Livestock Science 187 (2016) 61–67

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values

Hui Yang [a], Guosheng Su [b,*]

[a] Department of animal science, Fujian Vocational College of Agriculture, 350007 Fuzhou, China
[b] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

ARTICLE INFO

Article history:
Received 28 September 2015
Received in revised form
28 February 2016
Accepted 1 March 2016

Keywords:
Additive genetic variance
Breeding value
Pedigree information
Reliability
Phenotypic information

ABSTRACT

Reliability of estimated breeding values (EBV) depends on amount of information in both phenotypic data and pedigree data. This study investigated the impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values in populations of polytocous animals such as pig. Three populations which mimicked pig populations were generated. The animals were selected on a selection index including two traits (h^2=0.35 for trait 1 and h^2=0.10 for trait 2). Variance components and breeding values were estimated using different phenotypic data and pedigree data which cover different numbers of previous generations. Impact of phenotypic information of previous generations and depth of pedigree on EBV was evaluated by reliability of EBV for animals in the last generation (as candidates for breeding). Phenotypic data or pedigree data covering various numbers of previous generations led to different estimates of additive genetic variance (larger variance when earlier generations were included), especially for trait 1, because conceptual base populations were different in different data sets. Information from previous generations was more important for reliability of EBV of trait 2 than that of trait 1, and the contribution of previous phenotypic information to reliability of EBV was larger in the population with smallest sib group than the population with the largest sib group. According to the current study, phenotypic data including recent three to five generations and pedigree data added one to three ancestral generations will lead to asymptotic maximum reliability of EBV for animals in current population, dependent on population structure and heritability of trait. The results indicate: (1) In general, three generations of records plus two ancestral generations in pedigree is enough for predicting breeding values with regard to accuracy of EBV; (2) Variance components used for genetic evaluation should be estimated from the data set that is consistent with the data set for predicting breeding values with regard to conceptual base population; (3) In order to increase reliability of EBV in a polytocous animal population such as pig, the efficient approach is to measure the animals available within generation as many as possible so as to fully use information from sibs, instead of more data from early generations.

1. Introduction

Selective breeding is the most common approach to genetic improvement of livestock populations. Genetic gain by selection depends on accuracy of genetic evaluations for breeding candidates (Falconer and Mackay, 1996). The accuracy of estimated breeding value (EBV) for an animal is determined by the phenotypic information which has a contribution to the EBV (Henderson, 1975; Mrode, 2005). The more phenotypic information is used, the more accurate EBV can be obtained.

Best linear unbiased prediction (BLUP) of breeding values (Henderson, 1975) is a well-established methodology which integrates pedigree-based additive genetic relationship matrix and allows using phenotypic information of all relatives in the data for predicting breeding values. The more relatives of an individual have phenotypic records in the data, the more accurate EBV of the individual is. However, additive genetic relationships in livestock populations reduce 50% in each meiosis event. This means the information of earlier generations is less useful. The phenotypic records in the generations far from current generation may be not useful for predicting breeding values of breeding candidates, but increases computational demand. On the other hand, many studies have reported that selected data have an impact on estimation of genetic parameters and prediction of breeding values, and unbiased estimation can be obtained using the data including all information that led to the data in current generation (Goffinet, 1983; Sorensen and Kennedy, 1984; Van Der Werf and De Boer, 1990).

* Correspondene to: AU-Foulum, Blichers Alle 20, DK 8830 Tjele, Denmark.
E-mail address: guosheng.su@mbg.au.dk (G. Su).

http://dx.doi.org/10.1016/j.livsci.2016.03.001
1871-1413/

In addition to phenotypic data, pedigree data are required for genetic evaluation. In a BLUP model, the contribution of relatives' phenotypic information to EBV of an individual is realized though additive genetic relationship which is conventionally constructed using pedigree (Henderson, 1975; Mrode, 2005). In principle, more complete pedigree will lead to more efficient use of phenotypic information of relatives. But the distant ancestors have very little impact on the construction of relationships between candidates and their relatives with records. Thus it may be not necessary to keep distant ancestors in pedigree data. Moreover, the depth of pedigree could have a big impact on estimates of genetic parameters. Genetic parameters estimated using a linear mixed model reflect the genetic parameters of conceptual base population in which there is no genetic relationship between individuals (Kennedy and Moxley, 1975; Van Der Werf and De Boer, 1990; Bijma et al., 1997). When pedigree is traced back to distant ancestors, the base population will shift to an early generation, and the inbreeding coefficient in the resulting relationship matrix will increase. This implies that in a closed population, additive genetic variance estimated using a deeper pedigree could be larger than that obtained using a shallow pedigree.

The importance of information from relatives for accuracy of EBV depends on heritability of traits. According to the theory of quantitative genetics, information of relatives is more important for the traits with low heritability than those with high heritability (Falconer and Mackay, 1996). Furthermore, the importance of information from early generations is related to population structure. Compared with the population with large number of records per generation, information from previous generations is more important for the populations with small number of records per generation, because of relatively small number of contemporary relatives. So far, there is very little information in the literature (Mehrabani-Yeganeh et al., 1999; Lourenco et al., 2014) in terms of how depth of pedigree and how old phenotypic data being efficient for estimating genetic parameters and breeding values for animals in current generation.

The objective of this study was to assess the impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values in various scenarios. The study was based on the data simulated by mimicking pig populations, considering different traits and population structures with regard to population size and proportions of sires, dams and offspring.

2. Materials and methods

2.1. Data

The data used in this study were simulated by mimicking pig populations. To simplify the simulation procedure but without loss of generality, data were generated in a relatively simple procedure. The data of two correlated traits were generated simultaneously. Phenotypic value for a trait was generated as $y=\mu+a+e$, where μ was population mean in base generation, a was additive genetic effect (i.e., breeding value), and e was random residual. In base generation, breeding value for trait 1 (a_1) and trait 2 (a_2) was sampled from a bivariate normal distribution $\begin{bmatrix}a_1\\a_2\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2_{a1} & \sigma_{a1a2}\\ \sigma_{a1a2} & \sigma^2_{a2}\end{bmatrix}\right)$. In the following generations, breeding value of a trait for an animal was generated as $a=0.5(a_s+a_d)+a_m$, where a_s and a_d were breeding values of sire and dam, respectively, and a_m was Mendelian sampling deviation. Mendelian sampling deviations of the two traits for an animal were sampled from a bivariate normal distribution, $\begin{bmatrix}a_{m1}\\a_{m2}\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2_{m1} & \sigma_{m1m2}\\ \sigma_{m1m2} & \sigma^2_{m2}\end{bmatrix}\right)$, where the Mendelian sampling variance was defined as $\sigma^2_{mi}=0.5(1-F_p)\sigma^2_{ai}$, $(i=1, 2)$, where F_P was the average inbreeding coefficient of sire and dam (Mrode, 2005), and the covariance σ_{m1m2} was determined according to genetic correlation between the two traits, i.e., $\sigma_{m1m2}=r_{a1a2}\sqrt{\sigma^2_{m1}\sigma^2_{m2}}$. In each generation, random residuals of the two traits for an animals were sampled from a bivariate normal distribution $\begin{bmatrix}e_1\\e_2\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2_{e1} & \sigma_{e1e2}\\ \sigma_{e1e2} & \sigma^2_{e2}\end{bmatrix}\right)$.

Populations with non-overlapping generations were generated. In each generation, the sex of an animal was randomly assigned according to sex ratio 1:1. The breeding animals were selected according to an index $I=0.5y_1+0.5y_2$, where y_1 and y_2 were phenotypic values of trait 1 and trait 2. Selection based on phenotypic values instead of EBV was to simplify simulation procedure. This simplification would not lose generality with regard to the objective of this study.

The parameters for base population used in the simulation were as follows. For trait 1, population mean $\mu_1=100$, coefficient of variation $CV_1=15$, heritability $h^2_1=0.35$. Correspondingly, additive genetic variance $\sigma^2_{a1}=78.75$, and residual variance $\sigma^2_{e1}=146.25$. For trait 2, $\mu_2=100$, $CV_2=20$, $h^2_2=0.10$. Thus, $\sigma^2_{a2}=40$, and residual variance $\sigma^2_{e2}=360$. Genetic correlation and residual correlation between the two traits were set to -0.30.

The simulation included three scenarios of population structure. (1) Population S50D20N10 had 50 sires per generation, each sire mated 20 dams and each dam had one litter with 10 individuals having phenotypic records, corresponding 10,000 animals with records in each generation. (2) Population S100D20N5 had 100 sires per generation, each sire mated 20 dams and each dam had one litter with 5 individuals having phenotypic records. (3) Population S100D5N3 had 100 sires per generation, each sire mated 5 dams and each dam had one litter with 3 individuals having phenotypic records. All populations consisted of 25 generations of animals among which generation 0 (base generation) did not have phenotype records and generation 1 to generation 24 had phenotypic records. For each population, 50 replicates were generated.

2.2. Statistical analysis

For each population, genetic parameters and breeding values were estimated using various datasets which differed in number of generations in phenotypic data and pedigree data. Six phenotypic data sets were created; each included recent 1, 2, 3, 5, 10 or 20 generations of records. The pedigree datasets included the generations in phenotypic data and plus 1, 2, 3, 5, 10 or all ancestral generations, depending on the corresponding phenotypic data set. For example, for phenotypic data set comprising 20 generations, the pedigree data was possible to include 5 more generations at maximum, since the total number of generations available was 25. For each scenario of data set, 50 replicates were analyzed.

A two-trait linear mixed model was used to estimate variance components and predict breeding values (BLUP approach) of the two traits. The model is,

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix}=\begin{bmatrix}\mathbf{1}\mu_1\\ \mathbf{1}\mu_2\end{bmatrix}+\begin{bmatrix}\mathbf{Z_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z_2}\end{bmatrix}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix}+\begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix}$$

where $\mathbf{y_1}$ and $\mathbf{y_2}$ are the vectors of observations for trait 1 and trait 2, μ_1 and μ_2 are the intercepts, $\mathbf{1}$ is a vector of ones, $\mathbf{a_1}$ and $\mathbf{a_2}$ are the vectors of additive genetic effects, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are incidence matrices linking $\mathbf{a_1}$ and $\mathbf{a_2}$ to $\mathbf{y_1}$ and $\mathbf{y_2}$, $\mathbf{e_1}$ and $\mathbf{e_2}$ are the vectors of