Livestock Science 204 (2017) 98–103

Contents lists available at ScienceDirect

Livestock Science

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

The influence of natural selection in breeding programs: A simulation study

Silvia García-Ballesteros[a], Juan Pablo Gutiérrez[a], Luis Varona[b,c], Jesús Fernández[d,*]

[a] Departamento de Producción Animal, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, E-28040 Madrid, Spain
[b] Unidad de Genética Cuantitativa y Mejora Animal, Universidad de Zaragoza, E-50013 Zaragoza, Spain
[c] Instituto Agroalimentario de Aragón (IA2), E-50013 Zaragoza, Spain
[d] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Genetic diversity
Fitness
Stabilizing selection
Selection programs

ABSTRACT

Natural selection, acting directly on fitness or through stabilizing selection for other traits, has always been present in natural populations. In principle, this force will reduce the genetic variability, which is required for artificial selection programs. The genetic improvement of a trait with economic interest by selection programs depends on the amount of genetic diversity for that trait in a particular population. Most studies on the development of breeding programs account only for the genetic variation for the target trait itself. The objective in this work is determining, through computer simulations, the consequences for the evolution of selection programs, considering not only the variability for the trait of interest but also for fitness under different mutational models. Additionally, a scenario where the trait was subjected to stabilizing selection was also studied. Different parameters as effective population size, phenotypic mean, phenotypic variance and heterozygosity were used to monitor the performance in the different scenarios. In conclusion, considering the action of direct natural selection does not lead to lower levels of genetic variability for neutral traits, and thus it does not reduce the ability of populations to respond to artificial selection regardless of the mutational model used. On the other hand, stabilizing selection penalizes individuals with an extreme phenotype (which is the major objective in the artificial selection), reducing their fitness. Therefore, the artificial selection was ineffective in improving traits subjected to stabilizing selection.

1. Introduction

The increment (or modification) of the phenotypic mean for traits of interest is the main objective in genetic selection programs. The magnitude of the response to artificial selection is proportional to the amount of genetic variability for that trait in the population. Therefore, the genetic diversity present at the beginning of the process will be crucial, because it determines the ability to evolve or to be genetically improved. If we only consider the neutral variability, its level depends on the historical effective population size (N_e), being the genetic drift the process determining the rate of loss of genetic variability.

The genetic variability could be described as the existence of different alleles and genotypes and the first action before applying a selection program to a population should be to estimate its initial genetic diversity, and to predict the changes in it for the planned selection intensity. In general, the most frequent measures of genetic diversity based in locus are the observed heterozygosity (OH; proportion of heterozygous individuals), expected heterozygosity (EH; the heterozygosity in a population in Hardy-Weinberg equilibrium with the same allelic frequencies) and the allelic diversity (AD; number of different alleles for a locus). For quantitative traits, diversity is usually measured through the additive genetic variance (V_A) and the heritability (h^2) defined as the proportion of phenotypic variance explained by the additive genetic effects (Falconer and Mackay, 1996).

However, the fate of any population, whatever subjected to a selection program or not, also depends on the fitness of the individuals belonging to it (Reed and Frankham, 2003). Fitness is the capacity of survival, adaptation and reproduction of populations (Roff and Mousseau, 1987). Natural selection directly acts on fitness of population, shaping its genetic diversity in an equilibrium with the input occurring due to mutation. Natural selection may also interacts with the genetic diversity for quantitative (productive) traits through two mechanisms: i) natural selection reduces the census population size (N) leading indirectly to the loss of genetic diversity for any trait; ii) it has a direct effect due to the genetic correlation between fitness and the traits.

A particular way of interaction between fitness and productive traits is the so called stabilizing selection. Stabilizing selection favors

* Corresponding author.
E-mail address: jmj@inia.es (J. Fernández).

http://dx.doi.org/10.1016/j.livsci.2017.08.017
Received 1 February 2017; Received in revised form 8 June 2017; Accepted 18 August 2017

individuals with phenotypes close to an optimum value (i.e. they have higher fitness) and penalizes individuals far from it (Kingsolver et al., 2001; García-Dorado et al., 2007). Studies as Johnson and Barton (2005) have proven that stabilizing selection leads to the reduction of genetic diversity for the trait.

It is also well known that artificial selection implies the reduction in genetic variability for the target trait but also for fitness due to the reduction in N_e. This may lead to the fixation of deleterious mutations threatening the survival of the population (Lynch et al., 1995).

The objective in this work was to determine, through stochastic simulation, how the combination of natural directional and stabilizing selection affects the response to selection for a trait of interest. The viability of implementing selection programs under different scenarios was evaluated.

2. Materials and methods

2.1. Model assumptions and simulations

Diploid individuals with 20 chromosomes of 1 Morgan each were simulated. The genome comprised two types of biallelic loci: selective marker loci with effect on fitness and neutral marker loci. Each chromosome carried 1000 and 3000 selective and neutral loci, respectively. The selective loci were randomly allocated within the chromosome. One hundred neutral loci were randomly selected to control a quantitative trait, which was the objective of the breeding program. No loci affected fitness and the quantitative trait at the same time and, therefore, in the basic scenario no genetic correlation existed between them, besides the one arising from linkage disequilibrium between loci affecting fitness and the trait, respectively.

Our simulations for the different scenarios can be divided in two steps: first, we generated base populations in mutation-drift or mutation-selection-drift equilibrium; second, we implemented classic artificial selection on the trait of interest.

We simulated 5000 discrete generations for a randomly mating population of constant size ($N = 100$), where half the population were females and half males. Genotypes for the neutral loci were generated in the initial population by randomly choosing alleles between two possibilities (i.e. initial frequencies were 0.5). Selective loci were all fixed for the 'wild' type allele. We verified that the mutation-drift or mutation-drift-selection equilibrium has been achieved, by detecting the stabilization of the allelic frequencies and the expected heterozygosity. The rationale behind this first step was getting a sensible distribution of allelic frequencies (not a high proportion of fixed loci) and a reasonable linkage disequilibrium between all types of loci.

Every generation, gametes were created allowing for recombination between the pairs of homologous. Accordingly with the simulated chromosome length, the actual number of recombinations per chromosome was obtained from a Poisson distribution with $\lambda = 1$ and randomly allocated along the chromosome.

The neutral loci mutation rate was 2.5×10^{-3} by locus and it could occur in both alleles, switching allele 0 to allele 1 and vice versa. Lower mutation rates would imply lower levels of heterozygosity at the equilibrium and few segregating loci. This could be compensated by enlarging the population size and the number of generations in the first step. But general results would not change substantially (Gómez-Romano et al., 2013).

2.2. Fitness mutational models

For selective loci mutations, we consider two contrasting models regarding the mutation rate and the distribution of the effects of mutations:

- Mukai et al. (1972), numerous mutations with little effect occur in populations. They are difficult to be eliminated by natural selection, as only individuals with many mutations do not survive. Mutation rate used was $\mu = 0.5$ per haploid genome and generation.
- CGD (García-Dorado and Caballero, 2000) the mutations are few but with great effect on the individual fitness. They are easier to remove, because carriers die and do not transmit the deleterious mutations (natural selection is more effective in the removal). Mutation rate was $\mu = 0.03$ per haploid genome and generation (see selection coefficients below).

The number of selective loci mutations per individual were also obtained from a Poisson distribution with mean equal to twice the haploid mutation rate (i.e. 1 and 0.06 for Mukai and CGD models, respectively), and randomly allocated along the whole genome. The fitness of an individual was calculated as the product of the values for each selective *locus* (Lynch and O'Hely, 2001), being 1 for 'wild' type homozygotes, (1 - s) for 'mutant' type homozygotes and (1 - hs) for heterozygotes. The initial fitness was 1 (without deleterious 'mutant' type alleles) and consequently, the more mutations accumulated in selective loci the lower the individual fitness. Backward mutation was not allowed, thus, only changes from 'wild' type to deleterious 'mutant' type alleles were allowed.

The mean effect s and the mean dominance coefficient h of mutations were $s = 0.05$ and $h = 0.36$ under Mukai's model and $s = 0.26$ and $h = 0.20$ for CGD. The effect of each selective locus was sampled from a gamma distribution (with parameters α and β, being β the shape parameter). The values for each of the models were $\beta = 1$ and $\alpha = 0.05$ for Mukai and $\beta = 2.3$ and $\alpha = 0.61$ for CGD model. The dominance coefficient of mutations was obtained from a uniform distribution between 0 and exp(-ks), where k is a constant allowing the mean dominance coefficient to be the desired one. Actual values used were $k = 991$ and $k = 26$ for Mukai and CGD, respectively. The Mukai model parameters corresponds to an exponential distribution of gene effects that has been proposed as a likely distribution (see, e.g. Schultz and Lynch, 1997). The CGD model parameters have been obtained through minimum distance estimation of Drosophila data (García-Dorado et al., 1999). The particular combinations of alpha and beta used to define both models are the same as found in de Cara et al. (2013).

Offspring for the next generation were created from random chosen males and females. To decide if a newborn individual survived, we drew a random number to simulate environmental stochasticity from a uniform distribution between 0 and 1 and compared with the fitness for that individual (see the calculations of fitness above). If the fitness was less than this random value, the individual died and a new offspring was generated from new randomly chosen parents. This process was repeated until 100 offspring survived. In this study fitness was expressed as differences in viability, with no effect on fecundity or fertility. In general, larger number of matings (attempts) were needed to maintain the population size as the generations passed.

2.3. Quantitative trait

Genetic value (G_i) for the target trait was the summation of the individual values of each locus along the 100 neutral loci with effect on the trait (which were randomly chosen among the neutral loci), following the expression:

$$G_i = \sum_{j=1}^{j=n} a_{ij} \quad (1)$$

where a_{ij} is the genetic value of individual i for locus j. This implies the trait was additive across loci (i.e., no epistasis). The value of each locus was assigned through an allele coding system (Strandén and Christensen, 2011), being 1 for the homozygotes (11), 0 for the heterozygote, and −1 for the opposite homozygotes (00). The trait was also additive within locus. Therefore, the maximum genetic value for the trait was 100, when all loci were fixed for the allele 1, and the minimum −100, in the case of every loci homozygous for the allele 0.