Dominance in stochastic simulations of animal breeding programs

R.C. Gaynor, G. Gorjanc & J.M. Hickey

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Research Centre, Midlothian EH25 9RG, UK
chris.gaynor@roslin.ed.ac.uk (Corresponding Author)

Summary

The exploitation of heterosis and/or the avoidance of inbreeding are fundamental considerations in the design of animal breeding programs. However, stochastic simulations modeling these breeding programs are often limited in their ability to simply and clearly simulate either. To address this issue, we present formulae for inbreeding depression, heterosis and purebred-crossbred correlation in the context of adding dominance to stochastic simulations. We show how these formulae can be used to tune simulation parameters to achieve targeted levels for each of these genetic properties. This framework will make stochastic simulations better equipped to model complex breeding programs and aid in the design of entirely new breeding programs.

Keywords: animal breeding, heterosis, stochastic simulation

Introduction

Many animal breeding programs are designed specifically to exploit the beneficial aspects of heterosis through the use of crossbred animals, and presumably all breeding programs take steps to avoid the detrimental effects of inbreeding depression. Given the significance of these genetic properties it is somewhat surprising that many stochastic simulations fail to model dominance gene action, an important mechanism that gives rise to both heterosis and inbreeding depression. We believe that the reason for this is not due to a failure of researchers to realize the importance of dominance, but rather a widespread uncertainty in how to realistically model it. Failing to include dominance in these simulations has limited their ability to model important genetic properties such as heterosis and inbreeding depression, but also purebred-crossbred correlation.

Heterosis and inbreeding depression can be thought of as two sides of the same coin. Falconer and Mackay (1996) refer to heterosis as the converse of inbreeding depression because both can be explained as a byproduct of dominance gene action. This explanation comes from one of two competing theories: the dominance theory and the over-dominance theory (Lamkey & Edwards, 1999). Of these theories, the dominance theory has become the more widely accepted. Other mechanisms, such as epistasis can play a role, but since dominance alone can model heterosis and inbreeding depression we will focus only on modeling dominance.

The genetic correlation between purebred and crossbred performance is an important parameter for breeding programs that produce crossbred animals. The magnitude of this value is important for determining the relative efficiencies of breeding programs using pure-line...
selection, crossbred selection, or combined crossbred and purebred selection (Wei & Werf, 1994). Its value likely also plays a role in determining the value of adopting new breeding strategies to exploit heterosis, such as reciprocal recurrent genomic selection (Kinghorn et al., 2010).

Wellmann and Bennewitz (2011) provide an overview of the quantitative genetics of dominance gene action. They also suggest how to design stochastic simulations to include dominance gene action. However, they do not directly address crossbreeding. To address this important aspect of dominance, we present here our own guide for incorporating dominance in simulations. We do this by presenting formulae for inbreeding depression, heterosis and purebred-crossbred correlation. We have also made an additional simplifying assumption not used by Wellmann and Bennewitz to allow for easier, if slightly less realistic, construction of simulations based on the derived formulae.

Theory

This section presents formulae pertaining to inbreeding depression, heterosis and purebred-crossbred correlation in the context of a stochastic simulation. The structure of these simulations match those used by a software package that we have developed called AlphaSimR (available at: https://bitbucket.org/hickeyjohnteam/alphasimr/). The key features of these simulations are:

1. A coalescent simulator that does not model selection is used to simulate initial haplotypes that are paired to form an initial population in Hardy-Weinberg Equilibrium (HWE). Note that the coalescent model does not accommodate selection.
2. All quantitative trait nucleotides (QTN) are biallelic and have additive ($\alpha$) and dominance ($d$) effects.
3. Dominance effects are calculated by multiplying the absolute value of the additive effects by dominance coefficients ($\delta$).
4. Dominance coefficients are independent of additive effects and sampled from a normal distribution whose mean is $\mu_\delta$ and variance is $\sigma_\delta^2$.
5. Additive effects are initially sampled from a normal distribution with mean zero and variance one.
6. The additive effects are finally scaled linearly to achieve a desired value of additive genetic variance ($\sigma_A^2$) the initial animals in the first purebred line.

The final scaling of additive effects results in the variance of additive effects ($\sigma_A^2$) depending on multiple parameters within the simulation. This dependency is shown in formula (1) using the expectation of additive effect variance conditional on the number of QTN and their allele frequencies. To simplify this formula, we have used an approximation that ignores the contribution of linkage disequilibrium.

$$E(\sigma_A^2) \approx \frac{\sigma_A^2}{2\sum_{l=1}^{n_{\text{QTN}}} \frac{1}{p_l q_l [1 + (\mu_\delta^2 + \sigma_\delta^2)(p_l - q_l)^2]}}$$  (1)
Inbreeding depression

There are multiple formulae for expressing inbreeding depression. Falconer and Mackay (1996) expressed it as the change in mean due to a percentage change of the inbreeding coefficient. Wellmann and Bennewitz (2011) expressed inbreeding depression as the decrease in mean as the inbreeding coefficient goes from 0 to 1. This second formulation represents the difference between a population in HWE and one that is fully inbred. We have chosen to use this second formulation to describe inbreeding depression and present it below as formula (2).

\[ I_k = 2 \sum_{i=1}^{n} d_i p_i q_i = 2 \sum_{i=1}^{n} \delta_i |a_i| p_i q_i \]  

(2)

The expectation for inbreeding depression is more useful to researchers designing stochastic simulations. We show this with the expectation of inbreeding depression conditional on the number of QTN and QTN allele frequencies. Conditioning on these variables leaves the only unknown variables left as dominance coefficient and additive effect. Since we sample both from normal distributions, we can obtain their expectations using the properties of normal distributions. The resulting expectation is shown below as formula (3). Note that formula (3) is expressed in terms of additive genetic standard deviation by using formula (1).

\[ E(I_k) = \frac{\mu_s \sum_{i=1}^{n} p_i q_i}{\sqrt{2\pi \sum_{i=1}^{n} p_i q_i \left[ 1 + (\mu_s^2 + \sigma_s^2)(p_i - q_i)^2 \right]}} \]  

(3)

Formula (3) enables the impact of changing simulation parameters on inbreeding depression to be predicted. For example, formula (3) shows that changing additive genetic standard deviation results in a proportional change to inbreeding depression. Changes to the mean and variance of dominance coefficients result in more complicated changes to inbreeding depression. To more clearly show these changes, we have reduced formula (3) to the proportionality in formula (4).

\[ E(I_k) \propto \frac{\mu_s}{\sqrt{\varepsilon + \mu_s^2 + \sigma_s^2}} \]  

(4)

The proportionality given in formula (4) shows that inbreeding depression increases with an increase in mean dominance coefficient and decreases with an increase to dominance coefficient variance. The increase due to increasing mean of dominance coefficients is asymptotic due to the presence of the mean for dominance coefficients in the denominator. The denominator also contains a constant (\( \varepsilon \)) that depends on the number of QTN and their allele frequencies. Therefore the position of the asymptote depends on these parameters. The last component of the denominator is dominance coefficient variance. Its inclusion in the denominator leads to the decrease in inbreeding depression with an increase to dominance coefficient variance.
Formula (3) also shows how inbreeding depression changes as a function of the number of QTN. This is shown by examination of the two summations involving the number of QTN, one in the numerator and one in the denominator. In the numerator, changing the number of QTN results in a change proportional to the change in number of QTN. In the denominator, changing the number of QTN results in a change that is proportional to the square root of the change in the number of QTN. Taken together, this means changes in inbreeding depression due to changing the number of QTN are approximately proportional to the square root of the change in the number of QTN. Thus, a four times increase in the number of QTN is expected to approximately double the inbreeding depression when all other parameters are held constant.

**Single-cross heterosis**

To show how heterosis works in stochastic simulation we use the definition for single-cross heterosis given by Falconer and Mackay (1996). This measure of heterosis reflects the difference between the average performance of the crossbred animals and the average performance of their parental purebreds. The formula for single-cross heterosis is given in formula (5). Formula (5) contains terms for allele frequencies in each of the purebred lines. This is denoted with subscripts 1 and 2, where subscript 1 relates to the first purebred line that is used to set additive genetic variance.

\[
H = \sum_{i=1}^{n_{\text{QTN}}} q_i (p_{i1} - p_{i2})^2 = \sum_{i=1}^{n_{\text{QTN}}} q_i (p_{i1} - p_{i2})^2
\]  

(5)

Formula (5) shows that single-cross heterosis depends on both dominance and differences in allele frequencies. This means that both are needed at a locus for that locus to contribute to single-cross heterosis. As with inbreeding depression, greater insight can be gained by examining the expectation of this formula. We show this in formula (6), which is a function of the number of QTN and their allele frequencies.

\[
E(H) = \mu_0 \sigma^2 \sum_{i=1}^{n_{\text{QTN}}} (p_{i1} - p_{i2})^2 \\
\sqrt{\pi \sum_{i=1}^{n_{\text{QTN}}} p_{i1} p_{i2} [1 + \mu^2 + \sigma^2] (p_{i1} - q_{i1})^2}
\]

(6)

Formula (6) for expected heterosis follows some of the same trends as formula (3) for expected inbreeding depression. This means that changes to additive genetic standard deviation result in proportional changes to expected heterosis. Changes to the mean and variance of dominance coefficients results in changes to expected heterosis similar to the changes given in formula (4). Finally, expected heterosis also changes with changes to the number of QTN at a rate approximately proportional to the square root of the number of QTN.

**Purebred-crossbred correlation**

Purebred-crossbred genetic correlation measures association between the within breed breeding values for animals to their breeding values for crossbred performance for a given trait. Since breeding values depend on allele frequencies, the formula for purebred-crossbred
correlation depends on the allele frequencies of the parental breeds mated to form said crossbred animals. We present below the expectation of this correlation for single-cross crossbreds. The expectation pertains to purebred-crossbred correlation in the breed denoted by subscript 1.

\[
E(r_{PC}) = \frac{\sum_{i=1}^{\infty} p_i q_i \left[ 1 + \left( \mu_i^2 + \sigma_i^2 \right) (q_i - p_i) (q_i - p_i) \right]}{\sqrt{\sum_{i=1}^{\infty} p_i q_i \left[ 1 + \left( \mu_i^2 + \sigma_i^2 \right) (q_i - p_i) \right]^2} \sqrt{\sum_{i=1}^{\infty} p_i q_i \left[ 1 + \left( \mu_i^2 + \sigma_i^2 \right) (q_i - p_i) \right]^2}}
\]

(7)

The expectation given in formula (7) does not lead to as clear a set of trends as was observed for the previous expectations. This is due to dependencies on allele frequencies in both breeds. It is however clear that purebred-crossbred correlation is one if there is no dominance. Later in the paper we present an example simulation that more clearly illustrates the trend in formula (7) with regards to changes to the mean of dominance coefficients.

**Bulmer effect**

Formula (7) is for the expected purebred-crossbred correlation under an assumption of no selection. This assumption is clearly violated in animal breeding populations, because most are under selection for purebred performance. Thus, some allowance for selection must be made if the goal is to tune simulation parameters to match values of purebred-crossbred correlation observed in real populations. This can be accomplished by simulating one or more cycles of selection on purebred performance prior testing further scenarios. The cycles of selection will change the purebred-crossbred correlation by inducing genetic covariance in accordance with the Bulmer effect (Tallis, 1987).

The amount of covariance introduced by the Bulmer effect can be predicted to estimate purebred-crossbred correlation after one or more cycles of selection on purebred performance. We derived a formula for this based on formulae derived by Tallis (1987). These formulae all assume selection does not change allele frequency. Since this assumption is easily violated in a finite simulation, we have chosen to consider only one cycle of selection. In addition, we observe that additional cycles of selection do not lead to large changes in purebred-crossbred correlation. The formula for purebred-crossbred genetic correlation after one cycle of selection on purebred performance and followed by random mating is given in (8).

\[
r_{PC(0)} = \frac{r_{PC(0)} \sqrt{2 - k}}{\sqrt{2 - k r_{PC(0)}}}
\]

(8)

Formula (8) depends on the initial purebred-crossbred correlation (\(r_{PC(0)}\)) and the variable k. The value of k depends solely on selection intensity (Cochran, 1951; Bijma, 2012). The initial purebred-crossbred correlation can be predicted using the expectation in formula (7). This means that purebred-crossbred correlations can be predicted for a range of simulation parameters with known values for the number of QTN and their allele frequencies.

**Simulation**
The behavior of formula (7) for expected purebred-crossbred correlation was illustrated with a small simulation. The simulation started with 2000 genderless animals simulated by AlphaSimR using a population history similar to pigs as described in the manual for AlphaSim (Faux et al., 2016). A population split at 100 generations ago was added to form two separate breeds. The simulated species contained 19 chromosome pairs and 100 QTN per chromosome. The mean (0.193) and variance (0.097) of dominance coefficients were taken from Bennewitz and Meuwissen (2010). Note that Bennewitz and Meuwissen derived these values from detected QTL for meat quality and carcass traits, and these traits tend to show relatively low levels of heterosis. The additive genetic variance was arbitrarily set to 20 units.

The purebred-crossbred correlations in the simulated animals were measured both before and after a single cycle of selection and random mating. The top 5% of animals were selected based on true breeding values. The expected correlations both before and after selection were also calculated for the parameters used in the simulation and a range of alternative values for the mean dominance coefficient.

The simulation was repeated 10 times and the results are plotted in Figure 1. The plot shows good correspondence between the observed and predicted values. It also clearly shows that selection decreases the value of purebred-crossbred correlation and that increasing values for mean dominance coefficient results in lower purebred-crossbred correlations.

**Discussion**

In this paper we present formulae that can be used to tune parameters relating to dominance in stochastic simulations to more closely match real-world genetic parameters. We believe these formulae provide a framework for incorporating dominance gene action in stochastic simulations. Doing so will better equip simulations to manifest inbreeding depression and heterosis, which in turn can be used to design new breeding schemes that account for these important genetic parameters.

The formulae presented in this paper make several assumptions that are unlikely to be true. For example, we assumed dominance coefficients are independent of additive effects. For more realistic assumptions and a thorough discussion of those assumptions we direct the reader to Wellmann and Bennewitz (2011).

Several formulae in this paper are approximate, because they ignore the contribution of linkage disequilibrium. From theory and simulations not shown here, we expect the contribution from linkage disequilibrium to increase as the number of QTN and/or the mean dominance coefficient increases. When this occurs, we have observed that the formulae for heterosis and inbreeding depression remain fairly robust. However, the formula for purebred-crossbred correlation overestimates the actual correlation, potentially by a large amount. We plan to release a paper with formulae accounting for linkage disequilibrium in the future. Until such time, we suggest that the trends presented in the paper are of more valuable than the formulae themselves.

We specifically chose to use more simplified assumptions in our formulae to make designing stochastic simulations easier. The idea here being that a more simplified simulation
is more likely to be widely used and that it is better to use a simplified model of dominance than to not use dominance at all. After all, the infinitesimal model makes many simplifying assumptions and has been successfully used in many quantitative genetics experiments. It remains useful even though it clearly does not represent true genetic architecture.
Figure 1. Observed and expected purebred-crossbred correlations. The observed correlations prior to selection are represented with solid dots and the observed correlations after selection are represented with open dots. The solid line represents the average expected correlation before selection using formula (7) and the dashed line represents the average expected correlation after selection using formulae (7) and (8).
List of References


