Theor Appl Genet (1995) 90:49–62

L. L. Lo · R. L. Fernando · R. J. C. Cantet
M. Grossman

Theory for modelling means and covariances in a two-breed population with dominance inheritance

Received: 15 July 1993 / Accepted: 29 March 1994

Abstract This paper presents theory and methods to compute genotypic means and covariances in a two-breed population under dominance inheritance, assuming multiple unlinked loci. It is shown that the genotypic mean is a linear function of five location parameters and that the genotypic covariance between relatives is a linear function of 25 dispersion parameters. Recursive procedures are given to compute the necessary identity coefficients. In the absence of inbreeding, the number of parameters for the mean is reduced from five to three and the number for the covariance is reduced from 25 to 12. In a two-breed population, for traits exhibiting dominance, the theory presented here can be used to obtain genetic evaluations by best linear unbiased prediction and to estimate genetic parameters by maximum likelihood.

Key word Covariance between relatives · Dominance model · Recursion · Two-breed population

Introduction

Crossbreeding is used widely in animal production. One of the main purposes of crossbreeding is to take advantage of the heterosis that is often observed in crossbreds. The primary genetic mechanism for heterosis is directional dominance of favorable alleles at many loci (Falconer 1989). In the absence of inbreeding, theory is available to model the mean in crossbred populations (Dickerson 1973; Hill 1982; Eisen et al. 1983; Wei and Van der Werf 1993). However, due to lack of theory, genotypic variances and covariances have not been modelled exactly in crossbred populations under dominance inheritance (VanRaden 1992).

Alleles that are identical by descent (IBD) cause genotypic values between relatives to be correlated (Kempthorne 1954). Gillois (1964) and Harris (1964) defined 15 "identity modes" concerning the IBD states of four alleles in two individuals. Based on these identity modes, a set of five genetic parameters and its corresponding set of coefficients of identity were derived to compute genotypic covariance between purebred relatives under dominance inheritance (Gillois 1964; Harris 1964). Genotypic covariances between purebred relatives is a function of these five genetic parameters and their corresponding identity coefficients.

Harris (1964) gave recursive formulae to compute the identity coefficients. Without computing identity coefficients, Smith and Mäki-Tanila (1990) developed a recursive procedure to compute genotypic covariance directly for purebred populations. These papers have been discussed and extended by De Boer and Hoeschele (1993).

The objective of this paper is to present theory for modelling genotypic means and covariances in a population composed of two pure breeds, A and B, and any crosses involving these two breeds. Theory is derived under a model with dominance inheritance and multiple unlinked loci.

Supported in part by the Illinois Agricultural Experiment Station, Hatch Projects 35-0345 (R.L.F.) and 35-0367 (M.G.). A computer program implementing the methods described here is available upon request to R.L.F.

Communicated by E. J. Eisen

L. L. Lo[1] · R. L. Fernando (✉) · R. J. C. Cantet[2] · M. Grossman
Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1207 West Gregory Drive, Urbana, IL 61801, USA

Present addresses:
[1] Department of Animal Breeding, Animal Industry Research Institute, Taiwan Sugar Corporation, Chunan, Miao., Taiwan 35014, the Republic of China
[2] Departamento de Zootecnia, Facultad de Agronomia, Universidad de Buenos Aires, Capital Federal, Argentina

Theory

Consider a genotypic model with n unlinked loci in a random-mating two-breed population. It is assumed that the two pure breeds (A and B) are in gametic equilibrium. At locus t, let S_i^t be the random allele that individual i inherited from its sire j, and let D_i^t be the random allele that i inherited from its dam k. The paternal and maternal alleles of sire j at locus t are S_j^t and D_j^t; those of dam k are S_k^t and D_k^t. For convenience, consider the F_1 as a reference breed group (AB). The genotypic value of an individual in any breed group will be modelled using effects defined for the F_1. This model will be used to develop theory to compute genotypic means and covariances for a two-breed population in the presence of inbreeding.

Genotypic model

The genotypic value of individual i, G_i, is modelled as

$$G_i = \mu + \sum_{t=1}^{n} (\alpha_{S_i^t} + \alpha_{D_i^t} + \delta_{S_i^t D_i^t})$$

$$= \mu + \sum_{t=1}^{n} G_{S_i^t D_i^t} \tag{1}$$

where:

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

is the genotypic mean;

$$\alpha_{S_i^t} = \mathrm{E}(G_i | S_i^t) - \mu \tag{3}$$

is the additive effect of paternal allele S_i^t;

$$\alpha_{D_i^t} = \mathrm{E}(G_i | D_i^t) - \mu \tag{4}$$

is the additive effect of maternal allel D_i^t; and

$$\delta_{S_i^t D_i^t} = \mathrm{E}(G_i | S_i^t D_i^t) - \alpha_{S_i^t} - \alpha_{D_i^t} - \mu \tag{5}$$

is the dominance effect for genotype $S_i^t D_i^t$. The expectations in (2) through (5) are taken using allelic frequencies in the reference breed group. Thus, in the F_1, additive and dominance effects will have null expectations, and therefore $\underset{AB}{\mathrm{E}}(G_{S_i^t D_i^t}) = 0$. Similarly, the genotypic value for another individual i' can be modelled as

$$G_{i'} = \mu + \sum_{t=1}^{n} G_{S_{i'}^t D_{i'}^t} \tag{6}$$

where μ and $G_{S_{i'}^t D_{i'}^t}$ are defined as for i.

Means

The genotypic value for an individual i in any breed group, G_i, can be modelled as (1). The alleles S_i^t and D_i^t at locus t each comes from either breed A or B. Further, these alleles may be identical by descent (IBD) or not. Specifying the breed origin and identity states for alleles S_i^t and D_i^t results in a set of five two-breed identity modes for a single individual (TIMS):

N_1: $S_i^t \not\equiv D_i^t, S_i^t \in A, D_i^t \in A$

N_2: $S_i^t \not\equiv D_i^t, S_i^t \in A, D_i^t \in B$

N_2: $S_i^t \not\equiv D_i^t, S_i^t \in B, D_i^t \in A$

N_3: $S_i^t \not\equiv D_i^t, S_i^t \in B, D_i^t \in B$

N_4: $S_i^t \equiv D_i^t, S_i^t \in A, D_i^t \in A$

N_5: $S_i^t \equiv D_i^t, S_i^t \in B, D_i^t \in B$

where the symbol $\equiv$ denotes that alleles are IBD, $\not\equiv$ denotes that alleles are not IBD, and $\in$ is used to denote the breed of origin of alleles (Fig. 1). For example, N_1 is defined as the mode where alleles S_i^t and D_i^t are not IBD and are from breed A.

The expected value of G_i from (1) can be written as

$$\mathrm{E}(G_i) = \mu + \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})$$

$$= \mu + \sum_{t=1}^{n} \left[\sum_{r=1}^{5} E(G_{S_i^t D_i^t} | n_r) p_i^r \right] \tag{7}$$

where n_r is the event that the identity state and breed origin for the set of two alleles (S_i^t and D_i^t) belongs to identity mode N_r, and p_i^r is the probability of n_r. The p_i^rs will be referred to as TIMS coefficients. Rearranging (7), the genetic mean for individual i can be written as

$$\mathrm{E}(G_i) = \mu + \sum_{r=1}^{5} p_i^r \left[\sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} | n_r) \right]$$

$$= \mu + \sum_{r=1}^{5} p_i^r \mu_r$$

$$= \sum_{r=1}^{5} (\mu + \mu_r) p_i^r$$

$$= \sum_{r=1}^{5} \mu_r^* p_i^r \tag{8}$$

where $\mu_r^* = \mu + \mu_r = \mathrm{E}(G_i | n_r)$ for $r = 1, \ldots, 5$ are the location parameters for the two-breed population. Let A^* be a population where all individuals are homozygous at each locus with allelic frequencies of pure breed A. Similarly, B^* is defined to be a homozygous population with allelic frequencies of pure breed B. Now, μ_1^* is the genetic mean for pure breed A, μ_2^* is the genetic mean for the AB breed group, μ_3^* is the genetic mean for pure breed B, μ_4^* is the genetic mean for homozygous breed A^*, and μ_5^* is the genetic mean for homozygous breed B^*. Then, the genotypic mean is a linear function of five TIMS coefficients and the corresponding location parameters.

For inbred populations, these TIMS coefficients can be computed using a recursive procedure as shown later. For noninbred populations, n_1, n_2, and n_3 are functions of the breed composition of the parents as shown below. Note that identity modes N_4 and N_5 do not occur in noninbred populations.

Fig. 1 Identity modes N. Identical alleles are connected

Mode	Identity state		Breed origin	
N_r	S_i^t	D_i^t	S_i^t	D_i^t
N_1	○	○	A	A
N_2	○	○	A	B
N_2	○	○	B	A
N_3	○	○	B	B
N_4	○——	——○	A	A
N_5	○——	——○	B	B

The probability that alleles S_i^t and D_i^t belong to mode N_1 is

$$p_i^1 = f_A^S f_A^D \tag{9}$$

where f_A^S is the breed A composition for the sire and f_A^D is the breed A composition for the dam. Similarly, the probability that alleles S_i^t and D_i^t belong to mode N_2 is

$$p_i^2 = f_A^S f_B^D + f_B^S f_A^D \tag{10}$$

where f_B^D is the breed B composition for the dam and f_B^S is the breed B composition for the sire. The probability that alleles S_i^t and D_i^t belong to mode N_3 is

$$p_t^3 = f_B^S f_B^D. \tag{11}$$

Covariances

The genotypic covariance between individuals i and i' is

$$\mathrm{Cov}(G_i, G_{i'}) = \mathrm{E}(G_i G_{i'}) - \mathrm{E}(G_i)\mathrm{E}(G_{i'}). \tag{12}$$

The first term in (12), from (1) and (6), can be expressed as

$$\begin{aligned}\mathrm{E}(G_i G_{i'}) &= \mathrm{E}\left[\left(\mu + \sum_{t=1}^{n} \mathrm{G}_{S_i^t D_i^t}\right)\left(\mu + \sum_{t=1}^{n} G_{S_{i'}^t D_{i'}^t}\right)\right] \\ &= \mu^2 + \mu \sum_{t=1}^{n} [\mathrm{E}(G_{S_i^t D_i^t}) + \mathrm{E}(G_{S_{i'}^t D_{i'}^t})] + \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t}) \\ &\quad + \sum_{t=1}^{n} \sum_{\substack{t'=1 \\ t \neq t'}}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^{t'} D_{i'}^{t'}}).\end{aligned} \tag{13}$$

For the second term of (12),

$$\begin{aligned}\mathrm{E}(G_i)\mathrm{E}(G_{i'}) &= \left[\mathrm{E}\left(\mu + \sum_{t=1}^{n} \mathrm{G}_{S_i^t D_i^t}\right)\right]\left[\mathrm{E}\left(\mu + \sum_{t=1}^{n} G_{S_{i'}^t D_{i'}^t}\right)\right] \\ &= \left[\mu + \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\right]\left[\mu + \sum_{t=1}^{n} \mathrm{E}(G_{S_{i'}^t D_{i'}^t})\right] \\ &= \mu^2 + \mu \sum_{t=1}^{n} [\mathrm{E}(G_{S_i^t D_i^t}) + \mathrm{E}(G_{S_{i'}^t D_{i'}^t})] + \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}) \\ &\quad + \sum_{t=1}^{n} \sum_{\substack{t'=1 \\ t \neq t'}}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^{t'} D_{i'}^{t'}}).\end{aligned} \tag{14}$$

Note that because breeds A and B are assumed to be in gametic equilibrium and loci are unlinked, alleles at locus t in a crossbred are distributed independently of those alleles at locus t', i.e., $\Pr(S_i^t, S_i^{t'}) = \Pr(S_i^t)\Pr(S_i^{t'})$ (Lo et al. 1993). Thus, $\mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^{t'} D_{i'}^{t'}}) = \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^{t'} D_{i'}^{t'}})$, and (12), or the difference between (13) and (14), becomes

$$\mathrm{Cov}(G_i, G_{i'}) = \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t}) - \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}). \tag{15}$$

At locus t, individuals i and i' each has two alleles, S_i^t and D_i^t, and $S_{i'}^t$ and $D_{i'}^t$. Each allele comes from either breed A or B. To compute (15), we must know the identity state for each pair of alleles, i.e., whether or not the alleles are identical by descent (IBD), and the breed origin for each allele.

It is convenient to start by showing how the first term of (15) is computed when all four alleles are from the same population. For this situation, Gillois (1964) and Harris (1964) have defined 15 possible cases of identity by descent, called "identity modes", between alleles of i and i':

I_1: $S_i^t \equiv S_{i'}^t \not\equiv D_i^t \not\equiv D_{i'}^t, S_i^t \equiv S_{i'}^t \not\equiv D_{i'}^t$

I_2: $D_i^t \equiv D_{i'}^t \not\equiv S_i^t \not\equiv S_{i'}^t, D_i^t \equiv D_{i'}^t \not\equiv S_{i'}^t$

I_3: $S_i^t \equiv D_{i'}^t \not\equiv D_i^t \not\equiv S_{i'}^t, S_i^t \equiv D_{i'}^t \not\equiv S_{i'}^t$

I_4: $D_i^t \equiv S_{i'}^t \not\equiv S_i^t \not\equiv D_{i'}^t, D_i^t \equiv S_{i'}^t \not\equiv D_{i'}^t$

I_5: $S_i^t \equiv S_{i'}^t \not\equiv D_i^t \equiv D_{i'}^t$

I_6: $S_i^t \equiv D_{i'}^t \not\equiv D_i^t \equiv S_{i'}^t$

I_7: $S_i^t \equiv D_i^t \equiv S_{i'}^t \not\equiv D_{i'}^t$

I_8: $S_i^t \equiv D_i^t \equiv D_{i'}^t \not\equiv S_{i'}^t$

I_9: $S_i^t \equiv S_{i'}^t \equiv D_{i'}^t \not\equiv D_i^t$

I_{10}: $D_i^t \equiv S_{i'}^t \equiv D_{i'}^t \not\equiv S_i^t$

I_{11}: $S_i^t \equiv D_i^t \equiv S_{i'}^t \equiv D_{i'}^t$

I_{12}: $S_i^t \not\equiv D_i^t \not\equiv S_{i'}^t \not\equiv D_{i'}^t, S_i^t \not\equiv S_{i'}^t, D_i^t \not\equiv D_{i'}^t$

I_{13}: $\mathrm{S}_i^t \equiv \mathrm{D}_i^t \not\equiv S_{i'}^t \equiv D_{i'}^t$

I_{14}: $S_i^t \equiv D_i^t \not\equiv S_{i'}^t \not\equiv D_{i'}^t, S_i^t \equiv D_i^t \not\equiv D_{i'}^t$

I_{15}: $S_{i'}^t \equiv D_{i'}^t \not\equiv S_i^t \not\equiv D_i^t, S_{i'}^t \equiv D_{i'}^t \not\equiv D_i^t$

where, for example, I_1 is the mode where alleles S_i^t and $S_{i'}^t$ are IBD but not IBD to D_i^t and $D_{i'}^t$, and D_i^t and $D_{i'}^t$ are not IBD. This set of 15 I identity modes is represented graphically in Fig. 2 (Jacquard 1974). Gillois (1964) and Harris (1964) used these I identity modes to compute the genotypic covariance between relatives in purebred populations.

By ignoring the paternal or material origin of alleles, Harris (1964) and Jacquard (1974) grouped the 15 I identity modes into a set of nine J identity modes (Fig. 3):

J_1: I_1, I_2, I_3, I_4

J_2: I_5, I_6

J_3: I_7, I_8

J_4: I_9, I_{10}

J_5: I_{11}

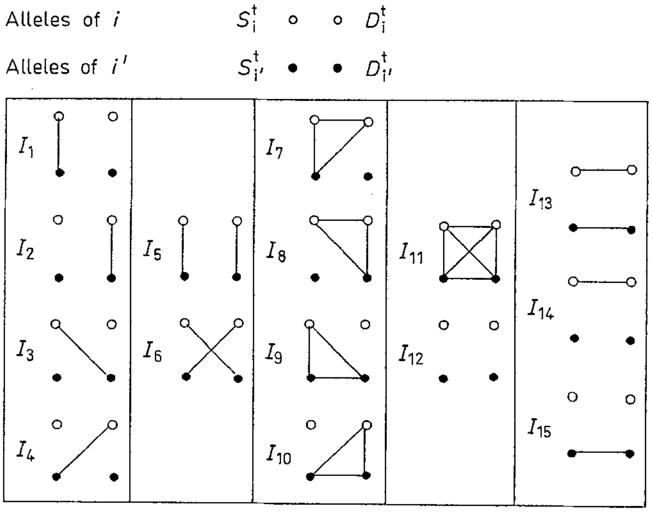
J_6: I_{12}

J_7: I_{13}

J_8: I_{14}

J_9: I_{15}.

Fig. 2 Identity modes I. Identical alleles are connected



Source : Jacquard (1974)

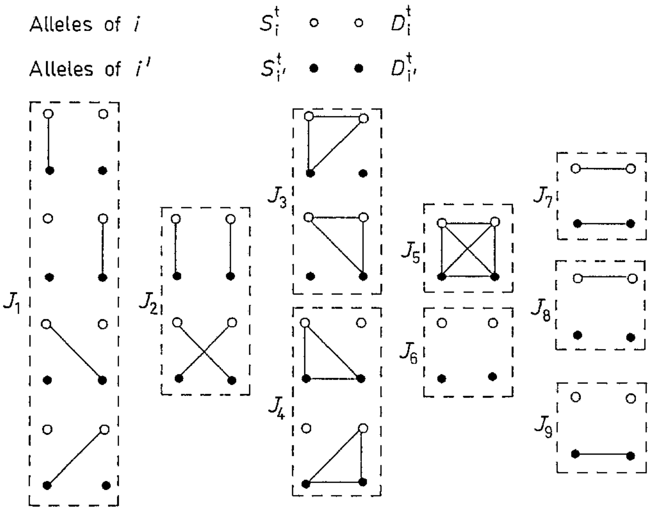


Fig. 3 Grouped identity modes J. Identical alleles are connected

At locus t, the first term of (15), $\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})$, can be written as

$$\begin{aligned}&\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})\\&=\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_1]\,\mathrm{Pr}[(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_1]+\cdots\\&\quad+\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_9]\,\mathrm{Pr}[(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_9].\end{aligned}$$

From Fig. 3, however, it can be seen that

$$\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t\in J_3]=\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_4]$$

and

$$\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_8]=\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in J_9].$$

Thus, for the purpose of computing $\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})$, it is not necessary to distinguish between J_3 and J_4, and between J_8 and J_9. Therefore, the nine J identity modes can be further reduced to seven K identity modes (Fig. 4):

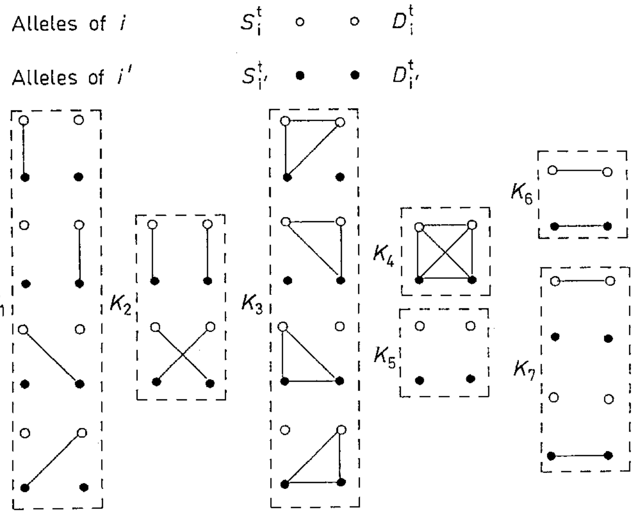
$K_1{:}J_1$ $K_3{:}J_3,J_4$

$K_2{:}J_2$ $K_4{:}J_5$

$K_5{:}J_6$ $K_7{:}J_8,J_9$

$K_6{:}J_7$

Fig. 4 Grouped identity modes K. Identical alleles are connected



Now, we can use these seven K modes to define two-breed identity modes for a pair of individuals (TIMPs) by specifying the breed origin for the alleles in addition to their identity states.

Consider K_1, which consists of four I identity modes, with two alleles IBD and two not IBD. The two IBD alleles are denoted C_{11}, and the two not IBD are denoted F_{11} and F_{12}. Specifying the breed origin for these alleles results in a set of eight L breed-specific identity modes for K_1:

L_1: $C_{11}\in A, F_{11}\in A, F_{12}\in A$

L_2: $C_{11}\in A, F_{11}\in A, F_{12}\in B$

L_3: $C_{11}\in A, F_{11}\in B, F_{12}\in A$

L_4: $C_{11}\in A, F_{11}\in B, F_{12}\in B$

L_5: $C_{11}\in B, F_{11}\in A, F_{12}\in A$

L_6: $C_{11}\in B, F_{11}\in A, F_{12}\in B$

L_7: $C_{11}\in B, F_{11}\in B, F_{12}\in A$

L_8: $C_{11}\in B, F_{11}\in B, F_{12}\in B$

which is summarized in Table 1. Note that by specifying the breed origin for each allele

$$\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in L_2]=\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in L_3]$$

and

$$\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in L_6]=\mathrm{E}[G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|(S_i^t,D_i^t,S_{i'}^t,D_{i'}^t)\in L_7].$$

To compute $\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})$, therefore, it is not necessary to distinguish between L_2 and L_3, and between L_6 and L_7. Hence, the set of eight L breed-specific identity modes, L_1 to L_8, can be reduced to a set of six TIMPs for K_1:

$M_1{:}L_1$

$M_2{:}L_2,L_3$

$M_3{:}L_4$

$M_4{:}L_5$

$M_5{:}L_6,L_7$

$M_6{:}L_8$

which is also summarized in Table 1. Similarly, specifying the breed origin for alleles IBD and alleles not IBD for K_2 through K_7 results in TIMPs M_7 through M_{30} (Table 2-7).

Table 1 Identity modes L and M resulting from K_1. The two IBD alleles are denoted C_{11}, and the two not IBD are denoted F_{11} and F_{12}

M	L	C_{11}	F_{11}	F_{12}
M_1	L_1	A	A	A
M_2	L_2	A	A	B
M_2	L_3	A	B	A
M_3	L_4	A	B	B
M_4	L_5	B	A	A
M_5	L_6	B	A	B
M_5	L_7	B	B	A
M_6	L_8	B	B	B

Table 2 Identity modes L and M resulting from K_2. The two pairs of IBD alleles are denoted C_{21} and C_{22}

M	L	C_{21}	C_{22}
M_7	L_9	A	A
M_8	L_{10}	A	B
M_8	L_{11}	B	A
M_9	L_{12}	B	B

Table 3 Identity modes L and M resulting from K_3. The three IBD alleles are denoted C_{31}, and the one not IBD is denoted F_{31}

M	L	C_{31}	F_{13}
M_{10}	L_{13}	A	A
M_{11}	L_{14}	A	B
M_{12}	L_{15}	B	A
M_{13}	L_{16}	B	B

Table 4 Identity modes L and M resulting from K_4. The four IBD alleles are denoted C_{41}

M	L	C_{41}
M_{14}	L_{17}	A
M_{15}	L_{18}	B

Table 5 Identity modes L and M resulting from K_5. In K_5, there are no alleles IBD. The two alleles from i are denoted F_{51} and F_{52} and the two alleles from i' are denoted F_{53} and F_{54}.

M	L	F_{51}	F_{52}	F_{53}	F_{54}
M_{16}	L_{19}	A	A	A	A
M_{17}	L_{20}	A	A	A	B
M_{17}	L_{21}	A	A	B	A
M_{18}	L_{22}	A	A	B	B
M_{17}	L_{23}	A	B	A	A
M_{19}	L_{24}	A	B	A	B
M_{19}	L_{25}	A	B	B	A
M_{20}	L_{26}	A	B	B	B
M_{17}	L_{27}	B	A	A	A
M_{19}	L_{28}	B	A	A	B
M_{19}	L_{29}	B	A	B	A
M_{20}	L_{30}	B	A	B	B
M_{18}	L_{31}	B	B	A	A
M_{20}	L_{32}	B	B	A	B
M_{20}	L_{33}	B	B	B	A
M_{21}	L_{34}	B	B	B	B

Now, for locus t, the first expectation term in (15) can be written in terms of M identity modes as

$$\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}) = \sum_{q=1}^{30} \mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|m_q)\mathrm{Pr}(m_q)$$

$$= \sum_{q=1}^{30} c_{ii'}^q \theta_q^t \tag{16}$$

Table 6 Identity modes L and M resulting from K_6. The two pairs of IBD alleles are denoted C_{61} and C_{62}

M	L	C_{61}	C_{62}
M_{22}	L_{35}	A	A
M_{23}	L_{36}	A	B
M_{23}	L_{37}	B	A
M_{24}	L_{38}	B	B

Table 7 Identity modes L and M resulting from K_7. The two IBD alleles are denoted C_{71}, and the two not IBD are denoted F_{71} and F_{72}

M	L	C_{71}	F_{71}	F_{72}
M_{25}	L_{39}	A	A	A
M_{26}	L_{40}	A	A	B
M_{26}	L_{41}	A	B	A
M_{27}	L_{42}	A	B	B
M_{28}	L_{43}	B	A	A
M_{29}	L_{44}	B	A	B
M_{29}	L_{45}	B	B	A
M_{30}	L_{46}	B	B	B

where m_q is the event that the breed origin-identity state (BO-IS) for the set of four alleles $(S_i^t, D_i^t, S_{i'}^t, D_{i'}^t)$ belongs to identity mode M_q, $(S_i^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q$, $c_{ii'}^q = \mathrm{Pr}(m_q)$, and

$$\theta_q^t = \mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|m_q).$$

The *c*s are called “TIMP coefficients”.

For n unliked loci, the first term in (15) can be written as

$$\sum_{t=1}^{n} \mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}) = \sum_{t=1}^{n}\sum_{q=1}^{30} c_{ii'}^q \theta_q^t$$

$$= \sum_{q=1}^{30} c_{ii'}^q \sum_{t=1}^{n} \theta_q^t$$

$$= \sum_{q=1}^{30} c_{ii'}^q \theta_q \tag{17}$$

where

$$\theta_q = \sum_{t=1}^{n} \theta_q^t.$$

The θs are the dispersion parameters for the two-breed population.

Note that for M_{16} through M_{30}, none of the alleles in i is IBD to an allele in i' (Table 5 through 7 and Fig. 4). Thus for $q = 16, \ldots, 30$

$$\sum_{t=1}^{n} \mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}|m_q) = \sum_{t=1}^{n} \mathrm{E}(G_{S_i^tD_i^t}|m_q)\mathrm{E}(G_{S_{i'}^tD_{i'}^t}|m_q).$$

Further, recall that $\underset{AB}{\mathrm{E}}(G_{S_i^tD_i^t}) = 0$ because F_1 was taken as the reference breed group. Therefore, from Tables 5 through 7, five θs are null:

$$\theta_{17} = \theta_{19} = \theta_{20} = \theta_{26} = \theta_{29} = 0. \tag{18}$$

The first term of (15) can thus be written as a linear combination of 25 θs. A recursive procedure can be used to compute the TIMP coefficients, as shown later. For noninbred populations, dispersion parameters can be reduced to a set of 12: θ_1 to θ_9, θ_{16}, θ_{18}, and θ_{21}.

For a two-breed population, the expectation term $\mathrm{E}(G_{S_i^t D_i^t})$ in (15) can be written in terms of TIMS as

$$\mathrm{E}(G_{S_i^t D_i^t}) = \sum_{r=1}^{5} \mathrm{E}[(G_{S_i^t D_i^t})|n_r]p_i^r$$

$$= \sum_{r=1}^{5} p_i^r \phi_r^t \tag{19}$$

where $\phi_r^t = \mathrm{E}[(G_{S_i^t D_i^t})|n_r]$.

For locus t, using (19), the second term of (15) is

$$\mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}) = \left(\sum_{r=1}^{5} p_i^r \phi_r^t\right)\left(\sum_{r'=1}^{5} p_{i'}^{r'} \phi_{r'}^t\right)$$

$$= \sum_{r=1}^{5}\sum_{r'=1}^{5} p_i^r p_{i'}^{r'} \phi_r^t \phi_{r'}^t. \tag{20}$$

Recall thatr $G_{S_i^t D_i^t}$ was defined for the F_1 breed group, where the relationship between alleles is given by N_2, so that terms involving ϕ_2^t in (20) are null. Thus, (20) can be written as

$$\begin{aligned}\mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(_{S_{i'}^t D_{i'}^t}) = {} & p_i^1 p_{i'}^1 (\phi_1^t)^2 + (p_i^1 p_{i'}^3 + p_i^3 p_{i'}^1)\phi_1^t \phi_3^t \\ & + p_i^3 p_{i'}^3 (\phi_3^t)^2 + p_i^4 p_{i'}^4 (\phi_4^t)^2 \\ & + (p_i^4 p_{i'}^5 + p_i^5 p_{i'}^4)\phi_4^t \phi_5^t + p_i^5 p_{i'}^5 (\phi_5^t)^2 \\ & + (p_i^1 p_{i'}^4 + p_i^4 p_{i'}^1)\phi_1^t \phi_4^t + (p_i^3 p_{i'}^4 + p_i^4 p_{i'}^3)\phi_3^t \phi_4^t \\ & + (p_i^1 p_{i'}^5 + p_i^5 p_{i'}^1)\phi_1^t \phi_5^t + (p_i^3 p_{i'}^5 + p_i^5 p_{i'}^3)\phi_3^t \phi_5^t. \end{aligned} \tag{21}$$

For n unlinked loci, the second term of (15) from (21) is

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}) = {} & p_i^1 p_{i'}^1 \sum_{t=1}^{n} (\phi_1^t)^2 + (p_i^1 p_{i'}^3 + p_i^3 p_{i'}^1) \sum_{t=1}^{n} \phi_1^t \phi_3^t \\ & + p_i^3 p_{i'}^3 \sum_{t=1}^{n} (\phi_3^t)^2 + p_i^4 p_{i'}^4 \sum_{t=1}^{n} (\phi_4^t)^2 \\ & + (p_i^4 p_{i'}^5 + p_i^5 p_{i'}^4) \sum_{t=1}^{n} \phi_4^t \phi_5^t + p_i^5 p_{i'}^5 \sum_{t=1}^{n} (\phi_5^t)^2 \\ & + (p_i^1 p_{i'}^4 + p_i^4 p_{i'}^1) \sum_{t=1}^{n} \phi_1^t \phi_4^t \\ & + (p_i^3 p_{i'}^4 + p_i^4 p_{i'}^3) \sum_{t=1}^{n} \phi_3^t \phi_4^t \\ & + (p_i^1 p_{i'}^5 + p_i^5 p_{i'}^1) \sum_{t=1}^{n} \phi_1^t \phi_5^t \\ & + (p_i^3 p_{i'}^5 + p_i^5 p_{i'}^3) \sum_{t=1}^{n} \phi_3^t \phi_5^t. \end{aligned} \tag{22}$$

Recall that M_{16} is the mode where each allele in individuals i and i' is independent and the alleles are from breed A (Table 5), and that N_1 is the mode where the two alleles in an individual are not IBD and both are from A. Thus $\mathrm{E}[(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t})|m_{16}] = \mathrm{E}[(G_{S_i^t D_i^t})|n_1]\mathrm{E}[(G_{S_{i'}^t D_{i'}^t})|n_1]$, so that $\sum_{t=1}^{n}(\phi_1^t)^2$ can be written as

$$\sum_{t=1}^{n} (\phi_1^t)^2 = \sum_{t=1}^{n} \mathrm{E}[(G_{S_i^t D_i^t})|n_1]\mathrm{E}[(G_{S_{i'}^t D_{i'}^t})|n_1]$$

$$= \sum_{t=1}^{n} \mathrm{E}[(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t})|m_{16}]$$

$$= \theta_{16} \tag{23}$$

which was defined before. Similarly,

$$\sum_{t=1}^{n} \phi_1^t \phi_3^t = \theta_{18} \tag{24}$$

$$\sum_{t=1}^{n} (\phi_3^t)^2 = \theta_{21} \tag{25}$$

$$\sum_{t=1}^{n} (\phi_4^t)^2 = \theta_{22} \tag{26}$$

$$\sum_{t=1}^{n} \phi_4^t \phi_5^t = \theta_{23} \tag{27}$$

$$\sum_{t=1}^{n} (\phi_5^t)^2 = \theta_{24} \tag{28}$$

$$\sum_{t=1}^{n} \phi_1^t \phi_4^t = \theta_{25} \tag{29}$$

$$\sum_{t=1}^{n} \phi_3^t \phi_4^t = \theta_{27} \tag{30}$$

$$\sum_{t=1}^{n} \phi_1^t \phi_5^t = \theta_{28} \tag{31}$$

$$\sum_{t=1}^{n} \phi_3^t \phi_5^t = \theta_{30} \tag{32}$$

where θ_qs were defined previously. Now, the second term of (15) can be expressed using the θ_qs as

$$\sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}) = \sum_{q=16}^{30} d_{ii'}^q \theta_q \tag{33}$$

where from (22), $d_{ii'}^q$s are products of TIMS coefficients:

$d_{ii'}^{16} = p_i^1 p_{i'}^1$ $\quad$ $d_{ii'}^{24} = p_i^5 p_{i'}^5$

$d_{ii'}^{18} = p_i^1 p_{i'}^3 + p_i^3 p_{i'}^1$ $\quad$ $d_{ii'}^{25} = p_i^1 p_{i'}^4 + p_i^4 p_{i'}^1$

$d_{ii'}^{21} = p_i^3 p_{i'}^3$ $\quad$ $d_{ii'}^{27} = p_i^3 p_{i'}^4 + p_i^4 p_{i'}^3$

$d_{ii'}^{22} = p_i^4 p_{i'}^4$ $\quad$ $d_{ii'}^{28} = p_i^1 p_{i'}^5 + p_i^5 p_{i'}^1$

$d_{ii'}^{23} = p_i^4 p_{i'}^5 + p_i^5 p_{i'}^4$ $\quad$ $d_{ii'}^{30} = p_i^3 p_{i'}^5 + p_i^5 p_{i'}^3$

and where $d_{ii'}^{17} = d_{ii'}^{19} = d_{ii'}^{20} = d_{ii'}^{26} = d_{ii'}^{29} = 0$.

Subtracting (33) from (17) yields the covariance between individuals i and i':

$$\begin{aligned}\mathrm{Cov}(G_i, G_{i'}) &= \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t}) - \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t})\mathrm{E}(G_{S_{i'}^t D_{i'}^t}) \\ &= \sum_{q=1}^{30} c_{ii'}^q \theta_q - \sum_{q=16}^{30} d_{ii'}^q \theta_q \\ &= \sum_{q=1}^{15} c_{ii'}^q \theta_q + \sum_{q=16}^{30} (c_{ii'}^q - d_{ii'}^q)\theta_q. \end{aligned} \tag{34}$$

Thus, covariance between relatives in a two-breed population is a function of TIMP coefficients, TIMS coefficients, and 30 dispersion parameters, five of which are null.

Variance

The variance, a special case of covariance, can be computed as

$$\mathrm{Var}(G_i) = \mathrm{Cov}(G_i, G_{i'})$$

$$= \sum_{t=1}^{n} \mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t}) - \sum_{t=1}^{n} [\mathrm{E}(G_{S_i^tD_i^t})]^2$$

$$= \sum_{q=1}^{15} c_{ii}^q\theta_q - (d_{ii'}^{16}\theta_{16} + d_{ii'}^{21}\theta_{21} + d_{ii'}^{22}\theta_{22} + d_{ii'}^{24}\theta_{24}) \qquad (35)$$

because $c_{ii}^q = 0$ for $q = 16,\ldots,30$ (Tables 5 through 7, Fig. 4) and $d_{ii'}^q = 0$ for $q \neq 16, 21, 22,$ or 24 (equation 18 and Fig. 1).

Recursive procedures

For purebred populations, identity coefficients between relatives have been computed using recursive formulae (Harris 1964). Also, covariance between relatives have been computed directly, without computing identity coefficients, using a recursive procedure (Smith and Mäki-Tanila 1990).

For crossbred populations, it is shown below how to obtain identity coefficients ($c_{ii'}^q$), coefficients of breed origin (p_i^r), and covariance between relatives using recursive procedures. It is assumed that individuals in the base populations are unrelated, noninbred, and of known pure breed.

TIMP coefficients

Consider TIMP coefficient q for two individuals i and i', $c_{ii'}^q$. This is the probability that alleles S_i^t, D_i^t, $S_{i'}^t$, and $D_{i'}^t$ belong to identity mode M_q. Note that for individual i, allele S_i^t is either allele S_j^t or D_j^t of sire j, with equal probability. Thus $c_{ii'}^q$ can be expressed recursively as

$$c_{ii'}^q = \Pr[(S_i^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q]$$
$$= \tfrac{1}{2}\Pr[(S_j^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q]$$
$$+ \tfrac{1}{2}\Pr[(D_j^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q]. \qquad (36)$$

This process will be referred to as "recursion of allele S_i^t". Similarly, allele D_i^t is either allele S_k^t or D_k^t of i's dam k, with equal probability. By recursion on allele D_i^t, each of the two probabilities in (36) becomes

$$\Pr[(S_j^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q] = \tfrac{1}{2}\Pr[(S_j^t, S_k^t, S_{i'}^t, D_{i'}^t) \in M_q]$$
$$+ \tfrac{1}{2}\Pr[(S_j^t, D_k^t, S_{i'}^t, D_{i'}^t) \in M_q] \qquad (37)$$

and

$$\Pr[(D_j^t, D_i^t, S_{i'}^t, D_{i'}^t) \in M_q] = \tfrac{1}{2}\Pr[(D_j^t, S_k^t, S_{i'}^t, D_{i'}^t) \in M_q]$$
$$+ \tfrac{1}{2}\Pr[(D_j^t, D_k^t, S_{i'}^t, D_{i'}^t) \in M_q] \qquad (38)$$

Suppose now that j' is the sire of i', with paternal and maternal alleles $S_{j'}^t$ and $D_{j'}^t$, and that k' is the dam of i', with paternal and maternal alleles $S_{k'}^t$ and $D_{k'}^t$. With equal probability, allele $S_{i'}^t$ is either $S_{j'}^t$ or $D_{j'}^t$ of sire j' and allele $D_{i'}^t$ is either $S_{k'}^t$ or $D_{k'}^t$ of dam k'. The probabilities in (37) and (38), therefore, can be processed further by recursion on alleles $S_{i'}^t$ and $D_{i'}^t$. This recursive process is continued until each of the four alleles is from a base population.

Recall that TIMPs specify the identity state and breed origin for alleles, and that each individual in the base population is assumed to be unrelated, noninbred, and of known pure breed; thus alleles of base individuals are not IBD. Therefore, the probability that alleles from base individuals belong to identity mode M_q is either unity or null, as can be seen from the following hypothetical example involving two individuals.

Let 1 and 2 be base individuals from breed A. Consider computing, say, $\Pr[(S_1^t, D_1^t, S_1^t, S_2^t) \in M_1]$. Note that one of the first two alleles is IBD to one of the second two alleles and that other alleles are not IBD. This BO-IS belongs to M_1 (Table 1 and Fig. 4), so that

$$\Pr[(S_1^t, D_1^t, S_1^t, S_2^t) \in M_q] = \begin{cases} 1 & \text{for } q = 1 \\ 0 & \text{for } q \neq 1. \end{cases}$$

Under Mendelian inheritance, alleles that an individual has in common with any ancestor must also be present in at least one of its parents. Following the rules given below ensures that the recursive process is consistent with Mendelian inheritance:

(1) Number individuals in ascending order from oldest to youngest.
(2) Always recurse on an allele of the youngest individual.
(3) Recurse on identical alleles simultaneously.

For example, consider the recursive process for $\Pr[(S_7^t, D_8^t, S_5^t, D_8^t) \in M_q]$. By Rule 2, recursion must be on allele D_8^t, and by Rule 3, recursion on identical alleles D_8^t must be done simulataneously. Suppose the dam of individual 8 is 3, with alleles S_3^t and D_3^t. Then

$$\Pr[(S_7^t, D_8^t, S_5^t, D_8^t) \in M_q] = \tfrac{1}{2}\Pr[(S_7^t, S_3^t, S_5^t, S_3^t) \in M_q]$$
$$+ \tfrac{1}{2}\Pr[(S_7^t, D_3^t, S_5^t, S_3^t) \in M_q].$$

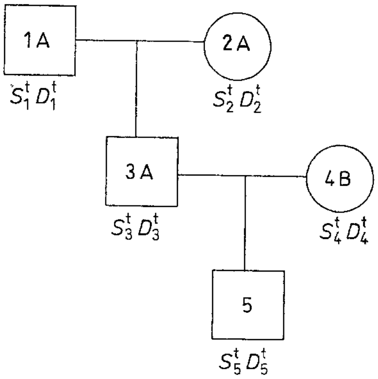
Now, to illustrate the recursive computation of TIMP coefficient, $c_{ii'}^q$, consider a pedigree of five individuals (Fig. 5) from a two-breed (A and B) population. Assume individuals 1 and 2 are base individuals from breed A and 4 is a base individual from breed B; thus they are unrelated and noninbred. Consider, for example, computation of c_{35}^2. With equal probability, allele S_5^t of individual 5 is either S_3^t or D_3^t of sire 3, and allele D_5^t is either S_4^t or D_4^t of dam 4. Thus, c_{35}^2 can be expressed recursively as

$$c_{35}^2 = \Pr[(S_3^t, D_3^t, S_5^t, D_5^t) \in M_2]$$
$$= \tfrac{1}{2}\{\Pr[(S_3^t, D_3^t, S_3^t, D_5^t) \in M_2] + \Pr[(S_3^t, D_3^t, D_3^t, D_5^t) \in M_2]\}$$
$$= \tfrac{1}{4}\{\Pr[(S_3^t, D_3^t, S_3^t, S_4^t) \in M_2] + \Pr[(S_3^t, D_3^t, S_3^t, D_4^t) \in M_2]$$
$$+ \Pr[(S_3^t, D_3^t, S_3^t, S_4^t) \in M_2] + \Pr[(S_3^t, D_3^t, D_3^t, D_4^t) \in M_2]. \qquad (39)$$

Note that alleles S_4^t and D_4^t are from base individual 4; thus recursion on these two alleles is not continued. Alleles S_3^t and D_3^t, however, are from non-base individual 3; thus recursion on these two alleles is continued. Finally, c_{35}^2 can be expressed as

$$c_{35}^2 = \tfrac{1}{16}\{\Pr[(S_1^t, S_2^t, S_1^t, S_4^t) \in M_2] + \Pr[(S_1^t, D_2^t, S_1^t, S_4^t) \in M_2]$$
$$+ \Pr[(D_1^t, S_2^t, D_1^t, S_4^t) \in M_2] + \Pr[(D_1^t, D_2^t, D_1^t, S_4^t) \in M_2]$$
$$+ \Pr[(S_1^t, S_2^t, S_1^t, D_4^t) \in M_2] + \Pr[(S_1^t, D_2^t, S_1^t, D_4^t) \in M_2]$$
$$+ \Pr[(D_1^t, S_2^t, D_1^t, D_4^t) \in M_2] + \Pr[(D_1^t, D_2^t, D_1^t, D_4^t) \in M_2]$$
$$+ \Pr[(S_1^t, S_2^t, S_2^t, S_4^t) \in M_2] + \Pr[(S_1^t, D_2^t, D_2^t, S_4^t) \in M_2]$$

Fig. 5 Pedigree of five individuals used in illustration of recursive procedures. *Circles* represent females, *squares* represent males. The breed of a founder is given by the *letter* within the square or circle



$$\begin{aligned}&+\Pr[(D_1^t,S_2^t,S_2^t,S_4^t)\in M_2]+\Pr[(D_1^t,D_2^t,D_2^t,S_4^t)\in M_2]\\&+\Pr[(S_1^t,S_2^t,S_2^t,D_4^t)\in M_2]+\Pr[(S_1^t,D_2^t,D_2^t,D_4^t)\in M_2]\\&+\Pr[(D_1^t,S_2^t,S_2^t,D_4^t)\in M_2]+\Pr[(D_1^t,D_2^t,D_2^t,D_4^t)\in M_2]\}.\end{aligned}\tag{40}$$

Recall that identity mode M_2 has one of the first two alleles IBD to one of the second two alleles and that the other two alleles are not IBD, with the two IBD coming from breed A and the two not IBD from breeds A and B (Table 1). Each of the 16 BO-IS combinations in (40) belongs to M_2. Thus, each probability in (40) is unity, and $c_{35}^2 = 1$.

Now for another example, consider computation of TIMP coefficient c_{15}^1. Note that alleles S_1^t and D_1^t are from base individual 1; thus recursion is not performed on these two alleles. Alleles S_5^t and D_5^t, however, are from non-base individual 5; thus recursion is performed on these two alleles. With equal probability, allele S_5^t is either S_3^t or D_3^t of its sire 3, and allele D_5^t is either S_4^t or D_4^t of its dam 4. Thus, c_{15}^1 can be processed recursively until all alleles are from base individuals:

$$\begin{aligned}c_{15}^1&=\Pr[(S_1^t,D_1^t,S_5^t,D_5^t)\in M_1]\\&=\tfrac{1}{2}\{\Pr[(S_1^t,D_1^t,S_3^t,D_5^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_3^t,D_5^t)\in M_1]\}\\&=\tfrac{1}{4}\{\Pr[(S_1^t,D_1^t,S_3^t,S_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,S_3^t,D_4^t)\in M_1]\\&\quad+\Pr[(S_1^t,D_1^t,D_3^t,S_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_3^t,D_4^t)\in M_1]\}\\&=\tfrac{1}{8}\{\Pr[(S_1^t,D_1^t,S_1^t,S_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_1^t,S_4^t)\in M_1]\\&\quad+\Pr[(S_1^t,D_1^t,S_1^t,D_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_1^t,D_4^t)\in M_1]\\&\quad+\Pr[(S_1^t,D_1^t,S_2^t,S_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_2^t,S_4^t)\in M_1]\\&\quad+\Pr[(S_1^t,D_1^t,S_2^t,D_4^t)\in M_1]+\Pr[(S_1^t,D_1^t,D_2^t,D_4^t)\in M_1]\}.\end{aligned}\tag{41}$$

None of the eight BO-IS combinations in (41) belongs to M_1. Thus, each probability in (41) is null, and $c_{15}^1 = 0$.

To compute c_{15}^2, note that each of the first four combinations in (41) belongs to M_2 and each of the last four combinations does not. Thus, each of the first four probabilities in (41) is unity, each of the last four probabilities is null, and $c_{15}^2 = \frac{1}{2}$.

TIMS coefficients

Recall that TIMS coefficients for individual i, p_i^r, is the probability that alleles S_i^t and D_i^t belong to identity mode N_r. With equal probability, allele S_i^t is either S_j^t or D_j^t of its sire j, and allele D_i^t is either S_k^t or D_k^t of its dam k. Thus, p_i^r can be expressed recursively as

$$\begin{aligned}p_i^r&=\Pr[(S_i^t,D_i^t)\in N_r]\\&=\tfrac{1}{2}\{[\Pr(S_j^t,D_i^t)\in N_r]+[\Pr(D_j^t,D_i^t)\in N_r]\}\\&=\tfrac{1}{4}\{[\Pr(S_j^t,S_k^t)\in N_r]+[\Pr(S_j^t,D_k^t)\in N_r]\\&\quad+[\Pr(D_j^t,S_k^t)\in N_r]+[\Pr(D_j^t,D_k^t)\in N_r]\}.\end{aligned}\tag{42}$$

This recursive process is continued until each allele is from a base population. Rules 1 through 3 for computing the TIMP coefficients must also be followed here.

To illustrate, consider the same pedigree that was used to compute TIMP coefficients (Fig. 5), with the same information and assumptions. Consider a coefficient of breed origin for individual 3, p_3^1. With equal probability, allele S_3^t is either S_1^t or D_1^t of its sire 1, and allele D_3^t is either S_2^t or D_2^t of its dam 2. Thus p_3^1 can be processed recursively until each allele is from a base population:

$$\begin{aligned}p_3^1&=\Pr[(S_3^t,D_3^t)\in N_1]\\&=\tfrac{1}{2}\{\Pr[(S_1^t,D_3^t)\in N_1]+\Pr[(D_1^t,D_3^t)\in N_1]\}\\&=\tfrac{1}{4}\{\Pr[(S_1^t,S_2^t)\in N_1]+\Pr[(S_1^t,D_2^t)\in N_1]\\&\quad+\Pr[(D_1^t,S_2^t)\in N_1]+\Pr[(D_1^t,D_2^t)\in N_1]\}.\end{aligned}\tag{43}$$

Each of the four BO-IS combinations in (43) belongs to identity mode N_1, where two alleles are not IBD and come from breed A (Fig. 1). Thus each probability in (43) is unity, and $p_3^1 = 1$.

Covariance between relatives

Covariance between relatives can be computed as in (34) using TIMP coefficients ($c_{ii'}$) and products of coefficients of breed origin ($d_{ii'}$) obtained by recursion. We now compute the covariance between relatives directly following a recursive procedure (Smith and Mäki-Tanila, 1990).

Consider the first expectation term of (15), $\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})$. With equal probability, allele S_i^t is either S_j^t or D_j^t of its sire j. Thus, $\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})$ can be expressed recursively as

$$\mathrm{E}(G_{S_i^tD_i^t}G_{S_{i'}^tD_{i'}^t})=\tfrac{1}{2}[\mathrm{E}(G_{S_j^tD_i^t}G_{S_{i'}^tD_{i'}^t})+\mathrm{E}(G_{D_j^tD_i^t}G_{S_{i'}^tD_{i'}^t})].\tag{44}$$

Similarly, with equal probability, allele D_i^t is either S_k^t or D_k^t of its dam k. Thus, by recursion on allele D_i^t, the two expectation terms in (44) can be expressed as

$$\mathrm{E}(G_{S_j^tD_i^t}G_{S_{i'}^tD_{i'}^t})=\tfrac{1}{2}[\mathrm{E}(G_{S_j^tS_k^t}G_{S_{i'}^tD_{i'}^t})+\mathrm{E}(G_{S_j^tD_k^t}G_{S_{i'}^tD_{i'}^t})]\tag{45}$$

and

$$\mathrm{E}(G_{D_j^tD_i^t}G_{S_{i'}^tD_{i'}^t})=\tfrac{1}{2}[\mathrm{E}(G_{D_j^tS_k^t}G_{S_{i'}^tD_{i'}^t})+\mathrm{E}(G_{D_j^tD_k^t}G_{S_{i'}^tD_{i'}^t})].\tag{46}$$

Each expectation term in (45) and (46) can be processed recursively as before until each allele is from a base population. Then, the BO-IS combination for the four alleles with belong to one of 30 TIMPs. If the BO-IS belongs to, say, M_q, then the expectation term, summed over n loci, can be replaced by θ_q.

Similarly, expectation terms $\mathrm{E}(G_{S_i^tD_i^t})\mathrm{E}(G_{S_{i'}^tD_{i'}^t})$ in (15) can also be processed by recursion on alleles S_i^t and D_i^t, and $S_{i'}^t$ and $D_{i'}^t$:

$$\begin{aligned}\mathrm{E}(G_{S_i^tD_i^t})\mathrm{E}(_{S_{i'}^tD_{i'}^t})=\tfrac{1}{16}&[\mathrm{E}(G_{S_j^tS_k^t})+\mathrm{E}(G_{S_j^tD_k^t})\\&+\mathrm{E}(G_{D_j^tS_k^t})+\mathrm{E}(G_{D_j^tD_k^t})]\\&\times[\mathrm{E}(G_{S_{j'}^tS_{k'}^t})+\mathrm{E}(G_{S_{j'}^tD_{k'}^t})\\&+\mathrm{E}(G_{D_{j'}^tS_{k'}^t})+\mathrm{E}(G_{D_{j'}^tD_{k'}^t})].\end{aligned}\tag{47}$$

The expectation terms in (47) can be processed recursively until each allele is from a base population. Then the BO-IS combination for the pair of alleles at a locus will belong to one of five N modes (Fig. 1), and the product of expectations summed over n loci can be replaced by θ_q as explained below.

Suppose, for example, the pair of alleles S_j^t and S_k^t are from base populations and belong to mode N_r, and the pair of alleles $S_{j'}^t$ and $S_{k'}^t$ are also from base populations and belong to $N_{r'}$. Then for $r = r' = 1$, the product of $\mathrm{E}[(G_{S_j^tD_k^t})|n_r]$ and $\mathrm{E}[(G_{S_{j'}^tS_{k'}^t})|n_{r'}]$ summed over n loci can be written as

$$\sum_{t=1}^{n}\mathrm{E}[(G_{S_j^tS_k^t})|n_r]\mathrm{E}[(G_{S_{j'}^tS_{k'}^t})|n_{r'}]=\sum_{t=1}^{n}\mathrm{E}[(G_{S_j^tS_k^t}G_{S_{j'}^tS_{k'}^t})|m_q]=\theta_q\tag{48}$$

with $q = 16$, because in M_{16} the four alleles are independent and are from breed A, and in N_1 the two alleles are not IBD and are from breed A. The relationship between all pairs of N modes and M modes is in Table 8. Thus, for example, if $r = 1$ and $r' = 3$, from Table 8, $q = 18$, and (48) in θ_{18}.

To illustrate, consider the same pedigree that was used to compute TIMP coefficients and TIMS coefficients (Fig. 5), with the same information and assumptions. For example, covariance between individuals 1 and 5 is

$$\mathrm{Cov}(G_{S_1^tD_1^t}G_{S_5^tD_5^t})=\mathrm{E}(G_{S_1^tD_1^t}G_{S_5^tD_5^t})-\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_5^tD_5^t}).\tag{49}$$

Alleles S_1^t and D_1^t are from a base individual; thus recursion is not performed on these two alleles. With equal probability, allele S_5^t is

Table 8 The relationship between pairs of N modes and M modes

	N_1	N_2	N_3	N_4	N_5
N_1	M_{16}	M_{17}	M_{18}	M_{25}	M_{28}
N_2	M_{17}	M_{19}	M_{20}	M_{26}	M_{29}
N_3	M_{18}	M_{20}	M_{21}	M_{27}	M_{30}
N_4	M_{25}	M_{26}	M_{27}	M_{22}	M_{23}
N_5	M_{28}	M_{29}	M_{30}	M_{23}	M_{24}

either S_3^t or D_3^t of its sire 3, and allele D_5^t is either S_4^t or D_4^t of its dam 4. Thus, the first expectation of (49) can be expressed recursively as

$$\mathrm{E}(G_{S_1^tD_1^t}G_{S_5^tD_5^t})=\tfrac{1}{4}[\mathrm{E}(G_{S_1^tD_1^t}G_{S_3^tS_4^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{S_3^tD_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t}G_{D_3^tS_4^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{D_3^tD_4^t})]. \tag{50}$$

Alleles S_4^t and D_4^t are from base individual 4. With equal probability, allele S_3^t is either S_1^t or D_1^t of its sire 1, and allele D_3^t is either allele S_2^t or D_2^t of its dam 2.
Thus expectations in (50) can be processed recursively on alleles S_3^t and D_3^t as

$$\mathrm{E}(G_{S_1^tD_1^t}G_{S_5^tD_5^t})=\tfrac{1}{8}[\mathrm{E}(G_{S_1^tD_1^t}G_{S_1^tS_1^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{D_1^tS_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t}G_{S_1^tD_4^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{D_1^tD_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t}G_{S_2^tS_4^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{D_2^tS_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t}G_{S_2^tD_4^t})+\mathrm{E}(G_{S_1^tD_1^t}G_{D_2^tD_4^t})]. \tag{51}$$

Each of the first four BO-IS combinations belongs to M_2, where one of the first two alleles is IBD to one of the second two alleles and the other two alleles are not IBD, with the two IBD coming from breed A and the two not IBD from breeds A and B (Table 1). Thus each of the first four expectations in (51), summed over n loci, is replaced by θ_2. Also, each of the last four BO-IS combinations belongs to M_{17}, where four alleles are not IBD, with one allele from breed B and the others from breed A. Thus each of the last four expectations, summed over the n loci, is replaced by θ_{17}. Hence

$$\sum_{t=1}^{n}\mathrm{E}(G_{S_1^tD_1^t}G_{S_5^tD_5^t})=\tfrac{1}{2}(\theta_2+\theta_{17})$$
$$=\tfrac{1}{2}\theta_2$$

because $\theta_{17}=0$ by (18).

Similarly, for the second term, $\mathrm{E}(G_{S_1^tD_1^t})$ in (49), alleles S_1^t and D_1^t are already from a base population, so there is no need to recurse on these alleles. For $\mathrm{E}(G_{S_5^tD_5^t})$, however, recursion on alleles of individuals 5 and 3 gives

$$\mathrm{E}(G_{S_5^tD_5^t})=\tfrac{1}{2}[\mathrm{E}(G_{S_3^tD_5^t})+\mathrm{E}(G_{D_3^tD_5^t})]$$
$$=\tfrac{1}{4}[\mathrm{E}(G_{S_3^tS_4^t})+\mathrm{E}(G_{S_3^tD_4^t})+\mathrm{E}(G_{D_3^tS_4^t})+\mathrm{E}(G_{D_3^tD_4^t})]$$
$$=\tfrac{1}{8}[\mathrm{E}(G_{S_1^tS_4^t})+\mathrm{E}(G_{D_1^tS_4^t})+\mathrm{E}(G_{S_1^tD_4^t})+\mathrm{E}(G_{D_1^tD_4^t})$$
$$+\mathrm{E}(G_{S_2^tS_4^t})+\mathrm{E}(G_{D_2^tS_4^t})+\mathrm{E}(G_{S_2^tD_4^t})+\mathrm{E}(G_{D_2^tD_4^t})]. \tag{52}$$

Thus, the product of $\mathrm{E}(G_{S_1^tD_1^t})$ and $\mathrm{E}(G_{S_5^tD_5^t})$ gives

$$\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_5^tD_5^t})=\tfrac{1}{8}[\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_1^tS_4^t})+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{D_1^tS_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_1^tD_4^t})+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{D_1^tD_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_2^tS_4^t})+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{D_2^tS_4^t})$$
$$+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{S_2^tD_4^t})+\mathrm{E}(G_{S_1^tD_1^t})\mathrm{E}(G_{D_2^tD_4^t})]. \tag{53}$$

For each product of expectations in (53) note that, because one of the pairs of alleles belongs to N_2, each expectation summed over n loci will be null. Finally, covariance between individuals 1 and 5 is

$$\mathrm{Cov}(G_1,G_5)=\tfrac{1}{2}\theta_2. \tag{54}$$

Genetic dispersion parameters

The dispersion parameters defined previously correspond directly to the identity modes. Therefore, this led to a recursive method to compute variances and covariances directly without computing identity coefficients. These parameters will be referred to as the "direct" dispersion parameters for a two-breed population. To give genetic meaning to the variances and covariances in a two-breed population, we show how to compute these using alternative parameters that have a more direct genetic interpretation. These alternative parameters will be referred to as the "genetic" dispersion parameters for a two-breed population.

Variance

Because loci are assumed to be unlinked and the pure breeds to be in gametic equilibrium, the variance of any individual i in the two-breed population can be written as

$$\mathrm{Var}(G_i)=\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}). \tag{55}$$

The above equation can be written as

$$\mathrm{Var}(G_i)=\sum_{t=1}^{n}[\mathop{\mathrm{E}}_{n_r}\mathrm{Var}(G_{S_i^tD_i^t}|n_r)+\mathop{\mathrm{Var}}_{n_r}\mathrm{E}(G_{S_i^tD_i^t}|n_r)] \tag{56}$$

(Kempthorne and Folks 1971). The first term of (56) can be expressed as

$$\mathop{\mathrm{E}}_{n_r}\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_r)=p_i^1\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_1)$$
$$+p_i^2\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_2)$$
$$\vdots$$
$$+p_i^5\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_5) \tag{57}$$

where, for example, $\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_1)=(V_A^A+V_D^A)$ is the genotypic variance in breed A, V_A^A is the additive variance in breed A, and V_D^A is the dominance variance in breed A. Similarly, the remaining variances in (57), summed over n loci, are genotypic variances for breed groups, AB, B, A^*, and B^*. Therefore, (57) is

$$\mathop{\mathrm{E}}_{n_r}\sum_{t=1}^{n}\mathrm{Var}(G_{S_i^tD_i^t}|n_r)=(V_A^A+V_D^A)p_i^1+(V_A^{AB}+V_D^{AB})p_i^2$$
$$+(V_A^B+V_D^B)p_i^3+(2V_A^A+4C_{AD}^{A^*}+V_D^{A^*})p_i^4$$
$$+(2V_A^B+4C_{AD}^{B^*}+V_D^{B^*})p_i^5 \tag{58}$$

where $V_{A_A}^{AB}+V_{A_B}^{AB}$ is the additive variance in the AB breed group, $V_{A_A}^{AB}$ is the additive variance of breed A alleles in the AB breed group, $V_{A_B}^{AB}$ is the additive variance of breed B alleles also in the AB breed group, V_D^{AB} is the dominance variance in the AB breed group. V_A^B is the additive variance in breed B, V_D^B is the dominance variance in breed B, $C_{AD}^{A^*}$ is the covariance between additive and dominance effects in breed A*, $V_D^{A^*}$ is the dominance variance in breed A*, $C_{AD}^{B^*}$ is the covariance between additive and dominance effects in breed B*, $V_D^{B^*}$ is the dominance variance in breed B*. These genetic dispersion parameters can be written in terms of direct dispersion parameters (θs) defined in the previous section. For example, consider the covariance between individuals i and i' given TIMP M_1. Recall that M_1 is the mode where only one allele in i is IBD to an allele in i' and all alleles are from breed A (Table 1). Thus, the covariance between i and i' is half the additive variance in breed A. Also, from the definition of covariance, this can be written as

$$\sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} | m_1) = \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} | m_1)$$

$$- \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} | m_1) \mathrm{E}(G_{S_{i'}^t D_{i'}^t} | m_1)$$

$$= \theta_1 - \theta_{16}.$$

Thus $V_A^A = 2(\theta_1 - \theta_{16})$. For another example, consider the covariance between i and i' given TIMP M_7. Recall that M_7 is the mode where both alleles in i are IBD to the alleles in i' and all alleles are from breed A (Table 2). Thus, this covariance is the sum of the additive and dominance variances in breed A. This can also be written as

$$\sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} | m_7) = \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} | m_7)$$

$$- \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} | m_7) \mathrm{E}(G_{S_{i'}^t D_{i'}^t} | m_7)$$

$$= \theta_7 - \theta_{16}.$$

Thus $V_D^A = \theta_7 + \theta_{16} - 2\theta_1$. The relationship between these genetic dispersion parameters and the direct dispersion parameters is given in Table 9.

Now, the second term of (56) can be written as

$$\sum_{t=1}^{n} \underset{n_r}{\mathrm{Var}} \mathrm{E}(G_{S_i^t D_i^t} | n_r) = \sum_{t=1}^{n} \underset{n_r}{\mathrm{Var}}(\phi_r^t)$$

$$= \sum_{t=1}^{n} \{\mathrm{E}[(\phi_r^t)^2] - [\mathrm{E}(\phi_r^t)]^2\}. \tag{59}$$

Recall that $\phi_r^t = \mathrm{E}(G_{S_i^t D_i^t} | n_r)$. Following Lo et al. (1993), (59) can be rearranged as

$$\sum_{t=1}^{n} \underset{n_r}{\mathrm{Var}} \mathrm{E}(G_{S_i^t D_i^t} | n_r) = p_i^1 p_i^2 \sum_{t=1}^{n} (\phi_1^t - \phi_2^t)^2$$

$$+ p_i^1 p_i^3 \sum_{t=1}^{n} (\phi_1^t - \phi_3^t)^2$$

$$\vdots$$

$$+ p_i^4 p_i^5 \sum_{t=1}^{n} (\phi_4^t - \phi_5^t)^2. \tag{60}$$

The first sum in (60) is the contribution to the variance due to differences in genotypic frequencies between breed groups A and AB. This will be denoted $V_S^{A,AB}$, and can be written in terms of the direct dispersion parameters as

$$V_S^{A,AB} = \sum_{t=1}^{n} (\phi_1^t - \phi_2^t)^2$$

$$= \sum_{t=1}^{n} (\phi_1^t)^2 = \theta_{16}. \tag{61}$$

Similarly, the second sum in (60) is the contribution to the variance due to differences in genotypic frequencies between breed groups A and B, and can be written as

$$V_S^{A,B} = \sum_{t=1}^{n} (\phi_1^t - \phi_3^t)^2$$

$$= \sum_{t=1}^{n} [(\phi_1^t)^2 - 2\phi_1^t \phi_3^t + (\phi_3^t)^2]$$

$$= \theta_{16} - 2\theta_{18} + \theta_{21}. \tag{62}$$

The remaining sums in (60) are contributions to the variance due to differences in genotypic frequencies between breed groups A^* and A ($V_S^{A^*,A}$), B^* and $A(V_S^{B^*,A})$, B and $AB(V_S^{B,AB})$, A^* and AB ($V_S^{A^*,AB}$), B^* and $AB\,(V_S^{B^*,AB})$, A^* and $B\,(V_S^{A^*,B})$, B^* and $B\,(V_S^{B^*,B})$, and A^* and B^* ($V_S^{A^*,B^*}$). Now, (60) can be written as

$$\sum_{t=1}^{n} \underset{n_r}{\mathrm{Var}} \mathrm{E}(G_{S_i^t D_i^t} | n_r) = V_S^{A,AB} p_i^1 p_i^2 + V_S^{A,B} p_i^1 p_i^3$$

$$+ V_S^{A^*,A} p_i^1 p_i^4 + V_S^{B^*,A} p_i^1 p_i^5$$

$$+ V_S^{B,AB} p_i^2 p_i^3 + V_S^{A^*,AB} p_i^2 p_i^4$$

$$+ V_S^{B^*,AB} p_i^2 p_i^5 + V_S^{A^*,B} p_i^3 p_i^4$$

$$+ V_S^{B^*,B} p_i^3 p_i^5 + V_S^{A^*,B^*} p_i^4 p_i^5. \tag{63}$$

The relationship between the genetic dispersion parameters in (63) and the direct dispersion parameters is given in Table 10.

Note that, in some special cases (e.g., two-breed terminal crossbreeding system), differences in genotypic frequencies between breed groups will not contribute to the genetic variance. For example, the coefficient $p_i^1 p_i^2$ for $V_S^{A,AB}$ is null in A because p_i^2 is null for individuals from the A breed group.

These genetic dispersion parameters that contribute to the second term of (56) result from differences in genotypic frequencies between breeds. Note that, under additive inheritance, there is an analogous contribution to the genetic variance due to differences in allelic frequencies between breeds (Lo et al. 1993).

Table 9 Genetic dispersion parameters (η_q) that contribute to the first term of (56), and their linear relationship to direct dispersion parameters (θ_q). The genetic identity coefficients ($r_{ii'}^q$) are also linearly related to $c_{ii'}^q$ as shown here

q	$r_{ii'}^q$	η_q	Relationship
1	$\frac{1}{2}c_{ii'}^1 + c_{ii'}^7 + c_{ii'}^{10} + 2c_{ii'}^{14}$	V_A^A	$2(\theta_1 - \theta_{16})$
2	$c_{ii'}^7$	V_D^A	$\theta_7 - 2\theta_1 + \theta_{16}$
3	$c_{ii'}^3 + c_{ii'}^8$	$V_{A_A}^{AB}$	θ_3
4	$c_{ii'}^5 + c_{ii'}^8$	$V_{A_B}^{AB}$	θ_4
5	$c_{ii'}^8$	V_D^{AB}	$\theta_8 - \theta_3 - \theta_4$
6	$\frac{1}{2}c_{ii'}^6 + c_{ii'}^9 + c_{ii'}^{13} + c_{ii'}^{15}$	V_A^B	$2(\theta_6 - \theta_{21})$
7	$c_{ii'}^9$	V_D^B	$\theta_9 - 2\theta_6 + \theta_{21}$
8	$c_{ii'}^{10} + 4c_{ii'}^{14}$	$C_{AD}^{A^*}$	$(\theta_{10} - \theta_{25}) - 2(\theta_1 - \theta_{16})$
9	$c_{ii'}^{14}$	$V_D^{A^*}$	$(\theta_{14} - \theta_{22}) + 4(\theta_1 - \theta_{16}) - 4(\theta_{10} - \theta_{25})$
10	$c_{ii'}^{13} + 4c_{ii'}^{15}$	$C_{AD}^{B^*}$	$(\theta_{13} - \theta_{30}) - 2(\theta_6 - \theta_{21})$
11	$c_{ii'}^{15}$	$V_D^{B^*}$	$(\theta_{15} - \theta_{24}) + 4(\theta_6 - \theta_{21}) - 4(\theta_{13} - \theta_{30})$

Table 10 Genetic dispersion parameters (η_q) that contribute to the second term of (56), and their linear relationship to direct dispersion parameters (θ_q). The genetic identity coefficients ($r_{ii'}^q$) are also linearly related to $c_{ii'}^q$ and $d_{ii'}^q$ as shown here

q	$r_{ii'}^q$	η_q	Relationship
12	$c_{ii'}^1 + c_{ii'}^7 + c_{ii'}^{16} - d_{ii'}^{16} - r_{ii'}^{13} - r_{ii'}^{14} - r_{ii'}^{15}$	$V_S^{A,AB}$	θ_{16}
13	$-\frac{1}{2}(c_{ii'}^{18} - d_{ii'}^{18})$	$V_S^{A,B}$	$\theta_{16} + \theta_{21} - 2\theta_{18}$
14	$\frac{1}{2}(d_{ii'}^{25} - c_{ii'}^{25} - c_{ii'}^{10})$	$V_S^{A,A*}$	$\theta_{22} + \theta_{16} - 2\theta_{25}$
15	$-\frac{1}{2}(c_{ii'}^{28} - d_{ii'}^{28})$	$V_S^{B*,A}$	$\theta_{24} + \theta_{16} - 2\theta_{28}$
16	$c_{ii'}^6 + c_{ii'}^9 + c_{ii'}^{21} - d_{ii'}^{21} - r_{ii'}^{13} - r_{ii'}^{19} - r_{ii'}^{20}$	$V_S^{B,AB}$	θ_{21}
17	$c_{ii'}^{14} + c_{ii'}^{22} + d_{ii'}^{22} - r_{ii'}^{14} - r_{ii'}^{19} - r_{ii'}^{21}$	$V_S^{A*,AB}$	θ_{22}
18	$c_{ii'}^{15} + c_{ii'}^{24} - d_{ii'}^{24} - r_{ii'}^{15} - r_{ii'}^{20} - r_{ii'}^{21}$	$V_S^{B*,AB}$	θ_{24}
19	$-\frac{1}{2}(c_{ii'}^{27} - d_{ii'}^{27})$	$V_S^{A*,B}$	$\theta_{22} + \theta_{21} - 2\theta_{27}$
20	$\frac{1}{2}(d_{ii'}^{30} - c_{ii'}^{30} - c_{ii'}^{13})$	$V_S^{B,B*}$	$\theta_{24} + \theta_{21} - 2\theta_{30}$
21	$-\frac{1}{2}(c_{ii'}^{23} - d_{ii'}^{23})$	$V_S^{A*,B*}$	$\theta_{22} + \theta_{24} - 2\theta_{23}$

Covariance

The covariance between individuals i and i' in a two-breed population can be written as

$$\begin{aligned}\mathrm{Cov}(G_i, G_{i'}) &= \sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t}) \\ &= \sum_{t=1}^{n} \underset{m_q}{\mathrm{E}}\, \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t} \mid m_q) \\ &\quad + \sum_{t=1}^{n} \underset{m_q}{\mathrm{Cov}}[\mathrm{E}(G_{S_i^t D_i^t} \mid m_q), \mathrm{E}(G_{S_{i'}^t D_{i'}^t} \mid m_q)]. \end{aligned} \tag{64}$$

The first sum of (64) can be written as

$$\begin{aligned}\sum_{t=1}^{n} \underset{m_q}{\mathrm{E}}\, \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t} \mid m_q) &= \Pr(m_1) \sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t} \mid m_1) \\ &\quad + \Pr(m_2) \sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t} \mid m_2) \\ &\quad\vdots \\ &\quad + \Pr(m_{30}) \sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t}, G_{S_{i'}^t D_{i'}^t} \mid m_{30}). \end{aligned} \tag{65}$$

Recall that for M_{16} through M_{30} no allele in i is IBD to an allele in i'. Thus, the last 15 sums in (65) are null. Each of the remaining sums in (65) is a linear function of genetic dispersion parameters. Some of these parameters also contribute to the variance in equation (58) and are given in Table 11. For example, the first sum in (65) is half the additive variance in breed A (V_A^A) as seen earlier.

The additional parameters that contribute only to the genotypic covariance are additive covariance between a breed A parent and a breed AB offspring ($C_A^{A,AB}$), additive covariance between a breed B parent and a breed AB offspring ($C_A^{B,AB}$), covariance between the dominance effect of a breed A* parent and the additive effect of an AB offspring ($C_{AD}^{AB,A*}$), and covariance between the dominance effect of a breed B* parent and the additive effect of an AB offspring ($C_{AD}^{AB,B*}$). For example, recall that M_2 is the mode where only one allele in i is IBD to an allele in i'. Also, for M_2 the IBD alleles and one that is not IBD are from breed A, and the remaining allele is from breed B. Thus, the second sum in (65) is the additive covariance between a breed A parent and a breed AB offspring ($C_A^{A,AB}$). Also, from the definition of covariance, this can be written as

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{Cov}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} \mid m_2) &= \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} G_{S_{i'}^t D_{i'}^t} \mid m_2) \\ &\quad - \sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} \mid m_2) \mathrm{E}(G_{S_{i'}^t D_{i'}^t} \mid m_2) \\ &= \theta_2 \end{aligned}$$

because, $\sum_{t=1}^{n} \mathrm{E}(G_{S_i^t D_i^t} \mid m_2) \mathrm{E}(G_{S_{i'}^t D_{i'}^t} \mid m_2) = 0$. Thus $C_A^{A,AB} = \theta_2$. The relationship between these genetic dispersion parameters that contribute only to the genotypic covariance and the direct dispersion parameters is given in Table 11.

The second term in (64) can be written as

$$\begin{aligned}\sum_{t=1}^{n} \underset{m_q}{\mathrm{Cov}}[\mathrm{E}(G_{S_i^t D_i^t} \mid m_q), \mathrm{E}(G_{S_{i'}^t D_{i'}^t} \mid m_q)] &= \sum_{t=1}^{n} [\underset{m_q}{\mathrm{Cov}}(\mathrm{E}(G_{S_i^t D_i^t} \mid n_{q_1}), E(G_{S_{i'}^t D_{i'}^t} \mid n_{q_2})] \\ &= \sum_{t=1}^{n} \underset{m_q}{\mathrm{Cov}}(\phi_{q_1}^t, \phi_{q_2}^t) \end{aligned} \tag{66}$$

where q_1 is the N mode that the set of alleles S_i^t and D_i^t belongs to and q_2 is the N mode that the set of alleles $S_{i'}^t$ and $D_{i'}^t$ belongs to when the set of the four alleles belongs to M mode M_q. For example, $q_1 = 1$ and $q_2 = 1$ when $m_q = 1$. From the definition of covariance, (66) can be written as

$$\begin{aligned}\sum_{t=1}^{n} \underset{n_q}{\mathrm{Cov}}(\phi_{q_1}^t, \phi_{q_2}^t) &= [\underset{ii'}{\Pr}(n_1, n_1) - \underset{i}{\Pr}(n_1) \underset{i'}{\Pr}(n_1)] \sum_{t=1}^{n} \{\phi_1^t \phi_1^t\} \\ &\quad\vdots \\ &\quad + \underset{ii'}{\Pr}(n_5, n_5) - \underset{i}{\Pr}(n_5) \underset{i'}{\Pr}(n_5)] \sum_{t=1}^{n} \{\phi_5^t \phi_5^t\} \end{aligned} \tag{67}$$

where, for example, $\Pr_{ii'}(n_1, n_1)$ is the joint probability that the set of alleles S_i^t and D_i^t belongs to N_1 and that the set S_i^t and $D_{i'}^t$, belongs to N_1, and $\Pr_i(n_1)$ is the marginal probability that the set S_i^t and D_i^t belongs to N_1. Each of the sums in (67) can be written in terms of the genetic dispersion parameters which are given in Table 10. For example, from (61), the first sum in (67) is equal to $V_S^{A,AB} = \theta_{16}$. Thus, no additional parameters are required for the second term of (64).

Now, the genotypic covariance between i and i' can be written in terms of genetic dispersion parameters as

$$\mathrm{Cov}(G_i, G_{i'}) = \sum_{q=1}^{25} r_{ii'}^q \eta_q \tag{68}$$

Table 11 Genetic dispersion parameters (η_q) that contribute only to the genotypic covariances, and their linear relationship to direct dispersion parameters (θ_q). The genetic identity coefficients ($r_{ii'}^q$) are also linearly related to $c_{ii'}^q$ as shown here

q	$r_{ii'}^q$	η_q	Relationship
22	$c_{ii'}^2 + 2c_{ii'}^{11}$	$C_A^{A,AB}$	θ_2
23	$c_{ii'}^5 + 2c_{ii'}^{12}$	$C_A^{B,AB}$	θ_5
24	$c_{ii'}^{11}$	$C_{AD}^{AB,A*}$	$\theta_{11} - 2\theta_2$
25	$c_{ii'}^{12}$	$C_{AD}^{AB,B*}$	$\theta_{12} - 2\theta_5$

where η_qs are the genetic dispersion parameters given in Tables 9–11, and $r^q_{ii'}$s are the genetic identity coefficients for a two-breed population. Because the η_qs are linear functions of θ_qs, $r^q_{ii'}$s are also linear functions of $c^q_{ii'}$s and $d^q_{ii'}$s (Tables 9–11).

Numerical example

The computation of genotypic means and covariances in a two-breed population is illustrated for a two-locus model using a pedigree of six individuals (Fig. 6). Individuals 1 and 3 are base individuals from breed *A*, and individual 2 is a base individual from breed *B*; thus they are assumed to be unrelated and noninbred.

Consider a 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 genotypes, genotypic values (g_{jklm}) and genotypic frequencies in breed groups *A*, *B*, and F_1 are in Table 12. The location parameters for this trait are $\mu_1^* = 5.91$, $\mu_2^* = 5.28$, $\mu_3^* = 4.64$, $\mu_4^* = 5.7$ and $\mu_5^* = 4.4$, and the direct dispersion parameters are in Table 13.

As the six individuals in this pedigree are from different breed groups, they have different genotypic means and variances. The genotypic means were computed directly by recursion without computing TIMS coefficients, using principles similar to those for computing the genotypic covariances directly. They are: 5.91, 4.64, 5.91, 5.28, 5.595 and 5.726, respectively, for the six individuals in Fig. 6. Genotypic means could also be computed from (8) using TIMS coefficients. Computation of genotypic variances and covariances is described below.

The non-zero $c_{ii'}$ and $d_{ii'}$ coefficients for the pedigree in Fig. 6 are in Tables 14, 15, and 16. Covariance between relatives can now be computed by using (34). For example, the covariance between individuals 1 and 5 is

$$\begin{aligned}\mathrm{Cov}(G_1, G_5) &= 0.5(\theta_1 - \theta_{16})\\ &= 0.5(1.0458 - 0.2789)\\ &= 0.3834\end{aligned}$$

because the non-zero identity coefficients are $c^1_{15} = d^{16}_{15} = 0.5$ from Tables 14 and 16.

Using (35), variance for inbred individual 6 is

$$\begin{aligned}\mathrm{Var}(G_6) &= \mathrm{Cov}(G_6, G_6)\\ &= 0.625\theta_7 + 0.25\theta_8 + 0.125\theta_{14} - 0.3906\theta_{16}\\ &\quad - 0.0156\theta_{22} + 0.1562\theta_{25}\\ &= 0.625(1.8568) + 0.25(1.6516) + 0.125(2.9244)\\ &\quad - 0.3906(0.2789) - 0.0156(0.3944)\\ &\quad + 0.1562(0.3146)\\ &= 1.7747\end{aligned}$$

Fig. 6 Pedigree of six individuals used in numerical example. *Circles* represent females, *squares* represent males. The breed of a founder is given by the *letter* within the square or circle

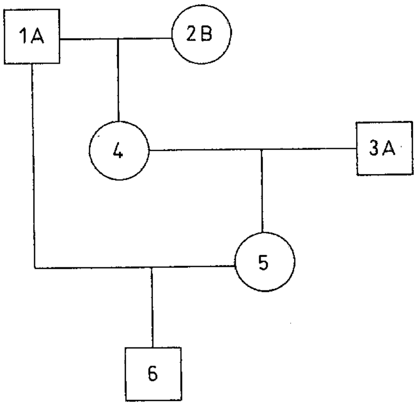


Table 12 Genotypes, genotypic values, and genotypic frequencies for the trait used in the numerical example. Frequency of U_1 is 0.3 in *A* and 0.4 in *B*, and frequency of V_1 is 0.2 in *A* and 0.7 in *B*. For each locus, the paternal allele is given first

Genotype	Genotypic value	Genotypic frequency		
$U_jU_kV_lV_m$	g_{jklm}	*A*	*B*	F_1
$U_1U_1V_1V_1$	2	0.0036	0.0784	0.0168
$U_1U_1V_1V_2$	3	0.0144	0.0336	0.0072
$U_1U_1V_2V_1$	3	0.0144	0.0336	0.0672
$U_1U_1V_2V_2$	4	0.0576	0.0144	0.0288
$U_1U_2V_1V_1$	4	0.0084	0.1176	0.0252
$U_1U_2V_1V_2$	5	0.0336	0.0504	0.0108
$U_1U_2V_2V_1$	5	0.0336	0.0504	0.1008
$U_1U_2V_2V_2$	6	0.1344	0.0216	0.0432
$U_2U_1V_1V_1$	4	0.0084	0.1176	0.0392
$U_2U_1V_1V_2$	5	0.0336	0.0504	0.0168
$U_2U_1V_2V_1$	5	0.0336	0.0504	0.1568
$U_2U_1V_2V_2$	6	0.1344	0.0216	0.0672
$U_2U_2V_1V_1$	5	0.0196	0.1764	0.0588
$U_2U_2V_1V_2$	6	0.0784	0.0756	0.0252
$U_2U_2V_2V_1$	6	0.0784	0.0756	0.2352
$U_2U_2V_2V_2$	7	0.3136	0.0324	0.1008

Table 13 Dispersion parameters (θ_q) for the trait used in the numerical example

q	θ_q	*q*	θ_q
1	0.7818	16	0.2669
2	0.5422	17	0.0000
3	0.5716	18	−0.2682
4	0.6156	19	0.0000
5	0.6468	20	0.0000
6	0.9500	21	0.2696
7	1.3408	22	0.2564
8	1.2376	23	−0.2196
9	1.6880	24	0.3944
10	1.3786	25	0.2396
11	1.2020	26	0.0000
12	1.3560	27	−0.2388
13	1.7312	28	−0.2994
14	2.7864	29	0.0000
15	3.3944	30	0.3032

Table 14 The non-zero $c^q_{ii'}$ coefficients for the pedigree in Fig. 6

ii'/q	1	2	5	7	9	10	16	18	27
11	0	0	0	1.000	0	0	0	0	0
12	0	0	0	0	0	0	0	1.000	0
13	0	0	0	0	0	0	1.000	0	0
14	0	1.000	0	0	0	0	0	0	0
15	0.500	0	0	0	0	0	0	0	0
16	0.500	0.250	0	0.125	0	0.125	0	0	0
22	0	0	0	0	1.000	0	0	0	0
23	0	0	0	0	0	0	0	1.000	0
24	0	0	1.000	0	0	0	0	0	0
25	0	0	0.500	0	0	0	0	0.500	0
26	0	0	0.250	0	0	0	0	0.625	0.125

Table 15 The non-zero $c^q_{ii'}$ coefficients for the pedigree in Fig. 6 (continued)

ii'/q	1	2	4	7	8	10	11	14	16	25
33	0	0	0	1.000	0	0	0	0	0	0
35	0.500	0.500	0	0	0	0	0	0	0	0
36	0.500	0	0	0	0	0	0	0	0.125	0.125
44	0	0	0	0	1.000	0	0	0	0	0
45	0	0.500	0.500	0	0	0	0	0	0	0
46	0	0.375	0.125	0	0.125	0	0.125	0	0	0
55	0	0	0	0.500	0.500	0	0	0	0	0
56	0.250	0.250	0.250	0.125	0	0.125	0	0	0	0
66	0	0	0	0.625	0.250	0	0	0.125	0	0

Table 16 The non-zero $d^q_{ii'}$ coefficients for the pedigree in Fig. 6

ii'/q	16	18	21	22	25	27
11	1.0000	0	0	0	0	0
12	0	1.0000	0	0	0	0
13	1.0000	0	0	0	0	0
15	0.5000	0	0	0	0	0
16	0.6250	0	0	0	0.1250	0
22	0	0	1.0000	0	0	0
23	0	1.0000	0	0	0	0
25	0	0.5000	0	0	0	0
26	0	0.6250	0	0	0	0.1250
33	1.0000	0	0	0	0	0
35	0.5000	0	0	0	0	0
36	0.6250	0	0	0	0.1250	0
55	0.2500	0	0	0	0	0
56	0.3125	0	0	0	0.0625	0
66	0.3906	0	0	0.0156	0.1562	0

Table 17 Genotypic covariance matrix for the pedigree in Fig. 6

Ind.	1	2	3	4	5	6
1	1.0739	0.0000	0.0000	0.5422	0.2574	0.6696
2	0.0000	1.4184	0.0000	0.6468	0.3234	0.1617
3	0.0000	0.0000	1.0739	0.0000	0.5285	0.2574
4	0.5422	0.6468	0.0000	1.2376	0.5789	0.5852
5	0.2574	0.3234	0.5285	0.5789	1.2225	0.7264
6	0.6696	0.1617	0.2574	0.5852	0.7264	1.3500

because the non-zero identity coefficients are $c^7_{66} = 0.625$, $c^8_{66} = 0.25$, $c^{14}_{66} = 0.125$, $d^{16}_{66} = 0.3906$, and $d^{22}_{66} = 0.0156$, and $d^{25}_{66} = 0.1562$ from Tables 15 and 16.

The matrix of genotypic covariances between individuals for the pedigree in Fig. 6 was computed directly by recursion, and is given in Table 17.

Conclusions

This paper describes theory and methods to compute genotypic means and covariances in a two-breed population for a dominance model multiple unlike loci. The genotypic mean is a linear function of five location parameters and the genotypic covariance between relatives is a linear function of 25 dispersion parameters. Recursive procedures are used to compute the necessary identity coefficients.

In the absence of inbreeding, the number of parameters for the mean is reduced from five to three and the number for the covariance is reduced from 25 to 12. The number of parameters required may be further reduced based on the structure of the population. Covariances in a two-breed terminal crossbreeding system, for example, are functions of nine genetic dispersion parameters. They are the additive variance in breed *A*, the dominance variance in breed *A*, the additive variance in breed *B*, the dominance variance in breed *B*, the additive variance in F_1 of alleles inherited from breed *A*, the additive variance in F_1 of alleles inherited from breed *B*, the dominance variance in F_1, the additive covariance between a breed *A* parent and an F_1 offspring, and the additive covariance between a breed *B* parent and an F_1 offspring. The covariance between purebreed *A* and F_1 halfsibs used by Wei et al. (1991 a,b) is one-half the additive covariance between a breed *A* parent and an F_1 offspring.

Assuming that genotypic values are normally distributed, the location and dispersion parameters can be estimated by maximum likelihood. Further research, however, is needed to determine the amount of data required to obtain useful estimates.

The theory presented here for computing genotypic means and covariances 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. Efficient methods to invert this matrix have been presented for an additive model in a purebred population (Henderson 1976) and in a multibreed population (Lo et al. 1993), and for a dominance model in a purebred population (Smith and Mäki-Tanila 1990; Hoeschele and VanRaden 1991). For a dominance model in a multibreed population, however, an efficient method to invert the covariance matrix has not been develoed. An alternative formulation for the mixed-model equations, which does not require the inverse of the covariance matrix, has been given by Harville (1976). This alternative formula may lead to a more efficient procedure to obtain BLUP for a dominance model in a two-breed population.

In deriving the theory presented here, it was assumed that loci are unlinked. Robustness of the covariance theory for violation of this assumption was examined by computer simulation (Lo 1993). Results suggest that the effect of linkage on covariances may be negligible.

The theory presented in this paper can be extended to include maternal effects and multiple traits, and for a multibreed population involving more than two breeds. If more traits and more breeds are involved, then additional parameters will be required.

References

De Boer IJM, Hoeschele I (1993) Genetic evaluation methods for populations with dominance and inbreeding. Theor Appl Genet 86:245–258

Dickerson GE (1973) Inbreeding and heterosis in animals. In: Anim Breed Genet Symp in Honor of Dr. J. L. Lush. Am Soc Anim Sci and Am Dairy Sci Assoc, Champaign Ill, pp 54–77

Eisen EJ, Horstghen-Schwark G, Saxton AM, Bandy TR (1983) Genetic interpretation and analysis of diallel crosses with animals. Theor Appl Genet 65:17–23

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, Inc., New York

Gillois M (1964) La relation d'identite en genetique, [Genetic identity relationship]. PhD thesis, Faculty of Science Paris. In: Jacquard A (1974) The genetic structure of populations. Springer-Verlag, Germany, pp 104–107

Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50:1319–1348

Harville DA (1976) Maximum likelihood appproaches to variance component estimation and to related problems. J Am Stat Assoc 72:320–340

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

Hill WG (1982) Dominance and epistasis as components of heterosis. J Anim Breed Genet 99:161–168

Hoeschele I, VanRaden PM (1991) Rapid method to compute inverses of dominance relationship matrices for noninbred populations by including sire by subclass effects. J Dairy Sci 74:557–569

Jacquard A (1974) The genetic structure of populations. Springer-Verlag, Germany

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

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

Lo LL (1993) Genetic evaluation and selection in multibreed populations. PhD thesis, University of Illinois, Urbana

Lo LL, Fernando RL, Grossman M (1993) Covariance between relatives in multibreed populations: Additive model. Theor Appl Genet 87:423–430

Smith SP, Mäki-Tanila A (1990) Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding. Genet Sel Evol 22:65–91

VanRaden PM (1992) Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J Dairy Sci 75: 3136–3144

Wei M, Van der Werf JHJ (1994) A mixed linear mixed model for combined crossbred and purebred selection. Anim Prod 59:400–413

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 purebreed and crossbred performance under the model with two loci. J. Anim Breed Genet 108:262–269