Contents lists available at ScienceDirect

Livestock Science
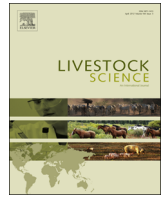
Accuracy and expected genetic gain under genetic or genomic evaluation in sheep flocks with different amounts of pedigree, genomic and phenotypic data

Joaquim Casellas [a,*], Jesús Piedrafita [b]

[a] Grup de Recerca en Millora Genètica Molecular Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
[b] Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ABSTRACTS

The socio-economic context of the sheep industry in multiple regions worldwide has prevented the systematic implementation of efficient genetic selection programs, although they are essential to optimize the viability of multiple sheep flocks. Moreover, little is known abut the minimum requirements to launch a selection program in sheep flocks as well as its expected profitability. This research focused on the simulation of sheep flocks with 100, 500, 1000, 5000 or 10,000 ewes that evolved during 10 years of pedigree, phenotypic, and even genomic data collection, and its subsequent evaluation by BLUP models. The evolution of the accuracy of predicted breeding values (**APBV**) as well as the expected genetic gain (**EGG**) were characterized year-by-year, also accounting for different availabilities of pedigree data (i.e., known or missing sires). As a whole, both parameters increased with the amount of pedigree and phenotypic data (or the number of years), although marginal increments were smaller with time. Other factors such as heritability and flock size also affected on the final trend, although the highest impact referred to pedigree/genomic data. The loss of accuracy and EGG when sire data was not collected was undoubful, as well as some relevant improvements when genomic data was available in large flocks. As a whole, pedigree and phenotypic data must be accumulated during a few years before performing the first genetic evaluations (the length of this preliminary stage depends on several parameters, although flock size is one of the most important), and completeness of pedigree data becomes essential to prevent unnecessary losses of APBV and EGG. The usefulness of of genomic selection seemed unclear, maybe with the only exception of largest flocks (10,000 ewes).

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Best linear unbiased prediction (**BLUP**) models were developed more than half a century ago (Henderson, 1950, 1973), and they revealed as a powerful tool to predict the genetic merit of the breeding stock. During the last decades, and even after the sudden emergence of genomic selection procedures (Meuwissen et al., 2001; Gianola et al., 2003), they were implemented in most of the predominant livestock breeds worldwide (Everett et al., 1983; Hunton, 1990; Rathje, 2000), although their impact on small ruminant species still remains limited (FAO, 2007). Animal breeding must be viewed as an under-utilized tool in the sheep industry in most countries (Pabiou et al., 2014), even though gains through genetic selection are clearly evident in other ruminant species such as beef (Koch et al., 2004; Enns and Nicoll, 2008) or dairy cattle (Hintz et al., 1978; Meland et al., 1982). This is mainly due to the socio-economic context of the sheep industry, although this kind of genetic initiatives must be enhanced, if possible, to optimize the economic viability of sheep flocks.

Even for single-flock designs, animal breeding programs may have complex logistics involving the (a) accurate collection of both phenotypic and pedigree data, (b) implementation of appropriate genetic evaluation models for the prediction of the genetic merit of the breeding stock, (c) planning of reproductive technologies in some cases, and (d) arrangement of mating, culling and selection decisions, apart from (e) some additional economic investment; profitability of any selection initiative must be double-checked in advance, in order to guarantee its success if moving forward. Within the context of the sheep species, there is a real shortage of scientific literature about the expected response to selection under variable productive scenarios and population structures; most of the available results focus on big populations where requisites for success in the genetic initiative are easily fulfilled (Roden, 1993;

* Correspondence to: Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
E-mail address: joaquim.casellas@uab.cat (J. Casellas).

Smulders et al., 2007; Valera et al., 2009). Nevertheless, those results cannot be generalized to small and local breeds as well as independent stockbreeders, whereas they have a remarkable contribution to the sheep industry from most of the countries (Carson et al., 2009; Blackburn et al., 2011; Hall, 2011). This becomes even more dramatic if additional genotyping costs are considered within the context of the animal breeding program to unravel paternities (Clarke et al., 2014) or, alternatively, to implement a genomic selection program (Meuwissen et al., 2001).

Given the lack of predictive tools for anticipating the expected genetic gain when artificial selection programs are implemented in sheep flocks with different sizes, the main objective of this manuscript was to simulate a broad range of sheep flocks and to evaluate the performance of the genetic or genomic evaluation models under different genetic backgrounds, flock sizes, and availabilities of pedigree and genomic data. This must be viewed as a basic tool for both the sheep industry and animal geneticists to anticipate the success (or failure) of selection initiatives in small-to-medium sheep flocks or even associative structures of stockbreeders.

2. Material and methods

Animal Care and Use Committee approval was not obtained for this study because all analyses were performed on simulated data sets. Neither real animals nor biological tissues from alive animals were involved in this research.

2.1. Simulated flocks

Stochastic processes implemented in this manuscript mimicked sheep flocks, and simulations focused on multiple levels of complexity, from sheep genome to flock management. Each individual was assumed to be diploid with 26 autosomal chromosomes of 100 cM each, this matching the real number of autosomal chromosomes in the sheep species, although with a slightly shorter genome length (Maddox et al., 2001). A total of 2000 single nucleotide polymorphisms (**SNP**) and 50 quantitative trait loci (**QTL**) were randomly distributed along each chromosome. Regardless of the size of the final flock to be analyzed, all simulations were founded with 50 males and 50 females that were heterozygous for all QTL and SNP. This base population evolved until generation 1000 under random mating, non-overlapping generations, and without modifying the effective population size. The aim of these preliminary generations was to reproduce the natural processes of genetic linkage and genetic drift across the sheep genome as previously described by Meuwissen et al. (2001); neither pedigree nor genomic data was retained from these preliminary generations.

After that, population expanded to the desired number of breeding stock, and evolved during 10 additional years as described below. Both recombination (Kosambi, 1943) and mutation phenomena were accounted for during the simulation process; mutation rates of 2.5×10^{-3} (SNP) and 2.5×10^{-5} (QTL; Meuwissen et al., 2001) were applied during the first 1000 preliminary generations (they switched the allele state from 1 to 2, or vice versa) and merged to 2.5×10^{-8} (Hickey and Gorjanc, 2012) afterwards. Only those populations with 250–500 polymorphic QTL and 30,000–35,000 polymorphic SNP in generation 1000 were kept for further analyses (Hayes and Goddard, 2001). Note that a genetic marker was assumed polymorphic if its minimum allele frequency (Chakraborty, 1992; Weir, 1992) exceeded 0.05. The additive allelic effect for the allele 1 of each QTL was fixed to 0 whereas the allelic effect for allele 2 was sampled from a standard Gaussian distribution; QTL effects were subsequently rescaled to generate an additive genetic variance (σ_a^2) of 1 after the first 1000 generations.

During the last 10 years of the simulation process, each flock expanded and maintained to the desired number of ewes, and the required number of rams was assigned by rounding up the 1:30 ram-to-ewe rate. The flock evolved under a once-a-year lambing system and 20% culling rate was assumed; required replacement ewe-lambs and ram-lambs were retained at random from the stock of lambs born every year, and pedigree data was store; genomic data was only retained for largest flocks (5000 and 10,000 ewes; see simulation scenarios below). Two phenotypic records were generated and stored every year for each alive ewe, one from a lowly heritable trait ($h^2 = 0.1$; e.g., litter size) and one from a moderately heritable trait ($h^2 = 0.3$; e.g., milk yield). These phenotypic traits were also influenced by two systematic sources of variation, the year effect and the ewe age with six levels (one, two, three, four, five, and six or more years); their effects were sampled from a uniform distribution between $-\sigma_e$ and σ_e (σ_e was the residual standard deviation of the trait). Moreover, a permanent environmental effect specific to each ewe was also accounted for during the generation of the phenotypic records. Its magnitude was sampled from a normal distribution with mean 0 and variance equal to σ_a^2; this effect must be viewed as a summary of any maternal contribution to the phenotype, as well as environmental effects permanently affecting the ewe during its productive life.

2.2. Simulation scenarios

This research focused on the implementation of a selection program in a domestic sheep flock and its evolution during the first 10 years in terms of statistical performance of the genetic evaluation models for a low- and a medium-heritable ewe-related phenotypic trait. Within this context, simulation scenarios varied in terms of flock size and pedigree/genomic data availability. A total of 100 replicates were simulated and analyzed for each specific scenario.

Note that five different flock sizes were assumed with 100, 500, 1000, 5000 and 10,000 ewe, and flock size was kept constant along the 10 years of pedigree, genomic and phenotypic data collection. On the other hand, different scenarios were created within each flock size in terms of pedigree/genomic data availability. The reference scenario assumed full knowledge of pedigree data and lack of genomic information, this illustrating a standard scheme under AI, controlled matings, or paternity identification testing. Nevertheless, this could not be realistic for most of the sheep flocks where multiple-sire mating systems are used and paternity testing is not systematically implemented. In this case, a secondary scenario was generated by removing sire data from the pedigree file.

In order to evaluate the impact of massive genomic information, additional scenarios were assumed for flocks with 1000 and 5000 ewes. In this case, pedigree was complemented with SNP data into the genomic evaluation models. Two different amounts of genomic data were considered by releasing genotypes of all rams and a percentage of ewes selected at random (20% and 50%). These two percentages were assumed in order to create different sizes of genotyped populations.

2.3. Genetic/genomic evaluation by BLUP

During the simulation process and for each flock, genetic (or genomic) evaluations where performed at the end of each year from year two to year ten (a total of nine independent genetic/genomic evaluations for each trait and flock), and restricting to pedigree, phenotypic and genomic data available at that point. Evaluations were performed within the context of BLUP models