

# Breeding Value and Variance Component Estimation From Data Containing Inbred Individuals: Application to Gynogenetic Families in Common Carp (*Cyprinus carpio* L.)

Piter Bijma, Johan A. M. Van Arendonk and Henk Bovenhuis

Department of Animal Breeding, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, Wageningen, The Netherlands

Manuscript received July 24, 1996

Accepted for publication January 16, 1997

## ABSTRACT

Under gynogenetic reproduction, offspring receive genes only from their dams and completely homozygous offspring are produced within one generation. When gynogenetic reproduction is applied to fully inbred individuals, homozygous clone lines are produced. A mixed model method was developed for breeding value and variance component estimation in gynogenetic families, which requires the inverse of the numerator relationship matrix. A general method for creating the inverse for a population with unusual relationships between animals is presented, which reduces to simple rules as is illustrated for gynogenetic populations. The presence of clones in gynogenetic populations causes singularity of the numerator relationship matrix. However, clones can be regarded as repeated observations of the same genotype, which can be accommodated by modifying the incidence matrix, and by considering only unique genotypes in the estimation procedure. Optimum gynogenetic sib family sizes for estimating heritabilities and estimates of their accuracy were derived and compared to those for conventional full-sib designs. This was done by means of a deterministic derivation and by stochastic simulation using Gibbs sampling. Optimum family sizes were smallest for gynogenetic families. Only for low heritabilities, there was a small advantage in accuracy under the gynogenetic design.

**I**N common carp, *Cyprinus carpio*, gynogenetic offspring can be produced by using irradiated sperm and by preventing cytoplasmic partition (KOMEN *et al.* 1991). Under gynogenetic reproduction, offspring receive genes only from their dams; paternal genes are not transmitted. The chromosomes in the maternal gametes are doubled, which results in 100% homozygous diploid offspring. However, within a group of gynogenetic offspring of a noninbred dam, there is still genetic variation (BONGERS *et al.* 1996). This is because the gametes produced by the dam differ due to Mendelian sampling. In this paper, a group of gynogenetic offspring of a non inbred dam is referred to as a gynogenetic sib family.

Gynogenetic reproduction in a female which is homozygous, *e.g.*, a female from a gynogenetic sib family, produces homozygous clone lines. All individuals in a homozygous clone line are genetically identical, because all gametes produced by a homozygous individual are identical. In this respect, gynogenesis is similar to double haploidy in plants. The preferred methods for estimating breeding values and variance components are based on mixed model equations (MEYER 1989). For these methods, the inverse of the numerator relationship matrix is required, which can be created using

the rules of HENDERSON (1976) and QUAAS (1976). However, under gynogenetic reproduction, these rules are no longer valid because of differences in transmission of genes from parents to offspring. When gynogenetic reproduction produces homozygous clone lines, the numerator relationship matrix will contain identical rows and columns. This means that the inverse of the numerator relationship matrix does not exist, and modifications are needed to solve the mixed model equations.

In this paper, effects of gynogenetic reproduction on the distribution of additive genetic variance within and between gynogenetic sib families are described. Subsequently, a general method for creating the inverse of the numerator relationship matrix under gynogenetic reproduction is described. To illustrate this method, an example for a small gynogenetic family is presented. Finally, standard errors of variance component estimates for gynogenetic sib family structures are compared to those for conventional full-sib designs.

## MATERIALS AND METHODS

**Distribution of variance:** Gynogenetic reproduction affects the distribution of additive genetic variance within and between gynogenetic sib families. In a population divided into families of equal size, the between- and within-family variance as a proportion of the variance in the original noninbred population can be calculated using the following (FALCONER 1989, p. 266):

Corresponding author: Piter Bijma, Department of Animal Breeding, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands. E-mail: piter.bijma@alg.vf.wau.nl

$$V_b = 2fV_a, \quad (1)$$

$$V_w = (1 + F - 2f)V_a, \quad (2)$$

where  $V_a$  is the additive genetic variance in the original, non-inbred population;  $V_b$  is the additive genetic variance between families;  $V_w$  is the additive genetic variance within families;  $F$  is the inbreeding coefficient of the members of a family; and  $f$  is the coefficient of coancestry between the members of a family, which is defined as the chance that a randomly chosen allele in an individual is identical by descent to a randomly chosen allele from the same locus in the other individual (FALCONER 1989 p. 89). The coefficient of coancestry is equal to half the coefficient of relationship between individuals. The total additive genetic variance in the population is  $V_t = V_b + V_w$ .

Consider a population consisting of homogeneous gynogenetic sib families descending from non inbred dams. For homozygous individuals, like gynogenetic sibs, the inbreeding coefficient  $F$  is equal to one. For diploids, the coefficient of coancestry between gynogenetic sibs is equal to 0.5, because there can be only two different alleles at any locus within a group of offspring from the same dam (BONGERS *et al.* 1996). Using  $F = 1$  and  $f = 0.5$ , it follows from (1) and (2) that both  $V_b$  and  $V_w$  are equal to the additive genetic variance in the noninbred dam population. This means that the total additive genetic variance in a population consisting of gynogenetic sib families is twice the additive genetic variance in the noninbred dam population (BONGERS *et al.* 1996).

An understanding of the mechanism that causes this doubling of the genetic variance can also be gained by looking at the gametic level. Because gametic variance equals  $1/2 V_a$ , the total additive genetic variance in a population consisting of homogeneous gynogenetic families equals

$$\text{var}(a_0) = \text{var}(g_d + g_d) = 2 \text{var}(g_d) + 2 \text{cov}(g_d, g_d) = 2V_a,$$

where  $g_d$  is a maternal gamete. Between-family variance originates from variance among genetic values of parents, whereas within-family variance originates from Mendelian sampling. The contribution of between- and within-family variance can be understood by partitioning the gametic value into half the genetic value of the dam and a Mendelian sampling contribution:

$$\begin{aligned} \text{var}(a_0) &= \text{var}(2g_d) = 4 \text{var}(\frac{1}{2}a_d + M_d) \\ &= \text{var}(a_d) + 4 \text{var}(M_d) = V_a + V_w = 2V_a. \end{aligned}$$

Narrow-sense heritability ( $h^2$ ) is the ratio of additive genetic variance over phenotypic variance (FALCONER 1989 p. 126). Phenotypic variance is the sum of additive genetic variance and residual variance ( $V_e$ ). Assuming a constant residual variance, heritability in a population consisting of gynogenetic sib families is equal to  $2V_a/(2V_a + V_e)$ , which results in

$$h_{gs}^2 = 2h^2/(h^2 + 1), \quad (3)$$

where  $h^2 = V_a/(V_a + V_e)$ , which is heritability in the noninbred dam population.

**Mixed model equations:** Observations can be represented by the following mixed model:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ . To obtain estimates for  $\boldsymbol{\beta}$  and predictions for  $\mathbf{u}$ , the mixed model equations (KENNEDY and MOXLEY 1975; HENDERSON 1976) can be used

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \quad (4)$$

where  $\mathbf{y}$  is a vector of phenotypic observations;  $\mathbf{X}$  is a design matrix relating nongenetic fixed effects to observations;  $\boldsymbol{\beta}$  is a

vector of nongenetic fixed effects;  $\mathbf{Z}$  is a design matrix relating additive genetic effects to observations;  $\mathbf{u}$  is a vector of random additive genetic effects;  $\mathbf{A}^{-1}$  is the inverse of the numerator relationship matrix;  $\mathbf{e}$  is a vector of random residual effects and  $\alpha$  is the variance ratio  $\sigma_e^2/\sigma_u^2$ , where  $\sigma_e^2$  is the variance of the residual effects and  $\sigma_u^2$  is the additive genetic variance in the base population (KENNEDY and MOXLEY 1975; HENDERSON 1976). Calculating  $\mathbf{A}^{-1}$  by direct inversion of  $\mathbf{A}$  is difficult because of the large dimensions. HENDERSON (1976) and QUAAS (1976) derived rules for obtaining  $\mathbf{A}^{-1}$  directly without having to invert  $\mathbf{A}$  explicitly. Rules for setting up  $\mathbf{A}^{-1}$  directly can be derived using partitioned matrix theory by considering the effect of adding one additional row to  $\mathbf{A}$ . Addition of the  $i$ th row gives

$$\mathbf{A}_i^{-1} = \begin{bmatrix} \mathbf{A}_{i-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + (\mathbf{a}_{ii} - \mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s}_i)^{-1} \begin{bmatrix} \mathbf{s}_i\mathbf{s}_i' & -\mathbf{s}_i \\ -\mathbf{s}_i' & 1 \end{bmatrix} \quad (5)$$

(B. TIER and J. SÖLKNER, unpublished; VAN ARENDONK *et al.* 1994). In this formula,  $a_{ii}$  is the diagonal element of  $\mathbf{A}$  for the  $i$ th individual and  $\mathbf{s}_i$  is a vector indicating the contributions of the parents of the  $i$ th individual. This formula requires the use of the diagonal elements of  $\mathbf{A}$ , which can be obtained from

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i-1} & \mathbf{A}_{i-1}\mathbf{s}_i \\ \mathbf{s}_i'\mathbf{A}_{i-1} & a_{ii} \end{bmatrix} \quad (6)$$

(VAN ARENDONK *et al.* 1994).

The term  $(a_{ii} - \mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s}_i)$  in (5) expresses the contribution of Mendelian sampling in the production of the  $i$ th genotype as a proportion of the total additive genetic variance in the base population and is referred to as  $\delta_{ii}$  in this paper. Under biparental reproduction,  $\mathbf{s}_i$  contains  $1/2$  at the positions corresponding to each parent and 0 elsewhere. In this case,  $\delta_{ii}$  becomes  $(a_{ii} - \mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s}_i) = (1 + 1/2 a_{a,d} - \mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s}_i) = (1 - 1/4 (a_{ss} + a_{dd})) = (1/2 - 1/4(F + F_d))$ , and (5) reduces to the result of HENDERSON (1976) (B. TIER and J. SÖLKNER, unpublished data).

In case of gynogenetic reproduction,  $\mathbf{s}_i$  contains a 1 only at the position of the dam. In this case, rules for building  $\mathbf{A}^{-1}$  directly (5) are simple. Individual  $i$ , produced gynogenetically from dam  $d$ , can be included in  $\mathbf{A}^{-1}$  by adding

$\delta_{ii}$  to the diagonal elements of  $\mathbf{A}^{-1}$  that correspond to dam  $(d, d)$  and individual  $(i, i)$ , and  $-\delta_{ii}$  to the off-diagonal elements that correspond to dam and individual, *i.e.*,  $(i, d)$  and  $(d, i)$ , where

$$\delta_{ii} = (2 - a_{dd})^{-1}. \quad (7)$$

Element  $a_{dd}$ , the diagonal element of  $\mathbf{A}$  for the dam, can be obtained from (6) or can be calculated as  $1 + F_d$ .

When the dam under gynogenetic reproduction or both parents under biparental reproduction are 100% homozygous ( $F = 1$ ), the genotypic value of the offspring produced can be represented as a linear combination of parental genotypic values. In this case, there is no Mendelian sampling variance within families, which is reflected in (5) by  $\delta_{ii}$  being zero. As a consequence,  $\mathbf{A}$  will be singular. For example, when gynogenetic reproduction is used to produce clones, as in the family shown in Figure 1, the  $\mathbf{A}$  matrix will be singular. In Figure 1, individual 1 is a noninbred dam that produced gynogenetic offspring 2, 3 and 4. Individual 5 is a gynogenetic offspring from individual 2. Because individual 2 is 100% homozygous, the genotype of 5 is identical to the genotype of 2, *i.e.*, 5 is a clone of 2. As a result  $\delta_{55}$  is equal to zero

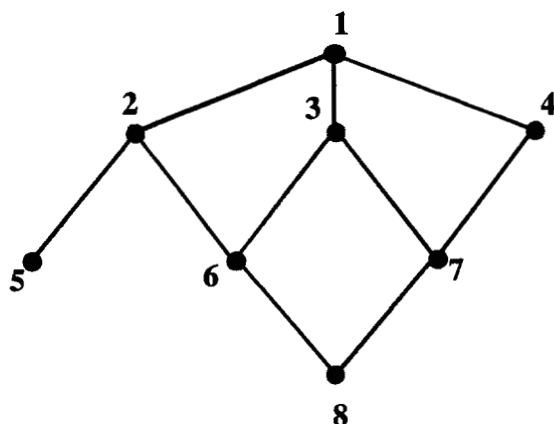


FIGURE 1.—Gynogenetic family structure with dam (1), gynogenetic sibs (2, 3, 4), clone (5),  $F_1$ 's (6, 7) and an  $F_2$  (8).

and  $\mathbf{A}$  is singular. However, because individuals 2 and 5 are genetically identical, the observation on individual 5 is a repeated observation on the genotype of individual 2. This information can be accommodated by modifying the incidence matrix  $\mathbf{Z}$  and the vector of breeding values  $\mathbf{u}$  (KENNEDY and SCHAEFFER 1990). As a result of this modification,  $\mathbf{u}$  contains breeding values only for unique genotypes and  $\mathbf{Z}$  contains a column for each unique genotype. Regarding clones as repeated observations solves the problem of singularity in this case.

Individuals 6 and 7 (Figure 1) are offspring of fully inbred parents, which entails  $\delta_{66}$  and  $\delta_{77}$  being zero. As a consequence, the genotypes of 6 and 7 are identical and equal to the parental average. Because only unique genotypes are included in  $\mathbf{A}^{-1}$ , the genotypes of individuals 6 and 7 do not appear in  $\mathbf{u}$ . However, observations on 6 and 7 provide information on the parental additive genetic effects, which can be incorporated in  $\mathbf{Z}$  by assigning half of the observation of 6 and 7 to the one parent, and the other half to the other parent. This can easily be done by adding  $1/2$  to the  $\mathbf{Z}$  matrix at the positions corresponding to the parents of 6 and 7. This generates off-diagonal elements in  $\mathbf{Z}'\mathbf{Z}$ .

Individual 8 (Figure 1) causes problems, because its parents do not directly correspond with a single unique genotype. In  $\mathbf{s}_i$ , only contributions of unique genotypes are included. The unique genotypes contributing to the genotype of individual 8 are the grandparents 2, 3 and 4, which contribute, respectively,  $1/4$ ,  $1/2$  and  $1/4$  of the genetic material. Using these values, (5) can be used to obtain  $\mathbf{A}^{-1}$ . Consider an individual  $i$  having fully inbred grandparents  $j$ ,  $k$ ,  $l$  and  $m$ , where  $j$  and  $k$  are the paternal grandparents and  $l$  and  $m$  are the maternal grandparents. In  $\mathbf{s}_i$ , the contribution of each grandparent is represented and (5) reduces to adding the following contributions to  $\mathbf{A}^{-1}$ :

$$\begin{aligned} 1/16 \delta^{ii} & \text{ to element } (x, y) & \text{ with } x = j, m \text{ and } y = j, m \\ -1/4 \delta^{ii} & \text{ to element } (i, x) \text{ and } (x, i) & \text{ with } x = j, m \\ \delta^{ii} & \text{ to diagonal element } (i, i) \end{aligned}$$

where

$$\delta^{ii} = (1/2 - 1/8(a_{j,k} + a_{l,m}))^{-1}. \quad (8)$$

The elements  $a_{j,k}$  and  $a_{l,m}$  are elements of  $\mathbf{A}$  and can be obtained from (6). In fact, those elements indicate the degree of inbreeding in the sire and dam of  $i$  and therefore the amount of Mendelian sampling variance in the production of the  $i$ th genotype.

*Example:* To illustrate the above rules, the mixed model equations for the small gynogenetic family in Figure 1 are given. Let the vector of observations be

$$\mathbf{y}' = (15 \ 18 \ 12 \ 13 \ 10 \ 17 \ 14 \ 15).$$

Individual 1 is a noninbred dam that produces the gynogenetic sibs 2, 3 and 4. Individual 5 is a gynogenetically produced clone of 2. Individuals 6 and 7 are referred to as  $F_1$ 's while 8 is referred to as an  $F_2$ . The genotypes of 5, 6 and 7 cause singularity of  $\mathbf{A}$ , so their genotypes are not included in the vector  $\mathbf{u}$ ; however their observations are in  $\mathbf{y}$ ,  $\mathbf{X}$  and  $\mathbf{Z}$ . For this family,  $\mathbf{A}^{-1}$  and  $\mathbf{Z}$  are as follows:

$$\mathbf{A}^{-1} = \begin{bmatrix} 4 & -1 & -1 & -1 & 0 \\ -1 & 1/4 & 1/2 & 1/4 & -1 \\ -1 & 1/2 & 2 & 1/2 & -2 \\ -1 & 1/4 & 1/2 & 1/4 & -1 \\ 0 & -1 & -2 & -1 & 4 \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Using these matrices, and a variance ratio  $\alpha$  equal to one, results in the following mixed model equations:

$$\begin{bmatrix} 8 & 1 & 2 1/2 & 2 & 1 1/2 & 1 \\ 1 & 5 & -1 & -1 & -1 & 0 \\ 2 1/2 & -1 & 3 1/2 & 3/4 & 1/4 & -1 \\ 2 & -1 & 3/4 & 3 1/2 & 3/4 & -2 \\ 1 1/2 & -1 & 1/4 & 3/4 & 2 1/2 & -1 \\ 1 & 0 & -1 & -2 & -1 & 5 \end{bmatrix} \begin{bmatrix} \beta \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_8 \end{bmatrix} = \begin{bmatrix} 114 \\ 15 \\ 36.5 \\ 27.5 \\ 20 \\ 15 \end{bmatrix}.$$

Solutions for the mixed model equations are as follows:  $\beta = 14.36$ ,  $\mathbf{u}' = [0.00 \ 0.26 \ -0.32 \ -0.57 \ -0.06]$ . Additive genetic effects ( $ebu$ ) for individuals 5, 6 and 7 can be calculated as  $\mathbf{Zu}$ , resulting in  $ebu_5 = 0.26$ ,  $ebu_6 = 1/2(0.26 - 0.32)$ ,  $ebu_7 = 1/2(-0.32 - 0.57)$ .

Summarizing, the inverse of the numerator relationship matrix for all possible relationships can be built according to (5). Equation 5 requires the calculation of inbreeding coefficients of all animals, which can be obtained from (6). In (5), the  $\mathbf{s}$  vector represents the proportion of genes contributed by each unique genotype. Only unique genotypes should be included in  $\mathbf{u}$  and  $\mathbf{A}^{-1}$ . Genotypes that are not unique can be identified by calculating  $\delta_{ii}$ , with a value of zero indicating that the genotype is not unique.

**Optimum designs for estimating heritability:** Gynogenesis opens ways to new experimental designs for estimating variance components that can capitalize on the closer relationships between individuals. To investigate the potential value of this technique, optimum gynogenetic sib family sizes for estimating heritabilities and standard errors of heritability estimates will be derived. Optimum design refers to the design that shows the lowest standard error of the estimated heritability. The results will be compared to results for conventional full-sib designs, because both gynogenetic and conventional full-sib designs are subject to effects of common environment and to maternal effects. For a population consisting of  $N$

families of  $n$  individuals each, the optimum  $n$  will be derived, keeping the total number of animals  $T = Nn$  constant. Results from a deterministic derivation as well as from stochastic simulation using Gibbs sampling will be shown.

**Deterministic approach:** The optimum family size depends on the sampling variance of the intraclass correlation  $t$  between the members of a family:

$$V(\hat{t}) = 2[1 + (n-1)t]^2(1-t)^2/(n-1)(T-n) \quad (9)$$

(FISCHER 1941). When  $T$  is constant and large,  $V(\hat{t})$  as a function of  $n$  has a minimum for  $n = 1 + 1/t$ . Substituting this result in (9) leads to the following approximation of the sampling variance in the optimum design:

$$V(\hat{t}) \approx 8t(1-t)^2/T, \quad (10)$$

which is the same result as obtained by ROBERTSON (1959). Ignoring common environment and maternal effects, the intraclass correlation  $t$  between full-sibs equals twice the coefficient of coancestry ( $f$ ) times the heritability:  $t = 2fh^2 = 0.5h^2$ . Optimum family sizes for a conventional full-sib design and the corresponding sampling variance of  $\hat{h}^2$  are

$$n = 1 + 2/h^2, \quad (11)$$

$$V(\hat{h}^2) \approx 16h^2(1 - \frac{1}{2}h^2)^2/T. \quad (12)$$

In a population consisting of gynogenetic sib families, the intraclass correlation is calculated as  $t = \text{cov}(P_1, P_2)/\sigma_{P1}\sigma_{P2}$ , where  $P_1$  and  $P_2$  are the phenotypes of two gynogenetic sibs. Ignoring common environment and maternal effects, this results in

$$t = \frac{1}{2}h_{gs}^2 = h^2/(h^2 + 1), \quad (13)$$

where  $h_{gs}^2$  is the heritability in the gynogenetic sib population and  $h^2$  is the heritability in the noninbred dam population. This result corresponds to  $t = 2f_{gs}^2$ , i.e., twice the coancestry times the heritability. We further wish to estimate  $h^2$  and its standard error. Combining (13) and  $n = 1 + 1/t$ , the optimum family size for a gynogenetic sib design is

$$n = 2 + 1/h^2. \quad (14)$$

To obtain the sampling variance of the estimated heritability for a gynogenetic design, the nonlinear relation between  $t$  and  $h^2$  in (13) is approximated by a linear relation, using a first order Taylor approximation (e.g., ENGLEFIELD 1989). This means that in (13) the relation between  $t$  and  $h^2$  is approximated by the tangent of the curve in the point of the estimate. The variance is then proportional to the quadratic of the slope of the tangent ( $dt/dh^2$ ). This results in  $V(t) = V(h^2)/(h^2 + 1)^4$ , which, when substituted into (10), leads to

$$V(\hat{h}^2) \approx 8h^2(h^2 + 1)/T. \quad (15)$$

The relative superiority of the gynogenetic design depends only on the heritability, because sampling variances in both (12) and (15) are inversely proportional to the total number of animals  $T$ .

**Stochastic approach:** To validate the deterministic theory, data from 1024 individuals were simulated under a gynogenetic sib design and under a conventional full-sib design with varying family size. The number of families for each design was calculated as  $N = 1024/n$ , where  $n$  is the family size.  $N$  was rounded to the nearest integer, so the total number of animals can show minor deviations from 1024. Simulated heritability and phenotypic variance were 0.4 and 100. Additive genetic values for sires and dams were chosen from  $N(0, \sigma_u^2)$ ; additive genetic values for gynogenetic offspring from  $N(u_d, \sigma_u^2)$ ; additive genetic values for full-sibs from  $N(\frac{1}{2}u_s + \frac{1}{2}u_d, \frac{1}{2}\sigma_u^2)$ ; and phenotypic values for the  $i$ th individual from

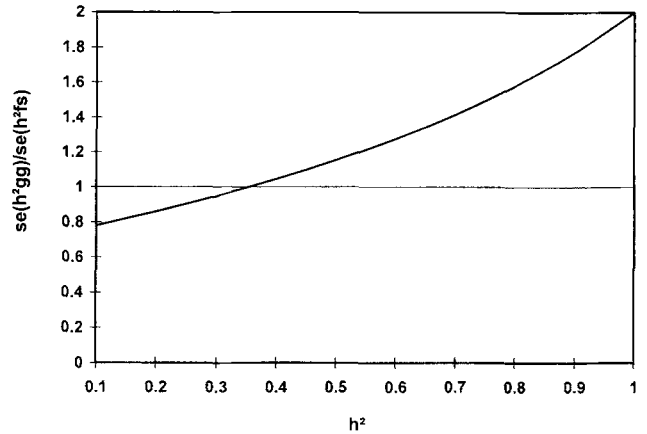


FIGURE 2.—Standard error of  $\hat{h}^2$  for an optimum gynogenetic design relative to an optimum conventional full-sib design, for different true heritabilities.

$N(\mu + u_i, \sigma_e^2)$ , where  $N$  denotes the normal distribution,  $\sigma_u^2$  is the additive genetic variance in the parental population,  $\sigma_e^2$  is the non-genetic variance,  $u_s$ ,  $u_d$  and  $u_i$  are the additive genetic value for sire, dam and individual, and  $\mu$  is the population mean. For each situation, 25 replicates were performed, except for the designs close to the deterministically determined optimum design, for which 75 replicates were performed.

Bayesian inferences on variance components were made using a Gibbs sampling algorithm. The model was

$$y = X\beta + Zu + e. \quad (16)$$

As fixed effect, only a general mean was included. For the conventional full-sib design, marginal posterior distributions were generated using the Gibbs sampling program maGgic developed by JANSS *et al.* (1995). For the gynogenetic sib design, the procedure is described in the APPENDIX. Flat prior distributions for variance components were used for both designs. For both the gynogenetic and the full-sib design, posterior distributions were based on a single chain of 40,000 samples for each replicate. Starting values for the variance components were equal to the simulated values.

The obtained Gibbs chain was thinned according to the approach suggested by RAFERTY and LEWIS (1992). From the obtained posterior distribution, the mode was used as an estimate of the parameter (i.e., variance component) and the posterior standard deviation was used as the equivalent of the standard error of the parameter estimate.

## RESULTS

**Deterministic approach:** Using (14) and (11), optimum family sizes for a heritability of 0.4 are 4.5 for a gynogenetic sib design and 6.0 for a conventional full-sib design. For a total of 1024 animals, the standard errors of  $\hat{h}^2$  are 0.066 and 0.063, respectively. In Figure 2, the standard error of the estimated heritability for an optimum gynogenetic sib design relative to an optimum conventional full-sib design is presented for different true heritabilities. This figure was obtained using (12) and (15). It can be seen that only for low heritabilities there is a small advantage for the gynogenetic sib design, while for high heritabilities there is a clear disadvantage. When  $h^2 = 0.35$ , optimum family size for the

TABLE 1  
Standard deviations of variance component estimates obtained from stochastic simulation

| Gynogenetic sib design |          |                        |                           |                        |                           | Full-sib design |          |                        |                           |                        |                           |
|------------------------|----------|------------------------|---------------------------|------------------------|---------------------------|-----------------|----------|------------------------|---------------------------|------------------------|---------------------------|
| $N*n$                  | No. repl | $\text{std}\sigma_e^2$ | $\text{std}\sigma_e^{2*}$ | $\text{std}\sigma_u^2$ | $\text{std}\sigma_u^{2*}$ | $N*n$           | No. repl | $\text{std}\sigma_e^2$ | $\text{std}\sigma_e^{2*}$ | $\text{std}\sigma_u^2$ | $\text{std}\sigma_u^{2*}$ |
| 16*64                  | 25       | 18.8                   | 12.3                      | 18.0                   | 12.0                      | 16*64           | 25       | 10.8                   | 7.5                       | 19.3                   | 12.5                      |
| 64*16                  | 25       | 10.1                   | 9.8                       | 9.0                    | 6.7                       | 64*16           | 25       | 6.6                    | 7.2                       | 10.5                   | 8.7                       |
| 128*8                  | 25       | 8.8                    | 9.2                       | 7.1                    | 5.7                       | 128*8           | 25       | 5.9                    | 5.9                       | 8.2                    | 6.7                       |
| 146*7                  | 25       | 8.6                    | 10.4                      | 6.6                    | 7.3                       | 146*7           | 75       | 5.9                    | 5.7                       | 7.9                    | 7.1                       |
| 171*6                  | 75       | 8.6                    | 8.5                       | 6.4                    | 5.6                       | 171*6           | 75       | 6.0                    | 5.9                       | 7.7                    | 7.6                       |
| 205*5                  | 75       | 8.4                    | 9.1                       | 6.1                    | 5.5                       | 205*5           | 75       | 6.1                    | 6.0                       | 7.7                    | 6.5                       |
| 256*4                  | 75       | 8.8                    | 9.4                       | 6.0                    | 5.6                       | 256*4           | 25       | 6.3                    | 7.8                       | 7.4                    | 8.5                       |
| 341*3                  | 25       | 9.2                    | 10.5                      | 6.1                    | 5.6                       | 341*3           | 25       | 6.9                    | 7.4                       | 8.1                    | 8.5                       |

No. repl, number of replicates;  $\text{std}\sigma_e^2$ , posterior standard deviation of  $\sigma_e^2$  averaged over replicates;  $\text{std}\sigma_e^{2*}$ , standard deviation between the modes of replicates for  $\sigma_e^2$ . Simulated heritability and phenotypic variance:  $h^2 = 0.40$ ,  $\sigma_p^2 = 100$ .

\* Calculated as follows:  $\sum (x_i - x_{avg})^2 / (\text{No. repl} - 1)$ , where  $x_i$  is the mode of the posterior distribution for the  $i$ th replicate and  $x_{avg}$  is the mode averaged over all replicates for this design.

gynogenetic sib design and for the conventional full-sib design have equal standard errors. From (11) and (14) it follows that the optimum family size is always smaller for a gynogenetic design. At low values of  $h^2$ , the optimum full-sib family size is approximately twice the optimum gynogenetic sib family size. When  $h^2$  approaches unity, the optimum family size for both designs approaches three.

**Stochastic approach:** Posterior standard deviations of the variance components obtained from stochastic simulation are presented in Table 1 for both designs. Besides the posterior standard deviations, the standard deviations between the modes of replicates are also in Table 1 (indicated by an asterisk). It can be seen, that for large family sizes the standard deviation between the modes of replicates ( $\text{std}\sigma_e^{2*}$ ) is smaller than the posterior standard deviation averaged over replicates ( $\text{std}\sigma_e^2$ ). When the family sizes are smaller and the estimates become more accurate, this difference disappears. For gynogenetic designs,  $\sigma_u^2$  is estimated more accurately than  $\sigma_e^2$ , whereas for full-sib designs this relation is reversed (see Table 1).

Modes and standard deviations of estimated heritabilities averaged over replicates are presented in Table 2 together with 95% confidence intervals for the standard deviations of the estimated heritability. For both designs, standard deviations of the heritability estimates rapidly decreased when family size was lowered from 64 to eight family members by increasing the number of families. The optimum family size for a gynogenetic design was around four to five family members. For the full-sib design, the optimum family size could not be determined very precisely from the confidence intervals, but was around five to six family members. These results are in agreement with the deterministically determined optimum family sizes. Standard errors for the optimum design obtained from the deterministic derivation were 0.066 for a gynogenetic design and 0.063 for a conventional full-sib design. Those standard errors

are within the confidence intervals for the standard deviations presented in Table 2. For a heritability of 0.4, no significant difference in accuracy of the estimated heritability between a gynogenetic and a full-sib design could be found from simulation, as was to be expected from Figure 2.

## DISCUSSION

Under gynogenetic reproduction, the additive genetic variance in the offspring population is doubled compared to the noninbred dam population and is equally distributed within and between gynogenetic sib families (BONGERS *et al.* 1996). A method for breeding value estimation for gynogenetic families using mixed model equations is presented. In the mixed model equations, the inverse of the numerator relationship matrix **A** is required. A general method for creating this inverse is presented. As a result, **A**<sup>-1</sup> can be built for any population that has unusual relationships. For the use of this method, the diagonal elements of **A** are required, which means that the inbreeding coefficients of individuals have to be calculated. Identical genotypes can be produced using gynogenesis. These identical genotypes cause singularity of the **A** matrix. However, observations on individuals having identical genotypes can be regarded as repeated observations on the same genotype, which can be accommodated by considering only the unique genotypes.

In this paper, attention was paid to the genetic analyses under gynogenetic reproduction. This work is also relevant for populations that contain fully inbred individuals or certain other forms of reproduction, including the production of recombinant double haploids in plants, and sex determination by ploidy or homozygosity in bees. (SMITH and ALLAIRE 1985; LAIDLAW and PAGE 1986). The analyses of data originating from a cross between inbred lines (FALCONER 1989 p. 270) is another example. Each founder line is completely in-

TABLE 2  
Estimated heritabilities

| Gynogenetic sib design |             |                  |                           | Full-sib design |             |                  |                           |
|------------------------|-------------|------------------|---------------------------|-----------------|-------------|------------------|---------------------------|
| $N*n$                  | $\hat{h}^2$ | $std(\hat{h}^2)$ | 95% C.I. $std(\hat{h}^2)$ | $N*n$           | $\hat{h}^2$ | $std(\hat{h}^2)$ | 95% C.I. $std(\hat{h}^2)$ |
| 16*64                  | 0.46        | 0.187            | 0.178; 0.196              | 32*64           | 0.42        | 0.130            | 0.122; 0.138              |
| 64*16                  | 0.41        | 0.092            | 0.086; 0.098              | 128*16          | 0.43        | 0.080            | 0.077; 0.083              |
| 128*8                  | 0.42        | 0.076            | 0.072; 0.080              | 256*8           | 0.41        | 0.067            | 0.065; 0.069              |
| 146*7                  | 0.39        | 0.071            | 0.067; 0.075              | 292*7           | 0.40        | 0.065            | 0.064; 0.066              |
| 171*6                  | 0.40        | 0.070            | 0.068; 0.072              | 342*6           | 0.41        | 0.064            | 0.063; 0.066              |
| 205*5                  | 0.39        | 0.067            | 0.065; 0.068              | 410*5           | 0.41        | 0.065            | 0.064; 0.066              |
| 256*4                  | 0.39        | 0.067            | 0.065; 0.068              | 512*4           | 0.41        | 0.064            | 0.062; 0.066              |
| 341*3                  | 0.42        | 0.071            | 0.068; 0.074              | 682*3           | 0.42        | 0.070            | 0.069; 0.071              |

$\hat{h}^2$ , mode of heritability estimates averaged over replicates;  $std(\hat{h}^2)$ , standard deviation of heritability estimates averaged over replicates; 95% C.I., 95% confidence interval. Simulated heritability and phenotypic variance:  $h^2 = 0.40$ ,  $\sigma_p^2 = 100$ .

bred and can be represented by a single genotype. The  $F_1$  generations can be represented as the average of the founder lines, *i.e.*, not unique. Individuals in subsequent generations have unique genotypes. The approach presented in this paper can be utilized to analyze data of such experiments and accommodates completely inbred founder lines.

Optimum family sizes for estimating heritabilities were found to be smaller for gynogenetic sib families than for conventional full-sib families. Compared to the biological capacity of fish, optimum family sizes are very small. Only for heritabilities below 0.35, there is a small advantage in accuracy of the estimated heritability for gynogenetic sib families. For higher heritabilities, there is a clear disadvantage.

Clone lines enable the measurement of different traits on the same genotype, for example, carcass quality at different ages. Clone lines might therefore be very useful for estimating the genetic correlations. Clone lines can also be useful for estimating genotype by environment interaction, because performance on the same genotype can be observed in different environments.

Because of the increased heritability and the doubled additive genetic variance in the offspring population, gynogenetic reproduction can be useful for selection purposes. However, inbred individuals show inbreeding depression. Furthermore, the phenotypic variance may increase due to embryonic damage (BONGERS *et al.* 1996) or due to an increased sensitivity of inbred individuals to environmental influences (FALCONER 1989). For these reasons, the increase in heritability can be lower than theoretically expected.

Cloning by means of gynogenesis offers the opportunity to capitalize on nonadditive genetic effects. It seems promising to produce crossbred clones to make optimum use of nonadditive effects. This can be done by using gynogenetically produced individuals from different lines as parents. In this way the additive genetic superiority of selected gynogenetic individuals can be combined with the effect of heterosis. A method to

estimate genetic values for selection of parents as developed here, is crucial in such schemes. Among these crossbreds there is no genetic variation, which would promote product uniformity.

We thank JACK C. M. DEKKERS for carefully reading this paper and giving useful comments. In addition we thank HANS KOMEN and GUUS B. J. BONGERS for introducing us into the world of gynogenesis.

#### LITERATURE CITED

- BONGERS, A. B. J., H. BOVENHUIS, A. C. VAN STOKKUM, G. F. WIEGERTJES, B. ZANDIEH-DOULABI *et al.* 1996 Distribution of genetic variance in gynogenetic and androgenetic families. *Aquaculture* (in press).
- ENGLEFIELD, M. J., 1989 *Mathematical Methods for Engineering and Science Students*. Edward Arnold, London.
- FALCONER, D. S., 1989 *Introduction to Quantitative Genetics*. Longman Group Ltd., Essex, UK.
- FISCHER, R. A., 1941 *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, UK.
- HENDERSON, C. R., 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69–83.
- JANSS, L. L. G., R. THOMPSON and J. A. M. VAN ARENDONK, 1995 Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theor. Appl. Genet.* **91**: 1137–1147.
- KENNEDY, B. W., and J. E. MOXLEY, 1975 Comparison of genetic group and relationship methods for mixed model sire evaluation. *J. Dairy Sci.* **58**: 1507–1514.
- KENNEDY, B. W., and L. R. SCHAEFFER, 1990 Reproductive technology and genetic evaluation, pp. 507–532 in *Advances in Statistical Methods for Genetic Improvement of Livestock*, edited by D. GIANOLA and K. HAMMOND. Springer-Verlag, Berlin.
- KOMEN, J., A. B. J. BONGERS, C. J. J. RICHTER, W. B. VAN MUISWINKEL and E. A. HUISMAN, 1991 Gynogenesis in common carp (*Cyprinus carpio* L.) II: The production of homozygous gynogenetic clones and F1 hybrids. *Aquaculture* **92**: 127–142.
- LAIDLAW, H. H., and R. E. PAGE, 1986 Mating designs, pp. 323–344 in *Bee Genetics and Breeding*, edited by T. E. RINDERER. Academic Press, New York.
- MEYER, K., 1989 Estimation of genetic parameters, pp. 161–167 in *Evolution and Animal Breeding*, edited by G. W. HILL and T. F. C. MACKAY. CAB International, Wallingford, UK.
- QUAAS, R. L., 1976 Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**: 949–953.
- RAFERTY, A. E., and S. M. LEWIS, 1992 How many iterates in the Gibbs sampler?, pp. 765–776 in *Bayesian Statistics*, edited by J. M. BERNARDO, J. O. BERGEN, A. P. DAVID, and A. F. M. SMITH. Oxford University Press, Oxford.

- ROBERTSON, A., 1959 Experimental design in the evaluation of genetic parameters. *Biometrics* **15**: 219–226.
- SMITH, S. P., and F. R. ALLAIRE, 1985 Efficient selection rules to increase non-linear merit: application to mate selection. *Genet. Sel. Evol.* **17**: 387–406.
- VAN ARENDONK, J. A. M., B. TIER and B. P. KINGHORN, 1994 Use of multiple genetic markers in prediction of breeding values. *Genetics* **137**: 319–329.
- WANG, C. S., J. J. RUTLEDGE and D. GIANOIA, 1994 Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* **26**: 91–115.

Communicating editor: M. K. UYENOYAMA

## APPENDIX

**Bayesian analyses implemented via Gibbs sampling:** In each cycle of Gibbs sampling, a sequence of realizations was computed conditional on the data and on the realizations of the previous cycle. Realizations for the  $t$ th cycle were sampled in the order  $\boldsymbol{\beta}^{[t]}$ ,  $\mathbf{u}^{[t]}$ ,  $\sigma_e^{2[t]}$ ,  $\sigma_u^{2[t]}$ . Conditional distributions used to sample the realizations were similar to JANSS *et al.* (1995). For illustration purposes, the conditional distributions are shown in Equation A2 for all unique genotypes in the family structure of Figure 1. In the analyses to determine the optimum family size, however, only dams and gynogenetic sibs were included and contributions from other kinds of individuals were omitted from their conditional equations. In Figure 1, different types of individuals can be distinguished. Individual 1 is referred to as a dam, individuals 2–4 are referred to as gynogenetic sibs, 6 and 7 are referred to as  $F_1$ 's, and individual 8 is referred to as an  $F_2$ . The genotypes of the individuals 5–7 are not included in  $\mathbf{u}$  because these genotypes are not unique. Notation is similar to JANSS *et al.* (1995).

As first step in a cycle of Gibbs sampling, the general mean was sampled as

$$\text{Sample } \boldsymbol{\beta}^{[t+1]} \text{ from } N(\check{\mathbf{y}}/n_b, \sigma_e^{2[t]}/n_t), \quad (\text{A1})$$

where  $N$  denotes the normal distribution;  $\check{\mathbf{y}}$  is the sum of all elements in  $\hat{\mathbf{y}} = \mathbf{y} - \mathbf{Z}\mathbf{u}^{[t]}$ ;  $\sigma_e^{2[t]}$  is the realization of the nongenetic variance from the previous cycle; and  $n_t$  is the total number of observations. So the fixed effect is sampled conditional on the realized genetic effects and on the non-genetic variance from the previous cycle.

Genetic effects were sampled starting with the earliest generation. The additive genetic effect of genotype  $i$  was updated as:

$$\text{Sample } u_i^{[t+1]} \text{ from } N(c_i/d_i, \sigma_e^{2[t]}/d_i), \quad (\text{A2})$$

where  $d_i u_i = c_i$  are the mixed model equations for the  $i$ th individual. Elements in (A2) are as follows.

For the dams:

$$c_i = \check{y}_i + \alpha \sum_k \delta^{kk} u_k^{[t]},$$

$$d_i = 1 + \alpha(\delta^{ii} + \sum_k \delta^{kk}),$$

where  $\check{y}_i$  is the  $i$ th element of  $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]})$ ;  $k$  indicates each gynogenetic sib descending from dam  $i$ ;  $\alpha$  is the variance ratio  $\sigma_e^2/\sigma_u^2$ ; and  $\delta^{ii}$  is the reciprocal of the Mendelian sampling term, as described in the mixed model section.

For the gynogenetic sibs:

$$c_i = \check{y}_i + \alpha \delta^{ii} u_d^{[t+1]} - \alpha \sum_k \delta^{kk} s_{k,i} \sum_{gpr} s_{k,gpr} u_{gpr}^{[t+1]} - 1/4 \sum_p u_p^{[t+1]} + \alpha \sum_k s_{k,i} \delta^{kk} u_k^{[t]},$$

$$d_i = z_{ii} + \alpha(\delta^{ii} + \sum_k s_{k,i}^2 \delta^{kk}),$$

where  $i$  is the gynogenetic sib;  $d$  is the dam of  $i$ ;  $k$  refers to an  $F_2$ ;  $gpr$  refers to the grandparents of the  $F_2$ , excluding  $i$  itself;  $p$  refers to the mates of  $i$  for the production of  $F_1$ 's; and  $z_{ii}$  in the  $i$ th diagonal element of the  $\mathbf{Z}'\mathbf{Z}$  matrix. The element  $s_{k,i}$  is the  $i$ th element of the  $\mathbf{s}$  vector for individual  $k$ , and indicates the contribution of  $i$  to the genotype of  $k$ . The same holds for  $s_{k,gpr}$ . For instance, for the family structure in Figure 1,  $s_{8,2} = 1/4$ , and  $s_{8,3} = 1/2$ . When it is unclear whether a genetic effect used is from state  $t$  or  $t+1$ , the state is not specified.

For the  $F_2$ :

$$c_i = \check{y}_i + \alpha \delta^{ii} \sum_{gpr} s_{i,gpr} u_{gpr}^{[t+1]},$$

$$d_i = 1 + \alpha \delta^{ii},$$

where  $\sum_{gpr} s_{i,gpr} u_{gpr}^{[t+1]}$  indicates summation of the genetic effects over all grandparental genotypes contributing to  $i$ , weighted by their contribution.

Residual and genetic variances were sampled from inverted chi-squared distributions. Those variance components were updated as follows:

$$\text{sample } \sigma_e^{2[t+1]} \text{ from } \mathbf{e}'\mathbf{e}/\chi_{(nt-2)}^2, \quad (\text{A3})$$

$$\text{sample } \sigma_u^{2[t+1]} \text{ from } \mathbf{u}'\mathbf{u}^{[t+1]}\mathbf{A}^{-1}\mathbf{u}^{[t+1]}/\chi_{(q-2)}^2, \quad (\text{A4})$$

where  $\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]} - \mathbf{Z}\mathbf{u}^{[t+1]})$ ;  $\mathbf{A}$  is the numerator relationship matrix;  $\chi_{(nt-2)}^2$  and  $\chi_{(q-2)}^2$  are random deviates from chi-squared distributions with  $(n_t - 2)$  and  $(q - 2)$  degrees of freedom;  $n_t$  is the number of observations; and  $q$  is the number of genotypes. Degrees of freedom  $(n_t - 2)$  and  $(q - 2)$  indicate a flat prior for variance components (WANG *et al.* 1994). A scalar computation of the quadratic  $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$  is  $\sum_i u_i^2 + \sum_j \delta^{jj} (u_j - 1/2 u_d - 1/2 u_s)^2$ , where indices  $d$  and  $s$  indicate dam and sire of  $j$  (QUAAS 1976). The first summation is over all base genotypes and the second over all nonbase genotypes. For the gynogenetic sibs, the second summation is modified to  $\sum_j \delta^{jj} (u_j - u_d)^2$ , and for the  $F_2$ 's to  $\sum_j \delta^{jj} (u_j - \sum_{gpr} s_{j,gpr} u_{gpr})^2$ , where indices  $gpr$  indicate grandparents of  $j$ . In each cycle of the Gibbs sampler, the heritability was calculated as  $h^{2[t+1]} = \sigma_u^{2[t+1]}/(\sigma_u^{2[t+1]} + \sigma_e^{2[t+1]})$ .