**J. Dairy Sci. 98:9051–9059**
**http://dx.doi.org/10.3168/jds.2015-9947**

# Including different groups of genotyped females for genomic prediction in a Nordic Jersey population

**H. Gao,*[1] P. Madsen,* U. S. Nielsen,† G. P. Aamand,‡ G. Su,* K. Byskov,† and J. Jensen***
*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark
†Seges, DK-8200 Aarhus N, Denmark
‡Nordic Cattle Genetic Evaluation, DK-8200 Aarhus N, Denmark

## ABSTRACT

Including genotyped females in a reference population (RP) is an obvious way to increase the RP in genomic selection, especially for dairy breeds of limited population size. However, the incorporation of these females must be conducted cautiously because of the potential preferential treatment of the genotyped cows and lower reliabilities of phenotypes compared with the proven pseudo-phenotypes of bulls. Breeding organizations in Denmark, Finland, and Sweden have implemented a female-genotyping project with the possibility of genotyping entire herds using the low-density (LD) chip. In the present study, 5 scenarios for building an RP were investigated in the Nordic Jersey population: (1) bulls only, (2) bulls with females from the LD project, (3) bulls with females from the LD project plus non-LD project females genotyped before their first calving, (4) bulls with females from the LD project plus non-LD project females genotyped after their first calving, and (5) bulls with all genotyped females. The genomically enhanced breeding value (GEBV) was predicted for 8 traits in the Nordic total merit index through a genomic BLUP model using deregressed proof (DRP) as the response variable in all scenarios. In addition, (daughter) yield deviation and raw phenotypic data were studied as response variables for comparison with the DRP, using stature as a model trait. The validation population was formed using a cut-off birth year of 2005 based on the genotyped Nordic Jersey bulls with DRP. The average increment in reliability of the GEBV across the 8 traits investigated was 1.9 to 4.5 percentage points compared with using only bulls in the RP (scenario 1). The addition of all the genotyped females to the RP resulted in the highest gain in reliability (scenario 5), followed by scenario 3, scenario 2, and scenario 4. All scenarios led to inflated GEBV because the regression coefficients are less than 1. However, scenario 2 and scenario 3 led to less bias of genomic predictions than scenario 5, with regression coefficients showing less deviation from scenario 1. For the study on stature, the daughter yield deviation/daughter yield deviation performed slightly better than the DRP as the response variable in the genomic BLUP (GBLUP) model. Therefore, adding unselected females in the RP could significantly improve the reliabilities and tended to reduce the prediction bias compared with adding selectively genotyped females. Although the DRP has performed robustly so far, the use of raw data is recommended with a single-step model as an optimal solution for future genomic evaluations.

**Key words:** genotyped cows, response variable, reliability, prediction bias

## INTRODUCTION

The size of the reference population (**RP**) is one of the important factors influencing the accuracy of genomic prediction (Goddard and Hayes, 2009). To date, RP mainly has consisted of proven bulls in national or international dairy cattle genomic selection programs (VanRaden et al., 2009; Harris and Johnson, 2010; Jorjani et al., 2010; Muir et al., 2010; Lund et al., 2011; Gao et al., 2013b). Due to the decreasing costs of genotyping and the increasing exchange of data on genotyped proven bulls, the prediction accuracies have been markedly enhanced due to the increased RP. This process has been beneficial for Holsteins, a widespread breed that is present in many countries. However, for the numerically smaller and geographically less widespread dairy breeds, such as Nordic Jersey or Nordic Red Cattle, the advantage of sharing reference data has been limited to international collaboration. An alternative solution to the problem of few proven bulls is to increase the RP by including genotyped females

---

Received June 11, 2015.
Accepted August 17, 2015.
[1] Corresponding author: hongding.gao@mbg.au.dk

(heifers and cows) even though the information from females is much less reliable compared with information from bulls that have been progeny tested using a large daughter group.

Prediction accuracy is expected to be enhanced by increasing the size of the RP. However, because the number of progeny from bulls being tested is shrinking due to the use of genomic selection as a pre-selection tool for young bulls entering the progeny-testing scheme, the RP will increase less rapidly over time (Schaeffer, 2006; Lillehammer et al., 2011). A simulation study by Thomasen et al. (2014) showed that the inclusion of genotyped cows in the RP was an efficient way to increase the genetic gain and would be a profitable investment for the breeding schemes of small breeds. Furthermore, since 2010, genotyping of females with a low-density (**LD**) chip has been implemented at a large scale in Holsteins in the United States.

Therefore, an appealing and cost-effective approach could be to genotype females using an LD chip such as the Illumina BovineLD BeadChip (http://support.illumina.com/array/array_kits/bovineld_dna_analysis_kit.html), followed by the imputation to higher density (Browning and Browning, 2009; Dassonneville et al., 2011).

The advantage of genotyping females in dairy cattle breeding has been reported in some previous studies. The first empirical study of the inclusion of cows in the RP was reported by Wiggans et al. (2011), where the records of the genotyped cows were pre-adjusted to be comparable with those of the genotyped bulls; these authors found an average gain in reliabilities of 3.5 and 0.9 percentage points in Holstein and Jersey populations, respectively. Pryce et al. (2012) demonstrated an improvement of 8 percentage points in the GEBV reliabilities by adding 10,000 genotyped cows to an RP consisting of approximately 3,000 bulls. Bapst et al. (2013) added approximately 1,236 genotyped cows to an existing RP consisting of 4,085 bulls in a Brown Swiss population but did not achieve a significant improvement in the accuracy of genomic prediction. In the United States, 30,852 genotyped Holstein cows were incorporated into the RP of 21,883 Holstein bulls, and an extra 0.4 percentage points of genomic reliability was observed when averaged across all traits (Cooper et al., 2014). In general, the outcomes of adding genotyped females to the RP appeared to vary among different implementations. The value of adding cows to the RP mainly appears to be dependent on the proportion of added genotyped cows and the size of the original bull RP.

Deregressed proof (**DRP**), which is a back-calculation of phenotypes from EBV using the reliabilities obtained from the traditional genetic evaluation, has been widely adopted nationally and internationally as the pseudo-phenotype of choice in genomic evaluation procedures. This method is used due to the advantages of simplification over the daughter yield deviation (**DYD**) and nonregressed property compared with EBV (Gao et al., 2013a). Therefore, DRP has worked fairly well when observations in the RP consisted of genotyped and progeny-tested bulls (Garrick et al., 2009; Lund et al., 2011; Gao et al., 2013a). However, for the genotyped females, the reliabilities of the DRP are much lower compared with the reliability of DRP for progeny-tested bulls. In such cases, the alternatives could be to use the yield deviation (**YD**) or raw phenotypic data in place of the DRP as the response variable for genotyped females because the YD is generated based only on the cows' own records and avoids the deregression procedure. An untested hypothesis that we address in this study regards the importance of the choice of the response variable for the genotyped females as a factor influencing the accuracy of genomic prediction.

The breeding organizations in Denmark, Finland, and Sweden (Viking Genetics) have initiated a female genotyping project with the chance of genotyping all heifers in entire selected herds using a LD chip. To assess the effect of including genotyped females in the RP, the first batch of data from this project was used. The purposes of this study were to (1) examine the effect of adding different sources of genotyped females to the RP for Nordic Jersey, and (2) explore the effect of different response variables for the genotyped females on prediction accuracy.

## MATERIALS AND METHODS

### *Data*

The animals used in this study consisted of 1,414 genotyped Nordic Jersey bulls born between 1981 and 2011 (with several individuals that were missing DRP). In addition, 1,154 proven Jersey bulls in the United States that were born between 1950 and 2009 were added to the RP through the collaboration for exchanging reference data to maximize the number of progeny-tested bulls in the RP (Su et al., 2014). A total of 4,251 genotyped females with DRP were classified into different subsets based on the genotyping strategy. Overall, 3,492 females born after 2010 were phenotyped and genotyped, along with the entire herd, using an LD chip through the LD project (hereafter referred as LD females). The remaining individuals, consisting of 759 genotyped and phenotyped females, were selected for genotyping by private breeders according to their individual breeding programs (hereafter referred as non-LD females). Depending on the genotyping date, among