J. Dairy Sci. 98:1–8
http://dx.doi.org/10.3168/jds.2015-9682

Estimation of genomic breeding values for milk yield in UK dairy goats

S. Mucha,*[1] R. Mrode,*† I. MacLaren-Lee,‡ M. Coffey,* and J. Conington*
*Animal and Veterinary Sciences, Scotland's Rural College, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom
†Animal Biosciences, International Livestock Institute, Nairobi 00100, Kenya
‡Illumina Cambridge Ltd., Great Abington, Cambridge CB21 6GP, United Kingdom

ABSTRACT

The objective of this study was to estimate genomic breeding values for milk yield in crossbred dairy goats. The research was based on data provided by 2 commercial goat farms in the UK comprising 590,409 milk yield records on 14,453 dairy goats kidding between 1987 and 2013. The population was created by crossing 3 breeds: Alpine, Saanen, and Toggenburg. In each generation the best performing animals were selected for breeding, and as a result, a synthetic breed was created. The pedigree file contained 30,139 individuals, of which 2,799 were founders. The data set contained test-day records of milk yield, lactation number, farm, age at kidding, and year and season of kidding. Data on milk composition was unavailable. In total 1,960 animals were genotyped with the Illumina 50K caprine chip. Two methods for estimation of genomic breeding value were compared—BLUP at the single nucleotide polymorphism level (BLUP-SNP) and single-step BLUP. The highest accuracy of 0.61 was obtained with single-step BLUP, and the lowest (0.36) with BLUP-SNP. Linkage disequilibrium (r^2, the squared correlation of the alleles at 2 loci) at 50 kb (distance between 2 SNP) was 0.18. This is the first attempt to implement genomic selection in UK dairy goats. Results indicate that the single-step method provides the highest accuracy for populations with a small number of genotyped individuals, where the number of genotyped males is low and females are predominant in the reference population.
Key words: genomic selection, single step, milk yield, dairy goat

INTRODUCTION

Genomic selection has become routine in many farmed livestock species such as dairy and beef cattle. This is mainly due to exchange of genotypes between countries, and reference populations for those species are now large, consisting of thousands of bulls with high reliability breeding values. This allows the prediction of genomic breeding values for young animals, which have no phenotypic records, with acceptably high accuracy. In the case of dairy goats, the breeding industry is not so well developed worldwide. Routine breeding value estimation is carried out in several countries such as Canada, France, United States, and Norway (Bélichon et al., 1999; Montaldo and Manfredi, 2002). Recently, because of the introduction of the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA; Tosser-Klopp et al., 2012), genomic tools became available for genetic improvement of goats. Currently, genomic selection in dairy goats has been introduced only in France (Carillier et al., 2013), with 2,810 genotyped Saanen and Alpine goats. In the UK, the number of genotyped goats is also relatively small, which poses certain restrictions with respect to the estimation of genomic breeding values. Accuracy of methods that use only phenotypes of the genotyped animals and ignore records of the nongenotyped part of the population (e.g., genomic BLUP, BLUP-SNP) is limited when the reference population is small. Therefore, an alternative approach was considered that integrates all of the available phenotypic, pedigree, and genomic information in a single-step procedure (Legarra et al., 2009; Misztal et al., 2009; Christensen and Lund, 2010). This approach has been regarded as computationally demanding in cases of large data sets with hundreds of thousands of genotyped animals. However, in goats the amount of data used in genetic evaluations is considerably lower than that of dairy cattle. Moreover, the method is easy to implement because it can use raw phenotypic records without the need to calculate deregressed proofs (**DRP**). It also allows evaluation of all animals (with and without genotypes) simultaneously.

The objective of this study was to evaluate BLUP-SNP and single-step approach for estimation of genomic breeding values in dairy goats. Additionally, level of linkage disequilibrium in the reference population was investigated.

Received April 8, 2015.
Accepted July 15, 2015.
[1]Corresponding author: sebastian.mucha@sruc.ac.uk

MATERIALS AND METHODS

Phenotypic Data

Lactation data were from 2 separate farm units in the UK owned by a single farming business. The data set comprised 590,409 records on 14,453 dairy goats kidding between 1987 and 2013. The population was created in 1985 by crossing 3 breeds: Alpine, Saanen, and Toggenburg. No particular crossing strategy existed. In each generation the best performing animals were selected for breeding, and as a result, a synthetic breed was created. The pedigree file contained 30,139 individuals, of which 2,799 were considered as founders. A total of 296 sires and 12,468 dams were in the pedigree. The data set contained test-day records of milk yield, along with information about lactation number (1 to 6), farm (2 farms), age at kidding (12 to 90 mo), and year (1987 to 2013) and season of kidding [summer (June to August), autumn (September to November), winter (December to February), and spring (March to May)]. Fat and protein content was not included in the analysis because it was not recorded on either of the farms contributing data. Only goats with more than 3 test-day observations were used for analysis. Additionally, the data set was restricted to groups having at least 10 records per level of herd test day, year-season and age at kidding. Test-day milk records below 0.2 and above 12.5 kg were removed from the data as error records. Lactation length was restricted to between 4 and 520 DIM because goats from the 2 farms are routinely milked for long lactations. Heritability of milk yield in the analyzed population was 0.56. For a detailed description of the analyzed population, see Mucha et al. (2014).

Pedigree Analysis

Pedigree completeness was assessed with complete generations equivalent. The number of equivalent generations traced was computed as the sum over all known ancestors of the terms $(1/2)^t$, where t is the ancestor's generation number, which is equal to 1 for the parents, 2 for the grandparents, and so on (Maignel et al., 1996).

Effective population size (realized N_e) based on individual increase in inbreeding (ΔF_i) was calculated following the approach proposed by Gutiérrez et al. (2009). The ΔF_i coefficients were computed as

$$\Delta F_i = 1 - \sqrt[t-1]{1 - F_i},$$

where F_i is the individual coefficient of inbreeding and t is the complete generations equivalent (Maignel et al., 1996). The effective population size (N_e) was obtained from $\overline{\Delta F}$, which was computed by averaging the ΔF_i of the n individuals included in a given reference subpopulation as $N_e = \frac{1}{2\overline{\Delta F}}$.

Genotypes

In total 1,960 animals were selected for genotyping. All of the available sires (150 individuals) were sampled, and subsequently, this set was supplemented by females. Selection of females for genotyping was based on 2 criteria: average daily lifetime yield and genetic relationship between the animals. The process was optimized in a way to select animals from the upper (group 1) and lower (group 2) tail of the distribution of average daily lifetime yield. Animals had been selected so that the relationship within the 2 groups was minimized and the relationship between the 2 groups was maximized. This was done with the software package Corona produced by Brian Kinghorn (University of New England, Armidale, Australia, personal communication).

Animals were genotyped commercially with a 50K Caprine Illumina SNP chip at Edinburgh Genomics (Edinburgh, UK). After filtering out SNP that were not in Hardy-Weinberg equilibrium, had minor allele frequency below 0.05, were monomorphic, had a call rate below 0.95, or had an Illumina GenCall (GC score) below 0.6, the data set contained 47,306 markers. Additionally, animals with a call rate below 0.9 were removed from further analyses. This resulted in 1,902 genotyped animals born between 2003 and 2012 being used. The SNP information was also used to correct the pedigree by means of sire and dam verification for animals with known parents and for parent discovery for animals with unknown pedigree. Finally, the genotype matrix was used for clustering based on principal component analysis, performed with SNP & Variation Suite v7.7.8 (Golden Helix Inc., Bozeman, MT). This analysis was done to investigate potential population stratification due to historical crossbreeding in the population.

Linkage Disequilibrium

Linkage disequilibrium (**LD**) was measured as r^2, which is the squared correlation of the alleles at 2 loci (Hill and Robertson, 1968):

$$r^2 = \frac{\left[f(\mathrm{AB}) - f(\mathrm{A})f(\mathrm{B})\right]^2}{f(\mathrm{A})f(\mathrm{a})f(\mathrm{B})f(\mathrm{b})},$$

where f(AB), f(A), f(a), f(B), and f(b) are observed frequencies of haplotype AB and of alleles A, a, B, and b, respectively. Linkage disequilibrium was calculated