An approach for the design of breeding programs using genomics

J.C.M. Dekkers[1*], H. Su[1], L. Kramer[1], and H. Yu[1]

[1] Iowa State University, 806 Stange Rd, Ames, IA, 50011; [*]jdekkers@iastate.edu

Abstract

Design and optimization of modern breeding programs requires methods and software to predict response to multiple-trait and multiple-stage breeding programs based on estimated breeding values and genomic predictions. Here, we introduce and describe the ShinySelAction software for these purposes. ShinySelAction is based on the widely used software SelAction for breeding programs with discrete generations but with a more exact implementation of multiple-stage selection programs and a more direct implementation of information from genomic predictions. This software can be used by breeders to reliably compare alternative breeding programs and for investment decisions for breeding programs that include genomic information.

Introduction

In the past decades, theory to predict selection response for quantitative traits in pedigree-based multiple-trait livestock breeding programs has been well developed, validated, and implemented, for example in the software SelAction (Rutten et al., 2002). SelAction uses to deterministic recursive equations that were developed based on standard quantitative genetic and selection index theory and pseudo best linear unbiased prediction (BLUP) methodology, to model changes of trait means and variance-covariance structures as a result of selection. It's aim is to predict equilibrium responses to multiple trait selection on estimated breeding values (EBV) that are obtained under the infinitesimal genetic model with consideration of the Bulmer effect but ignoring the effect of inbreeding on genetic (co-)variances. Nowadays, breeding programs are enhanced by genomic EBV (GEBV) based on genomic prediction, which provide more accurate EBV of animals at a younger age and can improve the effectiveness of breeding programs. Schrooten et al. (2005) and Dekkers (2007) showed how the GEBV for a trait can be incorporated in standard selection index programs such as SelAction by specifying an additional trait with heritability (close to) one and a genetic correlation with the original phenotype-based trait equal to the accuracy of the GEBV. Genetic and phenotypic correlations of the GEBV with other phenotype-based traits and with the GEBV of other traits were derived by Dekkers (2007). These approaches require estimates of the accuracy of GEBV as inputs, which have been derived by several. Recently, Dekkers et al. (2021) derived accuracies of GEBV by formulating the GEBV as an index of the standard pedigree-based EBV and a conceptual EBV based on genomic-deviated-from-pedigree relationships (DEBV). Combined with already available methods to model the accuracy of pedigree-based EBV (PEBV), this allows genomic information to be incorporated following Dekkers (2007) but using the accuracy of DEBV rather than of GEBV. Against this background, the purpose of this paper is to present and describe a computer program called ShinySelAction that extends the SelAction software to incorporate genomic information. In addition, a more exact derivation of predictions for multiple-stage selection programs is implemented.

Materials & Methods

Available program code for the discrete generations option of SelAction from Rutten et al. (2002) was translated into Python. The user interface was implemented in Shiny from R Studio (https://shiny.rstudio.com/).

Implementation of the Bulmer effect for multiple-stage selection breeding programs was updated. In SelAction, Bulmer equilibrium parameters were derived by iterating on information available in the final stage of selection, in order to avoid multi-variate normal integrals. Because the final stage has the highest accuracy, this leads to overestimation of the reductions in genetic (co-)variances and accuracies. In ShinySelAction, exact derivations of reductions in (co-)variances of and among the index used at each stage and of true and estimated breeding values at each stage are implemented based on the moment generation function of the n-dimensional truncated multi-normal distribution, following Tallis (1961).
Genomic information is incorporated as in Dekkers (2007) by specifying genomic predictions as correlated traits with heritability (close to) one. However, following Dekkers et al. (2021) genomic information was incorporated based on the accuracy of DEBV, rather than that of GEBV. In the following, parameters that are required for DEBV will first be described, followed by formulation of the genetic and phenotypic correlations among DEBV for multiple traits and their correlations with the phenotype-based traits.

DEBV parameters. In ShinySelAction, parameters for genomic information that must be provided include, for each trait, the accuracy of the DEBV in selection candidates (r_D) and the proportion of genetic variance captured by the markers (q^2). Both parameters depend on the effective number of chromosome segments, M_e, which is a population parameter that can be derived based on empirical accuracies of GEBV and of PEBV for an available reference data set for the population, as described by Dekkers et al. (2021) and implemented below. The required empirical accuracies of GEBV (r_{G_r}) and PEBV (r_{A_r}) in the reference can be estimated based on available phenotypic and genotype data by leave-one-out cross-validation, following Cheng et al. (2021). Based on these accuracies, an estimate of M_e for the population can then be derived in the following iterative manner (Dekkers et al. 2021, Appendix 1):

1) Compute the accuracy of DEBV in the reference population as: $r_{D_r} = \sqrt{\frac{r_{G_r}^2 - r_{A_r}^2}{1 + r_{A_r}^2(r_{G_r}^2 - 2)}}$.
2) Set the proportion of genetic variance captured by markers for DEBV equal to 1: $q^2 = 1$.
3) Compute θ_{D_r} in the reference population as: $\theta_{D_r} = \frac{r_{D_r}^2(1 - r_{D_r}^2 q^2 h^2)}{q^2 - r_{D_r}^2}$, where h^2 is heritability.
4) Compute M_e in the reference population as: $M_e = Nq^2h^2/\theta_{D_r}$, where N is the size of the reference population.
5) Update q^2 as: $q^2 = M/(M + M_e)$, where M is the number of genotyped markers.
6) Return to step 3) until stable values for q^2 and M_e are obtained.

Accuracies of GEBV and PEBV in the reference population (r_{G_r} and r_{A_r}) for use in step 1) of the above procedure to estimate M_e can be obtained based on data (phenotypes and genotypes) from a population that is under selection and for any trait (or as the average across traits) because M_e is a population parameter that is little affected by selection and heritability of the trait, as shown by Dekkers et al. (2021). If a suitable reference population is not available to estimate M_e, predictions of M_e by, e.g., Goddard (2009) or others can be used.
Based on the resulting population estimates of q^2 and M_e, the accuracy of DEBV in an unselected reference population, i.e. ignoring the Bulmer effect, for a trait with base population heritability h^2 and reference size N can then be derived as follows, which is based on the correction of Bijma and Dekkers (2021) (equation 3b) to Dekkers et al. (2021):

$$r_{D_r} = \sqrt{\left[1 + \theta_{D_r} - \sqrt{\left(1 + \theta_{D_r}\right)^2 - 4h^2q^2\theta_{D_r}}\right]/2h^2} \qquad \text{with} \quad \theta_{D_r} = Nq^2h^2/M_e$$

The accuracy of DEBV of selection candidates (=target population), r_D, depends on the accuracy of DEBV in the reference population, r_{D_r}, as derived above, and the loss of accuracy, p_{rt}, from the reference population to the target population due to recombination in each effective chromosome segment, and can be computed following Dekkers et al. (2021) as:
$r_D = r_{D_r} p_{rt} = r_{D_r}(1 - 2L/M_e)^{(l_p + l_m)}$
where L is the genome size in Morgans and l_p and l_m are the number of generations between the selection candidates and their closest paternal and maternal, respectively, ancestors in the reference population (= 1 if selection candidates are progeny of individuals in the reference population). Accuracy r_D is what is required to be entered into ShinySelAction for each trait with genomic information.

Adding genomic traits. Given the accuracy of the DEBV of selection candidates for each trait i, r_{Di}, and the proportion of genetic variance explained by markers for trait i, q_i^2 (typically assumed the same for each trait), genomic information is incorporated into ShinySelAction by adding the DEBV as genetic traits with heritability equal to 0.999. Following Dekkers (2007), the expanded genetic correlation matrix can then be derived as follows for a 3-trait example:

	BV1	**BV2**	**BV3**	**DEBV1**	**DEBV2**	**DEBV3**
BV1	1	r_{g12}	r_{g13}	r_{D1}	$r_{g12}\, r_{D2}$	$r_{g13}\, r_{D3}$
BV2	r_{g12}	1	r_{g23}	$r_{g12}\, r_{D1}$	r_{D2}	$r_{g23}\, r_{D3}$
BV3	r_{g13}	r_{g23}	1	$r_{g13}\, r_{D1}$	$r_{g23}\, r_{D2}$	r_{D3}
DEBV1	Symmetric			1	$r_{g12}\, r_{D1}\, r_{D2}/q_1 q_2$	$r_{g13}\, r_{D1}\, r_{D3}/q_1 q_3$
DEBV2				$r_{g12}\, r_{D2}\, r_{D1}/q_2 q_1$	1	$r_{g23}\, r_{D2}\, r_{D3}/q_2 q_3$
DEBV3				$r_{g13}\, r_{D3}\, r_{D1}/q_3 q_1$	$r_{g23}\, r_{D3}\, r_{D2}/q_3 q_2$	1

which can be computed using the following direct product (∘) of matrices:

$\begin{matrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{matrix}$	$\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \end{bmatrix}$
Symmetric	$\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & r_{D2}/q_2 & r_{D3}/q_3 \\ r_{D1}/q_1 & 1 & r_{D3}/q_3 \\ r_{D1}/q_1 & r_{D2}/q_2 & 1 \end{bmatrix}' \circ \begin{bmatrix} 1 & r_{D2}/q_2 & r_{D3}/q_3 \\ r_{D1}/q_1 & 1 & r_{D3}/q_3 \\ r_{D1}/q_1 & r_{D2}/q_2 & 1 \end{bmatrix}$

Similarly, the block of phenotypic correlations between BV and DEBV (upper-right block) can be computed as: $\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \end{bmatrix} \circ \begin{bmatrix} h_1 & h_1 & h_1 \\ h_2 & h_2 & h_2 \\ h_3 & h_3 & h_3 \end{bmatrix}$.
Note that phenotypic correlations among DEBV are the same as their genetic correlations.

Results and discussion

The approach described enables prediction of response to selection in multi-trait breeding programs with genomic prediction. This is essential for the design and optimization of most modern livestock breeding programs. The multi-variate normal infinitesimal genetic model is assumed but changes in genetic (co-)variances among traits as a result of the Bulmer effect are considered. The effects of inbreeding on genetic (co-)variances are, however, ignored, such that an equilibrium is reached in terms of genetic parameters and response to selection, following Rutten et al. (2002). This equilibrium response is a crucial outcome for comparison and

optimization of breeding programs, along with the rate of inbreeding that the breeding program generates. The approach is based on the methodology implemented in the SelAction software of Rutten et al. (2002) but with a more exact implementation of the Bulmer effect for multiple-stage selection.

The developed approach implements information from genomic predictions by considering them to be correlated phenotypes for genetic traits that are also assumed to follow the infinitesimal genetic model but with heritability equal to one, following Schrooten et al. (2005) and Dekkers (2007). Key input parameters for incorporating genomic predictions are their accuracy and the proportion of genetic variance explained by the markers. Both depend on M_e as a key population parameter, which can be derived as described by Dekkers et al. (2021) and summarized in this paper. The approach was validated using simulation in Dekkers et al. (2021).

In the proposed approach, the accuracy of DEBV is derived based on a pre-determined reference population size. This does not take into account that, in an ongoing breeding program, the reference population size increases over generations. Assuming a constant size of the reference population may, however, be realistic because there is evidence that data from older generations do not contribute much information to current GEBV and may even reduce their accuracy (Weng et al. 2016, Howard et al. 2018, Hidalgo et al. 2021).

Future developments include allowing for groups of selection candidates with different amounts of information, beyond male versus female selection candidates. This will allow the modelling of breeding programs where not all selection candidates are genotyped or not all selection candidates are phenotyped for all traits.

Acknowledgements

This research was funded by USDA-NIFA grant #2017-67015-26299. The authors thank Piter Bijma for providing the SelAction program code and Julius van der Werf for useful discussions.

References

Bijma, P., and Dekkers, J.C.M. (2022) Genetics Selection Evolution

Cheng J., Dekkers J.C., and Fernando R.L. (2021) J. Animal Breeding and Genetics 138(5):519-527 https://doi.org/10.1111/jbg.12545

Dekkers J.C.M. (2007) J. Animal Breeding and Genetics 124(6):331-341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Dekkers J.C.M., Su H., and Cheng J. (2021) Genetics Selection Evolution 53(1):1-23. https://doi.org/10.1186/s12711-021-00647-w

Goddard M. (2009) Genetica 136(2):245-57. https://doi.org/10.1007/s10709-008-9308-0

Hidalgo, J., Lourenco, D., Tsuruta, S., Masuda, Y., Breen, V., Hawken, R., Bermann, M. and Misztal, I., 2021. *J. Animal Sci*, *99*(9), p.skab239. https://doi.org/10.1093/jas/skab239

Howard, J.T., Rathje, T.A., Bruns, C.E., Wilson-Wells, D.F., Kachman, S.D. and Spangler, M.L., 2018. *J. Animal Breed.&Genet.*, *135*(4), pp.251-262. https://doi.org/10.1111/jbg.12334

Rutten M.J.M., Bijma P., Woolliams J.A., and Van Arendonk J.A.M. (2002) J. Heredity 93(6):456-458. https://doi.org/10.1093/jhered/93.6.456

Schrooten C., Bovenhuis H., Van Arendonk J.A.M., and Bijma P. (2005) J. Dairy Sci. 88: 1569–1581. https://doi.org/10.3168/jds.S0022-0302(05)72826-5

Tallis G.M. (1961) J. Royal Statistical Society: Series B (Methodological) 23(1): 223-229. https://doi.org/10.1111/j.2517-6161.1961.tb00408.x

Weng, Z., Wolc, A., Shen, X., Fernando, R.L., Dekkers, J.C., Arango, J., Settar, P., Fulton, J.E., O'Sullivan, N.P. and Garrick, D.J., 2016. Genetics Selection Evolution, 48(1) 22. https://doi.org/10.1186/s12711-016-0198-9