Running title: Across-country genomic predictions in sheep

Across-country genomic predictions in Norwegian and New Zealand Composite sheep populations with similar development history

Oliveira, H. R.[1,2], McEwan, J. C.[3], Jakobsen, J. H.[4], Blichfeldt, T.[4], Meuwissen, T. H. E.[5], Pickering, N. K.[6], Clarke, S. M.[3], Brito, L. F.[1,2,*]

[1] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, 47907, United States of America

[2] Centre for Genetic Improvement of Livestock (CGIL), Department of Animal Biosciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

[3] AgResearch Limited, Invermay Agricultural Centre, 50034 Mosgiel, New Zealand

[4] The Norwegian Association of Sheep and Goat Breeders, Box 104, 1431 Ås, Norway

[5] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Arboretveien 6, 1432 Ås, Norway

[6] Focus Genetics, 17 Mahia St, Ahuriri, Napier 4144, New Zealand

***Correspondence:**

Dr. Luiz F. Brito

E-mail: britol@purdue.edu

Mailing address: Department of Animal Sciences, Purdue University, West Lafayette, Indiana, 47907, USA

Abstract

The goal of this study was to assess the feasibility of across-country genomic predictions in Norwegian White Sheep (NWS) and New Zealand Composite (NZC) sheep populations with similar development history. Different training populations were evaluated (i.e., including only NWS or NZC, or combining both populations). Predictions were performed using the actual phenotypes (normalized) and the single-step GBLUP via Bayesian inference. Genotyped NWS animals born in 2016 (N = 267) were used to assess the accuracy and bias of genomic estimated breeding values (GEBVs) predicted for birth weight (BW), weaning weight (WW), carcass weight (CW), EUROP carcass classification (EUC), and EUROP fat grading (EUF). The accuracy and bias of GEBVs differed across traits and training population used. For instance, the GEBV accuracies ranged from 0.13 (BW) to 0.44 (EUC) for GEBVs predicted including only NWS, from 0.06 (BW) to 0.15 (CW) when including only NZC, and from 0.10 (BW) to 0.41 (EUC) when including both NWS and NZC animals in the training population. The regression coefficients used to assess the spread of GEBVs (bias) ranged from 0.26 (BW) to 0.64 (EUF) for only NWS, 0.10 (EUC) to 0.52 (CW) for only NZC, and from 0.42 (WW) to 2.23 (EUC) for both NWS and NZC in the training population. Our findings suggest that across-country genomic predictions based on ssGBLUP might be possible for NWS and NZC, especially for novel traits.

Keywords: carcass, GBLUP, breeding value, Norwegian White sheep, single-step, weight

Introduction

In various livestock species, the number of genotyped animals has exponentially increased over the past few years (Koivula, Strandén, Aamand, & Mäntysaari, 2018; Misztal & Legarra,

2017). However, for some specific populations the number of genotyped animals is still limited, which can compromise the current success of genomic predictions (VanRaden et al., 2009). Furthermore, in the case of sheep, there is a much larger number of commonly-raised breeds, with larger effective population sizes compared to the main livestock species, i.e. dairy cattle, beef cattle, pigs, and poultry (Brito et al., 2017a; Kijas et al., 2012). For instance, only recently have producers started to genotype animals from the Norwegian White Sheep (NWS) breed, which has limited the availability of genotypes from this population. Increasing the size of the training population by including genetically similar animals with both genotype and phenotype is expected to increase the prediction accuracy of genomic breeding values (GEBVs; Daetwyler, Villanueva, & Woolliams, 2008; Oliveira et al., 2019a; Uemoto, Osawa, & Saburi, 2017).

In general, NWS lambs are produced similarly to lambs in New Zealand, i.e., dual purpose (meat and wool) composite sheep, where about 10-20% of lambs born are derived from terminal crosses. Moreover, the breeding objectives and evaluation methodology used in both countries have converged over the years, as both economic returns are primarily derived from lamb meat production and similar consumers preferences. Details regarding the current indices used in New Zealand are available at https://www.sil.co.nz/technical/technical-notes, and underlying bio-economic modelling methodology has been outlined most recently by Bryne et al. (2012). In contrast, trait weighting in Norway is currently based on desired gains after surveying farmers preferences (e.g., https://www.saueavl.nsg.no/vaer_list.cfm), although previously a bio-economic model has been used (Eikje et al. 2008).

As expected, in Norway early growth and increased carcass weight have a strong emphasis both via direct and maternal pathways. Moreover, there is also a weighting against carcass fat and for lean content. Historically, Norway and New Zealand had strong emphasis on number of lambs

born, but in recent years this has decreased as reproduction rates have risen. Similarly, the emphasis on wool type and weight is now negligible. Selection index weighting differences include a stronger emphasis in New Zealand on lamb survival, improved disease resistance, and penalties for increased adult live weight. Additionally, it is worth to mention that sheep in Norway are grazed extensively on unimproved pasture during the summer for around 100 days, indoors on conserved pasture consisting of timothy for about 200 days during the winter, and on improved pasture during spring and fall, while New Zealand sheep are grazed and lambed primarily on improved ryegrass white clover pasture all year long.

Regardless of the breeding objectives and production system used in Norway and New Zealand, a recent study showed that there are relatively high genetic similarities between NWS and several New Zealand composite (NZC) sheep populations, because they had similar development history as a consequence of overlapping founder breeds (Oliveira et al., 2020). Therefore, the authors suggested that a collaborative initiative among Norway and New Zealand could be a feasible alternative to increase the accuracy of genomic predictions for traits recorded in both countries, as well as to allow the prediction of GEBVs for traits recorded in only one of the countries (or populations). However, no previous attempts have been performed to test this hypothesis.

The successful use of a training population composed by various populations strongly depends on the statistical methods used to combine all the information. Studies have shown that simultaneously combining phenotypic records, pedigree, and genomic information in the single-step Genomic Best Linear Unbiased Prediction approach (ssGBLUP; Aguilar et al., 2010; Christensen & Lund, 2010; Misztal, Legarra, & Aguilar, 2009) can lead to more accurate and less biased GEBVs, even when combining different breeds or populations in the training population

(Carillier, Larroque, & Robert-Granié, 2014). However, when comparing two methods to perform genomic predictions in dairy cattle, Hayes et al. (2009) concluded that combining different breeds in the same training population was not as effective as the within-breed analyses, even though the GEBVs were more accurate than the traditional parent average (PA). In this context, the main objective of this study was to compare the performance of different approaches to perform genomic predictions for five traits from NWS (i.e., birth, weaning, and carcass weights, EUROP carcass classification, and EUROP fat grading), using the ssGBLUP and a high-density SNP panel (~ 606K SNPs). The approaches tested were: 1) including only NWS in the training population; 2) including NWS and the more genetically similar NZC sheep populations as reported by Oliveira et al. (2020; i.e., Primera, Lamb Supreme, and "Other Dual-purpose") in the training population; and 3) including only the mentioned NZC sheep populations in the training population. Additionally, when combining NWS and different NZC sheep populations in the training, the different populations were analyzed under a single-breed approach, and under a multiple-breed approach considering NWS and NZC as different populations.

Material and methods

The data used in this study are from the traditional routine genetic evaluations for breeding flocks performed in Norway by the Norwegian Association of Sheep and Goat Breeders (NSG; Ås, Norway), and in New Zealand by individual breeders as part of the FarmIQ project (AgResearch; Mosgiel, New Zealand). Therefore, no Animal Care Committee approval was necessary for the purposes of this study.

Phenotypic data

The phenotypic data included information from 2002 to 2019 for Norway, and from 1990 to 2015 for New Zealand. Norwegian animals were from the NWS breed, and NZC animals were from three different NZC sheep populations: Primera, Lamb Supreme, and "Other Dual-purpose". Genetic similarities among NWS and these NZC sheep populations are shown in Oliveira et al. (2020).

As the phenotypic recording systems in both countries are independent, different traits (and/or data collection methods) were used. The traits available for Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). The traits available from New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye-muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm). Details about these traits and data recording processes performed in Norway and New Zealand can be found in Eikje, Ådnøy, & Klemetsdal (2008) and Brito et al. (2017b), respectively.

Phenotypic quality control was performed independently for each trait/country. Phenotypes were discarded if they were lower or higher than the mean ± 3.0 standard deviations (SD) within contemporary group. Contemporary groups were defined by the combination of country, flock, year, sex, and litter size. In addition, it was required that each contemporary group contained at least five animals, and that all animals with phenotypic data had age recorded. Descriptive statistics for each trait, after the quality control, are shown in Table 1. Average (SD) age of animals at the weaning and slaughter date were 137.45 (13.43) and 156.50 (20.18) days in Norway; and 88.76 (14.38) and 158.10 (27.54) days in New Zealand, respectively.

Test of homogeneity. In order to assure the homogeneity of traits for the genomic predictions (i.e., to confirm that the same traits were measured in both countries), a bootstrapping method (with replacement) was used to test for significance. In this context, 1,000 independent samples of 50 animals (with records) were used in a two-tailed F-test (Snedecor & Cochran, 1989) to evaluate the homogeneity of variances among residuals from the NWS and NZC populations. Residuals were estimated after adjusting for the systematic (i.e., fixed) effects in the model, i.e., flock, year, sex, litter size, and age. In addition, an independent t-test (Snedecor & Cochran, 1989) was used to statistically determine differences among trait means in the two populations. The t-test performed considered either equal or unequal variances, depending on the results of the F-test for each trait comparison. The pair of traits compared were: 1) BW_{NO} and BW_{NZ}; 2) WW_{NO} and WW_{NZ}; 3) CW_{NO} and CW_{NZ}; 4) CW_{NO} and CCW_{NZ}; 5) CW_{NO} and XCW_{NZ}; 6) EUC_{NO} and EMD_{NZ}; 7) EUC_{NO} and BUT_{NZ}; 8) EUF_{NO} and FGR_{NZ}; and 9) EUF_{NO} and FDM_{NZ}.

For both F- and t-test, significance was assumed when more than 5% of the tests (performed in the 1,000 independent samples) showed p-value < 0.05. Traits were normalized (following the Student's t-statistic; Snedecor & Cochran, 1989) after defining the pair of traits in which the F- and/or t-test was significant. Therefore, the normalized trait in the animal i (t_i) was calculated as:

$$t_i = \frac{X_i - \overline{X}}{s},$$

in which X_i is the phenotyped measured in the animal i, $\overline{X}$ is the country trait mean, and s is the SD estimated for the trait within each country.

Pedigree and genomic data

The pedigree file contained 3,174,345 and 206,180 animals from Norway and New Zealand, respectively, which included up to 10 generations back from the phenotyped animals. From these animals, a total of 792 NWS and 16,912 NZC animals were genotyped using a high-density (HD) SNP panel (Ovine Infinium® HD SNP Beadchip; developed through the International Sheep Genome Consortium and the FarmIQ project, AgResearch, New Zealand). From the NZC genotyped animals, 8,554; 6,092; and 1,831 animals were from Primera, Lamb Supreme, and "Other Dual Purpose" populations, respectively.

Three different genotypic quality controls were performed: 1) individually for NWS; 2) for NZC plus 267 NWS animals used in the validation (explained in the "*Training and validation populations*" section under the "*Genomic predictions"* topic); and 3) considering all sheep populations together. During the quality control, SNPs with unknown genomic positions and/or located in the sexual chromosomes, minor allele frequency (MAF) lower than 0.05, sample or SNP call rate lower than 95%, and extreme departure from the Hardy Weinberg equilibrium (*p-value* < 10^{-15}) were excluded. Genotypic quality controls were performed using the preGSf90 software (Aguilar at al., 2014; Misztal et al., 2002). The total number of genotyped animals and SNPs that remained after the quality control were 792 and 486,945 considering only NWS; 16,171 and 508,026 considering NZC and NWS validation animals; and 16,696 and 508,856 considering all sheep populations together, respectively.

Genomic predictions

Training and validation populations. After the quality control, reduced datasets were created from the full datasets, which included all phenotypic information available until 2015. Thereafter, the reduced datasets were used to predict GEBVs (using ssGBLUP), and traditional

breeding values (EBVs; using BLUP) in the three different scenarios: considering only NWS animals, only NZC animals, and both NWS and NZC animals together in the training population. Details about the training population used for each trait are included in the Table S1 (Supplementary Material).

The full dataset from Norway, which included phenotypic data up to 2019, was used to predict current EBVs for NWS animals. The current EBVs were used as a benchmark to validate GEBVs and EBVs obtained from the reduced datasets by assessing the accuracy and bias (defined in the "*Accuracy and bias"* section) of genomic predictions in the validation population. Using a more accurate EBV, such as the current EBV (i.e., EBV predicted using animals' own phenotypes and/or phenotypes from their offspring), as a benchmark for the genomic predictions has been an usual practice in animal breeding to represent the unknown true breeding values (e.g., Hayes et al., 2009; Edel et al., 2011; Badke et al., 2014; Weller et al., 2015; Oliveira et al., 2019b). Genotyped NWS animals born in 2016 (N = 267 males) were used as validation population in all the analyses.

Systematic effects. Systematic effects included in the model were contemporary group (defined by the combination of country, flock, year, sex, and litter size; for all traits), age of the dam (for BW), and the age of the animal at the measurement (for all traits, except BW), as they were found to be significant (*p-value* < 0.05) for both countries. Ages (i.e., age of the dam and animal) were assumed as linear covariables in the model.

BLUP and variance components estimation. The EBVs were predicted using BLUP, based on single-trait models. In matrix notation, the general model used for all traits in this study is defined as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \quad [1]$$

where **y**, **b**, **u**, and **e** are the vectors of observations (normalized when needed), systematic effects (i.e., contemporary group and age, as previously described), additive genetic random effects, and random residuals, respectively. The **X** and **Z** are the incidence matrices for **b** and **u**, respectively.

The assumptions made for the model [1] under the single-breed approach (used when including only NWS, only NZC, or when including both NWS and NZC animals in the training population without any differentiation among them) are:

$$\mathbf{y}|\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Zu}, \sigma_e^2); \mathbf{b} \sim N(0, \mathbf{\Sigma_b}), \text{ and } \mathbf{u}|\sigma_u^2, \mathbf{A} \sim N(0, \sigma_u^2 \otimes \mathbf{A}), \quad [2]$$

in which σ_u^2 and σ_e^2 are the additive genetic and residual variances, respectively; $\mathbf{\Sigma_b}$ is a diagonal matrix with large values to represent vague prior knowledge; and **A** is the traditional pedigree-based relationship matrix. Under this approach, σ_u^2 and σ_e^2 were assumed a priori to follow an inverted chi-square distribution, such that $\sigma_u^2|v_u, S_u^2 \sim \chi^{-2}(v_u, S_u^2)$ and $\sigma_e^2|v_e, S_e^2 \sim \chi^{-2}(v_e, S_e^2)$. Variance components and EBVs were not estimated using the traditional pedigree-based relationship matrix under the multiple-breed approach because no covariance exists between NWS and NZC (i.e., pedigree files are independent).

All variance components were estimated using a Bayesian approach, using the Markov chain Monte Carlo (MCMC) framework and the Gibbs sampler algorithm available in the gibbs2f90 software (Misztal et al., 2002). A MCMC chain length of 250,000 cycles, considering a burn-in period of 50,000 cycles, and a sampling interval (thin) of 10 cycles were used in all analyses. The convergence was verified through graphical analysis and the Geweke criterion (Geweke, 1992), both available in the Bayesian Output Analysis package (Smith, 2007) of the R software (R Core Team, 2016). Variance components were estimated using the **A** matrix in the reduced and full datasets. Thereafter, variance components estimated based on the reduced datasets

were considered in the ssGBLUP analyses and used to estimate the genetic parameters (heritabilities and genetic correlations). Estimating variance components and genetic parameters based solely on the **A** matrix has been a strategy currently performed in several recent ssGBLUP studies (e.g., Kang et al., 2018; Oliveira et al., 2019b).

The single-breed ssGBLUP approach. The ssGBLUP was used to predict the GEBVs by jointly combining phenotypic, pedigree, and genotypic information using the blupf90 software (Misztal et al., 2002). The same statistical model and assumptions used in BLUP (i.e., models and assumptions showed in [1] and [2]) were made in the ssGBLUP. However, the **A** matrix was replaced by the **H** matrix (Aguilar et al., 2010; Christensen & Lund, 2010; Misztal et al., 2009). The **H** matrix is defined as:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (0.95\mathbf{G} + 0.05\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad [3]$$

where **A** is the pedigree-based relationship matrix, $\mathbf{A}_{22}$ is the portion of the **A** matrix related to the genotyped animals, and **G** is the genomic relationship matrix, which was created as follow (VanRaden, 2008):

$$\mathbf{G} = \frac{\mathbf{WW'}}{2\sum_{j=1}^{n} p_j(1-p_j)}, \quad [4]$$

in which $\mathbf{W} = (\mathbf{M} - \mathbf{P})$, **M** is a matrix containing the centered genotypes (−1, 0, and 1 representing AA, Aa, and aa, respectively), and **P** is a matrix that contains the allele frequency for the SNP j [i.e., $2(p_j - 0.5)$] in its jth column.

The multiple-breed ssGBLUP approach. As an attempt to estimate the additive genetic covariance among populations and improve the accuracy of genomic predictions, variance

components and GEBVs were also estimated using the **H** matrix using multiple breed groups. The assumptions made for the model [1] under the multiple-breed approach are:

$$\mathbf{y}|\mathbf{b},\mathbf{u},\mathbf{G_0},\mathbf{R_0} \sim N(\mathbf{Xb}+\mathbf{Zu},\mathbf{R_0}\otimes\mathbf{I}); \ \mathbf{b}\sim N(0,\mathbf{\Sigma_b}\otimes\mathbf{I}), \text{ and } \mathbf{u}|\mathbf{G_0},\mathbf{A}\sim N(0,\mathbf{G_0}\otimes\mathbf{H}); \quad [5]$$

in which $\mathbf{G_0}$ and $\mathbf{R_0}$ are the additive genetic and residual (co)variance matrices, respectively; **I** is an identity matrix; and all the other terms were previously defined. The $\mathbf{G_0}$ and $\mathbf{R_0}$ matrices were assumed to follow an inverted Wishart distribution, $\mathbf{G_0}|v_u,\mathbf{V}_u^2 \sim IW(v_u,\mathbf{V}_u^2)$ and $\mathbf{R_0}|v_e,\mathbf{V}_e^2 \sim IW(v_e,\mathbf{V}_e^2)$; and they can be described as:

$$\mathbf{G_0}=\begin{bmatrix}\sigma^2_{u_{NWS}} & \sigma_{u_{NWS,NZ}}\\ \sigma_{u_{NWS,NZ}} & \sigma^2_{u_{NZ}}\end{bmatrix}, \text{ and } \mathbf{R_0}=\begin{bmatrix}\sigma^2_{e_{NWS}} & 0\\ 0 & \sigma^2_{e_{NZ}}\end{bmatrix}, \quad [6]$$

where $\sigma^2_{u_{NWS}}$ and $\sigma^2_{u_{NZ}}$ are the additive genetic variances for NWS and the NZC populations, respectively, and $\sigma_{u_{NWS,NZ}}$ is the additive genetic covariance between NWS and the NZC populations. The $\sigma^2_{e_{NWS}}$ and $\sigma^2_{e_{NZ}}$ are the residual variances for NWS and the NZC populations, respectively.

Accuracy and bias. For each trait, the accuracy of genomic prediction was estimated as the Pearson correlation coefficient calculated between GEBVs predicted using the reduced dataset and the EBVs predicted using the full dataset, for the validation population. In addition, the bias of GEBVs of validation animals was assessed using the regression coefficient estimated using a linear regression of GEBVs (predicted using the reduced datasets) on the EBVs predicted using the full datasets (i.e., b_1 obtained from $EBV_{full} = b_0 + b_1 \times GEBV$). In order to estimate the changes in accuracy and bias due to the use of genomic information, accuracies and bias were also estimated for the EBVs predicted using the reduced datasets (using the Pearson correlation and $EBV_{full} = b_0 + b_1 \times EBV$, respectively), for the validation animals.

Results

Traits comparison

In order to assure the homogeneity of phenotypes for the genomic predictions, nine pairs of traits were compared: 1) BW_{NO} and BW_{NZ}; 2) WW_{NO} and WW_{NZ}; 3) CW_{NO} and CW_{NZ}; 4) CW_{NO} and CCW_{NZ}; 5) CW_{NO} and XCW_{NZ}; 6) EUC_{NO} and EMD_{NZ}; 7) EUC_{NO} and BUT_{NZ}; 8) EUF_{NO} and FGR_{NZ}; and 9) EUF_{NO} and FDM_{NZ}. The percentage of statistical tests with p-value < 0.05, for both F- and t-tests, are shown in Table 2.

Based on the results from the F- and t-test (Table 2), all the pair of traits compared have different variance and trait means (before trait normalization; except for BW_{NO} and BW_{NZ}, in which no significant difference among means was observed). These results suggest that the traits are recorded differently between countries, and they indicate that data normalization is needed before combining the data from Norway and New Zealand in the same genomic evaluation. Therefore, genetic parameters and estimation of breeding values performed in this study used normalized phenotypes for all traits.

Heritabilities

Heritabilities estimated for each trait under the single-breed approach, in each scenario using the reduced datasets, are shown in Table 3. Posterior means and 95% highest posterior density intervals (inside brackets) for heritability estimated for NWS using the full datasets were 0.48 [0.477-0.478], 0.40 [0.404-0.408], 0.53 [0.531-0.533], 0.45 [0.452-0.455], and 0.46 [0.460-0.463], for BW_{NO}, WW_{NO}, CW_{NO}, EUC_{NO}, and EUF_{NO}, respectively.

Similar heritabilities were estimated when including only NWS in the analysis, using either the full or reduced datasets. In general, the heritabilities estimated for NZC were lower than the estimates for NWS (except for BW_{NO} and BW_{NZ}, in which both had similar heritabilities). The heritabilities estimated when combining data from Norway and New Zealand under the single-breed approach tended to have intermediate values when compared to the use of these populations separately. However, especially for the carcass traits (i.e., CW_{NO}, CW_{NZ}, CCW_{NZ}, and XCW_{NZ}) combining both data yielded higher heritability estimates. Heritabilities estimated using the multiple-breed approach were similar to the heritabilities estimated using NWS and NZC separately, for each trait. Heritabilities and genetic correlations estimated for each trait under the multiple-breed approach are shown in Table S2 (Supplementary Material).

Genomic predictions

Accuracy and bias of EBVs and GEBVs predicted using the reduced datasets and the single-breed approach are shown in Figure 1. Corresponding results predicted using the multiple-breed approach are shown in Figure S2 (Supplementary Material). Accuracies and biases of predictions of GEBVs, were similar when using either the single- and multiple-breed approaches.

The lowest accuracies were found for BW (ranged from 0.06 to 0.13), regardless of the information included in the training population (Figure 1a). On the other hand, the highest accuracies tended to be estimated for EUC and EUF (from 0.10 to 0.56 for EUC, and from 0.09 to 0.40 for EUF). However, different patterns were observed among traits. For instance, EBVs tended to be more accurately predicted than GEBVs for EUC (regardless of the genomic information included in the training population), and less accurately predicted than GEBVs when including NWS in the training population (either alone or combined with NZC data) for EUF. Accuracies

estimated for WW (0.08 to 0.27) and CW (0.13 to 0.25) were intermediate (higher than accuracies estimated for BW and lower than accuracies estimated for EUC and EUF), and similar for EBVs and GEBVs predicted using NWS data (alone or combined with NZC). Predictions using only NZC yielded the lowest accuracies for all traits (it ranged from 0.06 to 0.15 among traits). Including genomic information from NWS in the training population seems to increase the accuracies compared to EBVs for BW.

In general, EBVs and GEBVs predicted for BW using data from both NWS and NZC in the training population were strongly deflated (regression coefficient equal to 1.76), while EBVs and GEBVs predicted using either NWS or NZC alone were strongly inflated (regression coefficients ranged from 0.19 to 0.29; Figure 1b). For WW, GEBVs predicted using the different approaches were inflated (regression coefficients ranged from 0.35 to 0.42). For CW, combining NWS and NZC in the training population yielded the least biased predictions (i.e., regression coefficients ranged from 0.93 to 0.95) when compared to EBVs and GEBVs using the other training populations (regression coefficients ranged from 0.31 to 0.52). For EUC, regression coefficients closer to one were obtained for EBVs (0.78) and GEBVs predicted using only NWS in the training population (0.68). However, strongly biased predictions were obtained for EUC when using only NZC (regression coefficients were 0.11 and 0.17 when using BUT and EMD, respectively), or both NWS and NZC (1.51 for BUT and 2.23 for EMD) in the training population, regardless of the phenotype measured in New Zealand. For EUF, less biased predictions were obtained for GEBVs predicted using only NWS (0.65) or both NWS and NZC (regression coefficients ranged from 0.44 to 0.53) in the training population, when compared to EBVs (0.25) and GEBVs predicted using only NZC (regression coefficients were 0.19 to 0.24 when using FGR and FDM, respectively).

Discussion

The performance of genomic prediction relies, among other factors, on the size of the training population (Daetwyler, Villanueva, & Woolliams, 2008; Oliveira et al., 2019a; Uemoto, Osawa, & Saburi, 2017). In this context, some countries have agreed on sharing data to enlarge the training population for the genomic predictions in some livestock species. For instance, the SMARTER project (https://www.smarterproject.eu) has brought together data from multiple countries. This is an recent initiative that aims to improve resilience and efficiency in small ruminants across a range of different environments. However, official attempts to perform across-country genomic evaluations for sheep are still scarce in the literature. For instance, Legarra et al. (2014) performed within- and across-breed genomic predictions for various Western Pyrenees dairy sheep breeds raised in France and Spain, using single- and multiple-breed predictions. The authors concluded that genomic evaluations are more accurate than pedigree-based ones, but that no advantages were observed when combining all data together for the genomic predictions.

Official genomic multiple across-country evaluations (GMACE) have been routinely performed for dairy cattle by Interbull (VanRaden & Sullivan, 2010). In general, GMACE use the Mendelian sampling deviations as dependent variables in the evaluation model, which is calculated from the difference between GEBVs predicted inside the country and the parent average predicted using the traditional (pedigree-based) multiple across-country evaluation (MACE) method proposed by Schaeffer (1994). Thus, the success of the genomic prediction relies on the performance of the traditional MACE, which requires the use of pseudo-phenotypes (i.e., deregressed proofs) to assure that the phenotypes are independent (Fragomeni et al., 2019; Vandenplas et al., 2017).

It is important to highlight that several challenges have been faced when performing GMACE, such as: 1) high number of steps needed before performing the evaluations; 2) differences in the genetic and genomic evaluation systems performed inside each country; 3) estimation of deregressed proofs and their potential differences in accuracies depending on the deregression method used in each country (Oliveira et al., 2018; Oliveira et al., 2019a); 4) bias generated due to the preselection (Masuda, VanRaden, Misztal, & Lawlor, 2018); and 5) it usually requires strong trade in semen/animals between countries. In order to overcome similar challenges observed in multiple-step genomic evaluations, the ssGBLUP method has been recommended for within-country genetic evaluations (Aguilar et al., 2010; Misztal et al., 2009). However, to our best knowledge, few studies have used ssGBLUP for across-country genomic evaluations. In this context it is worth noting that the method used in this study (based on actual phenotypes) is different from the methods previously published (e.g., Legarra et al., 2014; Vandenplas et al., 2017; Colinet et al., 2018). For instance, previous studies have focused on the use of pseudo-phenotypes (i.e., daughter yield deviations; Legarra et al., 2014), or in the correction of the system of equations using EBVs and their associated reliabilities, without any explicit deregression step (Colinet et al., 2018; Vandenplas et al., 2017). In this context, Legarra et al. (2014) commented that using pseudo-phenotypes usually brings bias in the genomic predictions.

Even though the methods used by Vandenplas et al. (2017) and Colinet et al. (2018) have contributed to reduce the potential bias generated due to the previous inability to include foreign data in national evaluations (i.e., incomplete data) and overcome the challenges of pseudo-phenotypes, strong connections between the populations is preferred, and the estimation of EBVs is still required. In this context, including actual phenotypes from different countries in the same genomic evaluation using ssGBLUP might provide reasonably accurate GEBV predictions based

388 on phenotypes even if no recent connection between pedigrees exists (i.e., if the populations have
389 similar development history and genetic connectedness). Among the reasons for the lack of studies
390 in the literature using actual phenotypes from different countries is the difficulty to assure that the
391 same traits have been measured, because the data recording is usually independent among
392 countries. As an attempt to evaluate if the same selected traits have been recorded between Norway
393 and New Zealand, similarities between country trait means and variances were investigated for
394 nine pre-defined pair of traits (Table 2). Differences among trait means and variances indicate the
395 need for data normalization before combining data from Norway and New Zealand in ssGBLUP.
396 Heritabilities estimated in this study for the traits measured in Norway are, in general,
397 higher than the heritabilities estimated in the official genetic evaluations performed by NSG for
398 NWS (Blichfeldt & Lewis, 2015; NSG, 2020). These differences are likely a consequence of the
399 phenotypes used, as well as the different effects included in the statistical models. For instance,
400 official evaluations performed by NSG usually include genetic and permanent environment effects
401 of the biological (for BW) and foster (for WW, CW, EUC, and EUF) dams, which were not
402 accounted for in this study. In order to have the same statistical model for both countries (required
403 by the single-breed approach), maternal effects were not included in this study due to limited data
404 structure from New Zealand. Liu et al. (2015) has reported and explained this inflation in
405 heritability estimates when not accounting for important maternal effects in the statistical model,
406 using simulated data. The inflated heritabilities estimated in this study might result in
407 overestimated genetic gains for the traits measured in Norway, which are not realistic and will
408 likely disappear when the optimal statistical models are used to estimate the variance components.
409 Jia (2017) has shown alternatives to control de overfitting of heritabilities in genomic evaluations
410 using cross-validation.

411 Heritabilities estimated for WW_{NZ}, CW_{NZ}, EMD_{NZ}, and FGR_{NZ} were similar to the
412 heritabilities estimated by Brito et al. (2017c) for a New Zealand sheep population composed of
413 several pure or composite breeds, in which the main contributing ones were Primera, Texel, Lamb
414 Supreme, Coopworth, Romney and East Friesian. However, slightly higher heritabilities were
415 found for XCW_{NZ}, BUT_{NZ}, FDM_{NZ} compared to the authors [Brito et al. (2017c) reported
416 0.17±0.02, 0.25±0.03, 0.28±0.03 for XCW_{NZ}, BUT_{NZ}, FDM_{NZ}, respectively]. Heritabilities
417 estimated for BW_{NZ} in this study were higher than heritabilities reported by McRae et al. (2016)
418 in a New Zealand sheep population composed mainly of Romney, Coopworth, Perendale, and
419 Texel (0.24 ± 0.04). These differences in heritability estimates are likely related to the different
420 populations included in the analyses, as well as the different effects included in the statistical
421 models.

422 Similar heritabilities, accuracies and biases were estimated using both the single- and
423 multiple-breed approaches, for all analyzed traits (Tables 3 and S2, and Figure S1). These similar
424 results are likely due to the fact that the additive genetic covariance estimated between NWS and
425 NZC (i.e., $\sigma_{u_{NWS,NZ}}$) was based exclusively on the genotyped animals (i.e., small proportion of
426 data), as the pedigree files from both countries were not connected. The uncertainty of $\sigma_{u_{NWS,NZ}}$
427 can be inferred based on the large 95% highest posterior density intervals observed for the genetic
428 correlations estimated between both countries (Table S2, Supplementary Material). In this context,
429 it is worth to highlight that the multiple-breed approach has several theoretical advantages
430 compared to the single-breed approach, such as the use of different statistical models and variance
431 components for each population, and that these advantages were not fully explored due to the weak
432 relationship between animals from the two countries. Moreover, the multiple-breed approach can
433 allow us to make inferences regarding genotype by environment interaction (G×E), as G×E is often

analyzed as the genetic correlation estimated between one trait recorded at different environments (Falconer and Mackay, 1996). Thus, if this relationship increases, or if the proportion of genotyped animals increases in both countries, reliable G×E estimates and higher gains in accuracies might be expected when using the multiple-breed approach. This hypothesis is in agreement with Legarra et al. (2014), who commented that advantages of across-country genetic evaluations are only clearly shown when the populations are interbreed often, or when the traits evaluated are under control of quantitative trait loci (QTLs) with large effects. Similar results were also reported by Carillier, Larroque, & Robert-Granié (2014), who compared the variance components estimation and the prediction of GEBVs using similar approaches (i.e., single- and multiple-breed) in Alpine and Saanen goats.

The accuracy of genomic predictions in combined training populations depends on the size of training population and similarities between the breeds grouped. Moreover, Daetwyler et al. (2012) suggest that across-breed genomic predictions might be limited with the 50k SNP chip. In our case, similarities between NWS and the NZC populations used in this study have been previously reported in Oliveira et al. (2020), using several genetic diversity metrics such as consistency of gametic phase, runs of homozygosity, signatures of selection, and admixture. As conclusions, these authors commented that there is relatively high genetic diversity within each sheep population, and that NWS is more genetically related to the Primera, Lamb Supreme and "Other Dual Purpose" populations raised in New Zealand due to the high number of common ancestral breeds used in their development (Oliveira et al., 2020). This relatively high genetic diversity among some NZC and NWS might indicate that they can likely contribute in the genomic evaluations, specially using HD SNP panels.

In this study, the performance of the genomic predictions (in terms of accuracy) varied across traits (Figure 1a), which might be related to the number of phenotypes available for each trait (Table 1), trait heritability (Table 2), and number of animals with both phenotypes and genotypes in the training population (Table S1, Supplementary Material). However, in general, our results suggest that across-country genomic predictions based on ssGBLUP might be possible for NWS and NZC. Thus, even though the accuracies estimated when including only NZC in the training population were lower than the accuracies reported by Brito et al. (2017b), they show a promising opportunity to predict GEBV for traits not currently recorded in Norway. Differences between the accuracies reported by Brito et al. (2017b) and this study are due to the fact that Brito et al. (2017b) predicted GEBVs using a multiple-breed sheep population including only animals from New Zealand, and due to the fact that accuracies reported in this study were not adjusted by the square root of heritabilities. It is important to highlight that different gains in accuracy were observed by Legarra et al. (2014), depending on the breeds included in the analyses. However in their case, the authors commented that the improvement in accuracy observed when combining the Manech Tête Noire and Latxa Cara Negra Navarre breeds was unexpected, as there is low genetic relationship between these breeds. Similarly, Daetwyler et al. (2012) showed that the accuracy of genomic predictions for carcass and meat quality traits when combining data from several sheep breeds raised in Australia tended to be higher for the traits with the larger training population size. However, the mentioned authors also commented that using adjusted phenotypes as a benchmark for the GEBVs resulted in less variable accuracies compared to the use of unadjusted phenotypes (Daetwyler et al., 2012).

The expected benefit of genomic selection for NWS (i.e., difference between the black bars and diagonal strips in Figure 1) reported in this study was nearly small for the majority of traits,

which is related to both the relatively accurate EBVs predicted for this population and the small number of NWS animals with genotypes and phenotypes. In this context, a stronger impact of inclusion of genomic information might be observed when increasing the number of genotyped animals (mainly progeny tested sires) in the training population. However, it is important to highlight that our results suggest an opportunity for genetic improvement of novel traits (i.e., traits where the amount of phenotypes recorded is still limited), if countries share both phenotypes and genotypes, especially if the connection between populations increases (Carillier, Larroque, & Robert-Granié, 2014). However, more studies are needed to validate this hypothesis. A greater exchange of genetic material across countries is also recommended to increase the genetic connectedness among populations from both countries.

The regression coefficients used to assess bias of GEBVs also varied across traits. However, for the majority of traits and scenarios, the GEBVs were biased, indicating that the use of optimal scaling factors to combine the $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ matrices should be investigated while performing across-country genomic evaluations based on ssGBLUP. The use of optimal scaling factors have the potential to reduce bias in the ssGBLUP analysis (Misztal et al., 2013; Tsuruta, Misztal, Aguilar, & Lawlor, 2011). In addition, Tsuruta et al. (2019) showed that strong selection amplifies inflation in genomic predictions, and that using approximated inbreeding coefficients considering unknown parents group (**UPG**) can potentially reduce the bias. In this regard, Macedo et al. (2020) suggested that the use of metafounders should be preferred instead UPG for sheep, as they yield less biased genomic predictions in a population with only genotyped males. Using only genotyped males might yield more biased estimates due to the stronger selection performed for them compared to females. Anyhow, similar range in the regression coefficients were reported by Colinet et al. (2018) and Carillier, Larroque, & Robert-Granié (2014), while analyzing cattle and

goats, respectively. However, Carillier, Larroque, & Robert-Granié (2014) commented that using a within-breed model provided better dispersion of GEBVs. No clear difference was observed in this study (results not shown).

Future studies comparing the predictive performance of different statistical methods (such as GMACE and single-step correcting the system of equations using EBVs and their associated reliabilities) with the method reported in this study are recommended. Moreover, optimal statistical models and scaling factors for the **H** matrix, as well as the use of UPG and metafounders, should be investigated before implementing official genomic evaluations combining NWS and NZC.

Conclusions

Our findings support the feasibility of across-country genomic predictions based on ssGBLUP for birth weight, weaning weight, carcass weight, EUROP carcass classification, and EUROP fat grading, using NWS and NZC. Moreover, accuracies and biases estimated using only NZC in the training population show a promising opportunity to predict GEBV for novel traits or traits not currently recorded in Norway.

Acknowledgements

The authors gratefully acknowledge: FarmIQ (Ministry for Primary Industries' Primary Growth Partnership fund) for funding the work, Pāmu Farms and Focus Genetics for undertaking the progeny test evaluation, and Silver Fern Farms for processing the animals. We would also like to recognize the significant and on-going contribution of AgResearch farm and technical staff in management, trait recording and genotyping. Financial contributions from the Norwegian Research Council are gratefully acknowledged.

Conflict of interest

All authors declare that they have no conflict of interest.

Data availability

The data supporting the results of this article are included within the article and in its Supplementary Material. The raw data cannot be made available, as it is property of the sheep producers in New Zealand and Norway and this information is commercially sensitive.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. *Journal of Dairy Science*, *93*(2), 743–752. https://doi.org/10.3168/jds.2009-2730

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., & Wang, H. (2014). PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*, (August). https://doi.org/10.13140/2.1.4801.5045

Badke, Y. M., Bates, R. O., Ernst, C. W., Fix, J., & Steibel, J. P. (2014). Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *G3: Genes, Genomes, Genetics*, 4(4), 623-631. 10.1534/g3.114.010504

Blichfeldt, T., & Lewis, R. (2015). *Sheep Breeding in Norway*. Retrieved from

548 https://www.nationalsheep.org.uk/workspace/pdfs/blichfeldtthor22112015103318.pdf
549 Brito, L. F., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., Dodds, K. G., … Clarke,
550 S. M. (2017a). Genetic diversity of a New Zealand multi-breed sheep population and
551 composite breeds' history revealed by a high-density SNP chip. *BMC Genetics*.
552 https://doi.org/10.1186/s12863-017-0492-8
553 Brito, L. F., Clarke, S. M., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., …
554 Schenkel, F. S. (2017b). Prediction of genomic breeding values for growth, carcass and
555 meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genetics*.
556 https://doi.org/10.1186/s12863-017-0476-8
557 Brito, L. F., McEwan, J. C., Miller, S., Bain, W., Lee, M., Dodds, K., … Clarke, S. (2017c).
558 Genetic parameters for various growth, carcass and meat quality traits in a New Zealand
559 sheep population. *Small Ruminant Research*, *154*, 81–91.
560 https://doi.org/10.1016/j.smallrumres.2017.07.011
561 Byrne, T. J., Ludemann, C. I., Amer, P. R., Young, M. J. (2012). Broadening breeding objectives
562 for maternal and terminal sheep. *Livestock Science*, *144*, 20–36.
563 https://doi.org/10.1016/j.livsci.2011.10.010
564 Carillier, C., Larroque, H., & Robert-Granié, C. (2014). Comparison of joint versus purebred
565 genomic evaluation in the French multi-breed dairy goat population. *Genetics Selection*
566 *Evolution*, *46*(1), 67. https://doi.org/10.1186/s12711-014-0067-3
567 Christensen, O. F., & Lund, M. S. (2010). Genomic relationship matrix when some animals are
568 not genotyped Genomic prediction models. *Genetics Selection Evolution*, *42*(2).
569 Colinet, F. G., Vandenplas, J., Vanderick, S., Hammami, H., Mota, R. R., Gillon, A., Hubin, X.,
570 Bertozzi, C., Gengler, N. (2018). Bayesian single-step genomic evaluations combining local

and foreign information in Walloon Holsteins. *Animal*. https://doi.org/10.1017/S1751731117002324

Daetwyler, H. D., Swan, A. A., van der Werf, J. H., & Hayes, B. J. (2012). Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution*, 44(1), 1-11.

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0003395

Edel, C., Hamann, H., Neuner, S., Emmerling, R., & Götz, K. U. (2011). The German-Austrian genomic evaluation system for Fleckvieh (Siemmental) cattle. *Interbull Bulletin*, 44.

Eikje, L. S., Ådnøy, T., & Klemetsdal, G. (2008). The Norwegian sheep breeding scheme: description, genetic and phenotypic change. *Animal*, *2*(2), 167–176. https://doi.org/10.1017/S1751731107001176

Falconer, D. S., and T. F. C. Mackay. "Introduction into quantitative genetics." Essex: Prentice Hall (1996).

Fragomeni, B., Masuda, Y., Bradford, H. L., Lourenco, D. A. L., & Misztal, I. (2019). International bull evaluations by genomic BLUP with a prediction population. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2018-15554

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics*.

Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., & Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, *41*(1), 51. https://doi.org/10.1186/1297-9686-41-51

Jia, Z. (2017). Controlling the overfitting of heritability in genomic selection through cross validation. *Scientific reports*, 7(1), 1-9. 10.1038/s41598-017-14070-z.

Kang, H., Ning, C., Zhou, L., Zhang, S., Yan, Q., & Liu, J.-F. (2018). Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. *Journal of Dairy Science*, *101*(12), 11143–11149. https://doi.org/10.3168/jds.2018-15090

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R., Cristobal, M. S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.1001258

Koivula, M., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Reducing bias in the dairy cattle single-step genomic evaluation by ignoring bulls without progeny. *Journal of Animal Breeding and Genetics*, *135*(2), 107–115. https://doi.org/10.1111/jbg.12318

Legarra, A., Baloche, G., Barillet, F., Astruc, J. M., Soulas, C., Aguerre, X., Arrese, F., Mintegi, L., Lasarte, M., Maeztu, F., Beltrán de Heredia, I., Ugarte, E. (2014). Within-and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. *Journal of Dairy Science, 97(5),* 3200-3212. https://doi.org/10.3168/jds.2013-7745

Liu, C., Dupuis, J., Larson, M. G., Cupples, L. A., Ordovas, J. M., Vasan, R. S., Meigs, J. B., Jacques, P. F. and Levy, D. (2015). Revisiting heritability accounting for shared environmental effects and maternal inheritance. *Human genetics*, 134(2), pp.169-179. 10.1007/s00439-014-1505-6

Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., & Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution*, 52(47). https://doi.org/10.1186/s12711-020-00567-1

Masuda, Y., VanRaden, P. M., Misztal, I., & Lawlor, T. J. (2018). Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2017-13310

McRae, K. M., Baird, H. J., Dodds, K. G., Bixley, M. J., & Clarke, S. M. (2016). Incidence and heritability of ovine pneumonia, and the relationship with production traits in New Zealand sheep. *Small Ruminant Research*. https://doi.org/10.1016/j.smallrumres.2016.11.003

Misztal, I., & Legarra, A. (2017). Invited review: efficient computation strategies in genomic selection. *Animal*, *11*(5), 731–736. https://doi.org/10.1017/S1751731116002366

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, *92*(9), 4648–4655. https://doi.org/10.3168/jds.2009-2064

Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., VanRaden, P. M., & Lawlor, T. J. (2013). Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2012-5656

Misztal, I., Tsuruta, S., Strabel, T., Druet, T., & Lee, D. (2002). BLUPF90 and related programs (BGF90). *Proc. 7th World Congr. Genet. Appl. to Livest. Prod.*

NSG (2020). Referat fra møte nr 2/2020 i Avlsrådet for sau. Retrieved from https://www.nsg.no/getfile.php/13136489-1594690051/_NSG-PDF-

filer/Sau/Møtereferat%20Avlsrådet%20for%20sau/Avlsrådsmøte%2017062020%20-%20Referat.pdf

Oliveira, H. R., Brito, L. F., Sargolzaei, M., e Silva, F. F., Jamrozik, J., Lourenco, D. A. L., Schenkel, F. S. (2019a). Impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations in dairy cattle: A simulation study. *Journal of Animal Breeding and Genetics*. https://doi.org/10.1111/jbg.12407

Oliveira, H. R., Lourenco, D. A. L., Masuda, Y., Misztal, I., Tsuruta, S., Jamrozik, J., … Schenkel, F. S. (2019b). Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2018-15466

Oliveira, H. R., McEwan, J. C., Jakobsen, J., Blichfeldt, T., Meuwissen, T., Pickering, N., … Brito, L. F. (2020). Genetic Connectedness Between Norwegian White Sheep and New Zealand Composite Sheep Populations With Similar Development History. *Frontiers in Genetics*, *11*, 371. https://doi.org/10.3389/fgene.2020.00371

Oliveira, H. R., Silva, F. F., Brito, L. F., Guarini, A. R., Jamrozik, J., & Schenkel, F. S. (2018). Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. *Journal of Animal Breeding and Genetics*. https://doi.org/10.1111/jbg.12317

R Core Team. (2016). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Schaeffer, L. R. (1994). Multiple-Country Comparison of Dairy Sires. *Journal of Dairy Science*. https://doi.org/10.3168/jds.S0022-0302(94)77209-X

Smith, B. J. (2007). boa : An R Package for MCMC Output Convergence. *Journal of Statistical Software*. https://doi.org/http://dx.doi.org/10.18637/jss.v021.i11

Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods, 8thEdn. *Ames: Iowa State Univ. Press Iowa*.

Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Misztal, I., & Lawlor, T. J. (2019). Controlling bias in genomic breeding values for young genotyped bulls. *Journal of Dairy Science*, *102*(11), 9956–9970. https://doi.org/10.3168/JDS.2019-16789

Tsuruta, S., Misztal, I., Aguilar, I., & Lawlor, T. J. (2011). Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *Journal of Dairy Science*, *94*(8), 4198–4204. https://doi.org/10.3168/jds.2011-4256

Uemoto, Y., Osawa, T., & Saburi, J. (2017). Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. *Animal*, *11*(3), 382–393. https://doi.org/10.1017/S1751731116001762

Vandenplas, J., Spehar, M., Potocnik, K., Gengler, N., & Gorjanc, G. (2017). National single-step genomic method that integrates multi-national genomic information. *Journal of Dairy Science*. https://doi.org/10.3168/jds.2016-11733

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, *92*(1), 16–24. https://doi.org/10.3168/jds.2008-1514

VanRaden, P. M., & Sullivan, P. G. (2010). International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution*. https://doi.org/10.1186/1297-9686-42-7

Weller, J. I., Stoop, W. M., Eding, H., Schrooten, C., & Ezra, E. (2015). Genomic evaluation of a

relatively small dairy cattle population by combination with a larger population. *Journal of Dairy Science*, 98(7), 4945-4955. http://dx.doi.org/10.3168/jds.2014-9086

TABLES AND FIGURES

Table 1. Descriptive statistics for the traits analyzed in this study.

Trait[1]	N	Mean (SD)	Minimum	Maximum	CV (%)
Norway					
BW_{NO}	2,241,222	4.87 (1.02)	1.00	9.00	21.05
WW_{NO}	3,425,100	43.89 (7.9)	14.00	89.00	18.00
CW_{NO}	2,395,830	20.16 (3.02)	6.10	41.40	14.97
EUC_{NO}	2,395,170	8.43 (1.65)	1.00	15.00	19.58
EUF_{NO}	2,399,072	5.78 (1.49)	1.00	15.00	25.73
New Zealand					
BW_{NZ}	68,717	4.85 (1.05)	0.50	9.30	21.60
WW_{NZ}	195,932	29.85 (7.4)	6.00	65.50	24.78
CW_{NZ}	34,970	17.21 (3.28)	7.90	31.70	19.05
CCW_{NZ}	16,845	17.85 (3.32)	7.70	30.90	18.58
XCW_{NZ}	16,439	17.74 (3.32)	7.73	31.51	18.72
EMD_{NZ}	120,048	26.28 (3.16)	12.00	39.00	12.01
BUT_{NZ}	19,800	65.15 (3.34)	52.50	77.00	5.12
FGR_{NZ}	37,703	5.43 (3.48)	1.00	22.00	64.14
FDM_{NZ}	120,237	2.51 (1.14)	0.50	9.00	45.29

[1]The traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm). N: number of records. SD: standard deviation. CV: coefficient of variation.

Table 2. Percentage of tests with p-value lower than 0.05, for the F- and t-test used to compare the homogeneity among traits recorded in the different countries.

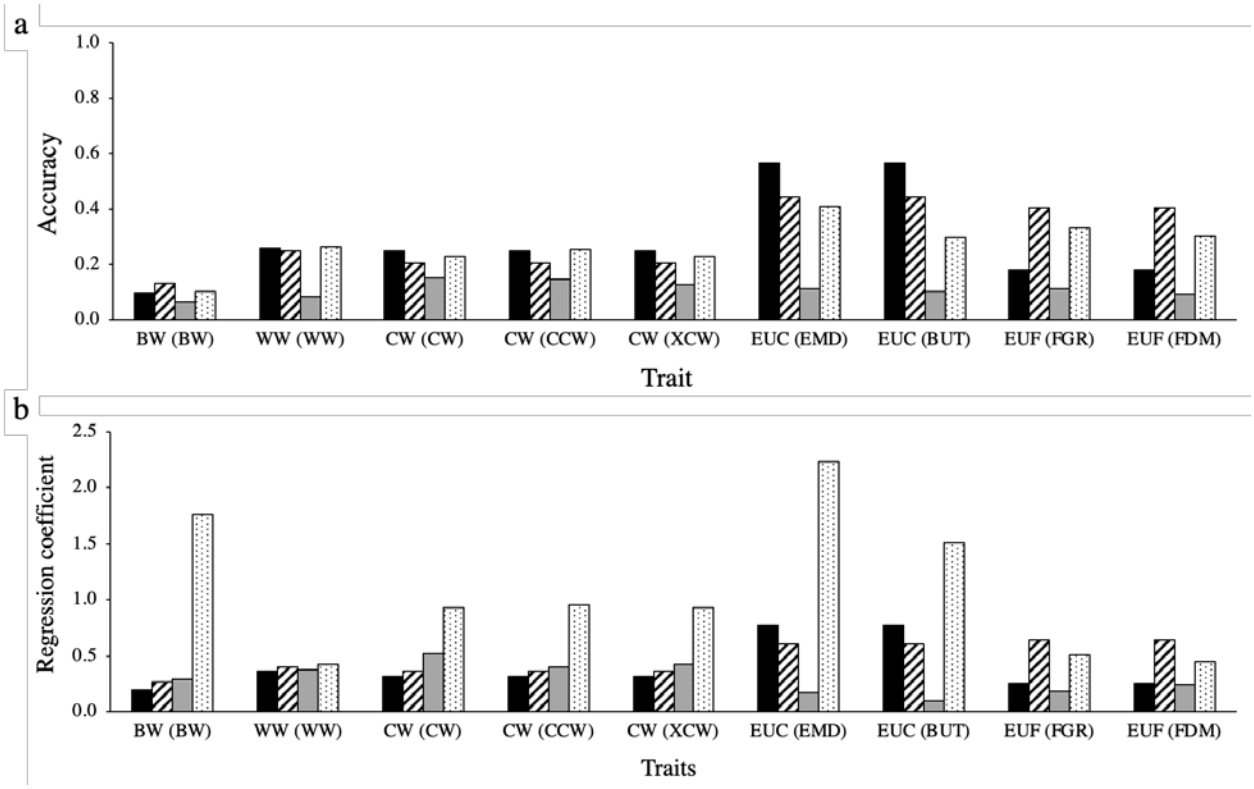
Comparison[1]	F (%)	t (%)
BW_{NO} and BW_{NZ}	6.4	4.9
WW_{NO} and WW_{NZ}	7.7	100.0
CW_{NO} and CW_{NZ}	13.7	98.8
CW_{NO} and CCW_{NZ}	13.4	92.7
CW_{NO} and XCW_{NZ}	14.9	94.5
EUC_{NO} and EMD_{NZ}	98.6	100.0
EUC_{NO} and BUT_{NZ}	99.7	100.0
EUF_{NO} and FGR_{NZ}	100.0	12.9
EUF_{NO} and FDM_{NZ}	51.9	100.0

[1]Traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). Traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm).

720 **Table 3.** Posterior means and 95% highest posterior density intervals (inside brackets) for
721 heritabilities (h^2) estimated in the different scenarios, using the reduced datasets and the single-
722 breed approach.

Scenario	Traits[1]	h^2
Only Norway	BW_{NO}	0.47 [0.472-0.474]
	WW_{NO}	0.41 [0.409-0.411]
	CW_{NO}	0.53 [0.529-0.531]
	EUC_{NO}	0.44 [0.442-0.445]
	EUF_{NO}	0.46 [0.457-0.461]
Both Norway and New Zealand	BW_{NO} and BW_{NZ}	0.47 [0.472-0.473]
	WW_{NO} and WW_{NZ}	0.39 [0.391-0.392]
	CW_{NO} and CW_{NZ}	0.55 [0.548-0.551]
	CW_{NO} and CCW_{NZ}	0.55 [0.549-0.551]
	CW_{NO} and XCW_{NZ}	0.55 [0.549-0.552]
	EUC_{NO} and EMD_{NZ}	0.39 [0.388-0.390]
	EUC_{NO} and BUT_{NZ}	0.40 [0.402-0.406]
	EUF_{NO} and FGR_{NZ}	0.41 [0.414-0.418]
	EUF_{NO} and FDM_{NZ}	0.44 [0.446-0.448]
Only New Zealand	BW_{NZ}	0.46 [0.457-0.467]
	WW_{NZ}	0.31 [0.300-0.313]
	CW_{NZ}	0.22 [0.196-0.234]
	CCW_{NZ}	0.30 [0.274-0.329]
	XCW_{NZ}	0.27 [0.239-0.296]
	EMD_{NZ}	0.37 [0.363-0.375]
	BUT_{NZ}	0.37 [0.350-0.392]
	FGR_{NZ}	0.26 [0.243-0.282]
	FDM_{NZ}	0.39 [0.379-0.391]

723 [1]The traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg),
724 carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading
725 (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight
726 (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated
727 carcass weight (XCW_{NZ}; kg), ultrasound eye-muscle depth (EMD_{NZ}; mm), butt circumference
728 (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm).
729



730

731 **Figure 1.** Accuracy (a) and bias (assessed based on the regression coefficient; b) of traditional
732 breeding values (black bars) and genomic breeding values predicted using only Norwegian
733 (diagonal strips), only New Zealand (grey), and both Norwegian and New Zealand (dotted)
734 animals in the training population. The traits measured in Norway were: birth weight (BW_{NO}; kg),
735 weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}),
736 and EUROP fat grading (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ};
737 kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg),
738 x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye-muscle depth (EMD_{NZ}; mm), butt
739 circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth
740 (FDM_{NZ}; mm).

741

742

SUPPLEMENTARY MATERIAL

Table S1. Number of animals with both genotypes and phenotypes in the training population of each trait.

Scenario	Traits[1]	N	Noffspring
Only Norway	BW_{NO}	133	191,560
	WW_{NO}	300	298,555
	CW_{NO}	0	134,970
	EUC_{NO}	0	134,938
	EUF_{NO}	0	134,960
Only New Zealand	BW_{NZ}	1,513	7,157
	WW_{NZ}	15,870	36,541
	CW_{NZ}	14,894	8,496
	CCW_{NZ}	14,841	8,760
	XCW_{NZ}	14,470	8,878
	EMD_{NZ}	7,846	23,774
	BUT_{NZ}	14,841	8,690
	FGR_{NZ}	14,880	8,834
	FDM_{NZ}	7,848	23,762

The traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm). N: number of animals with both genotypes and phenotypes in the training population. Noffspring: number of offspring from genotyped animals that have phenotypes in the training population.

Table S2. Posterior means and 95% highest posterior density intervals (inside brackets) for heritabilities estimated for different traits measured in Norway (h^2_{NO}) and New Zealand (h^2_{NZ}), and their genetic correlation (r_g), using the reduced datasets and the multiple-breed approach.

Traits	h^2_{NO}	h^2_{NZ}	r_g
BW_{NO} and BW_{NZ}	0.47 [0.470-0.472]	0.46 [0.451-0.460]	0.21 [0.038-0.337]
WW_{NO} and WW_{NZ}	0.41 [0.410-0.414]	0.33 [0.328-0.335]	0.04 [-0.005-0.084]
CW_{NO} and CW_{NZ}	0.53 [0.529-0.531]	0.23 [0.215-0.237]	0.15 [0.036-0.268]
CW_{NO} and CCW_{NZ}	0.53 [0.529-0.530]	0.28 [0.269-0.290]	-0.07 [-0.194-0.079]
CW_{NO} and XCW_{NZ}	0.53 [0.529-0.531]	0.26 [0.237-0.262]	0.11 [-0.002-0.198]
EUC_{NO} and EMD_{NZ}	0.44 [0.442-0.445]	0.35 [0.343-0.355]	0.32 [0.174-0.435]
EUC_{NO} and BUT_{NZ}	0.44 [0.441-0.445]	0.32 [0.318-0.330]	0.45 [0.362-0.510]
EUF_{NO} and FGR_{NZ}	0.46 [0.457-0.460]	0.19 [0.180-0.201]	0.12 [-0.034-0.217]
EUF_{NO} and FDM_{NZ}	0.46 [0.457-0.460]	0.35 [0.347-0.360]	0.00 [-0.080-0.077]

The traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm).

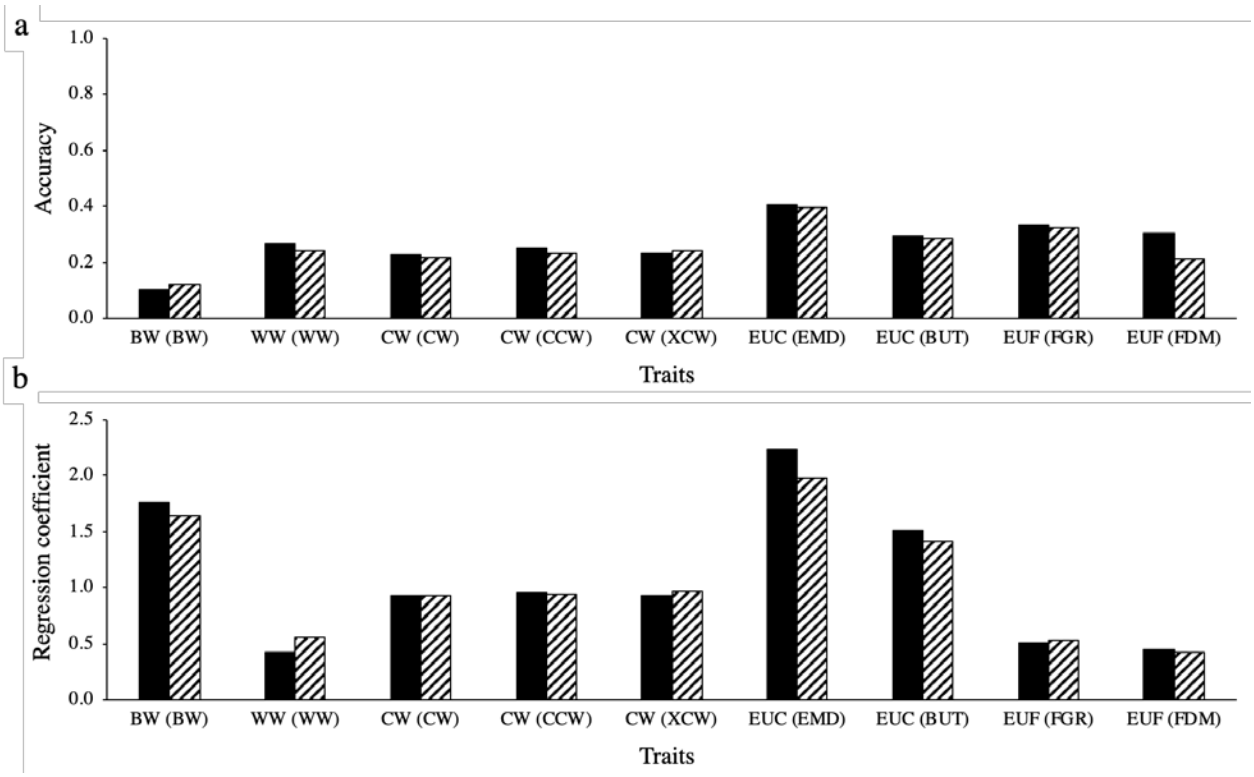


Figure S1. Accuracy (a) and bias (assessed using a regression coefficient; b) of genomic breeding values predicted using Norwegian and New Zealand animals in the training population under the single-breed (black bars) and multiple-breed (diagonal strips) approaches. The traits measured in Norway were: birth weight (BW_{NO}; kg), weaning weight (WW_{NO}; kg), carcass weight (CW_{NO}; kg), EUROP carcass classification (EUC_{NO}), and EUROP fat grading (EUF_{NO}). The traits measured in New Zealand were: birth weight (BW_{NZ}; kg), weaning weight (WW_{NZ}; kg), carcass weight (CW_{NZ}; kg), cold carcass weight (CCW_{NZ}; kg), x-ray estimated carcass weight (XCW_{NZ}; kg), ultrasound eye muscle depth (EMD_{NZ}; mm), butt circumference (BUT_{NZ}; cm), carcass fatness at the GR site (FGR_{NZ}; mm), and ultrasound fat depth (FDM_{NZ}; mm).