# University of Nebraska - Lincoln [DigitalCommons@University of Nebraska - Lincoln](https://digitalcommons.unl.edu/)

[Faculty Papers and Publications in Animal](https://digitalcommons.unl.edu/animalscifacpub) 

**Animal [Science](https://digitalcommons.unl.edu/animalscifacpub) Department** 

December 1997

# Calculation of breed direct and maternal genetic fractions and breed specific direct and maternal heterozygosity for crossbreeding data

L. Dale Van Vleck University of Nebraska-Lincoln, dvan-vleck1@unl.edu

Follow this and additional works at: [https://digitalcommons.unl.edu/animalscifacpub](https://digitalcommons.unl.edu/animalscifacpub?utm_source=digitalcommons.unl.edu%2Fanimalscifacpub%2F312&utm_medium=PDF&utm_campaign=PDFCoverPages)

**Part of the Animal Sciences Commons** 

Van Vleck, L. Dale, "Calculation of breed direct and maternal genetic fractions and breed specific direct and maternal heterozygosity for crossbreeding data" (1997). Faculty Papers and Publications in Animal Science. 312.

[https://digitalcommons.unl.edu/animalscifacpub/312](https://digitalcommons.unl.edu/animalscifacpub/312?utm_source=digitalcommons.unl.edu%2Fanimalscifacpub%2F312&utm_medium=PDF&utm_campaign=PDFCoverPages)

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



# **Brazilian Journal of Genetics**

# **Print ISSN 0100-8455**

# **Braz. J. Genet. vol. 20 no. 4 Ribeirão Preto Dec. 1997**

**doi: 10.1590/S0100-84551997000400013** 

# **Calculation of breed direct and maternal genetic fractions and breed specific direct and maternal heterozygosity for crossbreeding data**

*L.D. Van Vleck*

Roman L. Hruska U.S. Meat Animal Research Center. A218 Animal Sciences, University of Nebraska, Lincoln, NE 68583-0908, USA. Phone: 402/472-6010. Fax: 402/472-6362. E.mail: ansc418@unlvm.unl.edu.

#### **ABSTRACT**

Teaching, research, and herd breeding applications may require calculation of breed additive contributions for direct and maternal

genetic effects and fractions of heterozygosity associated with breed specific direct and maternal heterosis effects. These coefficients can be obtained from the first NB rows of a pseudo numerator relationship matrix where the first NB rows represent fractional contributions by breed to each animal or group representing a specific breed cross. The table begins with an *NB* x *NB* identity matrix representing pure breeds. Initial animals or representative crosses must be purebreds or two-breed crosses. Parents of initial purebreds are represented by the corresponding column and initial two-breed cross progeny by the two corresponding columns of the identity matrix. After that, usual rules are used to calculate the *NB* column entries corresponding to breeds for each animal. The *NB* entries are fractions of genes expected to be contributed by each of the pure breeds and correspond to the breed additive direct fractions. Entries in the column corresponding to the dam represent breed additive maternal fractions. Breed specific direct heterozygosity coefficients are entries of an *NB* x *NB* matrix formed by the outer product of the two *NB* by *1* columns associated with sire and dam of the animal. One minus sum of the diagonals represents total direct heterozygosity. Similarly, the *NB* x *NB* matrix formed by the outer product of columns associated with sire of dam and dam of dam contains breed specific maternal heterozygosity coefficients. These steps can be programmed to create covariates to merge with data. If **X** represents these coefficients for all unique breed crosses, then the reduced row echelon form function of MATLAB or SAS can be used on **X** to determine estimable functions of additive breed direct and maternal effects and breed specific direct and maternal heterosis effects.

## **INTRODUCTION**

Several methods are available to model breed additive and interaction effects for records of crossbred animals. With designed experiments, the coefficients are the same for groups by generation. In the more general cases of composites or unstructured breeding plans, calculation of the coefficients is time consuming at best. The purpose of this note is to outline computational procedures to simplify calculation of fractions of inheritance from ancestral breeds as well as fractions of breed specific heterozygosity for an animal and its dam. An additional section will outline a simple way to determine what functions of breed and heterozygosity effects are estimable from statistical solutions.

## **MATERIAL AND METHODS**

Breed effects can be modelled as breed combinations with linear contrasts used to separate breed effects and breed interactions (e.g., Dickerson, 1969; Wyatt and Franke, 1986). Breed effects can also be modelled as covariates with fractional contributions (e.g., Robison *et al*.,

1981). In that case, Westells rules for genetic groups can be used equivalently with breeds corresponding to the proxy parents (Westell *et al*., 1984, 1988). Breeding values from using Westells rules incorporate breed solutions weighted automatically by their fractional contributions to the animal plus a genetic deviation from that function. With Westell groups effects due to heterozygosity would need to be modelled separately, probably as covariates.

The equivalent model with breed effects as covariates would result in solutions for regression coefficients and for random genetic deviations. Predicted breeding values would then be constructed from linear functions of the estimated regression coefficients for breeds with weighting by fraction of inheritance by breed plus the predicted random genetic deviation for the animal. The additional effects of breed crosses are usually modelled as interaction effects which are equivalent to expected heterozygosity. Heterozygosity effects can be expressed as general (the breeds in the cross all interact equally) or specific (each breed cross has a different effect). Heterozygosity effects can be modelled as covariates in either case (or in intermediate cases where, for example, continental by continental crosses have the same effect, continental by English breeds may have a different effect, and English by indicus breeds may have still a different effect). In this note calculation of general and breed specific heterozygosity coefficients will be described with special cases requiring obvious modifications.

The methods for calculating breed fractions and heterozygosity fractions will be described with a simple example which will illustrate the concepts. Sketches of programming steps to do the calculations will also be given.

Pedigree files will be assumed available tracing back to the first crosses between breeds or even further. At some point the parents will be assumed unknown but the breeds of the parents will be known. Animals and parents will have unique identification with parent identification numbers smaller than their progeny (necessary only for use of standard methods of calculating inverse of numerator relationship matrix).

The example will be for three animals with records:

Animal 60 with sire 40 and dam 50 Animal 70 with sire 60 and dam 50 Animal 80 with sire 60 and dam 70

Sire 40 is breed A. Dam 50 is a cross of breeds B and C. (The procedure requires originating parents be a purebred or a two-breed cross).

The method will involve calculation of a pseudo tabular relationship matrix (Van Vleck, 1993). To facilitate that calculation, the original identification numbers need to be recoded to begin with 1. In fact, the first breed must be recoded to 1 followed by other breeds and then animals in identification order.

### **Recoding identification numbers**

The MTDFNRM program in the MTDFREML package of programs for estimating variance components (Boldman *et al*., 1995) can be used for the recoding. Assume the pedigree file is as follows:

i.e., identification numbers for animals 60, 70, 80 with sire and dam identification numbers. These fields can be in the data file or in a separate pedigree file. The MTDFNRM program would request the fields for animal, sire and dam (1, 2, 3 in this example) and also the number of groups (breeds) which would at this stage be answered as zero. The program would also ask if pedigree file with original and recoded identification is to be output (yes, and file MTDF13 is written). The output also includes the inbreeding coefficients of the animal, sire and dam which also might be added to the data file for analysis. The resulting file MTDF13 for the example is:

5



#### 5 3 4 80 60 70 .375 .000 .250

The number of animals (5) is in the first line of MTDF13 and that line must be deleted. Then numbers corresponding to the breeds of sire and dam of the originating animals must be substituted for the zeroes in fields 5 and 6 of MTDF13. The file then can be saved to another name, e.g., MTD13.PED, which for the example is:



MTD13.PED can be used as the pedigree file either to use Westell grouping for breed fractions or to compute breed fractions for use as covariates and for calculation of heterozygosity coefficients. Then MTDFNRM would be run using MTDF13.PED as the pedigree file with positions 4, 5, and 6 for animal, sire and dam fields and with 3 groups specified. Again, the recoding option is requested to match later with predicted breeding values or to be the basis for calculating breed fractions and heterozygosity coefficients. The output will again be called MTD13 but can be saved by copying to MTDF13.PDD. Note that the breeds are now recoded as "animals" 1, 2, and 3 with" unknown" parents.

If heterozygosity and inbreeding are not important, then the MTDFPREP and MTDFRUN programs of the MTDFREML package can be run to estimate variance components or to predict breeding values containing the appropriate fractions of the group (breed) effects. The solutions for "animals" 1, 2, and 3 will be the solutions for the breed effects.

If heterozygosity and inbreeding coefficients are needed (and also breed fractions to use as covariates) then the MTD13.PDD file can be used for

those calculations and for matching with the data file. For the example MTD13.PDD is:



The number of animals, on the first line of MTDF13.PDD, is now 8 because of the 3 groups and can be deleted depending on how the file is to be used. Probably it will be used with a program that will read the lines of MTD13.PDD to use for calculation of covariates. The basic fields to read will be the recoded fields 1, 2, and 3 plus field 4 (original animal ID) which will be used later for matching with the data file. If individual and maternal inbreeding are to be added to the data file, then fields 7 and 9 would also be read. These 4 (or 6) fields would be stored as vectors. The first 3 vectors will be used to calculate breed fractions for all recoded animals. After those calculations, field 4 will be used for

matching and fields 7 and 9 could be used for adding animal and dam inbreeding coefficients to the matched data file.

## **Calculation of breed fractions**

The usual rules for the tabular method (e.g., Van Vleck, 1993) will be used but only for rows corresponding to the breeds. Thus, the dimensions of the matrix of breed fractions will be the number of breeds by number of animals plus number of breeds.

The starting point of the program will be to put a 1 in the appropriate column for each of the breeds beginning for breed 1, a 1 in row 1; for breed 2, a 1 in row 2; and for breed 3, a 1 in row 3 as shown for the example. The numbers shown above the entries for the table are the recoded fields 1, 2, and 3 with entries from fields 2 and 3 identifying sires and dams (or breeds of sires and dams).

The usual rules for the tabular method will be used to add to each row of the column for each animal, 1/2 of the row entry for the sire plus 1/2 of the row entry for the dam beginning with the first animal beyond the columns for the breeds.

Let IA, IS, ID be the vectors of recoded identification numbers and NA and NB be the number of animals and breeds (3 in the example) and A(NB, NA + NB) be the matrix that will contain the breed fractions for each animal.

The computational steps are:

 $DO I = 1, NB$ 

 $A(I, I) = 1.$ 

ENDDO

```
DO J = NB + 1, NA + NB
```

```
JS = IS(J)
```

```
JD = ID(J)
```
 $DO K = 1, NB$ 

 $A(K, J) = (A(K, JS) + A(K, JD))/2.$ 

ENDDO

ENDDO

For the example on completion, A (3, 8) will be with headings added to

demonstrate the format of the matrix A:



The recoded IA, IS, and ID vectors are shown above A(NB, NA + NB). After this step, these fractions can be matched to the data file and used to calculate heterozygosity fractions. In fact, the vector IA is not needed as it is implied to be 1, 2, 3...,  $NA + NB$ . To demonstrate the matching step, let IO be the vector of original IDs read from field 4 of MTDF13.PDD and stored as ID. The steps for pulling out breed fractions for an animal and its dam (also their inbreeding coefficients) lead naturally to calculation of heterozygosity coefficients after the match is made.

Let a data line for an animal be read. Suppose the animal ID is 60. IO is (1, 2, 3, 40, 50, 60, 70, 80). Animal ID is used to search IO and at position 6 the match is made. Then A  $(1,6; 2,6; 3,6) = (.50, .25, .25)$ are the breed fractions for animal 60. The ID vector for position 6 contains the recoded dam ID for animal 60 which is 5. Thus, A(1,5; 2,5;  $3,5$ ) =  $(.00, .50, .50)$  are the breed fractions for the mother of animal 60. These two vectors, each of order number of breeds, can be added to the data line to be used as covariates. The inbreeding coefficients, if wanted, will come from AF  $(6) = 0$ , and DF  $(5) = 0$  with AF and DF the vectors corresponding to fields 7 and 9 of MTDF13.PDD.

## **Calculation of heterozygosity fractions**

Calculation of heterozygosity fractions will be illustrated for two cases: general and specific heterozygosity.

General heterozygosity for an animal which ignores breed specific heterozygosity is where  $f_S$  is the vector (column of A) corresponding to

sire of the animal and  $f<sub>D</sub>$  is the column corresponding to the dam of the animal (see Dickerson, 1969, 1973). For maternal heterozygosity the columns will be fsp and fpp corresponding to the sire of the dam and dam of the dam.

For example, suppose the data line for animal 70 is read and matched to position 7 of IO. Then  $IS(7) = 6$  and  $ID(7) = 5$  so that the columns 6 and 5 of A can be pulled out as  $f_6 = (8$  frac12;  $8$  frac14;  $&frac(14)$  and  $f_5 = (0 &&frac(12) &&frac(11)$ . Then the general heterozygosity for animal 70 is  $1 - (8$ frac12;) (0) -  $(8$ frac14;)  $(kfrac12) - (8frac14)$  (¼) (½) = ¾. The products to subtract from one can be calculated by summing the products of columns of A(.,6) and A(.,5). The general maternal heterozygosity fraction is calculated by finding  $ID(7) = 5$  for the dam of animal 70 and then finding the sire of recoded 5 from  $IS(5) = 2$  and dam of recoded 5 from  $ID(5) = 3$  so that  $f_2 = (0, 1, 0)$  and  $f_3 = (0, 0, 1)$  and general heterozygosity for the dam is  $1 - (0) (0) - (1) (0) - (0) (1) = 1$  as expected because the parents of recoded dam 5 are breeds 2 and 3.

Fields for the animal and maternal heterozygosity with values of 3/4 and 1 will be added to the data file, if general heterozygosity is being used in the statistical analysis.

Computation of specific heterozygosity requires calculation of the outer product of the same vectors that could be used for calculation of general heterozygosity. In symbolic terms let . The entries will correspond to fractions of genes from the same and from different breeds. For the example

$$
H = \begin{pmatrix} A, A & A, B & A, C \\ B, A & B, B & B, C \\ C, A & C, B & C, C \end{pmatrix}
$$

where diagonals correspond to the coefficients found for the calculation that was subtracted from 1 for general heterozygosity. The off-diagonal coefficients correspond to the fractions of specific heterozygosity. The corresponding symmetrical terms need to be added, i.e., heterozygosity fraction for A by C is  $H(A,C) + H(C,A)$ . For the example with animal 70 and recoded sire 6 and dam 5:

$$
f6f5 = \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix} \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0 & 1/4 & 1/4 \\ 0 & 1/8 & 1/8 \\ 0 & 1/8 & 1/8 \end{bmatrix}
$$

Thus, specific heterozygosity fractions for animal 70 are for A  $\times$  B, 1/4 + 0; for A x C,  $1/4 + 0$ , and for B x C,  $1/8 + 1/8$ .

For specific maternal heterozygosity

$$
f2f3 - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} [0 \quad 0 \quad 1) - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}
$$

so that specific maternal heterozygosity fractions are for A  $\times$  B, 0 + 0; A  $x C$ ,  $0 + 0$ ; and for B x C,  $1 + 0$ .

Program steps to calculate H and then the vector (HET) of specific heterozygosity values are shown next with JS and JD the recoded identification numbers of the sire and dam of the animal.

```
DO J = 1, NBX = A(J, JS)DO K = 1, NB
        H(J,K) = X * A(K, JD)ENDDO
ENDDO
IC = ODO I = 1, NB-1DO J = I + 1, NBIC = IC + 1HET(IC) = H(I,J) + H(J,I)ENDDO
```
## ENDDO

The vector, HET, can then be added to the data line for that animal.

The new data line with additions of animal and dam breed fractions, inbreeding coefficients, and either general or specific heterozygosity fractions for animal and dam can now be written for that animal.

## **Summary of Calculations**

1. Read pedigree file (animal; sire and dam if known).

2. Run MTDFNRM with no group option; enter numbers in the output file, MTDF13, to represent

breeds of sire and dam for foundation animals.

3. Run MTDFNRM with modified pedigree file from 2 with group option.

4. Read recoded animal, sire, dam, FA, FD and original animal identification vectors from 3.

5. Calculate matrix of pseudo-relationships of animals to foundation breeds  $(A_{NB \times (NA + NB)})$ .

6. Read data file, record of one animal at a time

a) Match animal ID from data file with position in original animal vector, IO

- b) Insert animal column of A in animals record
- c) Insert dam column of A in animals record
- d) Insert animal and dam inbreeding coefficients from FA and FD vectors
- e) Calculate and insert heterozygosity fractions
- i) Animal from columns for sire and dam of animal
- ii) Dam from columns for sire of dam and dam of dam
- f) Write modified record for animal
- 7. Repeat 6 until all records have been modified.

8. Use MTDFPREP and MTDFRUN (or other packages) for analysis with animal and dam breed fractions, animal and dam inbreeding coefficients, and animal and dam specific heterozygosity fractions as covariates.

## **Estimability of breed and heterozygosity effects**

Even with well designed crossbreeding experiments estimability of breed additive and specific heterozygosity (usually denoted as heterosis) effects is difficult to determine. With less systematic crossbreeding, determination of estimable functions is even more difficult. Fortunately, an easy way exists to determine estimable functions.

What is needed are the coefficients of the model matrix for breed and heterozygosity fractions for animals representing each unique combination of coefficients.

For example, if there are several foundation animals of a breed only coefficients for one are needed. Similarly, if there are several animals of a breed B by breed C cross, coefficients for only one of the animals are needed, but be sure these coefficients are not completely confounded with other fixed factors such as year or generation.

If other fixed effects are confounded with these coefficients then the model coefficients for each unique combination of genetic and other fixed factors would be needed. For the following example, such confounding will be assumed not to exist.

Assume animals 40, 50, 60, 70, 80 have records. With individual breed effects (a<sub>A</sub>, a<sub>B</sub>, a<sub>C</sub>), maternal breed effects (m<sub>A</sub>, m<sub>B</sub>, m<sub>C</sub>), individual general heterosis effect  $(h_a)$ , and maternal heterosis  $(h_m)$  the model matrix **X** is:

Obviously, with only five records and eight genetic effects, not all effects are estimable even in the absence of other fixed effects but this example can be used for illustration.

One approach is to use the expectations option in MTDFRUN to obtain the expected values of solutions



for the genetic effects in terms of the genetic parameters and other fixed effects. In that case, the program would be allowed to determine the necessary constraints. The expectation option describes the estimable functions.

The approach described here is to use the ECHELON function from SAS (1985) or the reduced row echelon form (RREF) function from MATLAB (1993). This procedure has been discussed by Eiswick Jr. *et al*. (1991).



With MATLAB, the command RREF(X) results in:

Which are coefficients of ( $a_A$  a<sub>B</sub> a<sub>C</sub> m<sub>A</sub> m<sub>B</sub> m<sub>C</sub> h<sub>a</sub> h<sub>m</sub>).

Any combination of these five functions is estimable. The only" clean" functions are:

 $m_A$  - m<sub>C</sub> from line 3 and h<sub>a</sub> from line 5.

## **RESUMO**

O ensino, a pesquisa e a criação de gado podem requerer o cálculo de contribuições aditivas de raça para os efeitos genéticos maternal e

direto e frações de heterozigosidade associadas com efeitos de heterose maternal e direta específicos para raça. Estes coeficientes podem ser obtidos a partir das primeiras linhas NB de uma pseudomatriz de parentesco, onde as primeiras linhas NB representam contribuições fracionais por raça a cada animal ou grupo representando um cruzamento reprodutor específico. A tabela se inicia com uma matriz de identidade NB x NB representando raças puras. Os animais ou cruzamentos representativos iniciais devem ser de raça pura ou cruzamentos entre duas raças. Os pais dos indivíduos de raça pura iniciais são representados pela coluna correspondente e progenie inicial dos cruzamentos entre duas raças pelas duas colunas correspondentes na matriz de identidade. A seguir, usam-se regras comuns para calcular as entradas das colunas NB correspondentes as raças para cada animal. As entradas NB seriam frações de genes que esperam serem contribuídos por cada uma das raças puras e correspondem às frações diretas aditivas de raças. As entradas na coluna correspondente à mãe representam frações maternais aditivas de raça. Os coeficientes de heterozigosidade direta específicos de raça são entradas de uma matriz NB x NB formadas pelo produto externo das duas NB por colunas associadas com o pai e a mãe do animal. Um menos a soma das diagonais representa a heterozigosidade direta total. De modo similar, a matriz NB x NB formada pelo produto externo das colunas associadas com o pai da mãe e a mãe da mãe contém os coeficientes de heterozigosidade maternos específicos para raça. Estes passos podem ser programados para criar covariantes para serem juntadas aos dados. Se X representa esses coeficientes para todos os cruzamentos únicos, então a "reduced row echelon form function" de MATLAB ou SAB pode ser usada em X para determinar as funções estimáveis dos efeitos maternal e direto de raça aditiva e os efeitos heteróticos de raça direto e maternal específicos.

#### **REFERENCES**

**Boldman, K.G., Kriese, L.A., Van Vleck, L.D., Van Tassell, C.P.** and **Kachman, S.D.** (1995). A manual for the use of MTDFREML. A set of programs to obtain estimates of variances and covariances (Draft). USDA-ARS-MARC, Clay Center, NE. [Links ]

**Dickerson, G.E.** (1969). Experimental approaches in utilizing breed resources. *Anim. Breed. Abstr. 37*: 191-202. [ Links ]

**Dickerson, G.E.** (1973). Inbreeding and heterosis in animals. In: *Proceedings of the Animal Breeding and Genetics Symposium in honor of Dr. Jay L. Lush*. ASAS, ADSA, PSA, Champaign, IL. pp. 54-77. [ Links ]

**Eiswick Jr., R.K., Gennings, C., Chinchilli, V.M.** and **Dawson, K.S.** (1991). A simple approach for finding estimable functions in linear models. Am. Stat. 45: 51-53. [Links ]

**MATLAB** (1993). *MATLAB* (*Version 4.0*). The MathWorks Inc., Natick, MA. [ Links ]

**Robison, O.W., McDaniel, B.T.** and **Rincon, E.J.** (1981). Estimation of direct and maternal additive and heterosis effects from crossbreeding experiments in animals. *J. Anim. Sci. 52*: 44-50. [Links ]

**SAS Institute Inc.** (1985). *SAS/IML Users Guide* (*Version 5*). Cary, NC. [ Links ]

**Van Vleck, L.D.** (1993). *Selection Index and Introduction to Mixed Model Methods*. CRC Press Inc., Boca Raton, FL, pp. 319-333. [ Links ]

**Westell, R.A., Quaas, R.L.** and **Van Vleck, L.D.** (1984). Genetic groups in an animal model. *J. Anim. Sci. 59* (Suppl. 1): 175 (Abstract). [ Links ]

**Westell, R.A., Quaas, R.L.** and **Van Vleck, L.D.** (1988). Genetic groups in an animal model. *J. Dairy Sci. 71*: 1310- 1318. [Links ]

**Wyatt, W.E.** and **Franke, D.E.** (1986). Estimation of direct and maternal additive and heterotic effects for preweaning growth traits in cattle breeds represented in the southern region. *Southern Cooperative Series, Bulletin 310, pp. 1-35.* [ Links ]

## **(Received May 26, 1997)**

© *2008 Sociedade Brasileira de Genética*

Rua Capitão Adelmio Norberto da Silva, 736 14025-670 Ribeirão Preto SP - Brazil Tel. / Fax: +55 16 621.8540, 620.1251, 620.1253

e-Mail sede@sgb.org.br