



Optimizing the design of small-sized nucleus breeding programs for dairy cattle with minimal performance recording

C. M. Kariuki,*†¹ H. Komen,† A. K. Kahi,‡ and J. A. M. van Arendonk†

*Department of Animal Sciences, Chuka University, PO Box 109-60400, Chuka, Kenya

†Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

‡Animal Breeding and Genomics Group, Department of Animal Sciences, Egerton University, PO Box 536-20115, Egerton, Kenya

ABSTRACT

Dairy cattle breeding programs in developing countries are constrained by minimal and erratic pedigree and performance recording on cows on commercial farms. Small-sized nucleus breeding programs offer a viable alternative. Deterministic simulations using selection index theory were performed to determine the optimum design for small-sized nucleus schemes for dairy cattle. The nucleus was made up of 197 bulls and 243 cows distributed in 8 non-overlapping age classes. Each year 10 sires and 100 dams were selected to produce the next generation of male and female selection candidates. Conception rates and sex ratio were fixed at 0.90 and 0.50, respectively, translating to 45 male and 45 female candidates joining the nucleus per year. Commercial recorded dams provided information for genetic evaluation of selection candidates (bulls) in the nucleus. Five strategies were defined: nucleus records only [within-nucleus dam performance (DP)], progeny records in addition to nucleus records [progeny testing (PT)], genomic information only [genomic selection (GS)], dam performance records in addition to genomic information (GS+DP), and progeny records in addition to genomic information (GS+PT). Alternative PT, GS, GS+DP, and GS+PT schemes differed in the number of progeny per sire and size of reference population. The maximum number of progeny records per sire was 30, and the maximum size of the reference population was 5,000. Results show that GS schemes had higher responses and lower accuracies compared with other strategies, with the higher response being due to shorter generation intervals. Compared with similar sized progeny-testing schemes, genomic-selection schemes would have lower accuracies but these are offset by higher responses per year, which might provide additional incentive for farmers to participate in recording.

Key words: minimal recording, breeding program, genomic selection, genetic gain

INTRODUCTION

Genetic improvement of livestock populations relies on the ability to create a selection differential and to disseminate superior germplasm to the commercial population. Breeding programs require a well-defined recording, selection, and dissemination structure. An impasse exists in the implementation of well-organized and effective breeding programs in Kenya and other developing countries due to technical and infrastructural constraints (Kosgey et al., 2006; Marshall et al., 2011; Rege et al., 2011). As a consequence, dairy farmers in Kenya need to rely on importation of semen for creating genetic improvement on their farms. The imported semen is produced by sires resulting from breeding programs in temperate countries. However, the response in Kenya might be disappointing due to genotype by environment interaction ($G \times E$). For example, Ojango and Pollott (2002) reported a genetic correlation of 0.49 between sires performing in both the United Kingdom and Kenya, indicating a substantial $G \times E$. Importation of semen from temperate regions, therefore, might not be the economically optimal strategy for improving dairy cattle production in Kenya (Okeno et al., 2010). An alternative strategy is the establishment of a local breeding program (Vargas and van Arendonk, 2004; Mulder et al., 2006). However, this requires an investment in infrastructure and performance recording.

The effectiveness of a local breeding program can be measured by its ability to generate genetic progress through identification of superior individuals to be used as parents of the next generation. Artificial insemination has greatly increased the reproductive capacities of sires. In dairy cattle breeding, the largest proportion of response to selection is generated through the selection of sires because they have a high reproductive capacity. The high reproductive capacity of sires offers opportunities for high selection intensities and high accuracies of selection. On the other hand, the accuracy of selection is dependent on the amount of performance

Received June 30, 2014.

Accepted August 28, 2014.

¹Corresponding author: charles.kariuki@wur.nl

and pedigree information available for genetic evaluation of selection candidates. Very high accuracies in the sire pathway have been achieved with progeny testing (PT)-based breeding programs. Attainment of these high accuracies requires each candidate sire to have 100 to 250 daughters for which performances are recorded, which translates to 10,000 to 25,000 daughters annually for a program involving 100 candidates (van Arendonk, 2011). However, participation of commercial producers in recording in Kenya is minimal and erratic (Wasike et al., 2011). Consequently, efforts to establish a local national dairy cattle improvement scheme based on progeny testing has been unsuccessful. Several studies looked into the opportunities of nucleus breeding schemes, which are less dependent on large-scale recording of pedigree and performance than progeny-testing schemes. Genomic selection (Meuwissen et al., 2001) has currently caused a major change in livestock improvement programs in developed countries. Genomic selection (GS) can greatly increase the response to selection, particularly in species with long generation intervals (Schaeffer, 2006). From a genetic point of view, GS also enables the establishment of more effective nucleus breeding programs in developing livestock industries (van Arendonk, 2011). However, the suitability of GS in a genetic improvement strategy for livestock sectors in developing countries is disputed and requires further research.

Adoption of innovations in technology requires careful consideration, comparing all possibilities to minimize risks while preventing loss of opportunity for improvement. Marshall et al. (2011) describe GS as a high-risk approach compared with PT for developing livestock sectors. Research on the technical and institutional factors influencing the successful establishment of GS breeding programs for developing livestock sectors is, therefore, crucial. The first step in the implementation of a sound breeding program is to optimize the genetic evaluation process, given the current limitations in performance and pedigree recording. The objective of the current study was, therefore, to compare GS and PT evaluation strategies on response to selection and accuracy when reliable pedigree and performance recording is minimal.

MATERIALS AND METHODS

The Population Structure

The Nucleus. The general structure of the population consisted of (1) a nucleus with selection candidates, elite dams and elite sires, which are used as parents for the next generation of selection candidates; (2) a

group of commercial recorded cows (CRC) to provide information for genetic evaluation; and (3) commercial nonrecorded cows, which benefitted from the selection efforts in the nucleus. The nucleus had cow and bull populations of 243 and 197 individuals, respectively, distributed into 8 non-overlapping age classes. The annual culling rate in the nucleus was fixed at 15%. Each year 100 cows and 10 bulls within the nucleus were selected as elite dams and active sires, respectively. The sex ratio at birth and conception rate within the nucleus were fixed at 0.50 and 0.90, respectively. Therefore, 45 male and 45 female new candidates were born each year.

Selection was done annually. All newborn males and females in the nucleus automatically become selection candidates. For all strategies, considered candidates were eligible for selection only when all information required for the specified selection strategy became available. Therefore, the number of candidates attaining selection age was affected by annual culling. For instance, when all information for selection of sires was available before 1 yr of age, the number of new male candidates was 43 (after accounting for culling) but when this information became available when bulls were 6 yr old, the number of new candidates was 22. All simulated schemes considered the same nucleus population. Within the nucleus, both males and females were simulated to have a maximal lifespan of 8 yr, and selection candidates attained sexual maturity when 1 yr old. The period between mating and calving was approximately 1 yr. Therefore, cows were modeled to have their first calf toward the end of their second year or early the third year, and had their first phenotypic record before they were 4 yr old.

Interaction Between the Nucleus and the Commercial Cow Population. Figure 1 presents the interaction between the nucleus and the commercial cow population. All selection was done in the nucleus. Selected sires were used on both nucleus and commercial dams. Selected dams were used only for production of new selection candidates in the nucleus. Commercial recorded dams provided information for the evaluation of young bulls, either as cows to produce test daughters under PT schemes or as the reference population in GS schemes. We wanted to look at a situation where the number of records on performance on pedigree recorded cows is limited. The current number of unique dam performance records submitted to the national data archive in Naivasha, Kenya, The Livestock Recording Centre, ranges between 200 and 1,000 per annum (T. M. Magothe, officer in charge, personal communication). We therefore investigated a range of 500 to 5,000 commercial recorded dams per year.

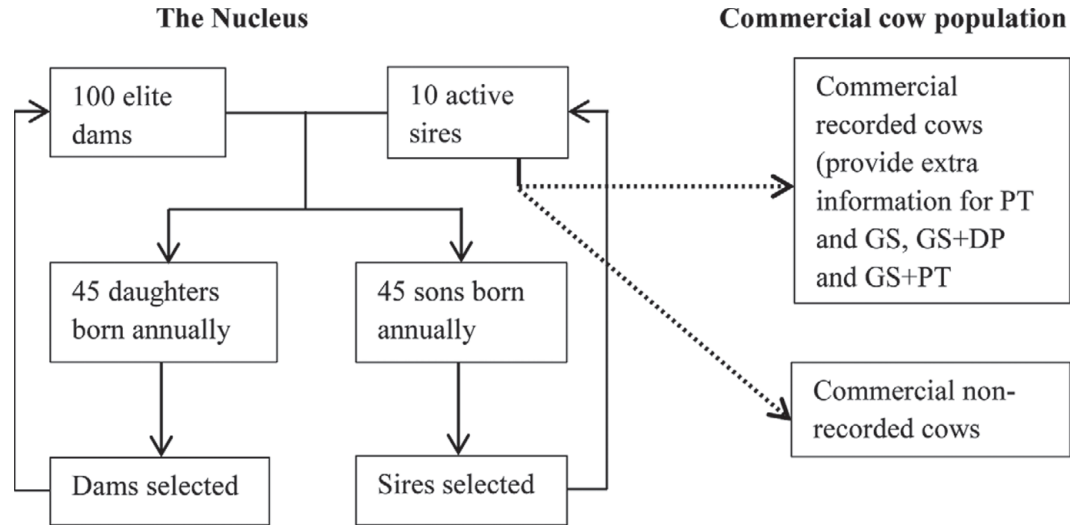


Figure 1. The interaction between the nucleus and the commercial cow population. All selection was in the nucleus and selected sires were used in the nucleus and the commercial cow population (thick dotted lines). The nucleus was closed (continuous lines). Information from the commercial recorded cows provided extra information for evaluation of selection candidates in progeny testing (PT), genomic selection (GS), genomic selection and dam performance in the nucleus (GS+DP), and genomic selection and progeny testing (GS+PT) schemes.

Alternative Breeding Schemes

Most economically important traits in dairy cattle are related to productivity of cows and, in most cases, initial records are available after the first lactation. Therefore, in this study, simulations considered a single trait recorded in dams only and records were available after the first lactation. Three levels of heritability of this single trait were used (0.15, 0.30, and 0.50) to determine the effect of heritability on response to selection and the optimum scheme.

Selection strategies were defined by the type of information used for genetic evaluation of male and female selection candidates. In all cases, we simulated an animal model genetic evaluation of the animals in the nucleus population, considering all genetic relationships. Table 1 details the sources of evaluation information for alternative selection strategies. Five strategies were defined. In the first strategy, evaluation was based on phenotypes recorded within the nucleus only [within-nucleus dam performance (DP); i.e., phenotype records for dams within the nucleus]. The second strategy was similar to DP, but with additional information for evaluation of bull candidates coming from daughter phenotypes in the commercial population [progeny testing (PT)]. In the third strategy, evaluation was exclusively on genomic information [genomic selection (GS); i.e., phenotypic information was used solely for estimation of allelic effects]. Strategy 4 was similar to GS, but nucleus cows/females had phenotypic records as additional information (GS+DP). The fifth strategy was similar to GS+DP, but bulls had daughter phenotypes

in the commercial population as additional information (GS+PT). The logic here is that the commercial dams that provide the progeny information can also be genotyped and then used as a reference population. In this case, the commercial dam information is used twice.

DP Schemes. The base (DP) schemes considered a situation where no performance recording occurred within the commercial cow population. Each year, 100 elite dams were mated to 10 active sires to produce 45 male and 45 female selection candidates. Therefore, each candidate bull had 5 half-sib sisters and each candidate dam had 4 half-sib sisters at birth. Under DP schemes, sires were selected only on pedigree information and performance records of half-sib sisters were ignored. This allowed selection of bulls when 1 yr old. Dams were selected after own performance records were available.

PT Schemes. Under PT schemes, phenotypic records were collected on progeny in the commercial population as an extra source of information for male candidates. The PT schemes, therefore, considered a situation where semen was collected from forty-five 1-yr-old bulls and used to inseminate cows in the commercial recorded group to produce test daughters. The number of cows in the commercial recorded group determined the number of test daughters that a candidate bull could produce. Two PT scenarios were simulated. In the first scenario (PT-15), schemes with 2,500 cows available annually in the commercial recorded group were simulated. Under the PT-15 schemes, young bulls were evaluated on performance records of 15 daughters in addition to the performance of relatives within

Table 1. Information sources and mean generation intervals for alternative within-nucleus dam performance (DP), progeny testing (PT), genomic selection (GS), genomic selection and within-nucleus dam performance (GS+DP), and genomic selection and progeny testing (GS+PT) schemes¹

Scheme	Information sources				Mean generation interval (yr)
	Males ²	Age when records were available (yr)	Females ²	Age when records were available (yr)	
DP schemes	Sire and dam BLUP EBV	1	Sire and dam BLUP EBV	1	4.2
			Own performance	4	
			2 HS sisters performance records	4	
			Mean BLUP EBV for dams of HS sisters	4	
PT schemes	Sire and dam BLUP EBV 3 HS sisters performance records Mean BLUP EBV for dams of HS sisters	1 4 4	Sire and dam BLUP EBV	1	6.2
			Own performance	4	
			2 HS sisters performance records	4	
			Mean BLUP EBV for dams of HS sisters	4	
GS schemes	Sire and dam BLUP genomic EBV Own genomic EBV	1 1	Sire and dam BLUP genomic EBV	1	3.0
			Own genomic EBV	1	
GS+DP schemes	Sire and dam BLUP genomic EBV Own genomic EBV 4 HS sisters genomic EBV Sire and dam BLUP EBV	1 1 1 1	Sire and dam BLUP genomic EBV	1	4.2
			Own genomic EBV	1	
			3 HS sisters genomic EBV	1	
			Sire and dam EBV	1	
			2 HS sisters performance records	4	
GS+PT schemes	Sire and dam BLUP genomic EBV Own genomic EBV 4 HS sisters genomic EBV Sire and dam BLUP EBV 3 HS sisters performance records Mean BLUP EBV for dams of HS sisters Daughter performance records	1 1 1 1 4 4 6	Sire and dam BLUP genomic EBV	1	6.2
			Own genomic EBV	1	
			3 HS sisters genomic EBV	1	
			Sire and dam BLUP EBV	1	
			Own performance	4	
			2 HS sisters performance records	4	
			Mean BLUP EBV for dams of HS sisters	4	
			Mean BLUP EBV for dams of HS sisters	4	

¹Alternative PT, GS, GS+DP, and GS+PT schemes differed in the number of daughters per sire and the size of the reference population and the mean age of selected parents.

²HS = half sib.

the nucleus. The second scenario (**PT-30**) simulated schemes with 5,000 cows available annually in the commercial recorded group. This resulted in each candidate bull within PT-30 schemes having 30 daughters with first-lactation records. The number of daughters per test bull was computed by fixing the conception rate and survival rates at 80% for the commercial recorded dam population. Bulls could be selected as parents for the next generation of nucleus animals from an age of 6 yr onwards.

GS Schemes. For GS, GS+DP, and GS+PT schemes, 4 sizes of reference populations (i.e., 500, 1,000, 2,500, and 5,000) were simulated. For comparison purposes, the number of selected sires and dams within the nucleus for genomic schemes were equal to those in DP and PT schemes. For GS+PT-500, GS+PT-1,000, GS+PT-2,500, and GS+PT-5,000 schemes, the number of progeny records per candidate bull was 3, 6, 15, and 30, respectively. In GS schemes, bulls and dams were selected at 1 yr of age, whereas in GS+DP schemes, bulls were selected at 1 yr of age and dams when 4 yr old. In GS+PT schemes, bulls were selected when 6 yr old and dams when 4 yr old.

Following the procedure of Dekkers (2007), genomic information was mimicked in the selection index calculations by including a correlated trait with heritability equal to unity. The reliability of genomic information was included as genetic correlation between this additional trait and primary trait. In all schemes where genomic information was a source of information, reference populations were constructed by genotyping and phenotyping the available commercial dams. Using a reference population consisting of animals that are genotyped and phenotyped will yield higher accuracies than a population with genotypes of sires with phenotyped daughters (Van Grevenhof et al., 2012).

The phenotypic and genetic correlations were calculated as hr_{gg} and r_{gg} , respectively, where h is the square root of the heritability of the trait and r_{gg} is the accuracy of the genomic EBV. The r_{gg} is determined by the size of the reference population (n_p), the effective number of loci in the base population (n_G), and the correlation of the true breeding value of the genotyped individuals and their phenotypes (r), and was computed as follows (Dekkers, 2007; Daetwyler et al., 2008):

$$r_{gg} = \sqrt{\frac{\lambda r^2}{\lambda r^2 + 1}}, \quad [1]$$

where $\lambda = n_p/n_G$; n_G depends on the historic effective size of the base population (N_E) and the size of the genome (L) in Morgan and was computed as $n_G = 2N_E L$. The N_E was taken to be 156 (Muasya et al., 2013) and L was equal to 30. Reference populations

were constructed by genotyping and phenotyping the same animals. Therefore, $r^2 = h^2$.

Generation Intervals. In dairy cattle breeding, selected sires and dams remain in the breeding population as active sires and dams for several years before culling. Thus, selected parents are in different age classes. For such situations, it is important to account for differences in the mean performance between the age-class cohorts due to selection (Ducrocq and Quaas, 1988). The current study, therefore, simulated selection by truncation with overlapping generations. The classical formula for determining the response to selection is modified to account for overlapping generations as follows (Ducrocq and Quaas, 1988):

$$\Delta G = \frac{\sum_{j=1}^{\text{no. of age classes}} i_j r_{IH,j} \sigma_{A,j} n_j / n_{\text{total}}}{L}, \quad [2]$$

where i_j is the selection intensity in age class j , $r_{IH,j}$ is the accuracy of selection for age class j , $\sigma_{A,j}$ is the additive genetic standard deviation for age class j , n_j is the number of selected parents from age class j , n_{total} is the total number of selected parents and L is the generation interval. The above equation is implemented in SelAction software (Rutten et al., 2002), which we used for our calculations. Age classes were defined by the year of birth so that animals of the same age belonged to the same class.

Investigations on the Effect of the Size of the Nucleus on Response to Selection

The effect of the size of the nucleus on the response to selection was investigated by simulating the response for a trait with heritability equal to 0.30. Nucleus sizes considered in the alternative schemes were 100, 200, 400, 600, 800, and 1,000 elite dams selected annually. For all schemes, 10 sires were selected each year. For simplicity, only the maximal size of the CRC group was considered. Therefore, the number of progeny records per sire for alternative PT and GS+PT schemes was fixed at 30, whereas the size of the reference population for GS, GS+DP, and GS+PT schemes was fixed at 5,000 dams. However, the number of selection candidates was proportionately adjusted, taking into account the size of the nucleus, conception and culling rates, and sex ratio in the nucleus, as explained earlier.

RESULTS

Response to Selection

Response to selection was influenced by the selection strategy, amount of information available, and herita-

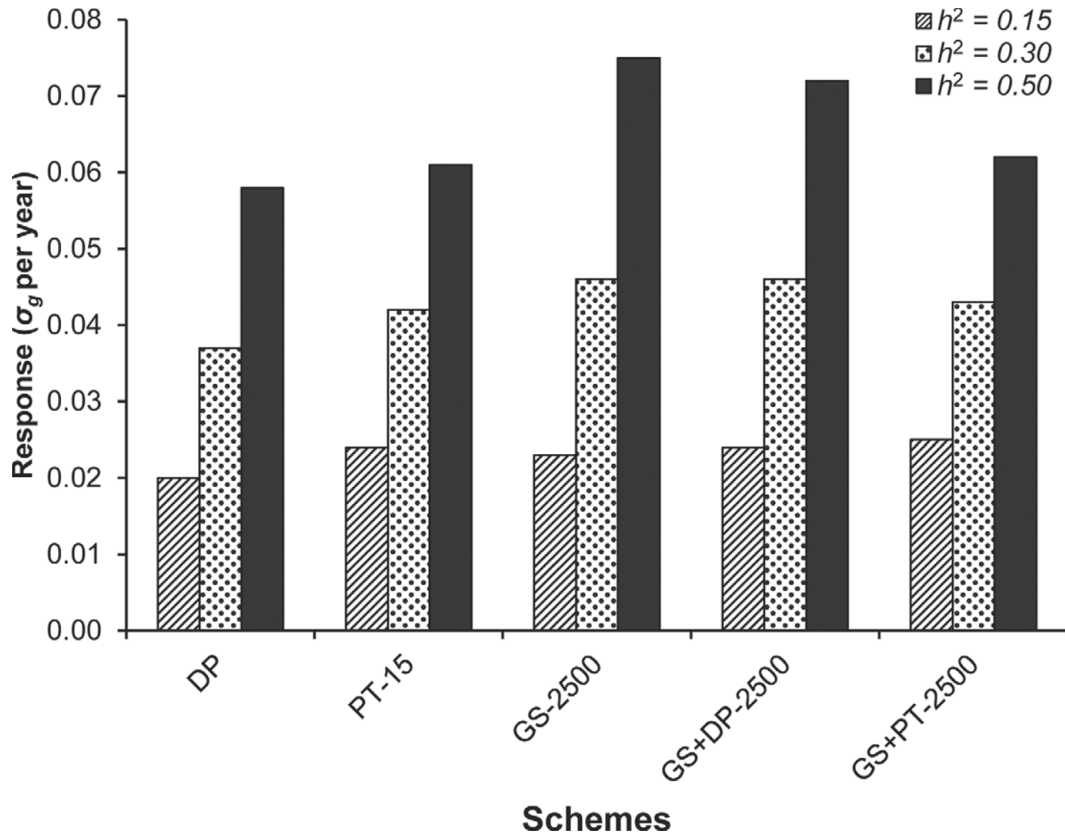


Figure 2. Effects of heritability on response to selection, in genetic SD (σ_g), for within-nucleus dam performance (DP), progeny testing (PT), genomic selection (GS), genomic selection and dam performance within the nucleus (GS+DP), and genomic selection and progeny testing (GS+PT) schemes for a situation with 2,500 commercial recorded dams and heritabilities (h^2) equal to 0.15, 0.30, and 0.50. PT-15 = young bulls were evaluated on performance records of 15 daughters in addition to the performance of relatives within the nucleus.

bility. Figure 2 compares the responses for alternative strategies for a situation with 2,500 CRC and heritability equal to 0.15, 0.30, and 0.50, expressed in genetic standard deviation (σ_g). With the heritability = 0.15, responses were approximately equal for all strategies ranging between 0.020 and 0.025 σ_g . Higher responses were observed for larger heritabilities. Increasing the heritability to 0.30 while maintaining the size of CRC at 2,500 increased the response in GS and GS+DP schemes by 7 and 9.5% compared with a PT scheme, respectively. Genomic selection schemes had the highest sensitivity to increase in heritability. For instance, GS schemes had an 100% increase in response when heritability was doubled from 0.15 to 0.30 compared with 92, 75, and 72% increases for GS+DP, PT, and GS+PT schemes, respectively, for a situation with 2,500 CRC. Increasing the heritability to 0.50 from 0.30 and the number of CRC 2,500 to 5,000 resulted in 59% increase in response for the GS scheme compared with 55, 42, and 43% increases for a similar for GS+DP, GS+PT, and PT schemes.

Figure 3 compares the effect of increasing the size of the CRC population from 2,500 to 5,000 for heritability equal to 0.30 and 0.50. Generally, increasing the size of CRC from 2,500 to 5,000 increased the response observed for all strategies. The highest increases were observed in the GS-only strategy. The GS+DP and GS+PT schemes had higher responses than comparable (in size) DP and PT schemes. The highest response was obtained with genomic selection with a CRC population of 5,000. However, GS schemes with a reference population of 500 and 1,000 cows had much lower responses than the base scheme (Table 2).

To standardize the comparison, the selection strategies under investigation were benchmarked against the basic scheme (i.e., DP). This was to determine the extra response per year in σ_g (expressed as a percentage) that could be achieved by adopting any of the alternative strategies and increasing the level of investment in recording, either of phenotypes or genotypes, from 2,500 to 5,000 CRC. A GS selection strategy had 15.0, 24.3, and 29.3% increase in response compared

Table 2. Accuracies (r_I) and extra response (ΔR) in genetic SD, expressed as a percentage when comparing alternative schemes to the basic within-nucleus dam performance (DP) scheme, for equal commercial recorded cows (CRC) and h^2

Scheme ¹	Parameter					
	$h^2 = 0.15$		$h^2 = 0.30$		$h^2 = 0.50$	
	ΔR (%)	r_I	ΔR (%)	r_I	ΔR (%)	r_I
PT-15	20.0	0.497	13.5	0.623	5.2	0.719
PT-30	40.0	0.604	27.0	0.727	15.5	0.811
GS-500	-45.0	0.069	-43.2	0.098	-39.7	0.126
GS-1000	-25.0	0.098	-18.9	0.137	-15.5	0.176
GS-2500	15.0	0.137	24.3	0.214	29.3	0.272
GS-5000	60.0	0.214	70.3	0.297	72.4	0.372
GS+DP-500	5.0	0.195	5.4	0.265	5.2	0.331
GS+DP-1000	10.0	0.207	10.8	0.282	12.1	0.353
GS+DP-2500	20.0	0.239	24.3	0.327	24.1	0.409
GS+DP-5000	40.0	0.283	43.2	0.386	41.4	0.479
GS+PT-500	-15.0	0.329	-13.5	0.435	-15.5	0.526
GS+PT-1000	0.0	0.392	-2.7	0.510	-6.9	0.608
GS+PT-2500	25.0	0.511	16.2	0.640	6.9	0.735
GS+PT-5000	45.0	0.622	29.7	0.744	17.2	0.825

¹PT = progeny testing; GS = genomic selection; GS+PT = genomic selection and within-nucleus dam performance; GS+PT = genomic selection and progeny testing. For PT-15 and PT-30 schemes, each candidate bull was evaluated on 15 and 30 daughter records, respectively; 500, 1,000, 2,500 and 5,000 represent the size of the reference population for the alternative schemes implementing genomic selection.

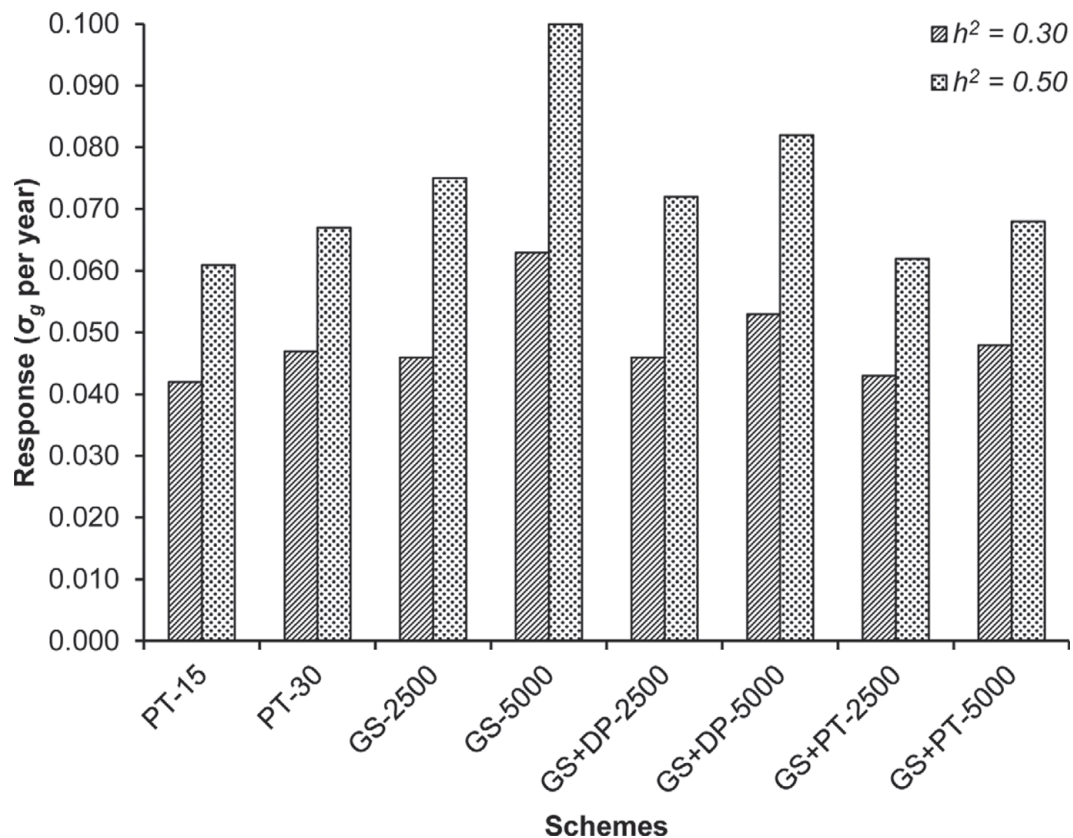


Figure 3. Effects of increasing the number of commercial recorded cows on response to selection, in genetic SD (σ_g), for alternative progeny testing (PT), genomic selection (GS), genomic selection and dam performance within the nucleus (GS+DP), and genomic selection and progeny testing (GS+PT) schemes for situations where the number of commercial recorded dams were 2,500 and 5,000 and heritabilities (h^2) equal to 0.30 and 0.50. PT-15 = young bulls were evaluated on performance records of 15 daughters in addition to the performance of relatives within the nucleus; PT-30 = each candidate bull had 30 daughters with first-lactation records.

with 20.0, 13.4, and 5.2% increase for PT selection strategy for a situation with 2,500 CRC and heritability equal to 0.15, 0.30, and 0.50, respectively (Table 2). When the CRC group had a population of 5,000 cows, increases in response for GS schemes were 60.0, 70.3, and 72.4% for heritability equal to 0.15, 0.30, and 0.50, respectively. When the CRC population was 500 or 1,000, GS+DP schemes had highest responses, followed by GS+PT schemes and GS schemes, for all heritabilities.

Accuracy of Index

Index accuracies for sires in DP, GS, GS+DP, and GS+PT schemes are presented in Table 2. The PT yielded the highest accuracies. Accuracies for PT and combined GS+PT selection were comparable, implying that the contribution of additional genomic information to a PT scheme on accuracy was negligible. The GS+DP selection had slightly higher accuracies than GS. Accuracy was sensitive to the amount of information available for evaluation for all strategies considered. This sensitivity was highest for GS and lowest for PT strategies, implying that GS could outperform PT when the reference population was sufficiently large.

Effect of Selection Strategy on Age Distribution of Selected Sires

For the different schemes, selection of parents for production of nucleus selection candidates was possible only after information from all sources was available. Inclusion of progeny phenotypes as sources of information (PT and GS+PT) resulted in a minimum age of selected sires of 6 yr. The age distribution of active sires is given in Figure 4. For GS schemes, more than 60% of the active bulls were 2 yr old and only 14% of the active bulls were more than 3 yr old. Generally, DP, GS, and GS+DP schemes had more intensive use of young bulls. The differences in mean generation interval are reflected in the differences in response between schemes (Figures 2 and 3).

Effect of the Nucleus Size on Response

Figure 5 presents the effects on response due to increasing the nucleus size from 100 to 1,000 elite dams while maintaining the number of active sires at 10, for a trait with heritability equal to 0.30. Response in all schemes increased with larger nucleus sizes. However, results show a diminishing increase as the nucleus size gets much larger. However, increasing the nucleus is likely to have a big effect on the rate of inbreeding, as it increases the selection intensity in the sires. Inbreeding was not considered in the current study as SelAction

software cannot calculate rates of inbreeding in overlapping generations.

DISCUSSION

Response to Selection

The aim of the current study was to optimize the design of small-sized nucleus genetic improvement dairy cattle schemes for situations where phenotype records are limited. Such schemes are important when producers' participation in pedigree and performance recording is minimal or for unique small-sized populations kept for specific products. Before adoption, it is important to predict the possible benefits of alternative selection strategies so as to determine the best alternative. We compared alternative selection strategies by benchmarking possible responses to schemes where recording was only within the nucleus and evaluation was on phenotypic information only (DP schemes). Our results show that GS schemes had higher annual responses for traits with moderate to high heritability and a moderately small population size (i.e., 2,500 to 5,000 CRC). In actual breeding schemes selection is not for a single trait but an aggregate of a group of traits (i.e., a total merit index). Heritability as expressed in this study can, therefore, be considered as the heritability index. A lower heritability index could reflect a more unfavorable genetic correlation between traits in the index.

Higher responses in GS were particularly due to the ability to greatly reduce the generation intervals for schemes using this selection strategy (Pryce et al., 2010). Shortening of generation intervals will result in more rounds of selection per unit time, hence increasing the achievable gains per period time. Genomic-based selection schemes are, therefore, highly attractive for situations where it is possible to reduce generation intervals (Van Grevenhof et al., 2012). All schemes would benefit from larger nuclei (Figure 5). A larger nucleus size would increase the number of selected parents and hence reduce accumulation of inbreeding.

Risks Associated with Adoption of Alternative Strategies

Risk is an important factor when deciding the selection strategy for adoption. Genomic selection has been described as a high-risk approach for creating genetic improvement in developing countries (Marshall et al., 2011). Components of risk for a breeding program are variance of the response and inbreeding.

Population size in our study referred to the number of animals on which phenotypic records were collected.

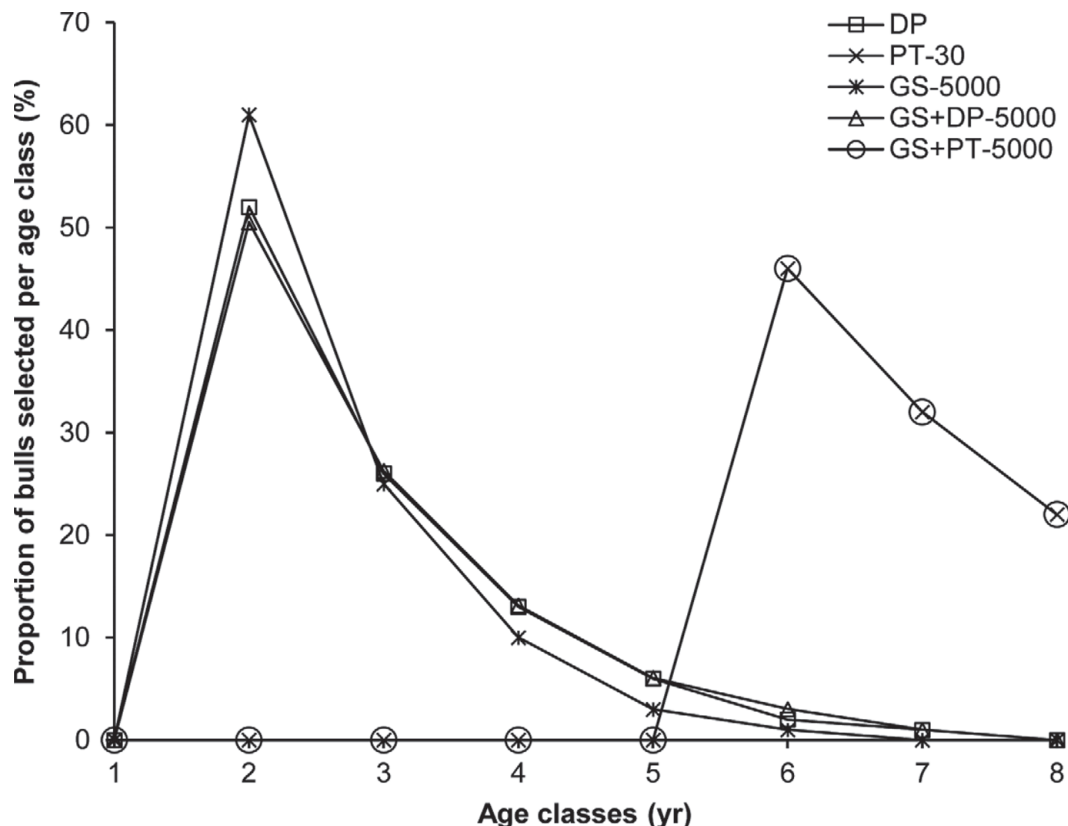


Figure 4. Effect of the selection criteria on proportions of active sires in different age classes for within nucleus dam performance (DP), progeny testing (PT), genomic selection (GS), genomic selection and dam performance within the nucleus (GS+DP), and genomic selection and progeny testing (GS+PT) schemes for a trait with heritability equal to 0.50 and 5,000 commercial recorded dams; PT-30 = each candidate bull had 30 daughters with records.

The mortality rates of calves in developing countries can be high due to infectious diseases or the harsh environment. The number of calves that need to be produced to reach the simulated population sizes is consequently a lot higher. The higher mortality will also increase the risk associated with the small-sized nucleus scheme.

Variance of Response. Deviation of the realized response from the predicted response is an attribute of the prediction error. The prediction error variance is determined by the accuracy of EBV and inbreeding (Meuwissen, 1991). Low accuracy implies larger variability of the estimated genetic value from the true value. Consequently, realized responses from breeding programs with lower selection accuracy are likely to deviate highly from the expectation. In addition, producers may be unwilling to buy semen from bulls with low reliabilities. Therefore, small-sized AI schemes will find it difficult to survive competition with larger schemes in a global dairy breeding market. However, for schemes aimed at breeding for “unique” products or qualities (e.g., breeding for disease and heat tolerance in the tropics), such competition is not likely to have

significant effects to the ability of these schemes to succeed, as they target different market niches.

Therefore, the most important risk for low-accuracy breeding programs is high variances in realized responses. In the present study, GS attained very low accuracies compared with PT. The use of genome-wide dense SNP (i.e., genomic selection) has increased the number of markers available for estimation of associations between markers and QTL (Meuwissen et al., 2001). However, sampling errors are determined by the sample size and heritability of the trait under consideration. Smaller sample sizes result in larger sampling errors, with the situation aggravated when concerned traits have low heritability (Lande and Thompson, 1990; Meuwissen et al., 2001). Consequently, a large proportion of the additive genetic variance may not be explained by markers when the reference population is small. Hence, the realized response is likely to be lower than predicted (Moghaddar and van der Werf, 2009). Variance in realized response due to low accuracies poses a major challenge for the GS strategy, as it is characterized by low accuracies (Pryce et al., 2010; Lillehammer et al.,

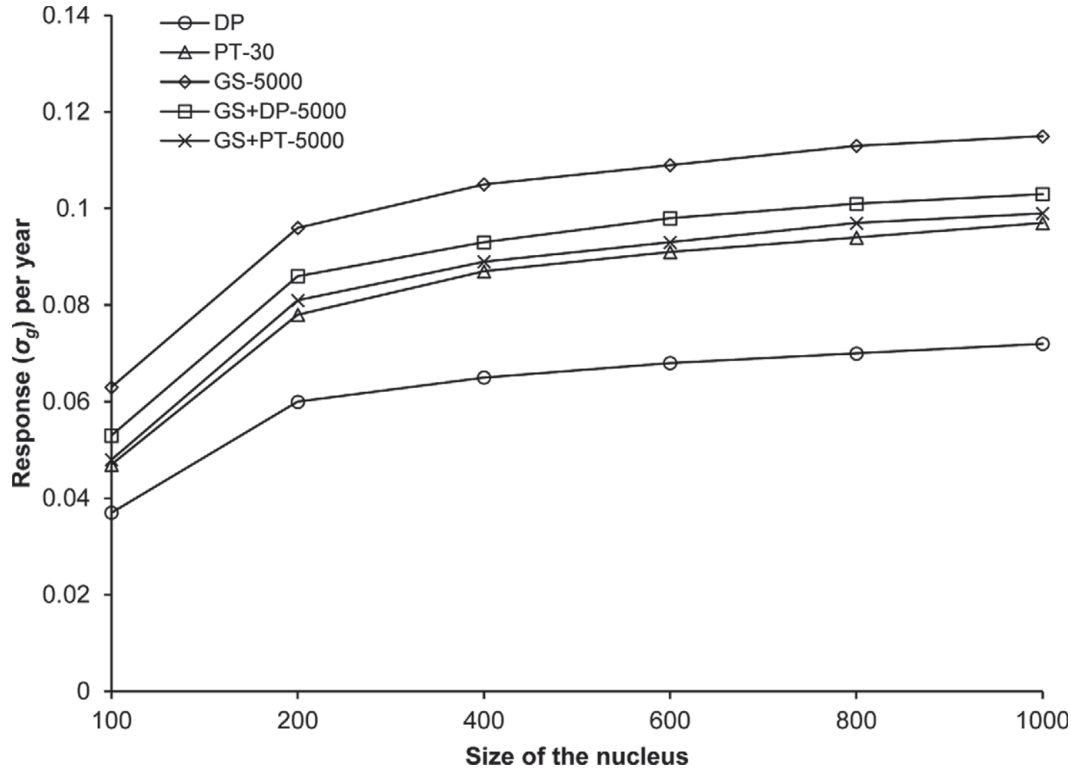


Figure 5. Effect of the nucleus size on response to selection for within-nucleus dam performance (DP), progeny testing (PT), genomic selection (GS), genomic selection and dam performance within the nucleus (GS+DP), and genomic selection and progeny testing (GS+PT) schemes for a trait with a heritability of 0.30 and 5,000 commercial recorded dams available. Nucleus sizes considered were 100, 200, 400, 600, 800, and 1,000 dams. σ_g = genetic SD; PT-30 = each candidate bull had 30 daughters with first-lactation records.

2011; Van Grevenhof et al., 2012). Such accuracies are generally much lower for smaller reference populations (Figure 4). A previous study showed that a population of approximately 100,000 dams with genotypes and phenotypes are required to attain an accuracy above 0.90 (Van Grevenhof et al., 2012).

Previous studies that have considered the potential of GS have assumed that such schemes will attain high accuracies (Schaeffer, 2006; König et al., 2009). Such accuracies are not achievable for small-sized schemes. Alternatives that involve a combination of information sources have been studied. Such alternatives have been preselection of sires and combined use of phenotypic and genotypic information for selection (Schrooten et al., 2005; Pryce et al., 2010). For small-sized nucleus schemes, preselection of sires will not be viable, as the available number of candidates is generally small; hence, all newborns need to be included as candidates. Results from the current study show that the combined use of progeny phenotypes with genomic information (GS+PT schemes) will have negligible effect on response compared with use of progeny phenotypes only (PT schemes; Figures 2 and 3). This is because all the benefits of reduced generation intervals are lost with

such schemes. A compromise would be to combine female phenotypic information from a nucleus dam (i.e., excluding progeny phenotypes) with genomic information. In this study, GS+DP-5000 schemes had relatively higher responses compared with PT and higher accuracies compared with GS only when heritability of the trait was 0.3 or higher.

Another option to maximize achievable accuracy within the constraint of a small-sized reference population is to adopt an improved genome-wide prediction method. Genome-wide prediction methods can be categorized based on how they model the variance of markers across the genome. Linear methods assume homogeneity of variance for all markers and include variants of genomic BLUP. Nonlinear Bayesian methods allow for heterogeneity of variance between markers. For small reference populations, previous simulation studies have shown that Bayesian predictions will yield higher accuracies compared with linear predictions (Meuwissen et al., 2001; Daetwyler et al., 2010).

Comparing schemes only on accuracies while ignoring responses may not give a balanced conclusion. In the present study, adoption of a GS scheme would result in 70.3% increase in response compared with the

basic scheme. Adoption of the GS+DP scheme would result in 43.2% extra response (Table 2). Therefore, despite the expected larger fluctuations in response for GS schemes, they will have higher mean performance at the population level, which compensates for the fluctuations.

Inbreeding. Inbreeding leads to increased homozygosity within the population, resulting in reduced genetic variance, inbreeding depression, and increase in problems associated with lethal genes. Future inbreeding is determined by the coancestry of parents. With small-sized nucleus schemes, where the number of parents is small, the rate of inbreeding is expected to be higher. This will have the effect of reduced response to selection in the long-term. Accumulation of inbreeding in future generations can be minimized by minimizing coancestry in the parents (Wray and Goddard, 1994; Brisbane and Gibson, 1995; Meuwissen, 1997). Methodologies have been developed to minimize inbreeding for small populations with overlapping generations (Meuwissen and Sonesson, 1998; Sonesson and Meuwissen, 2001).

The procedures of Meuwissen and Sonesson (1998) and Sonesson and Meuwissen (2001) were derived with best BLUP EBV selection in mind. The BLUP EBV of relatives are highly correlated, which can lead to more co-selection of relatives. The expected level of inbreeding accumulation in GS schemes should be equal to that in PT schemes, as inbreeding is determined mainly by the coancestry of selected parents. However, with GS, it is possible to make a distinction between EBV of full-sibs, as this strategy can explain some of the variance due to Mendelian sampling and, hence, reduce co-selection of siblings (Daetwyler et al., 2007).

General Discussion

Unfavorable G×E between sires performing in temperate and tropical countries necessitates the establishment of breeding programs within tropical production systems so as to economically optimize tropical dairy cattle industries. However, in developing countries (such as the case in Kenya) commercial herd pedigree and performance recording is minimal and inconsistent. Failure of the recording system to meet farmer expectations and to offer tangible returns are the main weaknesses in the pedigree and performance recording system in the Kenya dairy cattle (Wasike et al., 2011). This limits the opportunities to implement a genetic improvement program that relies on PT. An alternative approach for creating genetic improvement of dairy sector is, therefore, imperative. In addition, attractive and viable genetic improvement strategies are required

for cases with small unique populations with well-developed recording systems.

Alternative selection strategies (i.e., DP, PT, GS, GS+DP, and GS+PT) within the limitation of minimal recording were studied and the risks involved discussed. Adoption of a selection strategy that maximizes response to selection per period time would be the most beneficial. Our results show that this is achieved through use of GS. Risk associated with alternative strategies is an important factor to consider (Marshall et al., 2011). Although GS schemes are expected to have higher risks compared with PT schemes due to the lower accuracies, such schemes will generally have a higher mean performance. Therefore, we expect higher overall population mean performance if GS is adopted. In addition, GS are expected to have lower inbreeding accumulation over time compared with PT schemes due to minimized co-selection of siblings, as discussed earlier.

Other indirect effects of the selection strategy on the industry are worth consideration. Lack of tangible returns to producers for their recording efforts is a major deterrent from future participation (Wasike et al., 2011). Adoption of a strategy that would maximize the achievable response in the shortest possible time would be helpful in such a situation. Furthermore, with a system where recording structures are not well established, the accuracy of information coming from participating producers is likely to be low. As a consequence, reliabilities of estimates of population parameters and expected response will be low. Accuracy of such information can be verified by use of genomic information, in particular the accuracy of the pedigree information. Therefore, the overall risk associated with GS schemes may indeed be lower than that expected from PT strategy.

Reduction in the amount of recording required will have a monetary advantage. In the current study, we constructed the reference population by use of genotypes and phenotypes from the same animal. Genotyping need not be done annually but periodically. With well-stipulated genotyping intervals, phenotypic performances can be reduced to only the year previous to the genotyping period. This minimizes the amount of recording compared with PT strategy that require continuous recording. Another factor to consider is that small-sized PT dairy cattle selection schemes are constrained in flexibility in the number of selection candidates. The limited number of cows to produce test daughters implies that it is practically impossible to increase the information available per candidate sire or to increase the number of male candidates without lowering the number of evaluation records. On the other hand, potential exists of evaluating more candidates

(also from outside the nucleus) with GS, as the number of candidates evaluated is independent of the number of CRC.

The implementation of a breeding scheme is complicated and requires more details and considerations than could be addressed in the present study. However, the aim of this study was to investigate the potential for a small-sized nucleus dairy genetic improvement program. Our results show that the greatest potential for such programs lie in GS.

CONCLUSIONS

Feasibility exists for creating genetic gain through use of GS in small-sized nucleus dairy cattle breeding programs. An optimal breeding scheme would rely on annual genotyping of 5,000 commercially recorded cows in the genetic evaluation. Compared with similar-sized progeny-testing schemes, GS schemes would have lower accuracies but these are offset by higher responses per year, which might provide additional incentive for farmers to participate in recording.

ACKNOWLEDGMENTS

The authors are grateful to the Netherlands Organization for International Cooperation in Higher Education (NUFFIC; The Hague, the Netherlands) for financing this study. The first author thanks Elizabeth M. van Grevenhof (Animal Breeding and Genomics Centre, Wageningen University) for her assistance with the SelAction software.

REFERENCES

- Brisbane, J. R., and J. P. Gibson. 1995. Balancing selection response and rate of inbreeding by including relationships in selection decisions. *Theor. Appl. Genet.* 91:421–431.
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031.
- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 124:369–376.
- Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* 3:e3395.
- Dekkers, J. C. M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124:331–341.
- Ducrocq, V., and R. L. Quaas. 1988. Prediction of genetic response to truncation selection across generations. *J. Dairy Sci.* 71:2543–2553.
- König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs. *J. Dairy Sci.* 92:382–391.
- Kosgey, I. S., R. L. Baker, H. M. J. Udo, and J. A. M. van Arendonk. 2006. Successes and failures of small ruminant breeding programmes in the tropics: A review. *Small Rumin. Res.* 61:13–28.
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. *J. Dairy Sci.* 94:493–500.
- Marshall, K., C. Quiros-Campos, J. H. J. van der Werf, and B. Kinghorn. 2011. Marker-based selection within smallholder production systems in developing countries. *Livest. Sci.* 136:45–54.
- Meuwissen, T. H. E. 1991. Expectation and variance of genetic gain in open and closed nucleus and progeny testing schemes. *Anim. Prod.* 53:133–141.
- Meuwissen, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.* 75:934–940.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Meuwissen, T. H. E., and A. K. Sonesson. 1998. Maximizing the response of selection with a predefined rate of inbreeding: Overlapping generations. *J. Anim. Sci.* 76:2575–2583.
- Moghaddar, N., and J. H. J. van der Werf. 2009. Effect of the accuracy of an estimated QTL effect on response to marker-assisted selection. *Proc. Assoc. Adv. Anim. Breed. Genet.* 18:564–567.
- Muasya, T. K., K. J. Peters, and A. K. Kahi. 2013. Breeding structure and genetic variability of the Holstein Friesian dairy cattle population in Kenya. *Genet. Res.* 52:127–137.
- Mulder, H. A., R. F. Veerkamp, B. J. Ducro, J. A. M. van Arendonk, and P. Bijma. 2006. Optimization of dairy cattle breeding programs for different environments with genotype by environment interaction. *J. Dairy Sci.* 89:1740–1752.
- Ojango, J. M. K., and G. E. Pollott. 2002. The relationship between Holstein bull breeding values for milk yields derived in both the UK and Kenya. *Livest. Prod. Sci.* 74:1–12.
- Okeno, T. O., I. S. Kosgey, and A. K. Kahi. 2010. Genetic evaluation of breeding strategies for improvement of dairy cattle in Kenya. *Trop. Anim. Health Prod.* 42:1073–1079.
- Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. *J. Dairy Sci.* 93:5455–5466.
- Rege, J. E. O., K. Marshall, A. Notenbaert, J. M. K. Ojango, and A. M. Okeyo. 2011. Pro-poor animal improvement and breeding—What can science do? *Livest. Sci.* 136:15–28.
- Rutten, M. J. M., P. Bijma, J. A. Woolliams, and J. A. M. van Arendonk. 2002. SelAction: Software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Hered.* 93:456–458.
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Schrooten, C., H. Bovenhuis, J. A. M. van Arendonk, and P. Bijma. 2005. Genetic progress in multistage dairy cattle breeding schemes using genetic markers. *J. Dairy Sci.* 88:1569–1581.
- Sonesson, A. K., and T. H. E. Meuwissen. 2001. Minimization of rate of inbreeding for small populations with overlapping generations. *Genet. Res.* 77:285–292.
- van Arendonk, J. A. M. 2011. The role of reproductive technologies in breeding schemes for livestock populations in developing countries. *Livest. Sci.* 136:29–37.
- Van Grevenhof, E. M., J. A. M. Van Arendonk, and P. Bijma. 2012. Response to genomic selection: The Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genet. Sel. Evol.* 44:26.
- Vargas, B., and J. A. M. van Arendonk. 2004. Genetic comparison of breeding schemes based on semen importation and local breeding schemes: Framework and application to Costa Rica. *J. Dairy Sci.* 87:1496–1505.
- Wasike, C. B., T. M. Magothe, A. K. Kahi, and K. J. Peters. 2011. Factors that influence the efficiency of beef and dairy cattle recording system in Kenya: A SWOT-AHP analysis. *Trop. Anim. Health Prod.* 43:141–152.
- Wray, N. R., and M. E. Goddard. 1994. Increasing long-term response to selection. *Genet. Sel. Evol.* 26:431–451.