Genomic prediction of crossbred dairy cattle in Tanzania: A route to productivity gains in smallholder dairy systems

Mrode, R.; Ojango, J.; Ekine-Dzivenu, C.; Aliloo, H.; Gibson, J.; Okeyo, M.A.

Published in:
Journal of Dairy Science

DOI:
10.3168/jds.2020-20052

First published: 04/08/2021

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genomic prediction of crossbred dairy cattle in Tanzania: A route to productivity gains in smallholder dairy systems

R. Mrode,1,2* J. Ojango,1 C. Ekine-Dzivenu,1 H. Alllou,3 J. Gibson,3 and M. A. Okeyo1
1International Livestock Research Institute, Box 30709-01001 Nairobi, Kenya
2Scotland’s Rural College, Easter Bush, Midlothian, EH25 9RG, United Kingdom
3University of New England, Armidale 2350, Australia

ABSTRACT

Selection based on genomic predictions has become the method of choice for genetic improvement in dairy cattle. This offers huge opportunity for developing countries with little or no pedigree data, and preliminary studies have shown promising results. The African Dairy Genetic Gains (ADGG) project initiated a digital system of dairy performance data collection, accompanied by genotyping in Tanzania in 2016. Currently, ADGG has the largest body of dairy performance data generated in East Africa from a smallholder dairy system setting. This paper examines several models for the estimation of genetic parameters of daily milk yield and body weight.

Key words: smallholder dairy cattle, genomic selection, crossbreeds, body weight, milk yield

INTRODUCTION

Genomic selection has become the method of choice for improvement of dairy cattle in developed countries due to the accelerated rate of genetic progress, mostly because of reduction in generation interval and increased prediction accuracy. García-Ruiz et al. (2016) reported that genetic gains for milk, fat, and protein yields for registered cows in the United States were 50, 2.2, and 1.6 kg/yr before genomic selection was adopted but increased by more than 2-fold to 109, 6.0, and 4.1 kg/yr after genomic selection was introduced. In developing countries, with limited or no pedigree data, the use of genomic information offers a huge opportunity for prediction of the genetic merit of animals using the G matrix. In East Africa, preliminary application of genomic prediction for the estimation of genetic parameters for milk yield have been reported based on Dairy Genetics East Africa (DGEA) data, using various models such as fixed repeatability and random regression genomic best linear unbiased prediction (GBLUP; Brown et al., 2016; Ojango et al., 2019). The data were rather limited, with low validation accuracies reported (0.28–0.41).

In general, the calibration slopes for genomic prediction from small cattle populations are widely varied, either less than unity (0.29–0.97) or much higher (1.16–1.24; Mrode et al., 2019), indicating huge underpredictions for the study, with a regression coefficient of 0.29. However, the African Dairy Genetic Gains project (ADGG; https://www.ilri.org/research/projects/african-dairy-genetic-gains), which began initiating a digital system of dairy performance data collection in Tanzania in 2016, accompanied by genotyping, is currently the largest body of data generated in East Africa from a smallholder dairy system setting. This paper examines several models for the estimation of genetic parameters of daily milk yield and body weight,
and accuracy of genomic prediction using GBLUP and single-step (ss)GBLUP in Tanzania.

**MATERIALS AND METHODS**

**Genotypes**

Genotypic data for this study were from 5,268 cows and bulls genotyped with the GeneSeek Genomic Profiler Bovine 50K chip. About 47,843 SNP were returned from the laboratory, and, after the usual edits, 40,581 SNP were available for analysis. These were imputed to the Illumina HD chip using a reference population consisting of crossbred cattle from a previous East Africa Dairy Genetics Gain project and several European Holstein-Friesian, Jersey, Guernsey, and Ayrshire purebred animals (Aliloo et al., 2018). The sampled cows were crosses between indigenous East African breeds, most of the latter being ancient admixtures of African *Bos taurus* and *Bos indicus* (Kim et al., 2020), represented by N’dama and Small Eastern Africa Zebu, respectively, and 5 exotic *Bos taurus* dairy breeds. The percentages of indigenous and exotic genes for each cow were estimated by an admixture analysis (Aliloo et al., 2018). The exotic dairy percentage in each cow was then computed as the sum of the estimated percentage contributions of each of the 5 exotic dairy breeds. Four classes of animals were then created, based on the percentage of exotic genes, as follows: cows with >87.5%, 61–87.5%, 36–60%, and <36% exotic genes. These will account for effects of breed composition in all subsequent analyses.

**Performance Data**

Performance data consisting of test day milk yield and body weight were extracted from the International Livestock Research Institute database (https://www.adgg.ilri.org/uat/auth/auth/login) for cows registered in the ADGG project from November 2016 until May 2020 in Tanzania. The data consisted of 134,987 test day records for 14,741 cows from 8,735 farms or herds, with pedigree information comprising 63,889 records. The herds were nested within 673 villages, within 263 wards, within 34 districts, and within 12 regions of Tanzania. Wards in the study area are administrative units comprising several villages. The data were subjected to several edits, including the identification of either sire or dam; daily milk yields values being in the range of >1 kg/d to ≤45 kg/d; age at first calving of at least 18 mo; and DIM restricted to between 4 and 500 d. The upper limit was based on a preliminary analysis that showed a reduction in heritability for milk yield when records with DIM greater than 500 were included. Also, a lower limit of 305 d, for instance, does not reflect the real production pattern in the population due to tendency of farmers to milk cow longer for various reasons: to continue to provide a source of income, inability to buy a replacement cow on time, or failure of cow to become pregnant, due to fertility issues. After edits, 3,473 cows with 36,376 test day milk records were available for analysis, with lack of either sire or dam and genotypes accounting for 73% of lost records. No requirement for the presence of either parent was imposed in the editing procedure for cows that had genotypes. Body weight was captured at the time of milk recording using a heart girth tape, and these were converted to body weight using the formulas published for smallholder systems by Lukuyu et al. (2016).

For the purposes of parameter estimation, 2 data files were created: Data1, consisting of 1,906 cows and 19,015 test day records with genotypes and minimum of 3 test days in first recorded lactation, and Data2, which was similar to Data1 but included cows with at least 1 parent known, resulting in a total of 2,716 cows with 25,474 test day records and a pedigree with 6,550 animals.

A fixed regression model (FRM) was implemented for both data sets for daily milk yield and body weight using ASRemL (Gilmour et al., 2009). The FRM fitted for each trait, in matrix form, was as follows:

\[
\begin{align*}
\mathbf{y} &= \mathbf{X}\mathbf{\beta_1} + \Phi\mathbf{\beta_2} + \mathbf{Z_1hd} + \mathbf{Z_2a} + \mathbf{Z_3pe} + \mathbf{e},
\end{align*}
\]

where \(\mathbf{y}\) is the vector of observations (milk yield or body weight); \(\beta_1\) is the vector of fixed effects, consisting of wards, calving year-season, test year-month, the 4 classes for the proportion of exotic genes, and age effects nested within lactation class; \(\beta_2\) are vectors of parameters for the fixed lactation curves modeled by Legendre polynomials of order 2, which were nested within lactation class; \(\mathbf{hd}, \mathbf{a}, \text{ and } \mathbf{pe}\) represent the random effects of herd, animal, and permanent environmental effects, respectively. The design matrices \(\mathbf{X}, \mathbf{Z_1}, \mathbf{Z_2}\), and \(\mathbf{Z_3}\) related observations to fixed effects, random herd, pe, and animal effects, respectively; and \(\Phi\) was the matrix of Legendre polynomials of order 2 to model fixed lactation curves. The choice of Legendre polynomial of order 2 was based on a preliminary analysis that examined several orders (1 to 5) to discover which gave the best fit using the Akaike and Schwarz’s Bayesian information criteria (Wolfinger, 1993). Although 5 parities were represented in the data, parity effects were fitted as only 3 classes: parities 1 and 2 separately, and parities 3 to 5 pooled into the third class. Broadly 2 season subclasses (dry or wet) were created per year,
with January to June regarded as the dry seasons, but with some variations by regions. For Data1, model [1] was referred to as FRM-1, and the assumptions for the random effects were as follows: \(\text{var(herd)} = \sigma^2_{\text{ld}}\); \(\text{var(pe)} = \sigma^2_{\text{pe}}\); and \(\text{var(a)} = G\sigma^2\), where \(G\) was the genomic relationship matrix. The matrix \(G\) was computed using method 1 of VanRaden (2008), as

\[
G = MM^{-1}2\sum_{j}^n [p_j \{1 - p_j\}],
\]

where \(M\) is the matrix of SNP genotypes with elements centralized as \(z_{ij} - 2p_j\), where \(p_j\) is the allele frequency of the minor allele of the \(j\)th SNP. For Data2, model [1] was also fitted, and it is referred to as FRM-2, but with \(\text{var(a)} = H\sigma^2\), where the \(H\) matrix was computed from the \(G\) matrix and the pedigree of animals (Misztal et al., 2009). The inverse of \(H\) was computed as follows:

\[
H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \begin{pmatrix} G^{-1} + A_{22}^{-1} \end{pmatrix} \end{pmatrix}.
\]

In FRM-2, 6 classes of breed composition were fitted instead of 4 classes as in FRM-1. The 2 additional classes were created for the 810 cows in Data2 that had no genotypes but had at least 1 parent identified. The lack of genotypes implies that breed compositions computed from the admixture analysis were not available for these animals; therefore, the breed types assigned by farmers were used. The breed types assigned by farmers indicated that 635 of these cows were from crosses of indigenous cattle with exotic breeds such as Ayrshires, Holstein-Friesians, Jerseys, and Guernseys, and these were assigned to one group, whereas 175 were termed Small Eastern African Zebu, Maasai Zebu, or unknown, and were assigned to a second additional group.

In the various analyses detailed previously, herd effect was fitted as a random effect to accommodate the small sizes of herds (Powell et al., 2018). However, Ojango et al. (2019) demonstrated that the effects of herds are modeled either as fixed or random effects on estimates of heritability. To briefly investigate this influence, the FRM-1 model was implemented with herd effects fitted as a fixed effect nested within wards rather than a random effect.

In general, test day models for milk yield have been associated with higher estimates of heritabilities, due to the better correction of fixed and environmental factors relevant for each test day record and accounting for the shape of the lactation curve at the phenotypic and genetic levels. In an attempt to investigate the possible benefits of a test day model for milk yield in Tanzanian data, Data1 and Data2 were analyzed with a random regression model (RRM), which, in matrix form, was as follows:

\[
y = X\beta_1 + \Phi_2 + \text{hd} + \Phi_1a + \Phi_2pe + e, \tag{2}
\]

where vectors of solutions \(\beta_1, \beta_2, \text{and hd}\) and design matrices \(X\) and \(\Phi\) were as defined for model [1]; \(a\) and \(pe\) are vectors of animal and permanent environmental effects modeled with Legendre polynomials of orders 2 \((\Phi_1)\) and 1 \((\Phi_2)\), respectively. Attempts to fit Legendre polynomials of order 2 for both random effects did not converge, probably due to the limited data size. However, fitting Legendre polynomials of order 1 for both random effects gave a poorer fit based on Akaike or Schwarz’s Bayesian information criteria. For Data1, the RRM was termed RRM-1, and the assumptions were as follows: \(\text{var(pe)} = P = I\otimes P\), where \(P\) is a matrix of order 2 and \(\otimes\) is the Kronecker product; and \(\text{var(a)} = G\otimes D\), where \(D\) is a matrix of genetic covariance between random regression coefficients and a constant residual variance \(\{\sigma^2_e\}\) was assumed for all DIM. In the case of Data2, the only difference in terms of the assumptions was that \(\text{var(a)} = H\otimes M\). In all the RRM models, daily estimates of genetic variance \((va_i)\) and permanent environmental variance \((vpe_i)\) for day \(i\) were computed as \(va_i = \varphi_1M_1\varphi_1'\) and \(vpe_i = \varphi_2P_2\varphi_2'\), respectively, where \(\varphi_\beta\) is the vector of Legendre polynomials corresponding to the \(i\)th day for the \(j\)th variance component. Daily estimates of heritability for the \(i\)th day were then obtained as \(va_i/va_i + vpe_i + \sigma^2_{ld} + \sigma^2_e\).

Genomic breeding values \((gEBV_{ik})\) for day \(i\) for the kth animal was computed as \(gEBV_{ik} = \varphi_1a_k\), with \(a_k\) = vector of random regression coefficients for the kth animal.

The models for the analyses of Data1 (FRM-1 or RRM-1), as previously, have involved use of the \(G\) matrix and therefore are classified as GBLUP subsequently in the paper. Also, the models for the analyses of Data2 (FRM-2 and RRM-2) have been termed ssGBLUP, as the inverse of \(H\) is used in these analyses. These ssGBLUP analyses involved the use of pedigree information; therefore, attempts were made to verify the accuracy of the pedigree information collected under a smallholder system. Using genotype data, parent-offspring relationships identified in the pedigree were verified in terms of Mendelian inconsistencies in the genotypes of offspring given the genotypes of parents.
Validation of Genomic Prediction

Two types of validation were carried out for the genomic predictions. The first validation was an attempt to answer the question, “How accurate is the genomic system for predicting the genetic merit of candidate animals of different breed proportions, as classified by the 4 classes of animals by exotic genes?” A cross-validation was implemented that entails the exclusion of records for each class of cows of certain proportions of exotic genes and then undertaking genomic prediction. For the purposes of this cross-validation, yield deviations for all cows were computed as daily milk yield or body weight, corrected for all fixed and pe effects solutions from either model [1] or model [2] using the full data (no data excluded) and then averaged per cow. The accuracy of genomic predictions for the category of cows whose records were excluded from the prediction analysis was computed as the correlation of their gEBV with their yield deviations. Also, the degree of under- or overprediction was estimated by regressing their gEBV on their yield deviations. The cross-validation was implemented only for the fixed regression models FRM-1 and FRM-2 for milk yield and body weight.

The second validation was a forward validation, attempting to address the accuracy of predicting young or future candidate animals with only genotypes but no performance records. The forward validation was implemented by excluding the milk or body weight records of 270 cows born in 2014 and afterward in the genomic prediction system. Then the accuracy of genomic prediction and degree of under- or overprediction were computed as described for the cross-validation, using the gEBV and yield deviations for these 270 cows.

RESULTS AND DISCUSSION

General Data Summary

The overall means (±SD) for daily milk yield and body weight from Data1 were 8.3 ± 4.3 L and 299.9 ± 56.4 kg. The mean for daily milk yield for cows with >87.5%, 61–87.5%, 36–60%, and <36% exotic genes were 9.6 ± 4.2, 8.2 ± 4.2, 6.3 ± 3.6, and 4.5 ± 2.6 L, respectively, indicating a trend in milk production with increasing proportion of exotic genes. A similar trend was observed for body weight. Summaries of data for some variables in Data1 fitted in the various models are presented in Table 1. Also, simple plots of daily milk yield by weeks in milk for the 3 parity classes are shown in Figure 1. The plots for body weight have not been presented, as these showed very little variation by DIM.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of cows</th>
<th>Number of test days</th>
<th>Means ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Milk yield (L)</td>
</tr>
<tr>
<td>Number of test days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,026</td>
<td>5,846</td>
<td>7.4 ± 4.1</td>
</tr>
<tr>
<td>2</td>
<td>1,217</td>
<td>6,944</td>
<td>8.3 ± 4.2</td>
</tr>
<tr>
<td>3–5</td>
<td>787</td>
<td>6,225</td>
<td>9.2 ± 4.4</td>
</tr>
<tr>
<td>Percentage of exotic genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;87.5</td>
<td>740</td>
<td>7,359</td>
<td>9.6 ± 4.2</td>
</tr>
<tr>
<td>61–87.5</td>
<td>772</td>
<td>7,956</td>
<td>8.2 ± 4.2</td>
</tr>
<tr>
<td>36–60</td>
<td>310</td>
<td>3,005</td>
<td>6.3 ± 3.6</td>
</tr>
<tr>
<td>&lt;36</td>
<td>84</td>
<td>695</td>
<td>4.5 ± 2.6</td>
</tr>
</tbody>
</table>

Figure 1. Mean milk yield by weeks in milk for parities 1, 2, and 3–5 in 1,906 genotyped cows (Data1).
As expected, mean milk yield and body weight showed an increasing trend as the number of parities increased. The trends for mean milk yield and body weight for cows confirm the expected improvement in productivity with higher levels of exotic genes, but this was more pronounced for milk yield. The mean milk yield in cows with >87.5% exotic genes was twice that for cows with less than 36%. However, for body weight, this was only 25% higher in cows with >87.5% exotic genes compared with those with the least amount of exotic genes (<0.36%).

Heritability Estimates for Milk

The estimates of heritability for Data1 and Data2 from the fixed regression models (FRM-1 and FRM-2, respectively) are presented in Table 2. The heritability estimates for daily milk from FRM-1 and FRM-2 were essentially the same at 0.12 ± 0.03, indicating that GBLUP and ssGBLUP gave similar results for this data set. The estimates of total variance due to herds followed the same pattern but the total variance due to permanent environment (PE) was slightly higher for the ssGBLUP. Working with various fixed regression models on the East Africa Dairy Genetics Gain data in Kenya, Ojango et al. (2019) reported heritabilities that ranged from 0.05 ± 0.03 to 0.27 ± 0.05, but the estimate of heritability from the closest model to that fitted in this study was 0.09 ± 0.05, which is generally similar to the estimate obtained in this study. Generally, heritability estimates for daily test day milk from a fixed regression model are lacking in literature, because most estimates are based on 305-d milk yield or a random regression. It was expected that the use of ssGBLUP, which included more information, would have yielded higher heritability compared with GBLUP in this study, but similar estimates were obtained instead. Attempts were therefore made to understand the relationship between the genotyped and ungenotyped animals in the H matrix. First, examining the 1,906 cows that had genotypes and were used to construct the G matrix, the pedigree indicated that 12, 8, 206, and 680 cows had, respectively, both parents, only sire, only dam, and no parents identified. Moreover, nonzero additive genetic relationships from the pedigree were observed only among 81 cows of all genotyped cows, with an average additive genetic relationship of 0.457 and a range of 0.25 to 0.5. This clearly underlines the importance of the availability of genotypic information in enabling prediction of the genetic merit in smallholder systems, as the pedigree relationships are clearly inadequate. Of the 801 ungenotyped cows included in the ssGBLUP analysis, 62, 43, and 705 had, respectively, both parents, only sire, and only dam identified. However, nonzero additive genetic relationships occurred only among 251 of these cows, with an average additive relationship of 0.255, with a range of 0.062 to 0.5. In addition, only 56 nonzero additive genetic relationships occurred between the 1,906 genotyped and 801 of the ungenotyped cows, with an average additive relationship of 0.337, but with additive genetic relationships ranging from 0.125 to 0.5. It appears that the limited pedigree relationships in the H matrix in general could explain the lack of any improvements in genetic parameters in ssGBLUP compared with GBLUP in this study.

Although the pedigree information was scanty, as previously indicated, the recorded pedigrees were generally accurate. Using all available genotypes and pedigree information to investigate the accuracy of the pedigree relationships indicated that, for 7 progeny who were genotyped and whose parents both were known in the pedigree and also genotyped, Mendelian inconsistency varied from 0 to 0.3%. The same trend was observed for 4 genotyped progeny with only their sires known and genotyped. However, for 290 progeny that were genotyped as well as only their dams, 85% had Mendelian

| Table 2. Genetic parameters for daily milk yield (L) and BW (kg) from various fixed and random regression models (±SE) |
|----------------|-----------------------|-----------------|-----------------|-----------------|
| Trait          | Model1                | Heritability    | Variance due to PE2 | Variance due to herd | Phenotypic variance |
| Milk yield     | FRM-1                 | 0.12 ± 0.03     | 0.10 ± 0.03        | 0.23 ± 0.02        | 9.73 ± 0.19 |
|                | FRM-2                 | 0.12 ± 0.03     | 0.12 ± 0.03        | 0.22 ± 0.02        | 9.68 ± 0.16 |
|                | RRM-1                 | 0.22           | 0.14             | 0.21             | 9.76 |
|                | RRM-2                 | 0.24           | 0.21             | 0.21             | 9.72 |
| BW             | FRM-1                 | 0.24 ± 0.04     | 0.20 ± 0.04       | 0.22 ± 0.03       | 1.2876 ± 33.2 |
|                | FRM-2                 | 0.22 ± 0.04     | 0.22 ± 0.04       | 0.26 ± 0.03       | 1.3984 ± 29.9 |

1FRM-1 and FRM-2 = fixed regression model (model [1]: genomic BLUP) on 1,906 genotyped cows (Data1) and on 1,906 genotyped plus 810 ungenotyped cows (Data2), respectively. RRM-1 and RRM-2 = random regression (model [2]: single-step genomic BLUP) on Data1 and Data2, respectively.

2PE = permanent environment.
inconsistency of 0, and 4% and 11% had Mendelian inconsistencies of about 5 to 10% and 11 to 14% respectively. This level of accuracy is encouraging and indicates the need to support and prompt more farmers to record pedigree information, to avoid reliance on genotypic data, which is rather expensive for the prediction of genetic merit of animals, especially under the less commercially competitive smallholder systems. As farmers are both encouraged and supported to record more pedigree data over time, the availability of more pedigree information will help in designing effective strategic genotyping to help achieve increases in the accuracy of genetic predictions.

The corresponding heritability estimates for daily milk yield averaged over the whole lactation from RRM (Table 2) were about twice the estimates from the fixed regression model, but estimates of total variance due to PE or herds were generally similar for both models. The estimates for daily genetic, PE variances, and heritabilities for daily milk yield are shown in Figure 2. Apart from the heritability of 0.42 at 4 DIM, the estimates varied from 0.17 to 0.33, which is similar to trends reported for smallholder Holstein-Friesian cows in Ethiopia with a similar model (Meseret et al., 2015). In another study in Ethiopia, involving Horro, Boran, Friesian, Jersey, and Simmental breeds and their crosses from 3 research centers, Gebreyohannes et al. (2016) reported increasing heritability estimates from 0.17 at 30 DIM to 0.42 at 300 DIM from an RRM using the modified, incomplete, gamma function. However, the estimates reported by Ojango et al. (2019) for DGEA Kenya crossbred dairy cattle were higher at most of stages of lactations, with minimum of 0.25 at 60 DIM and maximum of 0.41 at 150 DIM. The last 2 studies did not model PE effects with any function, and this could partly account for differences between their results and those obtained in this study.

Estimates of genetic parameters for highly admixed populations, as in the population studied, are uncommon, but the trend in estimates of genetic variances and heritabilities are similar to those reported for small population of Holstein cows in Brazil (Cobuci et al., 2005). The higher estimates of genetic variances or heritabilities at the extremes of the stage of lactation has been referred to by Schaeffer and Jamrozik (2008) as an artifact of using Legendre polynomials in RRM. They concluded that the use of other functions such as Splines have not removed this artifact and indicated that the best method for removing the artifact would be to have a genetic effect for each day in milk over the lactation period. However, such a model will not be feasible in practice, as daily milk records for the whole lactation is not available on a national scale.

Heritability Estimates for Body Weight

The estimates of genetic parameters for body weight from the various models and data sets are presented in Table 2. The heritability of 0.24 ± 0.04 for FRM-1 from GBLUP is slightly higher than the estimates based on
The modeling of herd effects as fixed effects rather than random effects in the FRM-1 model for milk yield gave a higher heritability estimate of 0.23 ± 0.08, which is 92% higher than the estimate from model [1]; a slightly lower total variance due to PE effects of 0.08 ± 0.06, but total phenotypic variance was much reduced to 7.7 ± 0.24 compared with phenotypic variance of 9.73 ± 0.19 from model [1]. In the case of body weight, the estimate of heritability from modeling of herd as a fixed effect in the model FRM-1 was 0.33 ± 0.10, which is 67% higher than the estimate from the model with random herd effect; whereas variance due to PE increased by 50% to 0.24 ± 0.09, but total phenotypic variance was lower at 1,011.5 ± 46.9, compared with a phenotypic variance of 1,287.6 ± 33.2 from model [1]. In general, the sampling variances associated with the estimates for FRM-1 with fixed herd effects for both traits were larger compared with estimates from models with random herd effects. The increased heritability estimates from models with herd modeled as fixed rather than random, are due mainly to increases in genetic variances. This could indicate the difficulty of separating animal and herd effects when herd sizes are small (i.e., 1–3 cows per herd), particularly in cases where herds have only 1 cow and both herd and animal effects are therefore fully confounded. It seems that a random herd effect provides a better approach to modeling herd effects with genetic links across herds, due to the \( G \) matrix helping separate both effects (Powell et al., 2018)

**Genetic Correlation Estimates**

The genetic and phenotypic correlations among different DIM from the RRM are presented in Table 3. As expected, estimates of genetic correlation among different stages of lactation were high between adjacent DIM but decreased as intervals increased. The phenotypic correlations were low to medium in value, ranging from 0.07 to 0.50, and tended to follow the same pattern. The genetic correlations between early (4–84 DIM) and late (324–484 DIM) stages of lactation were slightly negative, but these correlations were generally very low (−0.17 to −0.05) and could essentially be regarded as zero. Ojango et al. (2019) reported similar results for the DGEA data in Kenya, but the range of estimates was larger, varying from −0.57 to −0.05. In their study, the limited data size implied that Legendre polynomials of order 2 and 0 were fitted for the additive genetic and PE effects, respectively. However, Meseret et al. (2015) also reported the problem of negative genetic correlations for Holstein-Friesian data in Ethiopia, from a model with Legendre polynomials of order 1 and 2 for additive genetic effects and PE effects, respectively; the problem was resolved by fitting an order of 2 for both additive genetic effects and PE effects. In this study Legendre polynomials of order 2 and 1 were fitted for additive genetic effects and PE effects, respectively, and problems of convergence occurred when higher orders were fitted. This illustrates the difficulty of choosing the best model for the analysis of the limited data observed in smallholder dairy systems. Therefore, the final model fitted may be based on what is feasible given the data structure, rather than what might be considered as theoretically optimum. In general, negative genetic correlation estimates between early and later milk test day records have been commonly reported for dairy animals raised in tropical environments (Rekaya et al., 1999; Bignardi et al., 2009; Singh et al., 2016). Although limited data structure could partly explain this, it has also been attributed to the rapid decline in milk (low persistency) generally associated with Zebu cows. As more data accumulates, the influence of levels of exotic dairy genes on these genetic correlations could be examined by splitting the data based on the level of breed proportion. However, the genetic correlations obtained from early test day yield could be used to predict yield up to about 320 DIM.
model were 0.20 ± 0.01 and 0.34 ± 0.11. Most studies examining the relationship between body weight and milk yield have been reported for herds in developed countries, usually from an RRM, as explained earlier. Therefore, such results are not really comparable with the estimates obtained in this study. However, Berry et al. (2002) reported a genetic correlation of 0.32 between body weight and milk yield at 60 DIM in Holstein-Friesian cows, but negative estimates were reported between milk yield and body weight for subsequent stages of lactation. The estimates they reported between cumulative milk yield at d 240 and body weight at various stages varied between 0.07 to 0.18, which is slightly lower than the estimates in this study.

In smallholder systems, feed is a major limiting factor; therefore, an important selection goal could be efficient cows that produce as much milk as possible with less feed, without a corresponding increase in body weight. Such a goal would favor highly producing but small to medium-sized cows. Although the aim in the long term is to develop a proper index that takes into consideration economic factors, the results from this study have been used to develop a restricted selection index for Tanzania using milk yield and body weight, with the goal of improving the rate of milk production but keeping body weight constant. Selection for milk yield will result in more efficient animals in feed utilization (Li et al., 2018), and the restriction in body weight will reduce energy requirements for body maintenance and some foot and leg diseases associated with heavy body weight. However, restricting body weight while selecting for production could also produce unfavorable genetic responses on cows’ fertility and health, due to body tissue mobilization for high yield. Therefore, the long-term goal is to include fertility and health traits in the index as well, to ensure the selection of more robust cows. Currently, ADGG is piloting the collection of somatic cell counts, and data on treatments of cows for various diseases and fertility are gradually accumulating.

Validation

The results of the cross-validation for milk yield and body weight based on the percentage of exotic genes are presented in Table 4. As expected, the accuracy of genomic prediction was higher for body weight than for milk yield, because body weight has a higher heritability. The accuracy estimates for body weight varied from 0.47 to 0.71 for the GBLUP and from 0.48 to 0.66 for ssGBLUP, indicating higher accuracies from the former, most likely due to the limited additional pedigree information used in ssGBLUP. The estimates of regression coefficients were generally close to unity, indicating very good calibration for the prediction of future animals of different breed proportions. However, the regression coefficients varied from 0.71 to 0.78 for the groups of cows with the highest amounts of exotic genes, indicating some degree of overprediction for these groups. The cross-validation accuracy for daily milk yield varied from 0.32 to 0.49 for GBLUP, and estimates from ssGBLUP were generally similar. The regression coefficients were only unity for the 2 group of cows with the highest proportion of exotic genes (>87.5% and 61–87.5%), but a high degree of underprediction occurred for cows with 36 to 60% exotic genes, and slight overprediction was observed for the cows with the least amount of exotic genes, again possibly due to the limited data size given the low heritability observed in this study.

Table 3. Genetic (below diagonal) and phenotypic (above diagonal) correlations among several DIM with heritabilities (bold) on the diagonal from the random regression model (model [2]) using 1,906 genotyped cows (Data1)
The average percentage of exotic genes for the 270 cows born in 2014 and afterward, which were excluded in the forward validation data, was 0.80, indicating a high degree of exotic genes in younger cows in Tanzania. The accuracies of genomic prediction for the forward validation for daily milk yield and body weight are presented in Table 5. Again, higher accuracies were obtained for body weight compared with milk yield, with values of 0.83 and 0.77 from GBLUP (FRM-1) and ssGBLUP (FRM-2), respectively, for body weight. Corresponding values for milk yield were 0.57 and 0.59, respectively, indicating an inconsistent pattern in terms of trend of estimates from GBLUP versus ssGBLUP. However, the estimates for accuracies from both methods are generally similar. Compared with the FRM, the accuracies for milk yield from the RRM were lower, at 0.55 (GBLUP) and 0.53 (ssGBLUP), but the difference was more pronounced for the latter. These slightly lower accuracies from the RRM could be due to the limited data structure, as indicated earlier, and this will be investigated further as more data accumulate in the project. The estimates of regression coefficients from the RRM varied from 0.92 to 1.0, indicating a little or no overprediction.

The small number of genotyped animals in genomic prediction studies in developing countries makes it difficult to clearly define separate reference and validation populations. Consequently, most validations are undertaken in test data sets created by either random or structured sampling from all genotyped animals, usually using a combination of both bulls and cows (Mrode et al., 2019).

The cross-classified accuracies reported in this study for milk yield are slightly higher than the estimates of 0.32 to 0.41 reported by Brown et al. (2016) for the DGEA data in Kenya but similar to the estimates for Gyr cattle in Brazil (Boison et al., 2017). The intermediate accuracies (0.60–0.71) for body weight for cows of 0.36 to 0.875 exotic gene composition indicate that candidate cows of these breed compositions could be selected with a good degree of confidence. The accuracies of 0.32 to 0.37 for milk yield from FRM-1 for cows with 61 to >87.5% and <36% exotic genes indicate that these can be used as guides to screen candidate individuals or select a team of bulls for breeding.

The estimates of accuracy from the forward validation are similarly higher than most of the estimates summarized by Mrode et al. (2019) for several traits from genomic studies in developing countries. Although these estimates are lower than estimates from developed countries for dairy production traits (0.70–0.90) with medium to high heritability (Moser et al., 2010; Wiggans et al., 2017), the estimates of 0.57 (milk yield)

### Table 4. Cross-validation results for daily milk yield (L) and BW (kg)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Model and parameters</th>
<th>Percentage of exotic genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&gt;0.875 (n = 740)</td>
</tr>
<tr>
<td>Daily milk yield</td>
<td>FRM-1 Corr. 0.37</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>FRM-1 Reg. 1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>FRM-2 Corr. 0.39</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>FRM-2 Reg. 1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>BW</td>
<td>FRM-1 Corr. 0.47</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>FRM-1 Reg. 0.71</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>FRM-2 Corr. 0.48</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>FRM-2 Reg. 0.78</td>
<td>1.2</td>
</tr>
</tbody>
</table>

*FRM-1 and FRM-2 = fixed regression model (model [1]; genomic BLUP) on 1,906 genotyped cows (Data1) and on 1,906 genotyped plus 810 ungenotyped cows (Data2), respectively. Corr. = correlations. Reg. = regression coefficients.

### Table 5. Forward validation results for daily milk yield (L) and BW (kg)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Method</th>
<th>Correlation</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk yield</td>
<td>FRM-1</td>
<td>0.57</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>FRM-2</td>
<td>0.59</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>RRM-1</td>
<td>0.55</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>RRM-2</td>
<td>0.53</td>
<td>0.92</td>
</tr>
<tr>
<td>BW</td>
<td>FRM-1</td>
<td>0.83</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>FRM-2</td>
<td>0.77</td>
<td>1.1</td>
</tr>
</tbody>
</table>

*FRM-1 and FRM-2 = fixed regression model (model [1]; genomic BLUP) on 1,906 genotyped cows (Data1) and on 1,906 genotyped plus 810 ungenotyped cows (Data2), respectively. RRM-1 and RRM-2 = random regression (model [2]; single-step genomic BLUP) on Data1 and Data2, respectively.
to 0.83 (body weight) from this study are encouraging and are expected to improve with time and provide the opportunity for selecting candidates animals in smallholder dairy systems, where currently no genetic improvement program is in place.

**CONCLUSIONS**

From these results, we can conclude that levels of additive genetic variation and, hence, heritability for both body weight and milk yield for crossbred populations in countries such as Tanzania are high enough to enable significant gains to be achieved through selection. This study provides useful parameters that can be used to inform the design and implementation of future breeding programs, in similar smallholder crossbred-based dairy systems. The importance of initially genotyping animals at the early stages of the breeding improvement program, to provide the basis of genetic prediction given the lack of pedigree in a smallholder dairy systems, followed by systematic procedures to subsequently record pedigree information, is clearly elucidated. The moderate to high levels of accuracy obtained indicate that selection on the basis of genomic prediction is not only feasible in smallholder dairy systems but is the most realistic pathway to realizing sustained genetic gains in these systems.

**ACKNOWLEDGMENTS**

The authors thank the Bill and Melinda Gates Foundation (Seattle, WA) for funding the African Dairy Genetic Gains project (ADGG, hosted at the International Livestock Research Institute, Nairobi, Kenya). We also acknowledge contributions by the Tanzania Livestock Research Institute (TALIRI; Dodoma, Tanzania), the University of New England (Armidale, NSW, Australia), and Scotland’s Rural College, (Edinburgh, Scotland) for cofinancing. We gratefully acknowledge contributions by TALIRI colleagues, as well as contributions from the other national partners and all the smallholder farmers in Tanzania who participated in the ADGG project and made their animals available for monitoring. The authors have not stated any conflicts of interest.

**REFERENCES**


