J. Dairy Sci. 97:1107–1116
http://dx.doi.org/10.3168/jds.2013-7135

Assessment of accuracy of genomic prediction for French Lacaune dairy sheep

G. Baloche,* A. Legarra,* G. Sallé,†‡ H. Larroque,* J.-M. Astruc,§ C. Robert-Granié,* and F. Barillet[*1]
*INRA, UR631, Station d'Amélioration Génétique des Animaux (SAGA), CS-52627, F-31326 Castanet-Tolosan, France
†INRA, UMR1282 Infectiologie et Santé Publique, F-37380 Nouzilly, France
‡Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, F-37000 Tours, France
§Institut de l'Elevage, 149 rue de Bercy, F-75595 Paris, France

ABSTRACT

Genomic selection in Lacaune dairy sheep was investigated based on genotypes from the OvineSNP50 BeadChip (Illumina Inc., San Diego, CA). Historical artificial insemination progeny-tested rams formed a population of 2,892 genotyped rams. Additional ungenotyped rams and females were included by single-step genomic BLUP (ssGBLUP). Three prediction strategies were tried: pseudo-BLUP (using all rams and daughter yield deviations), pseudo-ssGBLUP (using all rams and daughter yield deviations), and regular ssGBLUP (using all phenotypes and pedigree in an animal model). The population linkage disequilibrium was determined, with an average squared correlation coefficient of 0.11 for markers closer than 0.1 cM (lower than in dairy cattle). The estimated effective population is 370 individuals. Gain in accuracy of genomic selection over parent averages ranged from 0.10 to 0.20. Highest accuracies and lowest bias were found using regular ssGBLUP. Transition to a genomic breeding scheme is possible but costs need to be carefully evaluated.
Key words: genomic selection, Lacaune dairy sheep, accuracy

INTRODUCTION

Accurate genomic predictions that combine genotypic, phenotypic, and pedigree data available early in the life of livestock have the potential to reduce the generation interval for breeding schemes that focus on progeny testing (Meuwissen et al., 2001; Schaeffer, 2006). However, the potential reduction of generation interval for Lacaune dairy sheep is limited because the use of AI with fresh semen and a very rapid turnover of males result in an already short generation interval of 4.2 yr. However, early selection of fewer rams might decrease the maintenance costs of AI rams that are currently waiting for first-crop progeny tests.

The Lacaune dairy sheep breed was defined in the 1950s to 1960s by the pooling of several local breeds (Quittet and Franck, 1983). Genetic improvement programs for Lacaune dairy sheep started in the 1960s and have been managed by 2 AI companies [Confédération Générale de Roquefort (Millau, France) and Ovitest (Onet-le-Château, France)] since 1972. They progeny test 470 rams annually in a nucleus group of 368 commercial flocks (i.e., the companies are AI centers that do not own the flocks), with around 170,000 ewes recorded yearly for several traits. Those nucleus flocks, which are at the top of a pyramidal organization and include 20% of the Lacaune population (Barillet et al., 1986), benefit from extensive performance recording and AI. This is important because, contrary to dairy cattle, performance recording for sheep has high associated costs.

In practice, the existence of 2 companies implies 2 little-related subpopulations (which can be clearly seen using, for example, principal components analysis). Modeling the 2 subpopulations as a single one or as separate ones did not significantly change the results of this work (results not shown). That is, similar accuracies were observed when validation was performed within each subpopulation or the joint data set (results not shown).

Extensive AI for sheep is accomplished through the use of fresh semen, which requires the availability of numerous living rams to face seasonal AI demands that peak, in Lacaune, at 26,000 inseminations per week per company. Therefore, there is no storage of frozen semen. Thus, collection and storage of blood samples of AI rams began in 1995 in the view of QTL detection and localization and marker-assisted selection.

To develop a dense SNP chip for sheep, 20 countries organized the International Sheep Genomics Consortium (http://www.sheephapmap.org) in 2002. The OvineSNP50 BeadChip (Illumina Inc., San Diego, CA) was finally released in 2009.

Received June 13, 2013.
Accepted October 22, 2013.
[1]Corresponding author: francis.barillet@toulouse.inra.fr

Recently, a large set of Lacaune rams has been genotyped, with a focus on QTL detection and genomic selection. Duchemin et al. (2012) compared methods for genomic selection of Lacaune dairy sheep and reported an absolute increase in accuracy of 0.05 to 0.10 as assessed by forward validation compared with parent average. They used the same data for the reduced and full data sets as defined by Interbull (Mäntysaari et al., 2010), which overestimates accuracy (Olson et al., 2011) and makes the benefit of genomic selection for candidates unclear.

In addition, linkage disequilibrium (**LD**; the nonrandom association of alleles between 2 loci) was investigated because it strongly influences the power of QTL detection and the accuracy of genomic prediction (Pritchard and Przeworski, 2001; Ober et al., 2012). According to Meuwissen et al. (2001), genomic predictions reach an accuracy of 0.85 with a simulated LD of 0.20. Such a level of LD was hypothesized to be achieved with 35,000 multiallelic markers (Meadows et al., 2008) for meat and wool. However, Kijas et al. (2012) reported lower LD with the commercial OvineSNP50 BeadChip, which achieves such a density.

A factor related to LD is the effective population size: the smaller the population size, the higher the disequilibrium and the expected accuracy of linkage disequilibrium (Goddard, 2009). The latest estimates of effective population size in Lacaune sheep are pedigree based (Palhiere et al., 2000) and outdated.

The objectives of this study were to provide consolidated and updated LD measures across SNP markers and to determine the accuracy of genomic predictions through forward validation in the Lacaune breed. Concerning estimation of accuracy, this study differs from that of Duchemin et al. (2012) in the use of single-step genomic BLUP (**ssGBLUP**; Aguilar et al., 2010; Christensen and Lund, 2010), in the use of reduced and full data sets to assess accuracy by forward validation, and in a larger data set.

MATERIALS AND METHODS

Genomic Data

Currently, 5,000 progeny-tested rams with more than 10 daughters have DNA stored. Because exhaustive genotyping of rams was too expensive, the following genotyping strategy was chosen. Rams born from 2008 through 2009 and progeny tested were constituted as the validation population (see Table 1). Then, complete generations were (as much as possible) genotyped backward, including most rams born from 1998 through 2007 to form the training population. These included most ancestors (sires and maternal grandsires) of the validation population. No ram with less than 20 daughters in progeny testing was genotyped.

Table 1. Distribution of Lacaune dairy rams in the genomic selection test

Population	Rams (no.)	Birth year
Training	1,593	1999–2005
Excluded	707	2006–2007
Validation	592	2008–2009

Extraction of DNA from blood samples and genotyping was conducted by the LABOGENA laboratory (Jouy-en-Josas, France). Extracted DNA was available for 90% of the AI rams born from 2003 to 2009 and 25% for rams born from 1998 through 2002. Genotyped rams were roughly structured in 452 half-sib families, with a mean of 9 sons. Both breeding companies contributed to samples in accordance with the number of rams enrolled in their progeny-testing programs. The Illumina OvineSNP50 BeadChip developed through the International Sheep HapMap project (International Sheep Genomics Consortium, 2010) was used for genotyping. Illumina GenomeStudio software was used with default thresholds and cluster definitions for genotype calling.

From the set of SNP markers available, 92% were read for 99% of rams. Data for some SNP were discarded because of low minor allele frequency (<0.01), Hardy-Weinberg disequilibrium ($P < 10^{-5}$), or insufficient genotyping rate (<0.97). Mendelian inconsistencies were set to missing. Missing genotypes and genotyping errors (0.25%) were imputed using BEAGLE v3.4 software (Browning and Browning, 2007). The final data set included genotypes of 42,039 SNP from 2,892 rams born from 1996 through 2009 (Figure 1).

Phenotypes

For training and validation populations, phenotypes were daughter yield deviations (**DYD**; VanRaden and Wiggans, 1991), which were weighted by equivalent daughter contribution (**EDC**; Fikse and Banos, 2001; Sullivan, 2007). The DYD and EDC were used as the variable response to compute genomic predictions from a pseudo-ssGBLUP or pedigree-based prediction from a pseudo-BLUP that will be detailed later. Alternatively, observed performance measures (as used in official evaluation), were used in ssGBLUP. Training and validation populations were defined according to Interbull (Uppsala, Sweden) rules (Mäntysaari et al., 2010) for both methods. This defines a reduced data set and a full data set to fairly predict genetic merit of validation rams.