percentage exotic genes: cows with $> 87.5\%$, $61-87.5\%$, $36-60\%$, and $< 36\%$ exotic genes.

Models

Within Country Analyses

Initially, a within country GBLUP fixed regression model was implemented comprising ward, age nested parity, test-year-season, curves with Legendre polynomials nested within breed classes by parity as fixed effects plus the random effects of herd, animal and permanent environment. Heribilities of 0.11 ± 0.03 for Tanzania and of 0.05 ± 0.02 for Ethiopia, estimated from the within country analyses using ASREML (Gilmour, et al, 2009) were used in all within country subsequent genomic prediction models.

Validation data set comprised of 276 and 349 cows born after 2014 in Tanzania and Ethiopia respectively whose records were omitted from the analyses. The accuracy of prediction was computed as correlation between yield deviation (YD) of validation cows and their GEBV, where YD was MTD adjusted for all fixed and random effects apart from the random animal effect.

Analyses of combined data

Initially, the connectedness between both countries were examined using PCA analyses. This involved 3558 cows and 745 bulls with validated genotypes in both countries. The PCA analysis was done separately for cows and bulls. Then the genetic correlation for MTD between both countries was examined using a bivariate analysis with the YD from the within country analysis used as the response variable. The model for the bivariate analysis included country as fixed effect and random animal effects with the following assumptions:

$$\text{Var}(a) = \mathbf{G} \otimes \begin{pmatrix} \sigma_{a11}^2 & \sigma_{12}^2 \\ \sigma_{a21}^2 & \sigma_{22}^2 \end{pmatrix}$$

and

$$\text{Var}(e) = \mathbf{R} = \begin{pmatrix} r_{11} & 0 \\ 0 & r_{22} \end{pmatrix}$$

where Var(a) is the variance for genetic effects, $\mathbf{G}$ is the genomic relationship matrix

and σ^2_{ii} is the genetic variance for the ith country and σ^2_{ij} is the genetic covariance between country i and j; Var(e) is the residual variance with r_{ii} being the residual variance for i^{th} country.

In addition to the bivariate analysis, several other combined analyses of data from both countries were also undertaken. Joint analysis of both data sets combining all YDs and genotypes together which is equivalent to assuming a genetic correlation of unity. Moreover, several sub-setting strategies were also studied such as combining only the data and genotypes of the top 100 to 500 Tanzanian animals most related to the Ethiopia data with the Ethiopia data and vice versa, where the degree of relatedness computed was from the **G** matrix.   The validation data set for all combined analysis was as defined in the within country analysis.

Results & Discussion

The proportion of variance explained by the first and second principal components are shown in Figures 1 and 2 for cows and bulls respectively. The first principal component accounted for about 54% of the total variance for cows but lower for bulls at 45%. This indicates limited degree of genetic connectedness between both populations. This was confirmed by the very low genetic correlation estimate of 0.12±0.10 from the bivariate analysis.

The validation results from the within country analyses and various combined analyses are shown in Table 1. For Tanzania, the within country accuracy of genomic predictions was 0.58 but there was no increase when all the combined data was used. Li et al, (2015) similarly reported no increase in the reliability of GEBV for protein when genotyped French and Nordic Holstein bulls were included into genomic predictions in Brazil but reported a 2% and 45% increase for milk and fat respectively.

The inclusion of a sub-set of 250 or 500 related animals from Ethiopia in the Tanzania analysis only slightly increased the accuracy to 0.59. However, it resulted in estimates of regression coefficients which were closer to unity, thereby reducing the slight underprediction in the within country analysis.

Accuracy of the within country genomic prediction was lower for Ethiopia compared to Tanzania at 0.39, given the smaller data set and lower heritability. The combined analysis of data or including a sub-set of 500 related animals from Tanzania in the Ethiopia analysis only marginally increased the accuracy of genomic prediction by 2.5%, resulting in regression estimates which were closer to unity, thereby reducing the overprediction observed in the within country analysis. Similar marginal increases in reliability for cows were reported when foreign information from genotyped French and Nordic Holstein bulls were incorporated in genomic prediction in Brazil (Li et al, 2015).

Table 1. Validation results for milk yield from within country analysis for Tanzania and Ethiopia and analysis of combined data

	Tanzania		Ethiopia	
	Reg.	Corr.	Reg.	Corr.
Within Country Analysis	1.10	0.58	0.66	0.39
All data combined	1.17	0.58	0.67	0.40
Sub-setting				
Within country plus 250 cows from the other country	1.01	0.59	0.70	0.39
Within country plus 500 cows from the other country	1.01	0.59	0.70	0.40

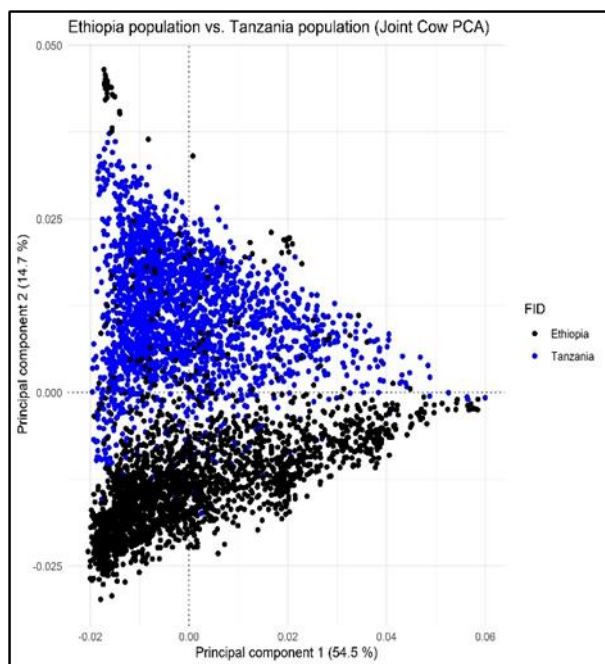Reg. = regressions; Corr. = correlations

Figure 1. PCA analysis of 3558 genotypes of cows from Tanzania and Ethiopia
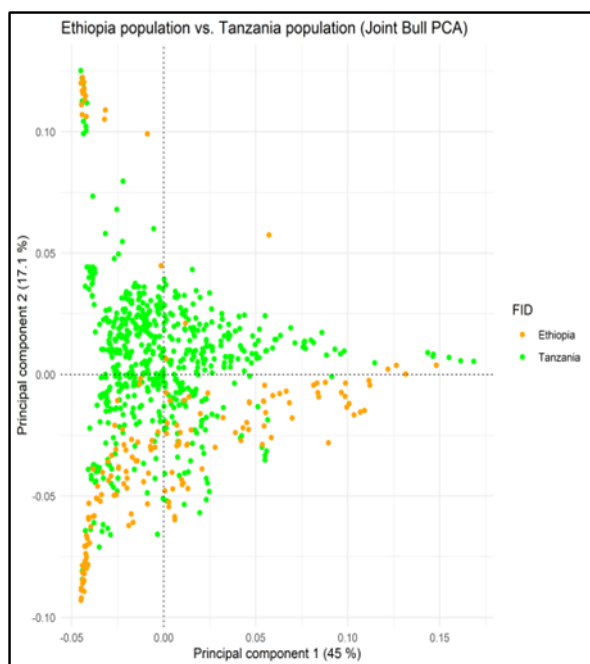

Figure 2. PCA analysis of 745 genotypes of bulls from Tanzania and Ethiopia

Conclusions

The little or no increase in accuracy of genomic prediction from the across country genomic prediction indicate no advantage from the analysis compared to the within country analysis given the current data structure. However, the study is useful in that the results will inform ADGG's subsequent across country evaluation designs, including, the critical next steps which may include exchange, between countries, of semen from top ranking bulls from ADGG within country evaluations. In addition, strategically expanding ADGG activities in Kenya which has historically sold many animals to other East Africa countries could improve the prospects of joint country analysis in future.

References

Aliloo H., Mrode R., Okeyo A.M., Ni G., Goddard M.E., Gibson J.P. 2018. The feasibility of using low density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. Journal Dairy Science 101,9108-9127.

Li X, Lund M.S, Zhang Q, Costa C.N, Ducrocq V, Su G. 2015. Short communication: Improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations. J. Dairy Sci. 99:4574–4579

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0.

Ojango, J. M. K., Marete A., Mujibi D., Rao J., Pool J., Rege J.E.O., Gondro C., W. M. S. P. Weerasinghe W.M.S.P., Gibson J.P, Okeyo A.M. 2014. A novel use of high density SNP assays to optimize choice of different crossbred dairy cattle genotypes in small-holder systems in East Africa. Pages 2–4 in Proceeding. 10th World Congress Genetics Applied to Livestock. Production.