Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 189-192

MULTI-TRAIT GENOMIC MODELS WITH TRAIT AND MARKER SPECIFIC WEIGHTS CAN SUBSTANTIALLY INCREASE THE ACCURACY OF GENOMIC BREEDING VALUES

T. Meuwissen[1] and V. Boerner[2]

[1] Faculty of Life Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway
[2] GHPC Consulting and Service Pty. Ltd., Armidale, NSW, 2350 Australia

SUMMARY

The GWABLUP (Genome-Wide Association based Best Linear Unbiased Prediction) approach used GWA analysis results to differentially weigh the SNPs in genomic prediction, and was found to improve the reliabilities of genomic predictions. However, the proposed multi-trait GWABLUP method assumed that the SNP weights were the same across the traits. Here we extended and validated the multi-trait GWABLUP method towards using trait-specific SNP weights. In a 3-trait dairy data set, multi-trait GWAS estimates of SNP effects and their standard errors were translated into trait-specific likelihood ratios for the SNPs having trait effects and posterior probabilities using the GWABLUP approach. This produced trait-specific prior (co)variance matrices for each SNP, which were applied in a SNP-BLUP model for genomic predictions, implemented in the APEX linear model suite. In a validation population, the trait-specific SNP weights resulted in more reliable predictions for all three traits. Especially, for somatic cell count, which was hardly related to the other traits, the use of the same weights across all traits was harming genomic predictions. The use of trait-specific SNP weights overcame this problem. In multi-trait GWABLUP analyses of ~30,000 reference population cows, trait-specific SNP weights resulted in up to 13% more reliable genomic predictions than unweighted SNP-BLUP, and improved genomic predictions for all three studied traits.

INTRODUCTION

A limitation of GBLUP and SNP-BLUP is their assumption that all SNPs contribute equally to the total genetic variance. Bayesian variable selection methods allocate more variance/weight to the most important SNPs but are complex and computationally demanding. Recently, GWABLUP was proposed which uses deterministic weights based on GWAS (Genome Wide Association Study) results (Meuwissen *et al*. 2024). Since the SNP weights are (pre)determined by the GWAS signals, GWABLUP is based either on a weighted SNP-BLUP or on a weighted G-matrix in GBLUP, which may both be extended to single step methods (ssGWABLUP). Also a multi-trait extension of GWABLUP was proposed assuming that SNP weights are equal across the traits (Meuwissen *et al*. 2024). The latter assumed that all traits are affected by the same QTL (Quantitative Trait Loci). But generally, different traits will be affected by different QTL, and the use of the same SNP weights across the traits is suboptimal. Our aim is here to extend the multi-trait GWABUP method to using SNP weights that are trait-specific and to compare the results to using equal weights across the traits. The methods are compared in the same dairy cattle data set as Meuwissen *et al*. (2024), and using the APEX linear models suite (Boerner 2024), which implements multi-trait SNP-BLUP with different (co)variance matrices per SNP and thus allows for different weights per SNP and per trait.

MATERIALS AND METHODS

The 3-trait dairy data set of Meuwissen *et al*. (2024) included the yield deviations (YD) of milk and protein yield and somatic cell count (SCC) and their reliability on 32,201 Norwegian Red cows, and was kindly provided by Geno SA (www.geno.no). Estimates of heritabilities, genetic and environmental correlations of the traits are depicted in Table 1.

Table 1. Heritabilities (diagonals), genetic correlations (below diagonals), and residual correlations (above diagonals) of the dairy traits

	Milk	Prot	SCC
Milk	0.26	0.96	-0.17
Protein	0.85	0.20	0.16
SCC	0.10	0.10	0.16

In Meuwissen *et al*. (2024), a canonically transformation of the 3 traits was performed (Ducrocq *et al*. 1993) which resulted in 3 genetically and environmentally independent canonical traits with standardised environmental variances of 1 and genetic variances of 0.16, 0.29 and 1.44, respectively. The data also included imputed HD genotypes on N_{snps}= 617,739 SNPs for all 32,201 cows (see Meuwissen *et al*. (2024) for details). Uniform SNP weights across the traits were obtained from the combined log-likelihood ratios of the GWAS of the 3 canonical traits as described in Meuwissen *et al*. (2024). The GWAS of the three canonical traits resulted also in estimates of their SNP effects and standard errors for each of the canonical traits. To obtain trait-specific SNP weights, these canonical trait SNP effects were back-transformed to original trait SNP effects together with their standard errors. Following Meuwissen *et al*. (2024), GWABLUP log-likelihood ratios per SNP and trait were calculated as $LR_{tj} = 0.5\,\widehat{b_{tJ}^2}/se_{tj}^2$, LR_{tj} where is the log-likelihood ratio, $\widehat{b_{tj}}$ is the GWAS estimate of the effect, and se_{tj} is the standard error of SNP j for trait t. The model averaging of MCMC Bayesian genomic prediction schemes is to some extend mimicked by calculating the moving average of the likelihood ratios, i.e. by averaging the likelihood ratios of the current SNP and two adjacent SNPs to the left and two adjacent SNPs to the right, resulting in $\overline{LR_{tj}}$. Following Meuwissen *et al*. (2024), posterior probabilities that the SNPs have non-zero effects, which serve as trait-specific weights for the SNPs, were calculated for trait t and SNP j as $\pi\, e^{\overline{LR_{tJ}}}/[\pi e^{\overline{LR_{tJ}}} + (1-\pi)]$ where the prior probability of non-zero SNP effects is $\pi = 0.001$.

The model for the multi-trait SNP-BLUP is $y = X\mu + Zb + e$ where y is a vector of yield deviations for the 3 traits (ordered by traits within cows); $X = 1_{n_a} \otimes I_3$ is a design matrix linking the records to their trait means (n_a is the number of animals); μ is a vector of estimates of trait means for the 3 traits; $Z = M \otimes I_3$ with M being a ($n_a \times N_{snps}$) matrix of centred allele counts of the cows; b is a vector of ($3 \times N_{snps}$) multi-trait SNP effects (ordered by traits within SNPs). The prior distribution of the residuals was $e \sim N(0, W \otimes R)$, where R is the residual (co)variance matrix of the traits (Table 1) and W is a ($n_a \times n_a$) diagonal matrix with the inverses of the weights of the yield deviations on the diagonal. For the analysis the multi-trait SNP-BLUP method of Liu *et al*. (2014) is used, where in our data all animals are genotyped and pedigree relationships are not used and thus set to an identity matrix. The (co)variances of the breeding values and SNP effects (traits within cows and SNPs) is modelled by $var(u) = Z\Theta Z' + \varepsilon I_{n_a}, var(b) = \Theta, cov(u,b) = Z\Theta$, where $u = Zb$ is a vector of multi-trait breeding values; ε is a small number (ε=0.01 was used here) added to regularise the matrix and making $var([u,b])$ non-singular; and Θ is a block-diagonal matrix of 3×3 blocks $\Theta = \sum_{\oplus} \; V_{SNP_j}$ where $\oplus$ denotes the direct sum. V_{SNP_j} is the 3×3 SNP specific (co)variance matrix across 3 traits, i.e. the SNP effects are a priori assumed unrelated with prior distributions $b_j \sim N(0, V_{SNP_j})$, where b_j is a 3×1 vector of effects of SNP j. Let $V_{SNP} = G/\sum_{k=1}^{N_{snps}} \; 2\,p_k(1-p_k)$ where G is the genetic (co)variance matrix of traits and p_k is the allele-frequency of SNP k. Further let P be an $N_{snps} \times 3$ matrix of posterior

probability of SNPs obtained the for 3 canonical traits (Meuwissen *et al*. 2024) with $P_{t,j} = \pi\, e^{\overline{LR_{tJ}}}/[\pi e^{\overline{LR_{tJ}}} + (1-\pi)]$, and h_t be a vector of 1 of length 3 and h_s be a vector of 1 of length N_{snp}. In case of regular multi-trait unweighted SNP-BLUP (SNP_{unw}-BLUP) $V_{SNP_{unw}} = I_{N_{snps}} \otimes V_{SNP}$, in case of equal weights across the traits (SNP_{eqw}-BLUP) $V_{SNP_j} = ((Ph_t)/((h_s/N_{snp})'P(h_t/3))) \otimes V_{SNP}$, and in case of SNP and trait-specific weights (SNP_{tsw}-BLUP), $V_{SNP_j} = \Omega(I_{N_{snps}} \otimes V_{SNP})\Omega, \Omega = I_{N_{snps}\times 3}\sqrt{\overrightarrow{(P/((h_s/N_{snp})'P(h_t/3)))'}}$. An efficient iterative double-preconditioned conjugate gradient algorithm (Vandenplas *et al*. 2019) was used to solve these equations as implemented in the APEX linear models suite (Boerner 2024). This analysis of the SNP_{unw}-BLUP, SNP_{eqw}-BLUP and SNP_{tsw}-BLUP models yielded multi-trait estimates of SNP effects and of animal genetic effects.

The records of cows born in 2018 (1988 cows) were used for validation, and their YDs were masked from the above data analyses. The remaining 30,213 cows were used for the training of the models, i.e. they were used for the GWAS analyses and the estimation of SNP effects. These SNP effects were used to obtain multi-trait breeding value estimates (EBVs) of the validation cows. The squared correlations between the EBVs of the 1988 validation cows and their YDs were used as an indicator of the reliabilities of the EBVs.

RESULTS AND DISCUSSION

Table 2 shows the reliabilities of the multi-trait genomic predictions measured as the squared correlations between GEBV and YDs for milk- and protein yields and SCC of the 1988 validation cows. The YD reliabilities for milk, protein and SCC are 0.409, 0.326, and 0.246, respectively (Meuwissen *et al*. 2024). When expressed relative to the YD reliabilities the reliabilities of SNP_{unw}-BLUP are 0.49 (=0.199/0.409), 0.54, and 0.68, for milk, protein and SCC, respectively.

The use of SNP weights in SNP_{eqw}-BLUP and SNP_{tsw}-BLUP models significantly improved genomic prediction reliabilities for milk and protein yields by 11-13%. SNP_{tsw}-BLUP obtained the highest reliability for all three traits, albeit the improvement for SCC was minor and not statistically significant.

Table 2. Reliabilities of genomic predictions measured as the squared correlations between GEBVs($\widehat{g_v}$) and yield deviations(y_v) of 1988 validation cows

SNP-BLUP Model	$cor(y_v, \hat{g}_v)^2$ *,**		
	Milk	**Protein**	**SCC**
SNP_{unw}-BLUP***	0.199[a]	0.178[a]	0.168[a]
SNP_{eqw}-BLUP***	0.223[b]	0.197[b]	0.160[a]
SNP_{tsw}-BLUP***	0.226[b]	0.201[b]	0.169[a]

*Standard errors of $cor(y_v, \hat{g}_v)^2$ are between 0.006 and 0.008.

**Different letters in the superscripts denote statistically significant differences ($P < 0.05$)

***Subscripts unw, teqw, and tsw mean unweighted, equal weights across the traits, and unequal weights across the traits, respectively.

Table 3 shows inflation biases of the multi-trait predictions measured as the regression coefficients of the yield deviations on the GEBVs for the 1988 validation cows. The inflation bias was only significant for the SCC analysis without SNP weights, where there was a deflation bias. The analyses that used SNP weights yielded virtually unbiased genomic predictions.

Table 3. Inflation bias of genomic predictions measured as the regression coefficient of yield deviations(y_v) on GEBVs($\widehat{g_v}$) for 1988 validation cows

SNP-BLUP Model	$b_{y_v,\hat{g}_v}$ *		
	Milk	Protein	SCC
SNPunw-BLUP**	1.08	1.07	1.19
SNPteqw-BLUP**	1.01	1.00	1.06
SNPtsw-BLUP**	0.99	0.98	1.01

*Standard errors of $b_{y_v,\hat{g}_v}$ are between 0.04 and 0.06.

**Subscripts unw, teqw, and tsw mean unweighted, equal weights across the traits, and unequal weights across the traits, respectively.

It seems natural to combine the GWAS signals across the traits by a multi-trait GWAS (Meuwissen *et al*. 2024), which makes optimal use of the data. But multi-trait GWAS analyses are computationally rather complicated, and simpler single-trait GWAS based approaches may be preferred. For SNP_{tsw}-BLUP models, single-trait GWAS analyses may be directly used to provide the trait-specific SNP weights. If the single-trait SNP analyses are not very powerful (do not result in clear genome-wide significant QTL signals), the use a multi-trait GWAS analysis may be worthwhile. The trait-specific SNP weights applied in SNP_{tsw}-BLUP adjust the prior variances of the traits on a per SNP basis, but not the correlations between the traits. A more flexible model would also estimate correlations of SNP effects on a per SNP basis. Gebreyesus *et al*. (2017) estimated correlations for groups of SNPs and obtained improved prediction reliabilities for milk composition traits. More research will be needed to investigate whether SNP specific correlations would increase the reliabilities of the genomic predictions.

CONCLUSION

The model with trait-specific SNP weights yielded EBVs with the highest reliability for all three traits analysed. For SCC, the model with identical SNP weights reduced the reliability of the EBV compared to unweighted SNP-BLUP. This problem was remedied by the use of trait-specific SNP weights. The multi-trait GWABLUP models yielded up to 13% more reliable EBV compared to unweighted multi-trait SNP-BLUP.

ACKNOWLEDGEMENTS

The authors are grateful to GENO SA (www.geno.no) for providing the data, to the Norwegian Research Council for funding of project number 309611.

REFERENCES

Boerner V. (2024) www.ghpc.ai
Ducrocq V. and Besbes B. J. (1993) *Anim. Breed. Genet*. **110**: 81.
Gebreyesus G., Lund M.S., Buitenhuis B., Bovenhuis H., Poulsen N.A. and Janss L.G. (2017) *Genet. Sel. Evol.* **49**: 1.
Liu Z., Goddard M.E., Reinhardt F. and Reents R. (2014) *J. Dairy. Sci.* **97**: 5833.
Meuwissen T., Eikje L.S. and Gjuvsland A.B. (2024) *Genet. Sel. Evol.* **56**: 1.
Vandenplas J., Calus M.P.L., Eding H. and Vuik C. (2019) *Genet. Sel. Evol.* **51**: 1.