Theor Appl Genet (1993) 87:423–430

Covariance between relatives in multibreed populations: additive model

L. L. Lo, R. L. Fernando, M. Grossman

Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1207 West Gregory Drive, Urbana, IL 61801, USA

Received: 25 January 1993 / Accepted: 20 April 1993

Abstract. Covariance between relatives in a multibreed population was derived for an additive model with multiple unlinked loci. An efficient algorithm to compute the inverse of the additive genetic covariance matrix is given. For an additive model, the variance for a crossbred individual is a function of the additive variances for the pure breeds, the covariance between parents, and segregation variances. Provided that the variance of a crossbred individual is computed as presented here, the covariance between crossbred relatives can be computed using formulae for purebred populations. For additive traits the inverse of the genotypic covariance matrix given here can be used both to obtain genetic evaluations by best linear unbiased prediction and to estimate genetic parameters by maximum likelihood in multibreed populations. For nonadditive traits, the procedure currently used to analyze multibreed data can be improved using the theory presented here to compute additive covariances together with a suitable approximation for nonadditive covariances.

Key words: Covariance – Additive model – Segregation variance – Genetic evaluation – Multibreed population

Introduction

The theory for covariance between relatives provides the basis to estimate genetic parameters and to use data from relatives in genetic evaluation. Covariance theory is relatively well understood for a purebred population (e.g., Kempthorne 1954). This is not the case, however, for a multibreed population, which is composed of several breed groups, each of which is made up of individuals belonging to a pure breed (e.g., Landrace of Duroc in swine) or a cross breed (e.g., F_1, F_2, or backcross).

The genetic value for an individual can be modelled as the sum of the additive and nonadditive components (Kempthorne 1954). In modelling the additive component, it is often assumed that variances are equal across breed groups (Hakim et al. 1990; Lo et al. 1992; Swan and Kinghorn 1992; Van Vleck et al. 1992). It has been shown, however, that this assumption is not appropriate for a multibreed population (Elzo 1983, 1990; Wei et al. 1991a, b).

To address the above problem, Elzo (1983, 1990) computed additive variance for a crossbred group as a weighted mean of additive variances of the pure breeds plus one half the covariance between parents, where each weight is the proportion of the corresponding pure breed in the crossbred. Mulibreed data have been analyzed using this approach in several studies (Elzo and Bradford 1985; Elzo and Famula 1985; Arnold et al. 1992). It can be shown, however, that this approach does not always lead to the correct additive variance for a crossbred group. The additive variance for the F_1, for example, is different from that for the F_2 even though the F_1 has the same breed composition as the F_2 crossbred (e.g., Lande 1981).

The objective of the present paper is to develop a general approach to compute genotypic variances and covariances between relatives in a multibreed population comprising an arbitrary number of pure breeds and all crosses involving these breeds, under a model with additive inheritance and multiple unlinked loci.

Supported in part by the Illinois Agricultural Experiment Station, Hatch Projects 35-0345 (RLF) and 35-0367 (MG)
Communicated by E. J. Eisen
Correspondence to: R. L. Fernando

Method

Consider a genotypic model with n unlinked loci in a random mating multibreed population. It is assumed that the pure breeds are in gametic equilibrium so that the covariance between alleles at any two loci will be null. At locus t, let S^t_i be the random allele that offspring i inherited from its sire j, and let D^t_i be the random allele that i inherited from its dam k. The paternal and maternal alleles of sire j will be denoted as S^t_j and D^t_j; those of dam k will be denoted as S^t_k and D^t_k (Fig. 1). The genotypic value of offspring i is denoted G_i.

Genotypic model

Consider as a reference breed group one that contains all alleles at each locus that are present in the multibreed population at that locus. In a multibreed population involving two pure breeds, for example, F_2 is a suitable reference breed group. In any breed group, the genotypic value of individual i, G_i, can be modelled as

$$G_i = \mu + \sum_{t=1}^{n} (\alpha_{S^t_i} + \alpha_{D^t_i}) \tag{1}$$

where

$$\mu = E(G_i) \tag{2}$$

is the genotypic mean in the reference breed group;

$$\alpha_{S^t_i} = E(G_i | S^t_i) - \mu \tag{3}$$

is the additive effect of paternal allele S^t_i in the reference breed group; and

$$\alpha_{D^t_i} = E(G_i | D^t_i) - \mu \tag{4}$$

is the additive effect of maternal allele D^t_i in the reference breed group. Note that $E(\alpha_{S^t_i}) = E(\alpha_{D^t_i}) = 0$ in the reference breed group.

Covariance

Covariance between relatives is the covariance between genotypic values of related individuals. The well-known tabular method used to construct the matrix of additive covariances in purebred populations (Henderson 1976; Quaas et al. 1984) is based on defining the additive covariance between individuals i and i' as the average covariance between i' and the parents of i, if i' is not a direct descendent of i (Chang et al. 1991). In the following development, we will show that the above result is also true for multibreed populations, under an additive model.

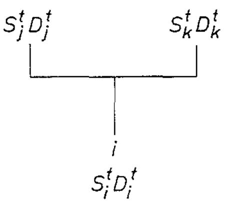

Fig. 1. Notation for alleles at locus t in individual i and in its parents, j and k

Let $G_{i'}$ be the genotypic value for individual i', where i' is not a direct descendent of individual i. From the genotypic model (1), the covariance between individuals i and i' is

$$\begin{aligned} \mathrm{Cov}(G_i, G_{i'}) &= \mathrm{Cov}\left(\sum_{t=1}^{n} \alpha_{S^t_i}, G_{i'}\right) + \mathrm{Cov}\left(\sum_{t=1}^{n} \alpha_{D^t_i}, G_{i'}\right) \\ &= \sum_{t=1}^{n} \mathrm{Cov}(\alpha_{S^t_i}, G_{i'}) + \sum_{t=1}^{n} \mathrm{Cov}(\alpha_{D^t_i}, G_{i'}) \end{aligned} \tag{5}$$

At locus t, random allele S^t_i of indivudual i is either allele S^t_j or D^t_j from its sire j. Thus, let Z be a random variable that is $S^t_i \Leftarrow S^t_j$ if S^t_j is a copy of S^t_j or is $S^t_j \Leftarrow D^t_j$ if S^t_i is a copy of D^t_j and $\Pr(S^t_i \Leftarrow S^t_j) = \Pr(S^t_i \Leftarrow D^t_j) = \frac{1}{2}$. The first term in (5), therefore, can be written as

$$\begin{aligned} \mathrm{Cov}(\alpha_{S^t_i}, G_{i'}) &= E[\mathrm{Cov}(\alpha_{S^t_i}, G_{i'} | Z)] \\ &\quad + \mathrm{Cov}[E(\alpha_{S^t_i} | Z), E(G_{i'} | Z)] \end{aligned} \tag{6}$$

(Kempthorne and Folks 1971).

Note that $G_{i'}$ is independent of Z because individual i' is not a direct descendent of individual i, so that $E(G_{i'} | Z) = E(G_{i'})$ and the second term in (6) is null.

For the first term in (6), if Z is $S^t_i \Leftarrow S^t_j$, then $\alpha_{S^t_i} = \alpha_{S^t_j}$; while if Z is $S^t_i \Leftarrow D^t_j$, then $\alpha_{S^t_i} = \alpha_{D^t_j}$. Thus, (6) can be written as

$$\begin{aligned} \mathrm{Cov}(\alpha_{S^t_i}, G_{i'}) &= \Pr(S^t_i \Leftarrow S^t_j)[\mathrm{Cov}(\alpha_{S^t_i}, G_{i'} | S^t_i \Leftarrow S^t_j)] \\ &\quad + \Pr(S^t_i \Leftarrow D^t_j)[\mathrm{Cov}(\alpha_{S^t_i}, G_{i'} | S^t_i \Leftarrow D^t_j)] \\ &= \tfrac{1}{2}[\mathrm{Cov}(\alpha_{S^t_j}, G_{i'}) + \mathrm{Cov}(\alpha_{D^t_j}, G_{i'})] \end{aligned} \tag{7}$$

Further, using (7), the covariance between the additive effect of the paternal allele, $\alpha_{S^t_i}$, and the genotypic value of individual i', $G_{i'}$, summed over n loci is

$$\begin{aligned} \sum_{t=1}^{n} \mathrm{Cov}(\alpha_{S^t_i}, G_{i'}) &= \frac{1}{2} \sum_{t=1}^{n} [\mathrm{Cov}(\alpha_{S^t_j}, G_{i'}) + \mathrm{Cov}(\alpha_{D^t_j}, G_{i'})] \\ &= \frac{1}{2} \mathrm{Cov}(G_j, G_{i'}) \end{aligned} \tag{8}$$

which is one half the covariance between genotypic values for individuals i' and j, the sire of i.

Similarly, the covariance between the additive effect of the maternal allele, $\alpha_{D^t_i}$, and the genotypic value of individual i', $G_{i'}$, summed over n loci is

$$\begin{aligned} \sum_{t=1}^{n} \mathrm{Cov}(\alpha_{D^t_i}, G_{i'}) &= \frac{1}{2} \sum_{t=1}^{n} [\mathrm{Cov}(\alpha_{S^t_k}, G_{i'}) + \mathrm{Cov}(\alpha_{D^t_k}, G_{i'})] \\ &= \frac{1}{2} \mathrm{Cov}(G_k, G_{i'}) \end{aligned} \tag{9}$$

which is one half the convariance between genotypic values for individuals i' and k, the dam of i.

Therefore, the covariance between genotypic values for individuals i and i' can be computed as

$$\mathrm{Cov}(G_i, G_{i'}) = \tfrac{1}{2}[\mathrm{Cov}(G_j, G_{i'}) + \mathrm{Cov}(G_k, G_{i'})] \tag{10}$$

which is the average covariance between genotypic values for individual i' and the parents j and k of indivudual i. This is the well-known result used in the tabular method to construct the matrix of additive covariances in purebred populations (Emik and Terrill 1949; Henderson 1976; Quaas et al. 1984), and has been used by Elzo (1983, 1990) to compute covariances in multibreed populations. These covariances, however, should not be computed using the usual tabular method because off-diagonal elements (covariances) are funcitons of diagonals (variances), and some of these variances are computed incorrectly, as shown below.

Variance

We will first present the theory for a multibreed population comprising three pure breeds (A, B, and C) and all crosses involving these breeds. The gametic disequilibrium for an arbitarary pair of alleles at loci t and t' (Δ) is the difference between the joint probability of the pair of alleles and the product of their marginal probabilities. The gametic disequlibrium of a cross (Δ^O) can be written in terms of the gametic disequilibria of the parental breed groups (Δ^S and Δ^D):

$$\begin{aligned}\Delta^O = {} & \tfrac{1}{2}(1-r)(\Delta^S + \Delta^D) \\ & + \tfrac{1}{4}(1-2r)(p_t^S - p_t^D)(p_{t'}^S - p_{t'}^D)\end{aligned} \tag{11}$$

(Nyquist 1990), where r is the probability of recombination between loci t and t', and, for example, p_t^S is the marginal probability of the allele at locus t in the sire breed group.

From (11), if parental breeds are in gametic equilibrium ($\Delta_{tt'}^S = \Delta_{tt'}^D = 0$) and loci are unlinked ($r = \tfrac{1}{2}$), then gametic equilibrium is maintained in the cross. Thus if the pure breeds are in gametic equilibrium and loci are unlinked, gametic equilibrium will be maintained in all the successive breed groups. Therefore, alleles at different loci are independent, and covariances between their additive effects are null, i.e., $\mathrm{Cov}(\alpha_{S_i^t}, \alpha_{S_i^{t'}}) = \mathrm{Cov}(\alpha_{S_i^t}, \alpha_{D_i^{t'}} = \mathrm{Cov}(\alpha_{D_i^t}, \alpha_{S_i^{t'}}) = \mathrm{Cov}(\alpha_{D_i^t}\, \alpha_{D_i^{t'}}) = 0$ for $t \neq t'$. The genotypic variance of individual i, from (1), is the sum of variances and covariances across the n unlinked loci and can be written as

$$\begin{aligned}\mathrm{Var}(G_i) = {} & \sum_{t=1}^{n} [\mathrm{Var}(\alpha_{S_i^t}) + \mathrm{Var}(\alpha_{D_i^t})] \\ & + 2\sum_{t=1}^{n} \mathrm{Cov}(\alpha_{S_i^t}, \alpha_{D_i^t})\end{aligned} \tag{12}$$

At locus t, let W be a random variable that is $S_i^t \in A$ if S_i^t is from pure breed A, is $S_i^t \in B$ if S_i^t is from pure breed B, or is $S_i^t \in C$ if S_i^t is from pure breed C. The variance of the additive effect of paternal allele S_i^t can be written as

$$\mathrm{Var}(\alpha_{S_i^t}) = E[\mathrm{Var}(\alpha_{S_i^t} \mid W)] + \mathrm{Var}[E(\alpha_{S_i^t} \mid W)] \tag{13}$$

(Kempthorne and Folks 1971). The first term in (13) can be written as

$$\begin{aligned}E[\mathrm{Var}(\alpha_{S_i^t} \mid W)] = {} & \Pr(S_i^t \in A)\mathrm{Var}(\alpha_{S_i^t} \mid S_i^t \in A) \\ & + \Pr(S_i^t \in B)\mathrm{Var}(\alpha_{S_i^t} \mid S_i^t \in B) \\ & + \Pr(S_i^t \in C)\mathrm{Var}(\alpha_{S_i^t} \mid S_i^t \in C) \\ = {} & f_A^S \underset{A}{\mathrm{Var}}(\alpha_{S_i^t}) + f_B^S \underset{B}{\mathrm{Var}}(\alpha_{S_i^t}) \\ & + f_C^S \underset{C}{\mathrm{Var}}(\alpha_{S_i^t})\end{aligned} \tag{14}$$

where, for example, f_A^S, the breed A composition for the sire, is the probability that paternal allele, S_i^t, is from pure breed A and $\underset{A}{\mathrm{Var}}$ is the variance using allelic frequencies of the breed A. Thus, in (14) $E[\mathrm{Var}(\alpha_{S_i^t} \mid W)]$ is the mean of the variances of the additive effects of paternal allele S_i^t, weighted by the proportion of each pure breed A, B, and C in the cross breed group.

The second term in (13), the variance of $E(\alpha_{S_i^t} \mid W)$, can be written as

$$\begin{aligned}\mathrm{Var}[E(\alpha_{S_i^t} \mid W)] = {} & E\{[E(\alpha_{S_i^t} \mid W)]^2\} \\ & - \{E[E(\alpha_{S_i^t} \mid W)]\}^2\end{aligned} \tag{15}$$

The expectation term $E[E(\alpha_{S_i^t} \mid W)]$ in (15) can be expressed as:

$$E[E(\alpha_{S_i^t} \mid W)] = f_A^S \varepsilon_t^A + f_B^S \varepsilon_t^B + f_C^S \varepsilon_t^C \tag{16}$$

where, for example,

$$\varepsilon_t^A = E(\alpha_{S_i^t} \mid S_i^t \in A) \tag{17}$$

is the conditional mean of $\alpha_{S_i^t}$ given that allele S_i^t is from pure breed A. Using (16) in (15) and rearranging yields

$$\begin{aligned}\mathrm{Var}[E(\alpha_{S_i^t} \mid W)] = {} & f_A^S(\varepsilon_t^A)^2 + f_B^S(\varepsilon_t^B)^2 + f_C^S(\varepsilon_t^C)^2 \\ & - (f_A^S \varepsilon_t^A + f_B^S \varepsilon_t^B + f_C^S \varepsilon_t^C)^2 \\ = {} & f_A^S f_B^S(\varepsilon_t^A - \varepsilon_t^B)^2 + f_A^S f_C^S(\varepsilon_t^A - \varepsilon_t^C)^2 \\ & + f_B^S f_C^S(\varepsilon_t^B - \varepsilon_t^C)^2\end{aligned} \tag{18}$$

Now, substituting (14) and (18) in (13), the variance of the additive effect of paternal allele S_i^t at locus t can be computed as

$$\begin{aligned}\mathrm{Var}(\alpha_{S_i^t}) = {} & f_A^S \underset{A}{\mathrm{Var}}(\alpha_{S_i^t}) + f_B^S \underset{B}{\mathrm{Var}}(\alpha_{S_i^t}) + f_C^S \underset{C}{\mathrm{Var}}(\alpha_{S_i^t}) \\ & + f_A^S f_B^S(\varepsilon_t^A - \varepsilon_t^B)^2 + f_A^S f_C^S(\varepsilon_t^A - \varepsilon_t^C)^2 \\ & + f_B^S f_C^S(\varepsilon_t^B - \varepsilon_t^C)^2\end{aligned} \tag{19}$$

Similarly, the variance of the additive effect of maternal allele D_i^t at locus t can be computed as

$$\mathrm{Var}(\alpha_{D_i^t}) = f_A^D \underset{A}{\mathrm{Var}}(\alpha_{D_i^t}) + f_B^D \underset{B}{\mathrm{Var}}(\alpha_{D_i^t}) + f_C^D\,{}^{\mathrm{Var}}_{C}(\alpha_{D_i^t})$$
$$+ f_A^D f_B^D(\varepsilon_t^A - \varepsilon_t^B)^2 + f_A^D f_C^D(\varepsilon_t^A - \varepsilon_t^C)^2$$
$$+ f_B^D f_C^D(\varepsilon_t^B - \varepsilon_t^C)^2 \qquad (20)$$

where, for example, f_A^D the breed A composition for the dam, is the probability that maternal allele, D_j^t, is from pure breed A. Now, the sum of (19) and (20) gives

$$\mathrm{Var}(\alpha_{S_i^t}) + \mathrm{Var}(\alpha_{D_i^t}) = f_A^S \underset{A}{\mathrm{Var}}(\alpha_{S_i^t}) + f_A^D \underset{A}{\mathrm{Var}}(\alpha_{D_i^t})$$
$$+ f_B^S \underset{B}{\mathrm{Var}}(\alpha_{S_i^t}) + f_B^D \underset{B}{\mathrm{Var}}(\alpha_{D_i^t})$$
$$+ f_C^S \underset{C}{\mathrm{Var}}(\alpha_{S_i^t}) + f_C^D \underset{C}{\mathrm{Var}}(\alpha_{D_i^t})$$
$$+ (f_A^S f_B^S + f_A^D f_B^D)(\varepsilon_t^A - \varepsilon_t^B)^2$$
$$+ (f_A^S f_C^S + f_A^D f_C^D)(\varepsilon_t^A - \varepsilon_t^C)^2$$
$$+ (f_B^S f_C^S + f_B^D f_C^D)(\varepsilon_t^B - \varepsilon_t^C)^2 \qquad (21)$$

In (21), note that $\underset{A}{\mathrm{Var}}(\alpha_{S_i^t})$ and $\underset{A}{\mathrm{Var}}(\alpha_{D_i^t})$ are identical in the pure breed A, so that $\underset{A}{\mathrm{Var}}(\alpha_{S_i^t}) = \underset{A}{\mathrm{Var}}(\alpha_{D_i^t}) = \sigma_{A_t}^2$. Similarly, $\underset{B}{\mathrm{Var}}(\alpha_{S_i^t}) = \underset{B}{\mathrm{Var}}(\alpha_{D_i^t}) = \sigma_{B_t}^2$, and $\underset{C}{\mathrm{Var}}(\alpha_{S_i^t}) = \underset{C}{\mathrm{Var}}(\alpha_{D_i^t}) = \sigma_{C_t}^2$. Then (21) can be written as

$$\mathrm{Var}(\alpha_{S_i^t}) + \mathrm{Var}(\alpha_{D_i^t}) = (f_A^S + f_A^D)\sigma_{A_t}^2 + (f_B^S + f_B^D)\sigma_{B_t}^2$$
$$+ (f_C^S + f_C^D)\sigma_{C_t}^2$$
$$+ (f_A^S f_B^S + f_A^D f_B^D)(\varepsilon_t^A - \varepsilon_t^B)^2$$
$$+ (f_A^S f_C^S + f_A^D f_C^D)(\varepsilon_t^A - \varepsilon_t^C)^2$$
$$+ (f_B^S f_C^S + f_B^D f_C^D)(\varepsilon_t^B - \varepsilon_t^C)^2 \qquad (22)$$

The breed composition for an offspring breed group is the average of the breed composition in the sire and the dam breed groups (Mather and Jinks 1971), e.g., $f_A^O = \frac{1}{2}(f_A^S + f_A^D)$, where f_A^O is the breed A composition for the offspring. Thus, (22) can be written as

$$\mathrm{Var}(\alpha_{S_i^t}) + \mathrm{Var}(\alpha_{D_i^t}) = 2f_A^O\sigma_{A_t}^2 + 2f_B^O\sigma_{B_t}^2 + 2f_C^O\sigma_{C_t}^2$$
$$+ (f_A^S f_B^S + f_A^D f_B^D)(\varepsilon_t^A - \varepsilon_t^B)^2$$
$$+ (f_A^S f_C^S + f_A^D f_C^D)(\varepsilon_t^A - \varepsilon_t^C)^2$$
$$+ (f_B^S f_C^S + f_B^D f_C^D)(\varepsilon_t^B - \varepsilon_t^C)^2 \qquad (23)$$

Now, for n unlinked loci, the additive variance of the pure breed A, σ_A^2, for example, is

$$\sigma_A^2 = 2\sum_{t=1}^{n} \sigma_{A_t}^2, \qquad (24)$$

and the genetic variance deriving from differences in allelic frequencies between the pure breeds A and B, δ_{AB}^2, for example, is

$$\delta_{AB}^2 = \sum_{t=1}^{n} (\varepsilon_t^A - \varepsilon_t^B)^2 \qquad (25)$$

Wright (1968) and Lande (1981) introduced the term segregation variance, σ_S^2, as the additional genetic variance segregating in the F_2 over that in the F_1. In (23), the term $(f_A^S f_B^S + f_A^D f_B^D)$, has value 0 if the offspring is an F_1 individual (i.e., $f_A^S = 1$, $f_B^S = 0$ and $f_A^D = 0$, $f_B^D = 1$) and has value $\frac{1}{2}$ if the offspring is an F_2 individual (i.e., $f_A^S = f_B^S = f_A^D = f_B^D = \frac{1}{2}$). Therefore, the additional genetic variance due to segregation from F_1 to F_2 is equal to $\frac{1}{2}\delta_{AB}^2$ as defined here (i.e., $\sigma_{S_{AB}}^2 = \frac{1}{2}\delta_{AB}^2$).

For n loci, the variance terms in (12), from (23) is

$$\sum_{t=1}^{n} [\mathrm{Var}(\alpha_{S_i^t}) + \mathrm{Var}(\alpha_{D_i^t})] = f_A^O\sigma_A^2 + f_B^O\sigma_B^2 + f_C^O\sigma_C^2$$
$$+ 2(f_A^S f_B^S + f_A^D f_B^D)\sigma_{S_{AB}}^2$$
$$+ 2(f_A^S f_C^S + f_A^D f_C^D)\sigma_{S_{AC}}^2$$
$$+ 2(f_B^S f_C^S + f_B^D f_C^D)\sigma_{S_{BC}}^2 \qquad (26)$$

To complete the computation of the variance for individual i using (12), we need the covariance between $\alpha_{S_i^t}$ and $\alpha_{D_i^t}$. Recall that allele S_i^t of individual i is either allele S_j^t or D_j^t from its sire j and that allele D_i^t of individual i is either allele S_k^t or D_k^t from its dam k (Fig. 1). The same reasoning used to obtain (8) from (6) can be used to compute the covariance between the additive effects of alleles S_i^t and D_i^t:

$$\mathrm{Cov}(\alpha_{S_i^t}, \alpha_{D_i^t}) = \tfrac{1}{4}[\mathrm{Cov}(\alpha_{S_j^t}^S, \alpha_{S_k^t}^D) + \mathrm{Cov}(\alpha_{S_j^t}^S, \alpha_{D_k^t}^D)$$
$$+ \mathrm{Cov}(\alpha_{D_j^t}^S, \alpha_{S_k^t}^D) + \mathrm{Cov}(\alpha_{D_j^t}^S, \alpha_{D_k^t}^D)]$$

The covariance between the additive effects of paternal and maternal alleles summed over n loci, is

$$\sum_{t=1}^{n} \mathrm{Cov}(\alpha_{S_i^t}, \alpha_{D_i^t}) = \frac{1}{4}\sum_{t=1}^{n} [\mathrm{Cov}(\alpha_{S_j^t}^S, \alpha_{S_k^t}^D) + \mathrm{Cov}(\alpha_{S_j^t}^S, \alpha_{D_k^t}^D)$$
$$+ \mathrm{Cov}(\alpha_{D_j^t}^S, \alpha_{S_k^t}^D) + \mathrm{Cov}(\alpha_{D_j^t}^D, \alpha_{D_k^t}^D)]$$
$$= \frac{1}{4}\mathrm{Cov}(G_j, G_k) \qquad (27)$$

Thus covariance between the additive effects of alleles S_i^t and D_i^t in individual i is the average covariance between the additive effects of alleles in indivuduals j and k, or one-fourth the covariance between genotypic values in individuals j and k, the parents of i. This covariance can be computed using (10), as shown previously. Now, substituting (26) and

(27) in (12), the additive variance over *n* loci of individual *i* is

$$\begin{aligned}\mathrm{Var}(G_i) = {} & f_A^O\sigma_A^2 + f_B^O\sigma_B^2 + f_C^O\sigma_C^2 + \tfrac{1}{2}\mathrm{Cov}(G_j, G_k) \\ & + 2(f_A^S f_B^S + f_A^D f_B^D)\sigma_{S_{AB}}^2 \\ & + 2(f_A^S f_C^S + f_A^D f_C^D)\sigma_{S_{AC}}^2 \\ & + 2(f_B^S f_C^S + f_B^D f_C^D)\sigma_{S_{BC}}^2 \qquad (28)\end{aligned}$$

Except for terms involving the segregation variances, (28) is equivalent to the expression given by Elzo (1983) for the additive variance of a crossbred individual. Thus, Elzo's formula is equal to (28) only if both parents are purebreds, as in the F_1 crossbred (e.g., $A \times B$ or $B \times A$). Therefore, (28) provides a general formula for additive variance for individuals in an arbitrary breed group from a multibreed population involving three pure breeds.

It can be shown algebraically that (14) generalizes to

$$E[\mathrm{Var}(\alpha_{S_i^t} | W)] = \sum_{p=1}^{P} f_p^S \underset{p}{\mathrm{Var}}(\alpha_{S_i^t}) \qquad (29)$$

and (18) generalizes to

$$\mathrm{Var}[E(\alpha_{S_i^t} | W)] = \sum_{p=1}^{P} \sum_{p'>p}^{P} f_p^S f_{p'}^S (\varepsilon_t^p - \varepsilon_t^{p'})^2 \qquad (30)$$

for a multibreed population comprising *P* pure breeds.

Table 1. The additive variances for various breed groups

Breed group	Variance
Pure breed	
A	σ_A^2
B	σ_B^2
C	σ_C^2
Two-way cross F_1	
$A \times B$	$\frac{1}{2}(\sigma_A^2 + \sigma_B^2)$
$A \times C$	$\frac{1}{2}(\sigma_A^2 + \sigma_C^2)$
$B \times C$	$\frac{1}{2}(\sigma_B^2 + \sigma_C^2)$
Backcross	
$AB \times A$	$\frac{3}{4}\sigma_A^2 + \frac{1}{4}\sigma_B^2 + \frac{1}{2}\sigma_{S_{AB}}^2$
$AB \times B$	$\frac{1}{4}\sigma_A^2 + \frac{3}{4}\sigma_B^2 + \frac{1}{2}\sigma_{S_{AB}}^2$
F_2	
$AB \times AB$	$\frac{1}{2}(\sigma_A^2 + \sigma_B^2) + \sigma_{S_{AB}}^2$
$AB \times AC$	$\frac{1}{2}(\sigma_A^2 + \sigma_C^2) + \sigma_{S_{AC}}^2$
$BC \times BC$	$\frac{1}{2}(\sigma_B^2 + \sigma_C^2) + \sigma_{S_{BC}}^2$
Three-way cross	
$A \times BC$	$\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_B^2 + \frac{1}{4}\sigma_C^2 + \frac{1}{2}\sigma_{S_{BC}}^2$
$A \times ABC$[a]	$\frac{3}{4}\sigma_A^2 + \frac{1}{8}\sigma_B^2 + \frac{1}{8}\sigma_C^2$ $+ \frac{1}{4}(\sigma_{S_{AB}}^2 + \sigma_{S_{AC}}^2) + \frac{1}{8}\sigma_{S_{BC}}^2$
General case	
	$\sum_{p=1}^{P} f_p^O \sigma_p^2$ $2\sum_{p=1}^{P}\sum_{p'>p}^{P}(f_p^S f_{p'}^S + f_p^D f_{p'}^D)\sigma_{S_{pp'}}^2$

[a] $ABC = A \times BC$

Thus, it follows that (28) generalizes to

$$\begin{aligned}\mathrm{Var}(G_i) = {} & \sum_{p=1}^{P} f_p^O \sigma_p^2 + \frac{1}{2}\mathrm{Cov}(G_j, G_k) \\ & + 2\sum_{p=1}^{p}\sum_{p'>p}^{P} (f_p^S f_{p'}^S + f_p^D f_{p'}^D)\sigma_{S_{pp'}}^2 \qquad (31)\end{aligned}$$

which is the additive variance of individual *i* in a multibreed population comprising an arbitrary number of pure breeds. In the case of noninbred populations, the additive variances of various breed groups are given in Table 1.

In summary, the additive covariance matrix for a multibreed population can be constructed by a tabular method using the theory presented here. The same rules as for purebred populations are used to obtain covariances, but rule (31) is used to obtain the additive variances. The only difference between this tabular method and that presented by Elzo (1983, 1990) is that his formula for the diagonal part did not include the term for the segregation variance.

Numerical example

The method to construct the additive covariance matrix for multibreed populations is illustrated using a simple pedigree of nine individuals (Fig. 2) in a two-breed (*A* and *B*) population. Individuals 1 and 3 are from pure breed *A* and individuals 2 and 4 are from pure breed *B*. A two-locus model is presented.

Consider a hypothetical trait determined by two loci (*U* and *V*) with two alleles at each locus (U_1 and U_2, V_1 and V_2). The genotypic values (g_{jklm}) and frequencies in breed groups *A*, *B*, F_1, and F_2 are given in Table 2. The F_2 was chosen to be the reference breed group. All individuals in breed *A* have the genotype $U_1U_1V_2V_2$, all those in breed *B* have the genotype $U_2U_2V_1V_1$, and all those in the F_1 have the genotype $U_1U_2V_2V_1$. Thus, additive variances of individuals 1

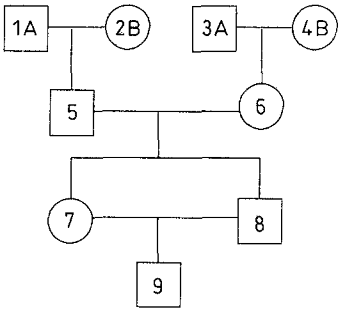

Fig. 2. Pedigree of nine individuals involving breeds *A* and *B*; *circles* represent females and *squares* represent males

through 6 are null. Covariances involving these individuals are also null. Any individual in the F_2 breed group, however, will have one of the 16 possible genotypes with equal probability (Table 2). Therefore, the additive variance of individuals 7 and 8 will not be null. The additive variance for individual 9, the offspring of 7 and 8, will also not be null. The covariances of individuals 7, 8, and 9 with individuals 1 through 6, however, are null.

To compute the variance for individual 7 using (31), we first compute the segregation variance, $\sigma^2_{S_{AB}}$, which requires computing ε_t^A and ε_t^B for each locus. At locus U, for example, the conditional mean of α_U, given that alleles U_1 and U_2 are from pure breed A, ε_U^A, is

$$\begin{aligned}\varepsilon_U^A &= E(\alpha_U \mid (U_1, U_2) \in A)\\ &= \alpha_{U_1} p^A_{U_1} + \alpha_{U_2} p^A_{U_2} \qquad (32)\end{aligned}$$

where, for example, $p^A_{U_1}$ is the frequency of allele U_1 in pure breed A. Recall that the F_2 was chosen as a reference breed group. Thus, the additive effect of allele U_1 can be written as

$$\alpha_{U_1} = E(G_i \mid U_1) - \mu \qquad (33)$$

where

$$E(G_i \mid U_1) = \sum_{k=1}^{2} \sum_{l=1}^{2} \sum_{m=1}^{2} g_{1klm} p^{F_1}_{U_k} p^{F_1}_{V_l} p^{F_1}_{V_m} \qquad (34)$$

is the conditional mean for G_i given U_1, where g_{1klm} is the value of an individual with genotype $U_1 U_k V_l V_m$, and, for example, $p^{F_1}_{U_k}$ is the frequency of allele k at locus U in the F_1, and $p^{F_1}_{V_l}$ is the frequency of allele l at locus V in the F_1. Also

$$\mu = \sum_{j=1}^{2} \sum_{k=1}^{2} \sum_{l=1}^{2} \sum_{m=1}^{2} g_{jklm} p^{F_1}_{U_j} p^{F_1}_{U_k} p^{F_1}_{V_l} p^{F_1}_{V_m} \qquad (35)$$

is the genotypic mean for the F_2. Thus, substituting the frequencies and genotypic values (Table 2) in (35), the genotypic mean for the F_2 is

$$\mu = \tfrac{1}{16}(0 + 5 + \cdots + 20) = 10$$

Table 2. Genotypic values and frequencies in breed groups A, B, F_1, and F_2

Genotype[a] $U_jU_kV_lV_m$	Genotypic value g_{jklm}	Genotypic frequency			
		A	B	F_1	F_2
$U_1U_1V_1V_1$	0	0	0	0	$\frac{1}{16}$
$U_1U_1V_1V_2$	5	0	0	0	$\frac{1}{16}$
$U_1U_1V_2V_1$	5	0	0	0	$\frac{1}{16}$
$U_1U_1V_2V_2$	10	1	0	0	$\frac{1}{16}$
$U_1U_2V_1V_1$	5	0	0	0	$\frac{1}{16}$
$U_1U_2V_1V_2$	10	0	0	0	$\frac{1}{16}$
$U_1U_2V_2V_1$	10	0	0	1	$\frac{1}{16}$
$U_1U_2V_2V_2$	15	0	0	0	$\frac{1}{16}$
$U_2U_1V_1V_1$	5	0	0	0	$\frac{1}{16}$
$U_2U_1V_1V_2$	10	0	0	0	$\frac{1}{16}$
$U_2U_1V_2V_1$	10	0	0	0	$\frac{1}{16}$
$U_2U_1V_2V_2$	15	0	0	0	$\frac{1}{16}$
$U_2U_2V_1V_1$	10	0	1	0	$\frac{1}{16}$
$U_2U_2V_1V_2$	15	0	0	0	$\frac{1}{16}$
$U_2U_2V_2V_1$	15	0	0	0	$\frac{1}{16}$
$U_2U_2V_2V_2$	20	0	0	0	$\frac{1}{16}$

[a] For each locus, the paternal allele is given first

The conditional mean for G_i given U_1 using (34) is

$$E(G_i \mid U_1) = \tfrac{1}{8}(0 + 5 + \cdots + 15) = 7.5$$

and the conditional mean for G_i given U_2 is

$$E(G_i \mid U_2) = \tfrac{1}{8}(5 + 10 + \cdots + 20) = 12.5$$

To compute ε_U^A using (32), we need to know $p^A_{U_1}$ and $p^A_{U_2}$. From Table 2, we see that $p^A_{U_1} = 1$ and $p^A_{U_2} = 0$. Thus, ε_U^A can be obtained using (33) in (32) as

$$\varepsilon_U^A = (7.5 - 10)(1) + (12.5 - 10)(0) = -2.5$$

Similarly, $\varepsilon_U^B = 2.5$, $\varepsilon_V^A = 2.5$, and $\varepsilon_V^B = -2.5$. Thus, the segregation variance can be computed by using the above values in (25), where $\sigma^2_{S_{AB}} = \frac{1}{2}\delta^2_{AB}$, as

$$\begin{aligned}\sigma^2_{S_{AB}} &= \tfrac{1}{2}[(\varepsilon_U^A - \varepsilon_U^B)^2 + (\varepsilon_V^A - \varepsilon_V^B)^2]\\ &= \tfrac{1}{2}[(-2.5 - 2.5)^2 + (2.5 + 2.5)^2] = 25\end{aligned}$$

In this example, recall that additive genetic variances for pure breeds are null and covariances involving individuals 1 through 6 are also null. Now, using (31), the additive genetic variance for individual 7 is

$$\begin{aligned}\mathrm{Var}(G_7) &= 2(f_A^5 f_B^5 + f_A^6 f_B^6)\sigma^2_{S_{AB}}\\ &= 2[(\tfrac{1}{2})(\tfrac{1}{2}) + (\tfrac{1}{2})(\tfrac{1}{2})]25 = 25\end{aligned}$$

Note that individuals 5 and 6 are parents of individual 7. If the additive variance for individual 7 were computed by Elzo's (1983) method, however, the result would be

$$\begin{aligned}\mathrm{Var}(G_7) &= \tfrac{1}{2}(\sigma_A^2 + \sigma_B^2) + \tfrac{1}{2}\mathrm{Cov}(G_5, G_6)\\ &= 0\end{aligned}$$

For this example, the additive variance for individual 7 can also be computed using genotypic frequencies as

$$\begin{aligned}\mathrm{Var}(G_7) &= E(G_7)^2 - [E(G_7)]^2\\ &= \frac{1}{16}[(0)^2 + (5)^2 + \cdots + (20)^2]\\ &\quad - \left[\frac{1}{16}(0 + 5 + \cdots + 20)\right]^2\\ &= 125 - 100 = 25\end{aligned}$$

which is equal to the variance computed using (31).

Individual 8 has the same additive genetic variance as individual 7 because they are full sibs. The covariance between individuals 7 and 8 is null because the covariances of 5 and 6 with 7 are null. For individual 9, using (31), the additive genetic variance is

$$\begin{aligned}\mathrm{Var}(G_9) &= 2(f^8_A f^8_B + f^7_A f^7_B)\sigma^2_{S_{AB}} \\ &= 2[(\tfrac{1}{2})(\tfrac{1}{2}) + (\tfrac{1}{2})(\tfrac{1}{2})]25 = 25\end{aligned}$$

because the covariance between the parents of individual 9 (7 and 8) is null and the additive variances of pure breeds are also null.

Now, the covariance between individuals 9 and 7 using (10) is

$$\begin{aligned}\mathrm{Cov}(G_9, G_7) &= \tfrac{1}{2}[\mathrm{Cov}(G_7, G_7) + \mathrm{Cov}(G_8, G_7)] \\ &= \tfrac{1}{2}(25 + 0) = 12.5\end{aligned}$$

Similarly, the covariance between individuals 9 and 8 is

$$\begin{aligned}\mathrm{Cov}(G_9, G_8) &= \tfrac{1}{2}[\mathrm{Cov}(G_7, G_8) + \mathrm{Cov}(G_8, G_8)] \\ &= \tfrac{1}{2}(0 + 25) = 12.5\end{aligned}$$

This numerical example shows that the variance computed using (31) is equal to that computed using genotypic frequencies. In practice, however, genotypic frequencies will not be known, and covariances will be computed as functions of additive genetic variances and segregation variances, which can be estimated by maximum likelihood.

The example also demonstrates that additive variances may be different for individuals with the same breed composition. In addition, it shows that the segregation variance in (31) is the difference between additive variances in F_2 and F_1 breed groups under an additive model, i.e., $\sigma^2_{S_{AB}} = \sigma^2_{F_2} - \sigma^2_{F_1} = 25 - 0 = 25$.

Inverse of additive covariance matrix

The theory for covariance between relatives in multibreed populations developed in the previous sections can be used to obtain genetic evaluations by best linear unbiased prediction (BLUP; Henderson 1973). The usual mixed model formulation for BLUP requires the inverse of the genotypic covariance matrix. We have shown that under additive inheritance and for a model with multiple unlinked loci, the covariance matrix ($\mathbf{G}_a$) for a multibreed population can be constructed by a tabular method using (10) and (31). Thus, $\mathbf{G}_a$ can be inverted efficiently using the approach of Henderson (1976) and Quaas (1988).

Following Quaas (1988), the inverse of $\mathbf{G}_a$ can be obtained as

$$\mathbf{G}_a^{-1} = (\mathbf{I} - \mathbf{P})'\mathbf{G}_\varepsilon^{-1}(\mathbf{I} - \mathbf{P}) \tag{36}$$

where $\mathbf{I}$ is an identity matrix of order n, $\mathbf{P}$ is a matrix relating progeny to parents, $\mathbf{G}_\varepsilon$ is the covariance matrix of Mendelian sampling residuals, and n is the number of individuals in the pedigree. It can be shown that $\mathbf{G}_\varepsilon$ is diagonal (e.g., Fernando and Grossman 1989), and element i on the diagonal (d_i) can be computed as

$$\begin{aligned}d_i = \mathrm{Var}(G_i) &- \tfrac{1}{4}[\mathrm{Var}(G_j) + \mathrm{Var}(G_k)] \\ &- \tfrac{1}{2}\mathrm{Cov}(G_j, G_k)\end{aligned} \tag{37}$$

where j and k are parents of i. For multibreed populations, the variance for each individual and the covariance between parents can be computed using (31) and (10). Now, let $\mathbf{Q} = (\mathbf{I} - \mathbf{P})' = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n)$, where $\mathbf{q}_i$ is a column vector of order n from $\mathbf{Q}$ corresponding to individual i. Because $\mathbf{G}_\varepsilon$ is diagonal, (36) can be written as

$$\mathbf{G}_a^{-1} = \mathbf{Q}\mathbf{G}_\varepsilon^{-1}\mathbf{Q}' = \sum_{i=1}^{n} \mathbf{q}_i \mathbf{d}_i^{-1} \mathbf{q}_i' \tag{38}$$

As shown by Quaas (1988), (38) leads to an efficient algorithm to obtian $\mathbf{G}_a^{-1}$.

Discussion

In the analysis of data from multibreed populations, the theory for additive genetic covariances developed by Elzo (1983, 1990) has been used to model the additive component of the genotypic value. It is shown here that Elzo's (1983) method to compute the additive variance for a crossbred individual does not give the correct results in general because it does not account for segregation variances. The segregation variance is due to the differences in allelic frequencies between the pure breeds, and is equal to the difference in additive variances between F_2 and F_1 breed groups (Lande 1981). The segregation of alleles will contribute to the variance of an individual only if one or both of its parents are crossbreds.

Provided that the variance of a crossbred individual is computed as described here, the covariance between relatives can be computed using formulae for purebred populations. For traits where nonadditive inheritance can be ignored, the inverse of the genotypic covariance matrix given here can be used in Henderson's (1973) mixed model equations to obtian genetic evaluations in multibreed populations by best linear unbiased prediction. Further, assuming a normal distribution, the theory presented here can be used to estimate by maximum likelihood the additive variance component for each pure breed and the segregation variance for each pair of pure breeds. The segregation variance can also be estimated simply as the difference between variances for the F_1 and F_2 breed groups as suggested by Lande (1981).

For traits where nonadditive inheritance can not be ignored, the procedure currently used to analyze multibreed data can be improved using the theory presented here to compute additive covariances together with a suitable approximation for nonadditive covariances. The approach presented here can be extended to accommodate multiple traits and maternal effects in multibreed populations.

References

Arnold JW, Bertrand, JK, Benyshek LL (1992) Animal model for genetic evaluation of multibreed data. J Anim Sci 70: 3322–3332

Chang HL, Fernando RL, Grossman M (1991) On the principle underlying the tabular method to compute coancestry. Theor Appl Genet 81:233–238

Elzo MA (1983) Multibreed sire evaluation within and across countries. PhD thesis, University of California, Davis

Elzo MA (1990) Recursive procedures to compute the inverse of the multiple trait additive genetic covariance matrix in inbred and noninbred multibreed populations. J Anim Sci 68: 1215–1228

Elzo MA, Bradford GE (1985) Multibreed sire evaluation procedures across countries. J Anim Sci 60:953–963

Elzo MA, Famula TR (1985) Multibreed sire evaluation procedures within a country. J Anim Sci 60:942–952

Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51–55

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 21: 467–477

Hakim AF, Van Vleck LK, Cundiff LV, Koch RM, Crouse JD, Boldman KG (1990) Estimates of genetic and phenotypic covariance among carcass and sensory characteristics of crossbred cattle with multiple traits animal model. J Anim Sci 68 (suppl 1):238–239

Henderson CR (1973) Sire evaluation and genetic trends. In: Anim Breed Genet Symp in Honor of Dr. J. L. Lush, Champaign, Illinois Am Soc Anim Sci and Am Dairy Sci Assoc, pp 10–41

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83

Kempthorne O (1954) The correlation between relatives in a random mating population. Proc Roy Soc London B143: 103–113

Kempthorne O, Folks L (1971) Probability, statistics, and data analysis. Iowa State Univ Press

Lande R (1981) The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99:541–553

Lo LL, McLaren DG, McKeith FK, Fernando RL, Novakofski J (1992) Genetic analyses of growth, real-time ultrasound, carcass, and pork quality traits in Duroc and Landrace pigs. II. Heritabilities and correlations. J Anim Sci 70:2387–2396

Mather K, Jinks JL (1971) Biometrical genetics, 2nd edn. Cornell Univ Press, Ithaca, New York

Nyquist WE (1990) Statistical Genetics. Unpublished class notes. Agronomy Department, Purdue University, West Lafayette, Indiana

Quaas RL (1988) Additive genetic model with groups and relationships. J Dairy Sci 71:1338–1345

Quaas RL, Anderson RD, Gilmour AR (1984) BLUP school handbook; use of mixed models for prediction and estimation of (co)variance components. Animal breeding and Genetics Unit, University of New England, NSW 2351, Australia

Swan AA, Kinghorn BP (1992) Evaluation and exploitation of crossbreeding in dairy cattle. J Dairy Sci 75:624–639

Van Vleck LD, Hakim AF, Cundiff LV, Koch RM, Crouse JD, Boldman KG (1992) Estimated breeding values for meat characteristics of crossbred cattle with an animal model. J Anim Sci 70:363–371

Wei M, Van der Steen HAM, Van der Werf JHJ (1991a) Relationship between purebred and crossbred parameters. I. Variances and covariances under the one-locus model. J Anim Breed Genet 108:253–261

Wei M, Van der Steen HAM, Van der Werf JHJ (1991b) Relationship between purebred and crossbred parameters. II. Genetic correlation between purebred and crossbred performance under the model with two loci. J Anim Breed Genet 108:262–269

Wright S (1968) Evolution and the genetics of population. vol 1. Genetics and biometrical foundations. Univ Chicago Press