J. Dairy Sci. 97:1–12
http://dx.doi.org/10.3168/jds.2013-7580

Genomic predictions based on a joint reference population for the Nordic Red cattle breeds

L. Zhou,* B. Heringstad,†[1] G. Su,*[1] B. Guldbrandtsen,* T. H. E. Meuwissen,† M. Svendsen,‡ H. Grove,† U. S. Nielsen,§ and M. S. Lund*
*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark
†Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway
‡Geno Breeding and AI Association, 1432 Ås, Norway
§Danish Agriculture Advisory Service, DK-8200 Aarhus N, Denmark

ABSTRACT

The main aim of this study was to compare accuracies of imputation and genomic predictions based on single and joint reference populations for Norwegian Red (NRF) and a composite breed (DFS) consisting of Danish Red, Finnish Ayrshire, and Swedish Red. The single nucleotide polymorphism (SNP) data for NRF consisted of 2 data sets: one including 25,000 markers (NRF25K) and the other including 50,000 markers (NRF50K). The NRF25K data set had 2,572 bulls, and the NRF50K data set had 1,128 bulls. Four hundred forty-two bulls were genotyped in both data sets (double-genotyped bulls). The DFS data set (DSF50K) included 50,000 markers of 13,472 individuals, of which around 4,700 were progeny-tested bulls. The NRF25K data set was imputed to 50,000 density using the software Beagle. The average error rate for the imputation of NRF25K decreased slightly from 0.023 to 0.021, and the correlation between observed and imputed genotypes changed from 0.935 to 0.936 when comparing the NRF50K reference and the NRF50K–DFS50K joint reference imputations. A genomic BLUP (GBLUP) model and a Bayesian 4-component mixture model were used to predict genomic breeding values for the NRF and DFS bulls based on the single and joint NRF and DFS reference populations. In the multiple population predictions, accuracies of genomic breeding values increased for the 3 production traits (milk, fat, and protein yields) for both NRF and DFS. Accuracies increased by 6 and 1.3 percentage points, on average, for the NRF and DFS bulls, respectively, using the GBLUP model, and by 9.3 and 1.3 percentage points, on average, using the Bayesian 4-component mixture model. However, accuracies for health or reproduction traits did not increase from the multiple population predictions. Among the 3 DFS populations, Swedish Red gained most in accuracies from the multiple population predictions, presumably because Swedish Red has a closer genetic relationship with NRF than Danish Red and Finnish Ayrshire. The Bayesian 4-component mixture model performed better than the GBLUP model for most production traits for both NRF and DFS, whereas no advantage was found for health or reproduction traits. In general, combining NRF and DFS reference populations was useful in genomic predictions for both the NRF and DFS bulls.

Key words: imputation, genomic BLUP, Bayesian 4-component mixture model, multiple population genomic prediction

INTRODUCTION

Simulation studies (e.g., de Roos et al., 2009) as well as analyses of real data (Brøndum et al., 2011; Lund et al., 2011; VanRaden et al., 2012) have shown that genomic predictions can work across different populations. By combining different populations of the same breed or related breeds in the reference population, more information is available for the estimation of marker effects. Hence, more accurate predicted breeding values will be obtained in genomic predictions. Accuracies increased when 3 related dairy cattle populations—Danish Red, Swedish Red, and Finnish Ayrshire—were combined into one reference population (Brøndum et al., 2011). Reliabilities increased by 10 percentage points, on average, when 4 European Holstein populations were combined in the reference population (Lund et al., 2011). Increases in reliabilities from 6 to 45 percentage points were achieved by combining 6 Brown Swiss populations (Jorjani et al., 2011). However, most of the predictive accuracy, at least with density of 50,000 SNP, most likely comes from predicting the effect of large segments of chromosome or relationships, rather than individual QTL effects (Daetwyler et al., 2012; Wientjes et al., 2013).

Received October 9, 2013.
Accepted March 13, 2014.
[1] Corresponding authors: bjorg.heringstad@nmbu.no and guosheng.su@agrsci.dk

Norwegian Red (**NRF**), with approximately 230,000 cows, has good performance in health, fertility, and milk production. It is the main dairy breed in Norway (95% of dairy cows). A previous study (Luan et al., 2009) reported that correlations of genomic EBV (**GEBV**) and daughter yield deviations for NRF varied widely between 0.12 and 0.62 for different traits and were low for health and reproduction traits. Danish Red (38,000 cows), Finnish Ayrshire (143,000 cows), and Swedish Red (116,000 cows) are important red dairy cattle populations in these Nordic countries. These 3 red cattle populations were merged into one composite breed (**DFS**), also named VikingRed, which currently has a joint genetic evaluation. The reliabilities of genomic predictions of DFS averaged 0.28 for 17 traits (Brøndum et al., 2011). The NRF breed is related to Swedish Red and Finnish Ayrshire (Olsen et al., 2011). Danish Red, Finnish Ayrshire, and NRF were also used in the Swedish Red breeding program (Bett et al., 2010). Because sires have been exchanged and used between these populations, some genetic links exist between NRF and DFS.

Combining these 2 related breeds provides an interesting approach to improve accuracies in their genomic predictions. Bayesian variable selection models have been shown to give a better persistence of genomic predictions (Gao et al., 2013). Because this joint data set includes many distant relationships across the breeds, we expect that these models would result in higher accuracies in genomic predictions than traditional genomic BLUP (**GBLUP**) models. In this study, our first objective was to investigate accuracies in imputation from 25,000 (**25K**) to 50,000 (**50K**) SNP for NRF bulls, using only the NRF data set or the NRF and DFS data sets together as the reference. The second objective was to investigate accuracies in genomic predictions for NRF and DFS using the single or joint NRF DFS reference population, and to compare the GBLUP model and a Bayesian 4-component mixture model in genomic predictions of different traits.

MATERIALS AND METHODS

Genotypic and Phenotypic Data

The SNP data for NRF consisted of 2 data sets of progeny-tested bulls: a data set with 2,572 bulls genotyped with 25,000-SNP chips (**NRF25K**; Affymetrix, Santa Clara, CA; Affymetrix, 2007) and a data set with 1,128 bulls genotyped with the 54,001 SNP of the BovineSNP50 chip (**NRF50K**; Illumina Inc., San Diego, CA; Matukumalli et al., 2009). A total of 442 NRF bulls were genotyped using both the 25K and 50K chips (double-genotyped bulls). The DFS data (**DFS50K**) included genotypes of BovineSNP50 chips on Danish Red, Finnish Ayrshire, and Swedish Red. A total of 13,427 genotyped animals were included in the DFS50K data set, of which around 4,700 were progeny-tested bulls, 3,440 were cows, and the rest were young bulls without progeny test results. The genetic correlations and number of common sires between DFS and NRF from Interbull international genetic evaluations are shown in Table 1 (http://interbull2.slu.se). According to the pedigree, 291 DFS bulls (242 Swedish Red, 44 Finnish Ayrshire, and 5 Danish Red) have been used in the NRF population, and 58 NRF bulls have been used in the DFS population. In our data, 18% (864/4,741) of the progeny-tested bulls in the DFS50K data set have common sires with the NRF bulls, and 14% (366/2,572) of the NRF bulls have common sires with the DFS bulls. The 864 DFS bulls that had common sires with NRF bulls represent 6% (58/911), 17% (394/2,344), and 28% (412/1,486) of the Danish Red, Finnish Ayrshire, and Swedish Red bulls, respectively.

The 3 data sets (NRF25K, NRF50K, and DFS50K) were edited by removing SNP with minor allele frequencies <0.001 and call rate (per locus) <0.1. After SNP editing, SNP common to both the NRF50K and DFS50K data sets were kept for further analyses. To impute the NRF from 25K to 50K, SNP in NRF25K that were not present in the 50K data set were excluded

Table 1. Genetic correlations between a composite breed (Danish Red, Finnish Ayrshire, and Swedish Red; DFS) and Norwegian Red (NRF), and number of bulls in common, for some production and fertility traits according to Interbull international genetic evaluation December 2013

Trait	Genetic correlation	No. of common bulls
Milk yield	0.91	66
Fat yield	0.90	67
Protein yield	0.89	67
56-d nonreturn rate of heifers	0.79	66
Interval from calving to first insemination	0.88	58
56-d nonreturn rate of cows	0.73	61
Cows' ability to conceive[1]	0.71	53
Calving interval	0.86	53

[1] Calving interval (NRF) and interval from first to last insemination (DFS).