Theor Appl Genet (1993) 85:868–872

THEORETICAL AND APPLIED GENETICS

Analysing gametic variation with an animal model

B. Tier and J. Sölkner[1]

Animal Genetics and Breeding Unit,* University of New England, Armidale 2351, Australia

Received June 1, 1992; Accepted June 26, 1992
Communicated by H. F. Linskens

Summary. A method is presented to treat gametes as homozygous diploid individuals allowing their inclusion into the relationship matrix between animals. In this way standard techniques developed for the analysis of individual genetic variation may be used to analyze gametic variation. An example is given for maternal gametic imprinting and equivalence is shown to a gametic model. The method may also be adopted for the analysis of species (like the honey bee) with one haploid sex.

Key words: Gametic models – Relationship Matrix

Introduction

Henderson's (1976) rules for obtaining the inverse of the numerator relationship matrix between individuals directly from a list of pedigrees and inbreeding coefficients have promoted the use of improved methods for the estimation of genetic variances and covariances and for the prediction of the breeding value of animals. Restricted maximum likelihood (REML) methods for variance component estimation (Meyer 1990) and models based on best linear unbiased prediction (BLUP) have now become standard tools in animal breeding. When the covariance structure between individuals is modelled by the numerator relationship matrix (**A**) in these estimation and prediction procedures this has become known as the 'animal' model.

More recently, the concept of the relationship matrix has been extended to the gametic relationship matrix (**G**) where paternal and maternal gametes of an individual are considered separately (Smith and Allaire 1985). The main use of the gametic relationship matrix has been as a basis for constructing the relationship matrix between dominance effects (Smith 1984; Schaeffer et al. 1989; Smith and Mäki-Tanila 1990). There have been suggestions, though, for the use of the gametic relationship matrix in its own right in analyses for species which, like the honey bee *Apis mellifera* (Smith and Allaire 1985), have both haploid and diploid sexes, as well as in the analysis of maternal gametic imprinting effects (Gibson et al. 1988; Schaeffer et al. 1989). Schaeffer et al. (1989) illustrated a method for combining animal and gametic models in the same analysis.

In this paper we present a method for treating haploids or gametes as diploid organisms, thus allowing their inclusion in **A** and the use of an animal model. The equivalence to the combined approach of Schaeffer et al. (1989) is shown.

Construction of A and $\mathbf{A}^{-1}$: the usual rules

The usual rules for building **A** recursively from an ordered list of pedigrees are as follows:

$$a_{ii} = 1 + F_i \tag{1}$$

$$a_{ji} = (a_{jp} + a_{jq})/2 \tag{2}$$

where a_{ii} is the diagonal element of **A** pertaining to individual i with parents p and q and a_{ji} is the off diagonal element between individual j and i. F_i is the

---

[1] *Present address:* Institut für Nutztierwissenschaften, Universität für Bodenkultur, Gregor-Mendel Strasse 33, A-1180 Vienna, Austria

* AGBU is a joint initiative of UNE and the NSW Department of Agriculture

Correspondence to: B. Tier

inbreeding coefficient of i, which is normally computed as half the relationship between its parents, $a_{pq}/2$.

Henderson (1976) showed that the inverse of the relationship matrix, $\mathbf{A}^{-1}$, may be set up recursively from a pedigree list by adding

$$d^{ii} \text{ to } a^{ii}, \tag{3}$$

$$-d^{ii}/2 \text{ to } a^{ip}, a^{iq}, a^{qi}, a^{pi} \tag{4}$$

$$d^{ii}/4 \text{ to } a^{pp}, a^{pq}, a^{qp}, a^{qq} \tag{5}$$

where

$$d^{ii} = 1/[1 - 0.25(a_{pp} + a_{pp})] \tag{6}$$

and a with superscripts relate to elements of $\mathbf{A}^{-1}$. Tier and Sölkner (private communication), dealing with unusual relationship structures, showed that d^{ii} as in (6) is a special case of a slightly more general formula, namely

$$d^{ii} = 1/[1 + F_i - 0.25(a_{pp} + a_{pq} + a_{qp} + a_{qq})]. \tag{7}$$

Under the diploid mode of inheritance and given the usual assumptions (deviations from which are listed in Tier and Sölkner 1992), $F_i = a_{pq}/2 = 0.25(a_{pq} + a_{qp})$ so that those terms cancel out and (6) applies.

Treating haploids or gametes as conceptual diploids

Haploid parents produce identical gametes. The only type of diploid organism that produces identical gametes is one that is fully homozygous (100% inbred). Therefore, conceptually haploids may be treated as homozygous diploids which are derived from a single parent. Similarly, we may visualize gametes as homozygous offspring of a single parent which in turn transfer their genes to produce a normal diploid offspring. The difference between haploids and gametes in this interpretation is that haploids may produce many offspring whereas a gamete has only a single offspring.

Rules (1) and (2) for building $\mathbf{A}$ apply for such individuals but may be written in the following special form:

$$a_{ii} = 2 \tag{8}$$

$$a_{ji} = a_{jp} \tag{9}$$

as we assume homozygosity ($F_i = 1$) and that individual i is derived from a single parent ($p = q$). In a similar manner, (7) is modified into

$$d^{ii} = 1/(2 - a_{pp}) \tag{10}$$

and for building $\mathbf{A}^{-1}$, add

$$d^{ii} \text{ to } a^{ii}, a^{pp} \text{ and} \tag{11}$$

$$-d^{ii} \text{ to } a^{ip}, a^{pi}. \tag{12}$$

Rules (8) to (12) should only be invoked when dealing with homozygous individuals. It should be noticed, and will be seen from an example later on, that inclusion of gametes in the proposed way does not alter the relationship structure between 'real' individuals.

Equivalence of the model including 'conceptual diploids' to the gametic model: an example

The genetic model

Schaeffer et al. (1989) present a model for data influenced by maternal gametic imprinting, i.e., where some genes are only expressed when received from the maternal gamete. For an extensive review and discussion of the phenomenon of parental imprinting see Solter (1988). They write the model for an observation y on the i^{th} individual [assuming that the only fixed effect is the mean (μ)] as

$$y_i = \mu + a_i + g_i^m + e_i \tag{13}$$

where a_i represents the additive genetic value inherited and expressed in the usual manner, g_i^m represents expression of additional additive genetic value inherited through the maternal gamete and e_i is the residual effect. The authors stress the fact that the genotype of the individual also includes the genetic effect of the paternal gamete (g_i^p), which is not expressed in the phenotype. The paternal genes may be expressed in later generations if they are subsequently transmitted through a female and, therefore, may not be omitted from analysis.

The additive + gametic model

Schaeffer et al. (1989), following Gibson et al. (1988), write their statistical model in matrix notation as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wg} + \mathbf{e} \tag{14}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{a}$ is the random vector of additive genetic effects, $\mathbf{g}$ is the random vector of additional genetic effects that are only expressed when derived from the maternal gamete, and $\mathbf{e}$ is the random vector of residuals. $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ are incidence matrices and the variance-covariance structure of the random variables assumed by Schaeffer et al. is

$$\mathbf{V}\begin{bmatrix}\mathbf{a}\\ \mathbf{g}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_a^2 & 0 & 0\\ 0 & \mathbf{G}\sigma_g^2 & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix} \tag{15}$$

where σ_a^2, σ_g^2 and σ_e^2 are the additive genetic, maternally imprinted genetic, and residual variances. Note that the gametic relationship matrix $\mathbf{G}$ includes both paternal and maternal gametes of an individual but the incidence matrix $\mathbf{W}$ links the observations only to the maternal gamete. Defining $\alpha_a = \sigma_e^2/\sigma_a^2$ and $\alpha_g = \sigma_e^2/\sigma_g^2$, the mixed model equations from (14) are

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\alpha_a & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W}+\mathbf{G}^{-1}\alpha_g \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix}. \tag{16}$$

The conceptual diploid model

An equivalent model to describe genetic model (13) may be written as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Hi} + \mathbf{e} \tag{17}$$

where **y**, **b**, **a**, **e**, **X** and **Z** are defined as before and **i** is the vector of maternal imprinted effects defined on an individual rather than gametic level with corresponding incidence matrix **H**. The variance-covariance matrix of random effects is

$$\mathbf{V}\begin{bmatrix} \mathbf{a} \\ \mathbf{i} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 & 0 \\ 0 & \mathbf{A}^*\sigma_i^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix} \tag{18}$$

where σ_i^2 is the imprinted variance as observed on the individual level. $\mathbf{A}^*$ is the relationship matrix between all individuals included in **A** augmented by the maternal gametes of individuals with records, treating those gametes as diploids as shown in Sect. 3. The mixed model equations for this model (with $\alpha_i = \sigma_e^2/\sigma_i^2$) are

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'H} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\alpha_a & \mathbf{Z'H} \\ \mathbf{H'X} & \mathbf{H'Z} & \mathbf{H'H}+\mathbf{A}^{*-1}\alpha_i \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{i}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{H'y} \end{bmatrix}. \tag{19}$$

Equivalence of models

Smith and Allaire (1985) describe and Jamrozik and Schaeffer (1991) show that **A** may be obtained from **G** as $\mathbf{A} = 0.5\mathbf{KGK}'$ with $\mathbf{K} = \mathbf{I}_N{*}\ [1\ 1]$, where N is the number of individuals and * denotes the direct (Kronecker) product of two matrices. $\mathbf{A}^*$ in (19) may be derived in a similar fashion as $\mathbf{A}^* = 0.5\mathbf{QGQ}'$. A row of **Q** pertaining to an animal consists of two elements with a value of unity relating the individual to its two parental gametes and zeroes elsewhere (as in **K**), whereas a row pertaining to a gamete treated as conceptual diploid consists of one element of value two relating the gamete to itself and zeroes elsewhere.

Because **i** is defined on the individual rather than the gametic level the relationship between **i** and **g** is given by

$$\mathbf{i} = 0.5\mathbf{Qg}. \tag{20}$$

Now

$$\begin{aligned} \mathbf{V(i)} &= \mathbf{V}(0.5\mathbf{Qg}) \\ &= 0.25\mathbf{QV(g)Q}' \\ &= 0.25\mathbf{QGQ}'\sigma_g^2 \\ &= 0.5\mathbf{A}^*\sigma_g^2. \end{aligned} \tag{21}$$

From (18) and (21)

$$\mathbf{V(i)} = \mathbf{A}^*\sigma_i^2 = 0.5\mathbf{A}^*\sigma_g^2. \tag{22}$$

Therefore

$$\sigma_i^2 = 0.5\sigma_g^2, \tag{23}$$

and

$$\alpha_i = \sigma_e^2/\sigma_i^2 = 2\sigma_e^2/\sigma_g^2 = 2\alpha_g. \tag{24}$$

The models are equivalent in the sense of Henderson (1985) in that both yield identical first and second moments of the data vector.

Numerical example

To demonstrate their method, Schaeffer et al. (1989) give a small numerical example including a total of six animals and three animals with records (Table 1). We shall use this example to demonstrate the equivalence of models (14) and (17). Matrices **A** and **G** and their inverses are given in the paper by Schaeffer et al. Tables 2 and 3 show relationship matrix $\mathbf{A}^*$ and its inverse where $\overrightarrow{AD}$, $\overrightarrow{DE}$ and $\overrightarrow{DF}$ denote 'diploid' maternal gametes transferring genes from mother to daughter.

The incidence matrices **Z**, **W** and **H** are

$$\mathbf{Z} = \left[\begin{array}{cccccc} A & B & C & D & E & F \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\mathbf{W} = \left[\begin{array}{cccccccccccc} A_1 & A_2 & B_1 & B_2 & C_1 & C_2 & D_1 & D_2 & E_1 & E_2 & F_1 & F_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\mathbf{H} = \left[\begin{array}{ccccccccc} A & B & \overrightarrow{BD} & C & D & \overrightarrow{DE} & \overrightarrow{DF} & E & F \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right].$$

where subscripts 1 and 2 in **W** relate to the paternal and maternal gametes of the animals. Schaeffer et al. (1989) assume that the heritability of the trait is 0.30 and that the maternally imprinted effects account for 5% of the total variation, so that $\alpha_a = 2.1667$, $\alpha_g = 13.0$ and [from (24)] $\alpha_i = 26.0$.

Table 1. Example data for maternal imprinting analysis (from Schaeffer et al. 1989)

Animal	Sire	Dam	Milk yield (kg)
D	*A*	*B*	6500
E	*A*	*D*	6000
F	*C*	*D*	5500

The solutions for μ, **a** and **g** from the additive + gametic model (14) are

$$\begin{bmatrix}\hat{\mu}\\ \hat{a}_A\\ \hat{a}_B\\ \hat{a}_C\\ \hat{a}_D\\ \hat{a}_E\\ \hat{a}_F\end{bmatrix}=\begin{bmatrix}5978.8\\ 39.7\\ 47.7\\ -87.4\\ 91.5\\ 57.5\\ -85.4\end{bmatrix} \text{ and: } \begin{bmatrix}\hat{g}_{A_1}\\ \hat{g}_{A_2}\\ \hat{g}_{B_1}\\ \hat{g}_{B_2}\\ \hat{g}_{C_1}\\ \hat{g}_{C_2}\\ \hat{g}_{D_1}\\ \hat{g}_{D_2}\\ \hat{g}_{E_1}\\ \hat{g}_{E_2}\\ \hat{g}_{F_1}\\ \hat{g}_{F_2}\end{bmatrix}=\begin{bmatrix}-8.0\\ -8.0\\ 8.0\\ 8.0\\ 0.0\\ 0.0\\ -15.9\\ 15.9\\ -8.0\\ -1.3\\ 0.0\\ -14.6\end{bmatrix}$$

Solutions for μ, **a** and **i** from the conceptual diploid model (17) are

$$\begin{bmatrix}\hat{\mu}\\ \hat{a}_A\\ \hat{a}_B\\ \hat{a}_C\\ \hat{a}_D\\ \hat{a}_E\\ \hat{a}_F\end{bmatrix}=\begin{bmatrix}5978.8\\ 39.7\\ 47.7\\ -87.4\\ 91.5\\ 57.5\\ -85.4\end{bmatrix} \text{ and: } \begin{bmatrix}\hat{i}_A\\ \hat{i}_B\\ \hat{i}_{\overrightarrow{BD}}\\ \hat{i}_C\\ \hat{i}_D\\ \hat{i}_{\overrightarrow{DE}}\\ \hat{i}_{\overrightarrow{DF}}\\ \hat{i}_E\\ \hat{i}_F\end{bmatrix}=\begin{bmatrix}-8.0\\ 8.0\\ 15.9\\ 0.0\\ 0.0\\ -1.3\\ -14.6\\ -4.7\\ -7.3\end{bmatrix}$$

Both models give the same solutions for the fixed and the additive genetic effects. The solutions for maternal gametes D_2, E_2 and F_2 are identical to the solutions for $\overrightarrow{BD}$, $\overrightarrow{DE}$ and $\overrightarrow{DF}$. The solutions for imprinted effects (**i**) on animals are identical to the average of paternal and maternal gametic effects (**g**) for each animal.

Table 2. Relationship matrix for pedigrees in Table 1 including maternal gametes treated as being diploid

Parents:	– –	– –	*B B*	– –	*A* $\overrightarrow{BD}$	*D D*	*D D*	A $\overrightarrow{DE}$	*C* $\overrightarrow{DF}$
Animals:	*A*	*B*	$\overrightarrow{BD}$	*C*	*D*	$\overrightarrow{DE}$	$\overrightarrow{DF}$	*E*	*F*
A	1	0	0	0	0.5	0.5	0.5	0.75	0.75
B	0	1	1	0	0.5	0.5	0.5	0.25	0.25
$\overrightarrow{BD}$	0	1	2	0	1	1	1	0.5	0.5
C	0	0	0	1	0	0	0	0	0.5
D	0.5	0.5	1	0	1	1	1	0.75	0.5
$\overrightarrow{DE}$	0.5	0.5	1	0	1	2	1	1.25	0.5
$\overrightarrow{DF}$	0.5	0.5	1	0	1	1	2	0.75	1
E	0.75	0.25	0.5	0	0.75	1.25	0.75	1.25	0.375
F	0.25	0.25	0.5	0.5	0.5	0.5	1	0.375	1

Table 3. Inverse of the relationship matrix in Table 2

Parents:	– –	– –	*B B*	– –	*A* $\overrightarrow{BD}$	*D D*	*D D*	*A* $\overrightarrow{DE}$	*C* $\overrightarrow{DF}$
Animals:	*A*	*B*	$\overrightarrow{BD}$	*C*	*D*	$\overrightarrow{DE}$	$\overrightarrow{DF}$	*E*	*F*
A	3	0	1	0	−2	1	0	−2	0
B	0	2	−1	0	0	0	0	0	0
$\overrightarrow{BD}$	1	−1	2	0	−2	0	0	0	0
C	0	0	0	2	0	0	1	0	−2
D	−2	0	−2	0	6	−1	−1	0	0
$\overrightarrow{DE}$	1	0	0	0	−1	2	0	−2	0
$\overrightarrow{DF}$	0	0	0	1	−1	0	2	0	−2
E	−2	0	0	0	0	−2	0	4	0
F	0	0	0	−2	0	0	−2	0	4

The matrix **Q** to transform **g** into **i** (as shown in 20) is for this example

$$\mathbf{Q} = \left[\begin{array}{c|cccccccccccc}
 & A_1 & A_2 & B_1 & B_2 & C_1 & C_2 & D_1 & D_2 & E_1 & E_2 & F_1 & F_2 \\
A & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
B & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\overrightarrow{BD} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
D & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
\overrightarrow{DE} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
\overrightarrow{DF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}\right]$$

Note that with inclusion of maternal imprinted effects, although additive, the individual merit and the breeding value of an animal are no longer identical. The individual merit is calculated by summing the additive effect of the animal and the imprinting effect of its maternal gamete (e.g., 56.2 for animal E) whereas the breeding value may be calculated by summing the additive effect and the imprinting effect of the individual (52.8 for animal E).

Discussion

Using conceptual diploids in an animal model to analyze gametic variation is appealing for several reasons. For one, it is very easy to adapt the current computer programs based on the animal model (e.g., DFREML, Meyer 1991) to include gametes in the way proposed.

It is also easy to include covariances between additive and gametic effects (assumed to be 0 in the previous example), by including the same relationship structure (with 'diploid' maternal gametes) on the additive animal level. This is important as the assumption of no covariance between additive and maternal gametic effects is based more on convenience than on biological reality. In some situations it may be convenient from a computational point of view to set up the same relationship structure on the additive and maternal genetic level and use **A*** for both, even without including a covariance between the **a** and **i** effects. The solutions for **a** and **i** as given above do not change in this case, but additional solutions of $\hat{a}$ are obtained for $\overrightarrow{BD}$, $\overrightarrow{DE}$ and $\overrightarrow{DF}$, namely 95.5, 83.4 and 4.0 which are the additive values of the maternal gametes of animals D, E and F respectively. The concept presented here may well be used for the analysis of populations with haploid and diploid sexes, although in the honey bee there are additional complications of multiple mating that may be accommodated using the approach of Tier and Sölkner (private communication).

Acknowledgements. We are grateful for the financial support of the Meat Research Corporation (Australia) and the Fonds zur Förderung der wissenschaftlichen Forschung (Austria). We thank Dr. Keith Hammond and colleagues for useful comments.

References

Gibson JP, Kennedy BW, Schaeffer LR, Southwood OI (1988) Gametic models for estimation of autosomally inherited effects that are expressed only when received from either male or female parent (abstract). J Dairy Sci 71 (suppl 1):143

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

Henderson CR (1985) Equivalent models to reduce computations. J Dairy Sci 68:2267–2277

Jamrozik J, Schaeffer LR (1991) An equivalent gametic model for animal dominance genetic linear model. J Anim Breed Genet 108:343–348

Meyer K (1990) Present status of knowledge about statistical procedures available to estimate variance and covariance components. In: Proc 4th World Congr Genet Appl Livest Prod., Edinburgh, UK, vol. 13, 407–418

Meyer K (1991) DFREML Programs to estimate variance components by restricted maximum likelihood using a derivative-free algorithm. User notes Version 2.0. AGBU, UNE, Armidale, Australia

Schaeffer LR, Kennedy BW, Gibson JP (1989) The inverse of the gametic relationship matrix. J Dairy Sci 72:1266–1272

Smith SP (1984) Dominance relationship matrix and inverse for an inbred population. Mimeo, Dep Dairy Sci, Ohio State Univ, Columbus, USA

Smith SP, Allaire FR (1985) Efficient selection rules to increase non-linear merit: application to mate selection. Genet Sel Evol 17:387–406

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

Solter D (1988) Differential imprinting and expression of maternal and paternal genomes. Annu Rev Genet 22:127–146