Liu, H., Milkevych, V., Su, G., Thomasen, J., Eiríksson, J., Karaman, E. (2026) Improving genomic prediction in crossbred dairy cattle using local ancestr, *Proceedings 13th World Congress on Genetics Applied to Livestock Production*: 24133 https://doi.org/10.31274/wcgalp.24133 ©2026 *Liu, et al.* CC BY-NC-ND

Improving genomic prediction in crossbred dairy cattle using local ancestry

H. Liu[1], V. Milkevych[1], G. Su[1], J.R. Thomasen[1,2], J.H. Eiríksson[3] and E. Karaman[1]

[1]Aarhus University, Aarhus, Denmark; [2]VikingGenetics, Assentoft, Denmark; [3]Agricultural University of Iceland, Hvanneyri, Borgarnes, Iceland [*]Corresponding author: hliu@qgg.au.dk

Introduction

Crossbreeding is increasingly used in dairy cattle to improve efficiency, robustness, and sustainability, and crossbred cows represent ~10% of the Danish dairy population (Sørensen et al., 2008). Delivering genomic breeding values for crossbreds remains challenging because single nucleotide polymorphism (SNP) effects and linkage disequilibrium can differ across ancestral breed backgrounds (Guillenea et al., 2022). Local ancestry inference assigns the ancestral breed origin of chromosomal segments in admixed individuals and can enable modelling breed-specific marker effects to improve accuracy and calibration in crossbreds (Eiríksson et al., 2022).

This study (1) compares two local ancestry-based genomic prediction approaches for production traits in Danish dairy cattle and (2) evaluates their performance relative to conventional single-step genomic evaluation. We evaluated the breed-of-origin of markers (BOM) approach and the breed-of-origin of alleles (BOA) approach and compared them with a pedigree-based best linear unbiased prediction (PBLUP) and a conventional single-step genomic best linear unbiased prediction (ssGBLUP).

Materials & Methods

Data. Three Danish dairy breeds were analysed: Holstein (HOL), Jersey (JER), and Red Dairy Cattle (RDC), for milk yield (MY). Phenotypes and genotypes of cows born before 2020-01-01 were used for training, and animals born after this date were used for validation. The BOM training set included 48,647 HOL, 27,505 JER, and 9,992 RDC animals with genotypes and phenotypes; for BOA, 6,517 genotyped crossbreds were added to training. Validation animals were 1,182 crossbreds with genotypes and phenotypes. For the single-step model (ssGBLUP), pedigrees of training and validation animals were traced for 3 generations in crossbreds, including ungenotyped but phenotyped crossbreds born after 2020-01-01, and for 2 generations in HOL, JER, and RDC, resulting in 135,896 genotyped animals, 92,658 phenotyped animals, and 438,939 animals in the pedigree. Single-step validation was performed in both 1,182 genotyped and 3,316 phenotyped-only crossbreds. As a baseline, we performed pedigree BLUP (PBLUP).

Phenotype correction. Corrected phenotypes were obtained by adjusting raw records for herd-year-lactation, calving-year-month, age at calving, and heterozygosity using a standard linear model. The corrected phenotypes were used to evaluate predictive ability and dispersion.

Local ancestry inference. Local ancestry (LA) was inferred using AllOr (Eiríksson et al., 2021) with a reference panel of purebred haplotypes from ancestors of crossbreds within five generations. Each genomic segment was assigned to an ancestral breed, providing the breed-of-origin information required for LA-based prediction.

BOM. Breed-specific SNP effects were estimated separately within each pure breed using a SNP-BLUP model, and genomic estimated breeding values (GEBV) for each animal were

computed by weighting breed-specific SNP effects according to the animal's inferred local ancestry (LA) along the genome. Marker effects were estimated using phenotypes and genotypes of the reference animals. For breed $b(b \in \{HOL, JER, RDC\})$, the model was:

$$\boldsymbol{y}_b = \mathbf{1}\mu_b + \boldsymbol{Z}_b\boldsymbol{\beta}_b + \boldsymbol{e}_b$$

Where $\boldsymbol{y}_b$ is the vector of phenotypes of all the n_b purebred animals for the specific pure breed; μ_b is the general mean (scalar); $\boldsymbol{Z}_b$ is the genotype matrix with for breed b coded as 0, 1, 2 for allele counts, $\boldsymbol{\beta}_b$ is the vector of random marker effects; $\boldsymbol{e}_b$ is the residual vector. Marker effects were fitted simultaneously, with $\beta_b \sim N(0, \boldsymbol{I}\sigma^2_{\beta_b})$. The same model was applied separately for HOL, JER and RDC. The GEBV for a crossbred animal i was calculated with the BOM method as follows (Eiríksson et al., 2021):

$$GEBV_{BOM,i} = \sum_{b=1}^{3} (\boldsymbol{z}'_{i,b}\widehat{\boldsymbol{\beta}}_b + p_{i,b} \cdot \hat{\mu}_b)$$

Where $\boldsymbol{z}_{i,b}$ is a vector of breed b allele contributions for the animal i, $\widehat{\boldsymbol{\beta}}_b$ is the vector of SNP marker effects estimated from the model for breed b; $p_{i,b}$ is the the genome-wide average proportion of alleles assigned to breed b in animal i; $\hat{\mu}_b$ is the breed-specific intercept estimated from the model for breed b. The term $\boldsymbol{p}_{i,b}\widehat{\boldsymbol{\mu}}_b$ accounts for breed-specific baseline differences, whereas $\boldsymbol{z}'_{i,b}\widehat{\boldsymbol{\beta}}_b$ captures within-breed marker effects.

BOA. The model to estimate breed-specific SNP effects using LA information is as follows (Karaman et al., 2021):

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X}\boldsymbol{b} + \boldsymbol{M}_{HOL}\boldsymbol{\beta}_{HOL} + \boldsymbol{M}_{JER}\boldsymbol{\beta}_{JER} + \boldsymbol{M}_{RDC}\boldsymbol{\beta}_{RDC} + \boldsymbol{e},$$

Where $\boldsymbol{y}$ is the $n \times 1$ vector of phenotypes for purebred and crossbred animals, where n is the total number animals: μ is the overall mean; $\boldsymbol{X}$ is the $n \times 3$ matrix of average breed proportions, computed identically to BOM, and its (i, b) entry equals $p_{i,b}$ in BOM (where i indexes animal and b denotes breed: HOL, JER and RDC); $\boldsymbol{b}$ is a 3×1 vector of fixed breed effects; $\boldsymbol{M}_{\boldsymbol{HOL}}$, $\boldsymbol{M}_{\boldsymbol{JER}}$ and $\boldsymbol{M}_{\boldsymbol{RDC}}$ are $n \times m$ matrices of SNP allele content specific to each breed, based on LA; $\boldsymbol{e}$ is the vector of residuals. Shrinkage of SNP effects was imposed through the prior distributions assigned to the SNP effects, following Karaman et al. (2021). Alleles were first assigned to breed-specific matrices, then centered using breed-specific allele frequencies with missing entries set to zero, and finally summed across the two alleles at each locus to obtain the genotype matrices $\boldsymbol{M}_{\boldsymbol{HOL}}$, $\boldsymbol{M}_{\boldsymbol{JER}}$ and $\boldsymbol{M}_{\boldsymbol{RDC}}$.

After estimating breed effects and breed-specific SNP marker effects, the GEBV of each animal was calculated as:

$$\boldsymbol{GEBV}_{\boldsymbol{BOA}} = \boldsymbol{X}\widehat{\boldsymbol{b}} + \boldsymbol{M}_{\boldsymbol{HOL}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{HOL}} + \boldsymbol{M}_{\boldsymbol{JER}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{JER}} + \boldsymbol{M}_{\boldsymbol{RDC}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{RDC}}$$

BOM and BOA were fitted in Julia using the NextGP.jl statistical analysis package (Karaman et al., 2023), using a Bayesian approach with Markov chain Monte Carlo for parameter estimation and genomic prediction.

Evaluation. Predictive ability was measured as the Pearson correlation between GEBVs and corrected phenotype in the validation set. Dispersion bias was assessed as the regression slope of corrected phenotype on estimated breeding values, where values close to 1 indicate good calibration.

ssGBLUP. To quantify the benefit of incorporating phenotypes from non-genotyped animals via pedigree links, we performed a ssGBLUP using DMU5 (Madsen, 2012). As a baseline, we also performed aPBLUP analysis using the same model but replacing the genomic relationships with pedigree-based relationships.

Results & Discussion

Average breed proportions of the crossbred reference population were 0.50 HOL, 0.16 JER, and 0.33 RDC. *Figure 1* compares breed proportions derived from pedigree and genomic information and shows broadly similar patterns across animals, supporting consistent breed composition estimates between the two information sources.

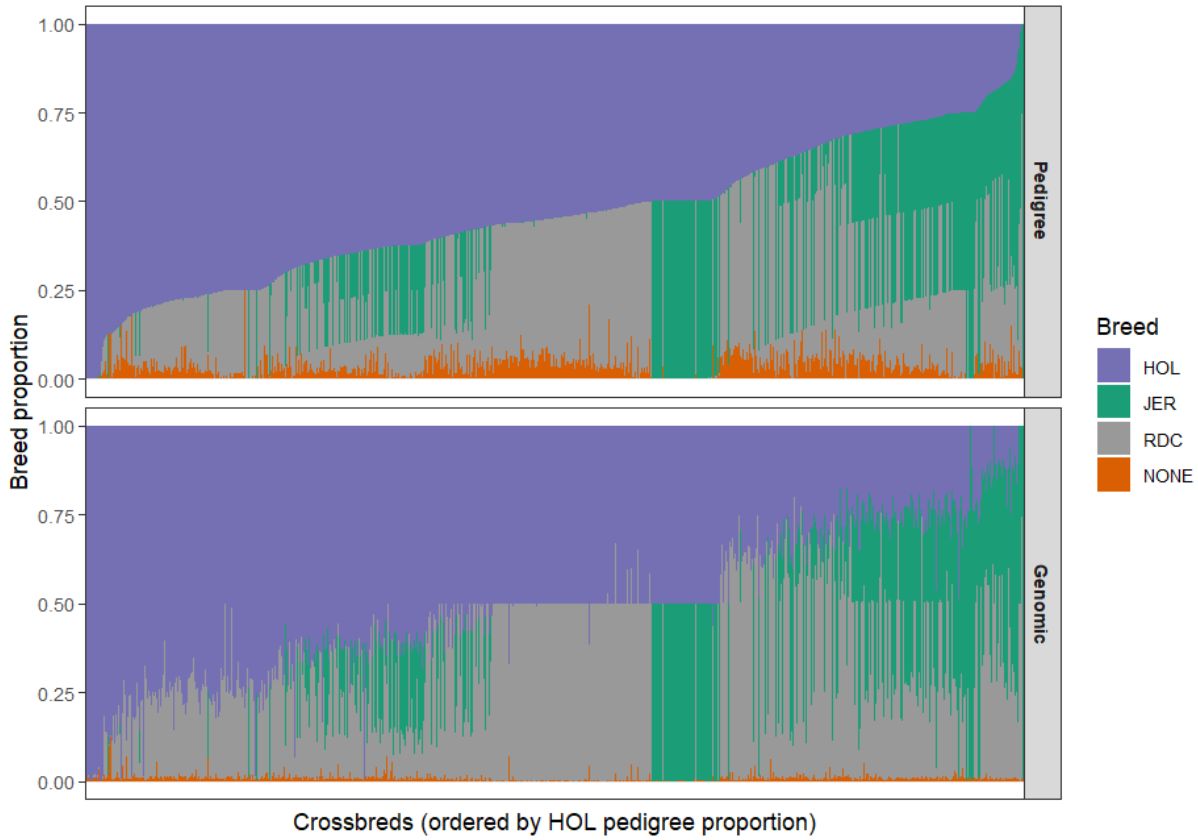

Figure 1. Breed proportions of crossbred animals inferred from pedigree and genomic information. Each vertical bar represents one crossbred animal, ordered by pedigree-based Holstein (HOL) proportion. The upper panel shows pedigree-based breed proportions, and the lower panel shows genomic breed proportions inferred from local ancestry by AllOr. Colors indicate HOL, Jersey (JER), Red Dairy Cattle (RDC), and unassigned proportion (NONE).

As presented in Table 1, predictive ability improved substantially when moving from PBLUP to genomic approaches. PBLUP showed the lowest predictive ability (0.447), whereas BOM and BOA increased predictive ability to 0.533 and 0.538, respectively, and ssGBLUP was highest overall (0.558). The predictive ability for phenotyped-only crossbreds were 0.458 from ssGBLUP.

Within the breed-of-origin framework, BOA provided only a small advantage over BOM by including of crossbred phenotypes. This may be because the training population was still dominated by purebred animals, limiting the additional impact of crossbred information. Moreover, any errors in local ancestry inference could reduce the ability of BOA more than BOM to fully capture breed-specific effects. Calibration also differed across methods: PBLUP had the highest regression slope (1.241), suggesting more deflated predictions, whereas ssGBLUP showed the best calibration among the genomic models (0.955). BOA also improved dispersion relative to BOM (0.913 vs. 0.733), indicating better calibrated predictions under BOA.

Importantly, ssGBLUP and breed-of-origin models exploit different information pathways: ssGBLUP gains from pedigree links and information from additional relatives, whereas BOA

and BOM aim to better represent admixed genetic architecture. Therefore, a natural next step is to develop single-step extensions of breed-of-origin models, where benefits may be partly additive by combining improved modelling of breed-specific effects with information from relatives.

Table 1. Predictive ability and dispersion (calibration) across methods. Predictive ability was the Pearson correlation between corrected phenotypes and estimated breeding values in the genotyped validation set. Dispersion was assessed as the regression slope of corrected phenotypes on breeding values.

	Accuracy	Bias
BLUP	0.447	1.241
BOM	0.533	0.733
BOA	0.538	0.913
ssGBLUP	0.558	0.955

Conclusions

Overall, genomic models clearly outperformed PBLUP, and BOA improved predictive ability and calibration relative to BOM. Conventional single-step genomic best linear unbiased prediction (ssGBLUP) provided the best overall calibration and predictive ability, making it useful for exploiting phenotypes from non-genotyped relatives through pedigree connections. These results motivate developing a single-step extension of local ancestry modelling (ssBOA), which could combine allele-level breed-of-origin modelling with the information gains from the full recorded population.

Acknowledgement

This work was supported by the Milk Levy Fund (Mælkeafgiftsfonden). Project information is available at https://qgg.au.dk/forskning/projekter/aktuelle-projekter/forbedrede-genomiske-vaerktoejer-for-krydsningskoeer. The authors declare no conflicts of interest.

References

Eiríksson, J. H., E. Karaman, G. Su, and O. F. Christensen. (2021). Genet. Sel. Evol. 53:84. https://doi.org/10.1186/s12711-021-00678-3.

Eiríksson, J. H., K. Byskov, G. Su, J. R. Thomasen, and O. F. Christensen. (2022). J. Dairy Sci. 105:5178–5191. https://doi.org/10.3168/jds.2021-21644.

Karaman, E., G. Su, I. Croué, and M. S. Lund. (2021). Genet. Sel. Evol. 53:46. https://doi.org/10.1186/s12711-021-00637-y.

Karaman, E., V. Milkevych, and L. Janss. (2023). Book of Abstracts of the 74th Annual Meeting of the European Federation of Animal Science. Wageningen Academic Publishers, p. 514. https://doi.org/10.3920/978-90-8686-936-7.

Guillenea, A., G. Su, M. S. Lund, and E. Karaman. (2022) J. Dairy Sci. 105:2426–2438. https://doi.org/10.3168/jds.2021-21173.

Madsen, P., and J. Jensen. (2012) A user's guide to DMU: A package for analyzing multivariate mixed models. Version 6, release 5.1. https://dmu.ghpc.au.dk/dmu/DMU/Doc/Previous/dmuv6_guide.5.1.pdf. Accessed Jan. 29, 2026.

Sørensen, M. K., E. Norberg, J. Pedersen, and L. G. Christensen. (2008). J. Dairy Sci. 91:4116–4128. https://doi.org/10.3168/jds.2008-1273.