J. Dairy Sci. 100:1–14
https://doi.org/10.3168/jds.2016-11479

Cow genotyping strategies for genomic selection in a small dairy cattle population

J. Jenko,*[1] G. R. Wiggans,† T. A. Cooper,† S. A. E. Eaglen,* W. G. de L. Luff,‡ M. Bichard,§ R. Pong-Wong,* and J. A. Woolliams*

*The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, UK
†Animal Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, USDA-ARS, Beltsville, MD 20705
‡World Guernsey Cattle Federation, The Hollyhocks, 10 Clos des Goddards, Rue des Goddards, Castel GY5 7JD, Guernsey
§English Guernsey Cattle Society, 12 Southgate Street, Launceston, Cornwall PL15 9DP, UK

ABSTRACT

This study compares how different cow genotyping strategies increase the accuracy of genomic estimated breeding values (EBV) in dairy cattle breeds with low numbers. In these breeds, few sires have progeny records, and genotyping cows can improve the accuracy of genomic EBV. The Guernsey breed is a small dairy cattle breed with approximately 14,000 recorded individuals worldwide. Predictions of phenotypes of milk yield, fat yield, protein yield, and calving interval were made for Guernsey cows from England and Guernsey Island using genomic EBV, with training sets including 197 de-regressed proofs of genotyped bulls, with cows selected from among 1,440 genotyped cows using different genotyping strategies. Accuracies of predictions were tested using 10-fold cross-validation among the cows. Genomic EBV were predicted using 4 different methods: (1) pedigree BLUP, (2) genomic BLUP using only bulls, (3) univariate genomic BLUP using bulls and cows, and (4) bivariate genomic BLUP. Genotyping cows with phenotypes and using their data for the prediction of single nucleotide polymorphism effects increased the correlation between genomic EBV and phenotypes compared with using only bulls by 0.163 ± 0.022 for milk yield, 0.111 ± 0.021 for fat yield, and 0.113 ± 0.018 for protein yield; a decrease of 0.014 ± 0.010 for calving interval from a low base was the only exception. Genetic correlation between phenotypes from bulls and cows were approximately 0.6 for all yield traits and significantly different from 1. Only a very small change occurred in correlation between genomic EBV and phenotypes when using the bivariate model. It was always better to genotype all the cows, but when only half of the cows were genotyped, a divergent selection strategy was better compared with the random or directional selection approach. Divergent selection of 30% of the cows remained superior for the yield traits in 8 of 10 folds.

Key words: genomic selection, genotyping cows, cow genotyping strategies, Guernsey

INTRODUCTION

Response to selection can be increased by changing the ratio of the accuracy of EBV to the generation interval, and an intermediate age exists where this ratio is maximized, thus defining the optimum selection age. For conventional evaluations based solely on pedigree and phenotypes, the accuracy of parent average EBV is too low, precluding the intense selection of young bulls at birth. For this purpose, bulls for widespread use are often selected only after the phenotypes of their first crop daughters are known, at around 5 yr of age. A benefit of genomic selection is its potential to increase the accuracy of EBV early in life. To achieve this, a sufficient number of individuals with phenotypes or progeny records needs to be genotyped (Meuwissen et al., 2001). Based on this training set of individuals, SNP effects are then estimated. These estimates can then be used for the calculation of genomic EBV of genotyped individuals without phenotypic observations on themselves, or lactating daughters in the case of young bulls. When the accuracy of a genomic EBV is high enough, the optimum selection age for the parents of a future generation can be lowered, reducing the generation interval. This might result in a doubling of the rate of genetic gain in dairy schemes compared with conventional breeding values (Schaeffer, 2006).

The accuracy of a genomic EBV will be higher when the number of genotyped individuals with own perfor-

Received May 17, 2016.
Accepted September 21, 2016.
[1]Corresponding author: janez.jenko@roslin.ed.ac.uk

mance or progeny records is large (Daetwyler et al., 2008, 2010; Goddard, 2009). In large populations, many sires have achieved very accurate progeny tests from large daughter groups, and have been genotyped. This has enabled the successful implementation of genomic selection in large populations of dairy cattle (VanRaden et al., 2009). However, for small cattle breeds genomic selection is still a challenge as their limited resources restrict the prediction accuracy, as either the number of sires with a large number of daughters is too small, or the progeny tests are weak. Three solutions are possible to overcome this problem. One is to include genotypes from the same breed but from the other country (Cooper et al., 2016), another is to combine the breed-specific reference population with other breeds (Hayes et al., 2009; Olson et al., 2012; Hozé et al., 2014), and the last is to include cows in the reference population (Pryce et al., 2012; Calus et al., 2013; Cooper at al., 2015).

The success of combining the reference population with another breed depends on the genetic distance between them, numbers of genotyped individuals, and SNP chip density. Genomic evaluation requires that the different populations are at least distantly related (Habier et al., 2010). To increase genetic gain, the reference population and selection candidates should share recent ancestors (Clark et al., 2012; Pszczola et al., 2012). This relationship is higher when genotypes from cows of the same breed are available compared with individuals from different breeds, but their accuracy is often smaller compared with de-regressed proofs of bulls from large breeds, and are typically expected to add less information per genotyped individual, although this difference depends on the heritability. de Roos (2011) estimated that the addition of 7 cows for a trait with a heritability of 0.1 gives the same gain as adding 1 bull with 100 tested progeny, whereas for the trait with a heritability of 0.5 this ratio decreased to 2 cows per bull. Simulations performed by Jiménez-Montero et al. (2012) showed that not only the number of cow genotypes but also the genotyping design can increase the accuracy of genomic EBV. The accuracy of divergent selection on yield or breeding value deviations was higher than when selecting at random or based on the extreme values in the upper tail.

The goal of this study was to estimate the benefit of using cow genotypes for genomic selection in a small dairy cattle population. An additional goal was to determine the effect of different cow genotyping strategies on the accuracy of selection. The Guernsey breed represented by bull and cow genotypes from England and Guernsey Island is a suitable population for this study. Guernsey is one of the smaller dairy breeds with approximately 14,000 recorded individuals worldwide, and of these, 2,000 are on Guernsey Island.

MATERIALS AND METHODS

Study Samples

A total of 1,637 genotypes from Guernsey cattle were available: 197 from bulls and 1,440 from cows. Of the bull samples, 29 were genotyped with the Illumina BovineHD Genotyping BeadChip (**777K**; Illumina Inc., San Diego, CA) and 168 with the GeneSeek Genomic Profiler HD BeadChip Version 1 (**75K**; Neogen Corp., Lexington, KY). All of the cow samples were genotyped with the GeneSeek Genomic Profiler for Dairy Cattle Version 3 (**25K**; Neogen Corp.).

Genotyped bulls were part of the AI program and were born between 1957 and 2013. Except for the most recent ones, they had daughters with records available and were included in genetic evaluations. One bull had both parents genotyped and 75 bulls had one parent genotyped. Cows with genotypes were a cohort of Guernsey cows present on the island in early 2014. They were born between 1997 and 2013 and were included in the milk recording scheme. One hundred thirty-three cows had both parents genotyped, and 705 cows had one parent genotyped.

Genotype Quality Check

Before the genotypes were checked for quality, 3 individuals were discovered to have been repeated, and the sample with the higher call rate was kept. For all 3 chips, SNP were checked for the position and name: 199 SNP had the same name but different positions, or had different names but with the same position as another and these were excluded. The SNP on the sex chromosomes were excluded from all the chips. Individuals were excluded when overall call rate was <0.85 or heterozygosity was outside the interval of mean ± 3 SD calculated for the relevant SNP chip. Altogether, 107 samples from the 25K chip, 1 from the 75K chip, and 1 from the 777K chip failed these criteria as shown in Appendix A Figures A1, A2, and A3. Then, SNP loci were excluded if call rate <0.85: 546 were excluded for the 25K chip, 1,327 for the 75K chip, and 12,712 for the 777K chip. For imputation, individuals genotyped with 777K were merged with 75K using only 72,679 SNP from the 75K chip. Finally, SNP with Hardy-Weinberg equilibrium test $P < 10^{-6}$ or minor allele frequency (**MAF**) <0.05 were removed, resulting in the availability of 64,657 and 17,716 SNP on the 75K and 25K chip, respectively.

The pedigree relationship was checked separately for duos and trios using PLINK (Purcell et al., 2007) by comparing the known genotypes of parents and offspring. Parent-offspring duos with more than 1% of