

## ESTIMACIÓN DE COMPONENTES DE VARIANZA EN POBLACIONES AUTOFECONDADAS DE CRUZAS ENTRE PROGENITORES HOMOCIGÓTICOS NO EMPARENTADOS

### ESTIMATION OF VARIANCE COMPONENTS WITH SELFED CROSSES AMONG INBRED UNRELATED PARENTS

Jaime Sahagún Castellanos

Programa Universitario de Investigación y Servicio en Olericultura. Departamento de Fitotecnia, Universidad Autónoma Chapingo. 56230 Chapingo, Edo. de México. Tel y Fax (01-595) 2-15-00 Ext. 6185. Correo electrónico: jsahagun@taurus1.chapingo.mx

#### RESUMEN

En las especies autóгамas la aplicación de los diseños de apareamiento para estimar componentes de varianza se ha visto restringida por las dificultades que se presentan para hacer las cruzas y producir semilla suficiente para la evaluación de campo. Una alternativa para generar semilla suficiente es la autofecundación de las cruzas. Sin embargo, el arreglo genotípico de las cruzas autofecundadas difiere del de las cruzas originales; además, los estudios teóricos que a este respecto se han hecho se han basado en el modelo de dos alelos por locus, que es insuficiente para poblaciones formadas a partir de tres o más progenitores. Este estudio se hizo para determinar el sesgo de los estimadores de la varianza aditiva  $\left(\sigma_A^2\right)$  y de dominancia  $\left(\sigma_D^2\right)$  derivados con base en el modelo de dos alelos con frecuencias iguales y en las cruzas autofecundadas de p progenitores. Respecto a  $\sigma_A^2$ , el sesgo fue  $2FD_1 + (1/4)F^2D_2$ , en donde F es el coeficiente de endogamia de las cruzas autofecundadas ( $F < 1$ ),  $D_1$  es la covarianza entre los efectos aditivos de los genes y las desviaciones de dominancia de los genotipos formados por dos genes idénticos por descendencia, y  $D_2$  es la varianza de estas desviaciones de dominancia. Este sesgo se hace cero para dos alelos con frecuencias génicas iguales y se incrementa cuando el gene más deseable tiene una mayor frecuencia que el otro. El estimador de  $\sigma_D^2$ , en cambio, es insesgado. Otro componente estudiado fue el cuadrado de la media de la depresión endogámica. El estimador propuesto es una función del cuadrado medio del error y de las medias de los progenitores y de sus cruzas.

**Palabras clave:** Depresión endogámica promedio, diseños de apareamiento, covarianzas entre parientes, varianza aditiva, varianza de dominancia.

#### SUMMARY

It is difficult to apply mating designs in self-pollinated crops to estimate variance components, because the obtention of crosses is not easy and the amount of seeds produced for field evaluation is frequently limited. An alternative to produce enough seed is by selfing

the crosses. The genotypic array of the selfed crosses, however, differs from the unselfed crosses, and the available theoretical studies are based on a two-loci model, which is not appropriate for populations formed by the crosses among three or more parents. This study was made to determine the estimator biases for the additive  $\left(\sigma_A^2\right)$  and dominance  $\left(\sigma_D^2\right)$  variances derived on the basis of a two-allele model and selfed crosses among p parents. For  $\sigma_A^2$  the bias was  $2FD_1 + (1/4)F^2D_2$ , where F is the inbreeding coefficient of the selfed crosses ( $F < 1$ ),  $D_1$  is the covariance between the gene additive effects and the dominance deviations of the genotypes carrying alleles, which are identical by descent, and  $D_2$  is the variance of these dominance deviations. For the two-allele case, this bias is zero if the allelic frequencies are 0.5, but it increases as  $p > q$  (p is the frequency of the most desirable allele). For  $\sigma_D^2$  the two-allele-model estimator is unbiased. An additional component is the square mean of the inbreeding depression; its proposed estimator is a function of the error mean square and the mean of the parents and crosses.

**Index words:** Average inbreeding depression, mating designs, covariance among relatives, additive variance, dominance variance.

#### INTRODUCCIÓN

Los diseños de apareamiento han sido muy extensamente utilizados para el estudio de parámetros relacionados con la acción génica en poblaciones objeto de mejoramiento genético, pero en ocasiones su aplicación se ha visto restringida por la difícil y laboriosa tarea de hacer las cruzas o por la reducida cantidad de semilla  $F_1$  que se obtiene para evaluación. Estos problemas, que se agudizan en forma muy particular en las especies autóгамas, ya habían sido puestos a consideración por Molina (1993); este autor incluyó en la problemática los dialélicos y los diseños I y II de Carolina del Norte, y para disponer de semilla

suficiente para la evaluación de campo propuso la obtención de semilla de generaciones avanzadas de autofecundación de las  $F_1$ . Benítez (1999) también se refirió a esta estrategia para incrementar la semilla para evaluación de las cruzas resultantes de la recombinación de los materiales seleccionados en programas de selección recurrente en trigo (*Triticum aestivum* L.).

Con base en la evaluación de cruzas autofecundadas, Molina (1993) propuso estimadores de los componentes de varianza genética aditiva y de dominancia, Márquez y Molina (1995) derivaron la varianza genotípica entre las familias autofecundadas de hermanos completos y de medios hermanos que se forman del apareamiento aleatorio entre  $N$  líneas homocigóticas, y Benítez (1999) derivó una fórmula para determinar el avance genético en un programa de selección recurrente. Sin embargo, el alcance de esos resultados es relativamente limitado toda vez que se basan en un modelo de dos alelos por *locus*, seguramente insuficiente para el estudio de una población formada por la cruce de tres o más progenitores de origen diferente.

Una población formada a partir de las cruzas entre tres o más progenitores cuyo coeficiente de endogamia es superior a cero tiene una varianza que, además de incluir las componentes aditiva y de dominancia, debe tener componentes asociados a la presencia de genotipos constituidos por dos genes idénticos por descendencia y a la existencia de tres o más alelos por *locus* (Cockerham, 1983; Cockerham y Matzinger, 1985). Así, estimar componentes de varianza genéticos o determinar la respuesta esperada a la selección debe producir sesgos que pueden ser motivo de la generación de falsas expectativas (por exceso o defecto) respecto al éxito que pueda producir el manejo genotécnico de una población.

En el contexto de la hipótesis referente a estrategias de estimación y predicción erróneas se planeó esta investigación con el objetivo principal de determinar el sesgo de los estimadores de los componentes aditivo y de dominancia de la varianza genética de una población formada a partir de las  $p^2$  cruzas entre  $p$  progenitores ( $p \geq 3$ ) completamente endogámicos y no emparentados. Los estimadores cuyos sesgos se pretende determinar son los generados con base en un modelo de dos alelos con frecuencias génicas iguales. El sesgo se determinó con respecto a un modelo con alelos múltiples con frecuencias génicas arbitrarias. Adicionalmente, para cada componente de la varianza de la población que forman las cruzas después de  $n$  generaciones de autofecundación se propone un estimador.

## MARCO TEÓRICO Y MÉTODOS

La población estudiada es una formada por individuos diploides y bisexuales, sin efectos maternos, mutación, migración, epistasis ni desequilibrio de ligamiento. Con equilibrio de ligamiento y sin epistasis, el estudio de un carácter cuantitativo puede hacerse con base en el modelo de un *locus*. Las varianzas, medias, etc., para todos los *loci* involucrados en el carácter objeto de estudio es la suma de las varianzas, medias, etc., respectivamente, derivadas para cada *locus*.

Se considerará que los  $p$  progenitores de la población no tienen relación de parentesco y que la frecuencia del alelo  $A_u$  es  $p_u$  ( $u=1, 2, \dots, a; a \leq p$ ). Así, una vez que se han cruzado todos los progenitores y que se han hecho  $n$  autofecundaciones de cada cruce, el arreglo genotípico de la población resultante es

$$\left[ 1 - (1/2)^n \right] \sum_u p_u A_u A_u + (1/2)^n \sum_u \sum_v p_u p_v A_u A_v \quad (\text{Ec. 1})$$

en donde el coeficiente de endogamia es  $F = 1 - (1/2)^n$  y  $1 - F = (1/2)^n$  (Kempthorne, 1969).

Cuando proceda, los resultados de la varianza genética correspondiente a la población descrita en la Ec.1, serán expresados en términos de los componentes de la varianza de la población, cuyo arreglo genotípico es el descrito en Ec. 1 para  $F = 0$ , población que ha sido muy estudiada (Kempthorne, 1969).

El valor genotípico codificado de  $A_i A_j (y_{ij})$  será descrito en términos de los efectos aditivos de  $A_i (\alpha_i)$  y  $A_j (\alpha_j)$  y de la desviación de dominancia de  $A_i A_j (\delta_{ij})$ . Así,  $y_{ij} = \alpha_i + \alpha_j + \delta_{ij}$ .

De acuerdo con esta notación, la varianza de los valores aditivos o varianza aditiva ( $\sigma_A^2$ ), y la varianza de las desviaciones de dominancia o varianza de dominancia ( $\sigma_D^2$ ), son expresables como (Kempthorne, 1969)

$$\sigma_A^2 = 2 \sum_u p_u \alpha_u^2 \quad (\text{Ec. 2})$$

$$\sigma_D^2 = \sum_u \sum_v p_u p_v \delta_{uv}^2 \quad (\text{Ec. 3})$$

Además,

$$\sum_u p_u \alpha_u = \sum_v p_v \alpha_v = \sum_u p_u \delta_{uv} = \sum_v p_v \delta_{uv} = \sum_u \sum_v p_u p_v \delta_{uv} = 0 \quad (\text{Ec. 4})$$

Con  $F > 0$ , la población cuyo arreglo genotípico se describe en Ec.1 incluye genotipos formados por genes idénticos por descendencia. La presencia de estos genotipos sugiere que la varianza genética debe incluir componentes de varianza asociados a los efectos de tales genotipos. Dos de estos efectos son (Cockerham, 1983)

$$D_1 = \sum_u p_u \alpha_u \delta_{uu} \quad (\text{Ec. 5})$$

$$D_2 = \sum_u p_u \delta_{uu}^2 - \left( \sum_u p_u \delta_{uu} \right)^2 \quad (\text{Ec. 6})$$

Weir y Cockerham (1976) derivaron la varianza genética de una población con un coeficiente de endogamia  $F$  ( $\sigma_{GT}^2$ ). Ésta fue

$$\sigma_{GT}^2 = [1 + F] \sigma_A^2 + [1 - F] \sigma_D^2 + 4FD_1 + FD_2 - F(1 - F)\bar{H} \quad (\text{Ec.7})$$

En donde  $\sigma_A^2$ ,  $\sigma_D^2$ ,  $D_1$  y  $D_2$  corresponden a las definiciones en las Ecuaciones 2, 3, 5 y 6, respectivamente, y  $\bar{H} = \left( \sum_u p_u \delta_{uu} \right)^2$ . Para la población que se estudia en este trabajo (Ec. 1), la varianza genética es entonces

$$\sigma_G^2 = \left[ 2 - (1/2)^n \right] \sigma_A^2 + (1/2)^n \sigma_D^2 + 4 \left[ 1 - (1/2)^n \right] D_1 + \left[ 1 - (1/2)^n \right] D_2 - \left[ (1/2)^n \right] \left[ 1 - (1/2)^n \right] \bar{H} \quad (\text{Ec. 8})$$

### RESULTADOS

La información de la evaluación de las  $p^2$  cruza correspondientes a la estrategia dialélica permite un análisis de varianza para grupos de familias de diferentes tipos, como progenitores, familias (como las obtenidas después de  $n$  generaciones de autofecundación de las cruza entre progenitores diferentes), etc. También se pueden estimar covarianzas entre parientes para estimar los componentes de la varianza de la población que produce la autofecundación, durante  $n$  generaciones, de las  $p^2$  cruza dialélicas. Para derivar la varianza de las medias de los valores geno-

típicos de las familias que genera la autofecundación de las  $p^2$  cruza dialélicas, debe considerarse que la cruza de los progenitores  $A_u A_u$  y  $A_v A_v$  después de ser sometida a un sistema recurrente de autofecundación durante  $n$  generaciones produce una familia cuyo arreglo y valor genotípico son, respectivamente,

$$(1/2) \left[ 1 - (1/2)^n \right] A_u A_u + (1/2)^n A_u A_v + (1/2) \left[ 1 - (1/2)^n \right] A_v A_v$$

y

$$(1/2) \left[ 1 - (1/2)^n \right] \left[ 2\alpha_u + \delta_{uu} \right] + (1/2)^n \left[ \alpha_u + \alpha_v + \delta_{uv} \right] + (1/2) \left[ 1 - (1/2)^n \right] \left[ 2\alpha_v + \delta_{vv} \right]$$

Como  $F = 1 - (1/2)^n$ , después de un poco de álgebra este valor genotípico ( $G_{uv}$ ) puede reexpresarse como

$$G_{uv} = \alpha_u + \alpha_v + (1/2)^n F \left[ \delta_{uu} + \delta_{vv} \right] + (1/2)^n \delta_{uv}$$

Así, por (4), la media genotípica de todas las familias para cualquier generación  $n$  de autofecundación es

$$\mu_{HCA} = \sum_u \sum_v p_u p_v \left\{ \alpha_u + \alpha_v + (1/2)^n F \left[ \delta_{uu} + \delta_{vv} \right] + (1/2)^n \delta_{uv} \right\} = (1/2)^n F \left[ \sum_u p_u \delta_{uu} + \sum_v p_v \delta_{vv} \right] = F \sum_u p_u \delta_{uu} \quad (\text{Ec. 9})$$

Además, la varianza entre medias de dichas familias es

$$\begin{aligned}
 SC(VG)_{HCA} &= \sum_u \sum_v p_u p_v \left\{ \alpha_u + \alpha_v + (1/2)F[\delta_{uu} + \delta_{vv}] \right. \\
 &\quad \left. + (1/2)^n \delta_{uv} \right\}^2 \\
 &= \sum_u \sum_v p_u p_v \left\{ \alpha_u^2 + \alpha_v^2 + (1/4)F^2(\delta_{uu}^2 + \delta_{vv}^2 + 2\delta_{uu}\delta_{vv}) \right. \\
 &\quad \left. + (1/2)^{2n} \delta_{uv}^2 + F\alpha_u\delta_{uu} + F\alpha_v\delta_{vv} \right\} \\
 &= 2\sum_u p_u \alpha_u^2 + (1/2)F^2 \sum_u p_u \delta_{uu}^2 + (1/2)F^2 \left( \sum_u p_u \delta_{uu} \right)^2 \\
 &\quad + (1/2)^{2n} \sum_u \sum_v p_u p_v \delta_{uv}^2 + 2F \sum_u p_u \alpha_u \delta_{uu} \\
 &= \sigma_A^2 + (1/2)^{2n} \sigma_D^2 + 2FD_1 + (1/2)F^2 \sum_u p_u \delta_{uu}^2 \\
 &\quad + (1/2)F^2 \left( \sum_u p_u \delta_{uu} \right)^2
 \end{aligned}$$

De acuerdo con la Ec. 9, la varianza de las medias genotípicas de las familias que producen todas las cruzas después de haber sido autofecundadas durante n generaciones es

$$\begin{aligned}
 \sigma_{HCA}^2 &= \sigma_A^2 + (1/2)^{2n} \sigma_D^2 + 2FD_1 + (1/2)F^2 \sum_u p_u \delta_{uu}^2 \\
 &\quad + (1/2)F^2 \left( \sum_u p_u \delta_{uu} \right)^2 - F^2 \left( \sum_u p_u \delta_{uu} \right)^2 \\
 &= \sigma_A^2 + (1/2)^{2n} \sigma_D^2 + 2FD_1 + (1/2)F^2 D_2 \quad (Ec. 10)
 \end{aligned}$$

Por otra parte, para progenitores homocigóticos cuyo coeficiente de endogamia es igual a 1, la varianza genética entre ellos es igual a la de todas las "cruzas" entre cada progenitor consigo mismo y, también, a la de las familias formadas por las n generaciones de autofecundación de estas cruzas. Como la media genotípica de los progenitores es

$$\sum_u p_u (2\alpha_u + \delta_{uu}) = \sum_u p_u \delta_{uu}$$

la varianza entre sus medias genotípicas es

$$\begin{aligned}
 \sigma_p^2 &= \sum_u p_u (2\alpha_u + \delta_{uu})^2 - \left[ \sum_u p_u \delta_{uu} \right]^2 \\
 &= 4\sum_u p_u \alpha_u^2 + 4\sum_u p_u \alpha_u \delta_{uu} + \sum_u p_u \delta_{uu}^2 - \left[ \sum_u p_u \delta_{uu} \right]^2 \\
 &= 2\sigma_A^2 + 4D_1 + D_2 \quad (Ec. 11)
 \end{aligned}$$

Otra fuente de información para la generación de estimadores de los componentes de la varianza genética en la Ec. 8 es la varianza de las familias de medios hermanos autofecundados. Un progenitor de genotipo  $A_u A_v$  genera, de sus cruzas con todos los progenitores, una familia de medios hermanos cuyo arreglo genotípico es  $\sum_u p_u A_u A_v$ . Después de n generaciones de autofecundación de esta familia el arreglo genotípico resultante es

$$(1/2)F \sum_u p_u A_u A_u + (1/2)^n \sum_u p_u A_u A_v + (1/2)F A_v A_v$$

Como la media genotípica de esta familia es  $(1/2)F \sum_u p_u \delta_{uu} + \alpha_v + (1/2)F \delta_{vv}$ , la media de todas las familias de este tipo, de acuerdo con (4), es

$$\mu_{MHA} = (1/2)F \sum_u p_u \delta_{uu} + (1/2)F \sum_v p_v \delta_{vv} = F \sum_u p_u \delta_{uu}$$

Así, la varianza de las familias de medios hermanos autofecundados durante n generaciones es

$$\begin{aligned}
 \sigma_{MHA}^2 &= \sum_v p_v \left\{ (1/2)F \sum_u p_u \delta_{uu} + \alpha_v + (1/2)F \delta_{vv} \right\}^2 - F^2 \left( \sum_u p_u \delta_{uu} \right)^2 \\
 &= (1/4)F^2 \left( \sum_u p_u \delta_{uu} \right)^2 + \sum_v p_v \alpha_v^2 + (1/4)F^2 \sum_v p_v \delta_{vv}^2 \\
 &\quad + (1/2)F^2 \left( \sum_u p_u \delta_{uu} \right)^2 + F \sum_v p_v \alpha_v \delta_{vv} - F^2 \left( \sum_u p_u \delta_{uu} \right)^2 \\
 &= (1/2)\sigma_A^2 + FD_1 + (1/4)F^2 D_2 \quad (Ec. 12)
 \end{aligned}$$

En virtud de que la varianza genética aquí estudiada Ec. 7 tiene cinco componentes, la estimación de éstos requiere dos ecuaciones independientes adicionales a las Ecuaciones 10, 11 y 12. La covarianza entre medias de pares de progenitores y las medias de las familias correspondientes de hermanos completos, que generan las cruza entre cada par de progenitores, es otra fuente de información que puede ser utilizada para la derivación de estimadores. Para la derivación de esta covarianza es necesario determinar la media de los progenitores  $A_u A_u$  y  $A_v A_v$  (P), la media de la familia que forma la cruza  $A_u A_u \times A_v A_v$ , seguida de autofecundación por n generaciones (FA) y los valores esperados de P y FA. En virtud de que  $P = \alpha_u + \alpha_v + (1/2)\delta_{uu} + (1/2)\delta_{vv}$ ,

$FA = \alpha_u + \alpha_v + (1/2)F[\delta_{uu} + \delta_{vv}] + (1/2)^n \delta_{uv}$ , y de que la media de P y FA son, respectivamente,

$$E(P) = \sum_u p_u \delta_{uu} \quad \text{y} \quad E(FA) = F \sum_u p_u \delta_{uu}$$

la covarianza entre P y FA es

$$\begin{aligned} Cov[P, FA] &= \sum_u \sum_v p_u p_v \left\{ [\alpha_u + \alpha_v + (1/2)\delta_{uu} + (1/2)\delta_{vv}] [\alpha_u + \alpha_v + (1/2)F(\delta_{uu} + \delta_{vv}) + (1/2)^n \delta_{uv}] - F \left[ \sum_u p_u \delta_{uu} \right]^2 \right\} \\ &= \sum_u \sum_v p_u p_v \left\{ \alpha_u^2 + (1/2)F\alpha_u \delta_{uu} + \alpha_v^2 + (1/2)F\alpha_v \delta_{vv} + (1/2)\alpha_u \delta_{uu} + (1/4)F\delta_{uu}^2 + (1/4)F\delta_{uu} \delta_{vv} + (1/2)\alpha_v \delta_{vv} + (1/4)F\delta_{uu} \delta_{vv} + (1/4)F\delta_{vv}^2 \right\} - F \left[ \sum_v p_v \delta_{vv} \right]^2 \\ &= 2 \sum_u p_u \alpha_u^2 + (1+F) \sum_u p_u \alpha_u \delta_{uu} + (1/2)F \sum_u p_u \delta_{uu}^2 + (1/2)F \left( \sum_u p_u \delta_{uu} \right)^2 - F \left( \sum_u p_u \delta_{uu} \right)^2 \\ &= \sigma_A^2 + (1+F)D_1 + (1/2)FD_2 \end{aligned} \quad (Ec. 13)$$

En las Ecs. 10, 11, 12 y 13 no se incluye al parámetro  $\tilde{H}$ . Para generar una ecuación que lo incluya, considérese que  $\bar{P}$  y  $\bar{H}$  representan la media fenotípica de todos los progenitores y de todas las cruza entre progenitores diferentes autofecundadas. La media de la diferencia  $\bar{P} - \bar{H}$  es sólo de origen genético, y como  $E[\bar{P}] = \sum_u p_u \delta_{uu}$  y  $E[\bar{H}] = F \sum_u p_u \delta_{uu}$ ,  $E[\bar{P} - \bar{H}] = (1-F) \sum_u p_u \delta_{uu}$ . Ahora, como lo que se desea estimar es  $\tilde{H} = \left( \sum_u p_u \delta_{uu} \right)^2$ , y como

$$E[\bar{P} - \bar{H}]^2 = (1-F)^2 \left( \sum_u p_u \delta_{uu} \right)^2 + Var[\bar{P} - \bar{H}],$$

entonces

$$E \left[ \frac{1}{(1-F)^2} (\bar{P} - \bar{H})^2 \right] = \left( \sum_u p_u \delta_{uu} \right)^2 + \frac{1}{(1-F)^2} Var[\bar{P} - \bar{H}]$$

Además, para un experimento de evaluación con r repeticiones, como todas las observaciones del experimento son independientes, si  $\sigma^2$  representa la varianza del error experimental,

$$Var[\bar{P} - \bar{H}] = Var(\bar{P}) + Var(\bar{H}) = \frac{\sigma^2}{pr} + \frac{\sigma^2}{p(p-1)r} = \sigma^2 / [r(p-1)]$$

Así,

$$E \left[ \frac{1}{(1-F)^2} (\bar{P} - \bar{H})^2 \right] = \left( \sum_u p_u \delta_{uu} \right)^2 + \frac{\sigma^2}{(1-F)^2 r(p-1)}$$

Por esta razón, un estimador insesgado de  $\tilde{H}$  basado en el cuadrado medio del error experimental (CME) es

$$\hat{H} = \frac{1}{(1-F)^2} (\bar{P} - \bar{H})^2 - CME / [r(p-1)(1-F)^2] \quad (Ec. 14)$$

Para generar estimadores insesgados de  $\sigma_A^2$ ,  $D_1$  y  $D_2$ , considérese el sistema formado por las Ecs. 11, 12 y 13:

$$2\sigma_A^2 + 4D_1 + D_2 = \sigma_p^2$$

$$\sigma_A^2 + (1 + F)D_1 + FD_2 / 2 = Cov[P, FA]$$

$$(1/2)\sigma_A^2 + FD_1 + (F^2 / 4)D_2 = \sigma_{MHA}^2$$

En virtud de que la matriz que forman los coeficientes en los lados izquierdos de estas ecuaciones y su inversa son, respectivamente,

$$\begin{bmatrix} 2 & 4 & 1 \\ 1 & 1 + F & F/2 \\ 1/2 & F & F^2/4 \end{bmatrix} \text{ y } \frac{1}{(F-1)^2} \begin{bmatrix} F^2/2 & -2F & 2 \\ -F/2 & 1 + F & -2 \\ 1 & -4 & 4 \end{bmatrix},$$

los estimadores insesgados de  $\sigma_A^2$ ,  $D_1$  y  $D_2$  son,

$$\hat{\sigma}_A^2 = (F^2 / 2)\hat{\sigma}_p^2 - 2F\hat{Cov}(P, FA) + 2\hat{\sigma}_{MHA}^2 \tag{Ec. 15}$$

$$\hat{D}_1 = -(F/2)\hat{\sigma}_p^2 + (1 + F)\hat{Cov}(P, FA) - 2\hat{\sigma}_{MHA}^2 \tag{Ec. 16}$$

$$\hat{D}_2 = \hat{\sigma}_p^2 - 4\hat{Cov}(P, FA) + 4\hat{\sigma}_{MHA}^2 \tag{Ec. 17}$$

Finalmente, de la Ec. 10, un estimador insesgado de  $\sigma_D^2$  es

$$\hat{\sigma}_D^2 = 4^n \left[ \hat{\sigma}_{HCA}^2 - \hat{\sigma}_A^2 - 2F\hat{D}_1 + (F^2 / 2)\hat{D}_2 \right] \tag{Ec. 18}$$

Como se puede apreciar, para aplicar los estimadores de  $\sigma_A^2$ ,  $\sigma_D^2$ ,  $\bar{H}$ ,  $D_1$  y  $D_2$  Ecs. 14 a 18, es necesario a su vez generar estimadores de  $\sigma_p^2$ ,  $\sigma_{HCA}^2$ ,  $Cov(P, FA)$ , etc. Los estimadores  $\hat{\sigma}_p^2$  y  $\hat{\sigma}_{HCA}^2$  pueden ser generados en el análisis de varianza de la información producida por la

evaluación de campo de las  $p^2$  familias de hermanos completos autofecundados (Cuadro 1). Para la estimación de  $Cov(P, FA)$  se cuenta con r observaciones para cada progenitor y para las familias correspondientes de medios hermanos. Con base en estos pr pares de valores experimentales se hace la estimación de la covarianza.

Cuadro 1. Análisis de varianza de las  $p^2$  familias de un dialélico completo de p progenitores endogámicos no emparentados para estimar componentes de varianza genética. Las familias se autofecundaron durante n generaciones.

Fuentes de variación	Cuadros medios	Esperanzas de cuadrados medios	Grados de libertad
Bloques			r-1
Familias	T	$\sigma^2 + r\sigma_{HCA}^2$	$p^2-1$
Progenitores (P)	P	$\sigma^2 + r\sigma_p^2$	p-1
Cruzas (H) <sup>†</sup>	H	$\sigma^2 + r\sigma_C^2$	p(p-1)-1
P vs H	P vs H	$\sigma^2 + r(p-1)(1-F)^2 \left( \sum_u p_{iu} \delta_{uu} \right)^2$	1
Error	E	$\sigma^2$	(r-1)(p <sup>2</sup> -1)

<sup>†</sup> Sólo las que involucran progenitores diferentes.

Respecto a  $\sigma_{MHA}^2$ , en virtud de que se trata de la varianza entre las medias de las familias de medios hermanos autofecundados se puede, sólo con el fin de estimar  $\sigma_{MHA}^2$ , hacer el análisis como si se tratara de un bloques al azar generalizado con p tratamientos. Las repeticiones del tratamiento asociado al progenitor i (i=1,2,..., p) dentro de un mismo bloque, son las p parcelas asignadas a las familias de hermanos completos autofecundados que tienen en común a tal progenitor i (Cuadro 2). El estimador  $\sigma_{MHA}^2$  es entonces de la forma

$$\hat{\sigma}_{MHA}^2 = \frac{T - BT}{rp}$$

en donde T y BT son los cuadrados medios debidos a los tratamientos y a la interacción entre bloques y tratamientos, respectivamente. Desde luego, otra alternativa para estimar  $\sigma_{MHA}^2$  es formar un compuesto balanceado de la semilla de las p familias que tiene cada progenitor común. Esta semilla corresponderá a una familia de medios hermanos que puede ser incluida en el experimento de evaluación. Sin embargo, en virtud de que la evaluación de p familias adicionales, además del incremento de costos de

investigación que esto demanda, es innecesaria, se considera que la visualización de la evaluación de las familias de medios hermanos en el experimento en un diseño en bloques al azar generalizado es una buena estrategia.

Cuadro 2. Análisis de varianza de las  $p^2$  cruzas de un dialélico completo autofecundadas durante  $n$  generaciones en el que las  $p$  cruzas directas de cada progenitor en cada bloque se consideran como  $p$  repeticiones de un mismo tratamiento.

Fuentes de variación	Grados de libertad	Cuadrados medios	Esperanzas de cuadrados medios
Bloques (B)	$r-1$	<b>B</b>	
Tratamientos (T)	$p-1$	<b>T</b>	$\sigma^2 + p\sigma_{bt}^2 + rp\sigma_{MHA}^2$
B x T	$(r-1)(p-1)$	<b>BT</b>	$\sigma^2 + p\sigma_{bt}^2$
Error	$rp(p-1)$	<b>E</b>	$\sigma^2$

### DISCUSIÓN

Molina (1993), en su estudio sobre el problema que aquí se investiga, y con base en el modelo de dos alelos por locus y frecuencias alélicas iguales a 0.5, generó sendos estimadores para  $\sigma_A^2$  y  $\sigma_D^2$ . En términos de la notación aquí utilizada, éstos fueron de la forma

$$\hat{\sigma}_A^2 = 2\hat{\sigma}_{MHA}^2 \quad \text{y} \quad \hat{\sigma}_D^2 = 4^n \left[ \hat{\sigma}_{HCA}^2 - 2\hat{\sigma}_{MHA}^2 \right]$$

Con base en el modelo de alelos múltiples aquí se ha encontrado que

$$\sigma_{MHA}^2 = (1/2)\sigma_A^2 + FD_1 + (1/4)F^2D_2$$

de manera que, en términos de este modelo, el estimador de  $\sigma_A^2$  generado por Molina (1993) tiene un valor esperado igual a  $\sigma_A^2 + 2FD_1 + (1/2)F^2D_2$ ; es decir, su sesgo es igual a  $2FD_1 + (1/2)F^2D_2$ . Desde luego, la magnitud de éste depende de las magnitudes de  $F$ ,  $D_1$  y  $D_2$ . Sin embargo, un análisis detallado del sesgo con base en el modelo de alelos múltiples es complicado por la multiplicidad de situaciones que se puede generar para tipos de acción y frecuencias génicas, además de que  $D_1$ , por ser una covarianza, puede tomar valores negativos o positivos. Con dos alelos por locus  $\alpha_1 = q\alpha$ ,  $\alpha_2 = -p\alpha$ ,  $\delta_{11} = -2q^2\mathbf{d}$  y  $\delta_{22} = -2p^2\mathbf{d}$ , en donde  $p$  y  $q$  son las frecuencias de  $A_1$  y  $A_2$ , y  $\alpha$  y  $\mathbf{d}$  son el efecto promedio de la sustitución de un gene y el valor genotípico codificado del genotipo  $A_1A_2$  (Falconer, 1989). Sahagún (2002), omitiendo pasos de la

derivación, mostró expresiones para  $D_1$  y  $D_2$ . En detalle, para dos alelos por locus,

$$\begin{aligned} D_1 &= \sum_{i=1}^2 p_i \alpha_i \delta_{ii} \\ &= -2pq^3 \alpha \mathbf{d} + 2p^3 q \alpha \mathbf{d} \\ &= 2pq(p^2 - q^2) \alpha \mathbf{d} \\ &= 2pq(p - q) \alpha \mathbf{d} \end{aligned}$$

y

$$\begin{aligned} D_2 &= \sum_i p_i \delta_{ii}^2 - \left( \sum_i p_i \delta_{ii} \right)^2 \\ &= 4pq^4 \mathbf{d}^2 + 4p^4 q \mathbf{d}^2 - \left[ -2pq^2 \mathbf{d} - 2p^2 q \mathbf{d} \right]^2 \\ &= 4pq(p - q)^2 \mathbf{d}^2 \end{aligned}$$

Por lo tanto, respecto al modelo de alelos múltiples el sesgo del estimador de  $\sigma_A^2$  derivado por Molina (1993) es  $2FD_1 + (1/4)F^2D_2 = 4Fpq(p - q)\alpha\mathbf{d} + (1/4)F^2 4pq(p - q)^2 \mathbf{d}^2$

Claramente, con  $p = q = 0.5$  este sesgo se hace cero, como ya lo había indicado Cockerham (1983) para  $D_1$  y  $D_2$ . Presumiblemente, por esta razón Molina (1993) propuso el estimador con base en un modelo de frecuencias iguales, particularmente aplicable a poblaciones derivadas de la cruce entre dos líneas puras. Por otra parte, para cualesquiera frecuencias génicas, la covarianza  $D_1$  puede ser negativa si y sólo si  $(p-q)\alpha\mathbf{d} < 0$ . Para el caso de interés aplicado, por su frecuente ocurrencia en el mejoramiento genético, en que  $0 < \mathbf{d} \leq \mathbf{a}$ ,  $(p-q)\alpha\mathbf{d}$ , y, en consecuencia,  $D_1$  es negativa siempre que la frecuencia del gene más deseable sea inferior que la de su alelo ( $q > p$ ). Por el contrario,  $D_1$  será positiva siempre que  $p > q$ ; es decir, en poblaciones élite  $D_1$  siempre será positiva y es en éstas en donde se debe esperar que  $D_1$  esté asociada con los mayores sesgos en la estimación de la varianza aditiva. Esto es así debido a que la otra componente del sesgo,  $D_2$ , por ser una varianza, nunca puede ser negativa.

Respecto a la varianza de dominancia, debido a que con el modelo de alelos múltiples y cualesquiera frecuencias génicas, según la Ec. 12,

$$\sigma_{MHA}^2 = (1/2)\sigma_A^2 + FD_1 + (1/4)F^2D_2$$

entonces, al considerar la Ec. 10 se obtiene que

$$\begin{aligned} E[\hat{\sigma}_D^2] &= E\left\{4^n \left[ \hat{\sigma}_{HCA}^2 - 2\hat{\sigma}_{MHA}^2 \right] \right\} \\ &= 4^n \left[ \sigma_A^2 + (1/2)^{2n} \sigma_D^2 + 2FD_1 + (1/2)F^2D_2 \right] \\ &\quad - \sigma_A^2 - 2FD_1 - (1/2)F^2D_2 \\ &= \sigma_D^2 \end{aligned} \quad (\text{Ec.19})$$

Esto significa que el estimador de  $\sigma_D^2$  propuesto por Molina (1993) para el caso del modelo de dos alelos por *locus* con frecuencias alélicas de 1/2 es insesgado. Interesantemente, este estimador también es insesgado para el caso de frecuencias alélicas desiguales y aún para alelos múltiples con cualesquiera frecuencias alélicas. Eso es lo que deja ver el resultado de la Ecuación 15. Éste es general en términos de número de alelos y en términos de frecuencias alélicas.

En seguimiento a la investigación de Molina (1993), Márquez y Molina (1995) determinaron la varianza genotípica entre familias autofecundadas de hermanos completos y de medios hermanos para frecuencias génicas arbitrarias, con base en un modelo de dos alelos, pero no derivaron estimadores de  $\sigma_A^2$  y  $\sigma_D^2$ .

Respecto a la magnitud del coeficiente de endogamia, cuando  $F = 1$  el sistema de componentes de varianza y sus estimadores correspondientes se trastoca. Con  $F = 1$ , la varianza genética de la población Ec. 1 se reduce a

$$\sigma_G^2 = 2\sigma_A^2 + 4D_1 + D_2$$

Es decir, han desaparecido explícitamente la varianza de dominancia y  $\bar{H}$  como componentes de la varianza genética de la población. La varianza de dominancia desaparece porque en una población con un coeficiente de endogamia igual a 1 sólo hay desviaciones de dominancia asociadas a genotipos cuyos dos genes son idénticos por descendencia, y la varianza de estas desviaciones de dominan-

cia no es  $\sigma_D^2$ , sino  $D_2$ . De hecho, en cada generación de autofecundación la proporción de individuos cuyos genotipos están constituidos por dos alelos que no son idénticos por descendencia se reduce en 50 %, lo que disminuye la información sobre la varianza de dominancia y conlleva a una reducción en la precisión de su estimación (Stuber, 1970). Ya con  $F = 1$  la población debe tener un arreglo genotípico igual al de los progenitores; de ahí que la varianza de esta población sea igual a la varianza derivada para los progenitores Ec. 11.

Respecto a  $\bar{H}$ , el cuadrado de la media de las desviaciones de dominancia de individuos portadores de alelos idénticos por descendencia, si bien es parte de  $D_2$ , con  $F = 1$  ya no aparece en forma aislada. De hecho, en cada generación se incrementa la proporción de individuos homocigóticos por portar alelos idénticos por descendencia; la autofecundación de genotipos portadores de alelos no idénticos por descendencia produce una progenie en la cual 50 % de los individuos son homocigóticos por portar alelos idénticos por descendencia.

En el otro extremo, con  $F=0$  por excluir autofecundaciones,  $\sigma_{MHA}^2$  y  $\sigma_{HCA}^2$  no serán parte del escenario y la varianza de la población Ec. 7 se reducirá a la expresión  $\sigma_{GT}^2 = \sigma_A^2 + \sigma_D^2$ , y los estimadores serán de la forma  $\hat{\sigma}_A^2 = 2\hat{\sigma}_{MH}^2$  y  $\hat{\sigma}_D^2 = \hat{\sigma}_{HC}^2 - 2\hat{\sigma}_{MH}^2$ , en donde  $\sigma_{MH}^2$  y  $\sigma_{HC}^2$  corresponden a la varianza entre familias de medios hermanos y a la varianza entre familias de hermanos completos, respectivamente.

Respecto a la población cuyos parámetros son objeto de estimación, habrá que considerar el origen de los progenitores; si éstos se obtienen de una muestra aleatoria de una población en apareamiento aleatorio, entonces las varianzas estimadas ( $\sigma_A^2$  y  $\sigma_D^2$ ) serán las de la población originalmente muestreada. Sahagún (1998, 2000, 2001) consideró este tipo de muestreos para determinar el efecto del nivel de endogamia de los progenitores en varios diseños de apareamiento. Por otra parte, la población cuyos componentes de varianza son objeto de estimación también puede ser la que genera el apareamiento aleatorio de los progenitores (Sahagún, 1999); en este caso los progenitores debieron haber sido seleccionados de acuerdo con criterios preestablecidos y no con una muestra aleatoria de alguna población. Molina (1993) y Márquez y Molina (1995) abundan sobre el tema.

## CONCLUSIONES

El estimador de la varianza aditiva ( $\sigma_A^2$ ) asociada a un dialélico completo de  $p$  progenitores endogámicos no emparentados en que las cruzas se autofecundan consecutivamente durante  $n$  generaciones es sesgado si se deriva con base en un modelo de dos alelos con frecuencias génicas iguales. El sesgo se establece con respecto a la  $\sigma_A^2$  de una población donde haya más de dos alelos con frecuencias génicas arbitrarias; tal sesgo es  $2FD_1 + (1/4)F^2D_2$ , en donde  $F$  es el coeficiente de endogamia  $\left[1 - (1/2)^n\right]$ ;  $D_1$  es la covarianza entre los efectos aditivos de los genes y las desviaciones de dominancia de los genotipos que son homocigóticos por estar formados por dos genes idénticos por descendencia, y  $D_2$  es la varianza de estas desviaciones de dominancia. En el caso de dos alelos cuyas frecuencias son iguales, el sesgo es igual a cero. Sin embargo, en la medida en que el gene más deseable tiene una mayor frecuencia que el otro la magnitud del sesgo se incrementa. Por otra parte, el estimador de la varianza de dominancia propuesto para el caso de dos alelos por *locus* siempre es insesgado. Finalmente, es de resaltar que en el caso  $F = 1$ , la varianza de la población generada por todas las cruzas de la  $p$  progenitores sólo incluye los componentes  $\sigma_A^2$ ,  $D_1$  y  $D_2$ , debido a que al no haber genotipos heterocigóticos no habrá las desviaciones de dominancia, lo que imposibilita la existencia de la varianza de dominancia; como el arreglo genotípico se reduce a sólo  $\sum p_u A_u A_u$ , con respecto a las  $\delta_{uu}$  sólo  $D_2$  puede ser definido.

## BIBLIOGRAFÍA

- Benítez R I (1999)** Respuesta esperada en líneas de trigo seleccionadas con base en el comportamiento de sus mestizos autofecundados. *Rev. Fitotec. Mex.* 22:121-131.
- Cockerham C C (1983)** Covariance of relatives from self-fertilization. *Crop Sci.* 23:1177-1180.
- Cockherman C C, D F Matzinger (1985)** Selection response based on selfed progenies. *Crop Sci.* 25:583-488.
- Falconer D S (1989)** *Introduction to Quantitative Genetics*. Third Ed. Longman. Londres, New York. 438 p.
- Kempthorne O (1969)** *An Introduction to Genetic Statistics*. John Wiley and Sons, New York. 545 p.
- Márquez S F, J D Molina G (1995)** Varianza genética aditiva en cruzas masiva, individuos y familias masivas  $F_3$  en autógamias. *Rev. Fitotec. Mex.* 18:16-24.
- Molina G J D (1993)** Varianza de familias de hermanos completos y medios hermanos en generaciones avanzadas de autofecundación y su uso en los diseños de apareamiento. I. Progenitores homocigóticos. *Agrociencia Serie Fitociencia* 4:93-104.
- Sahagún C J (1998)** Precisión de estimadores del diseño I con diferentes niveles endogámicos en diferentes asignaciones de hembras a machos. *Rev. Fitotec. Mex.* 21:69-81.
- Sahagún C J (1999)** Simultaneous estimation of genotypic mean and variance of synthetic varieties from inbred and unrelated parents. *J. Genet. Breed.* 53:293-300.
- Sahagún C J (2000)** Estimación de varianzas genéticas con medios hermanos maternos con diferentes niveles endogámicos y repeticiones. *Agrociencia* 34:21-34.
- Sahagún C J (2001)** Estimación de varianzas con progenitores endogámicos de familias independientes de hermanos completos. *Agrociencia* 35:169-179.
- Sahagún C J (2002)** Avance genético en aptitud combinatoria con mestizos autofecundados, multialelismo y probadores homocigóticos. *Agrociencia* 36:67-81.
- Stuber C W (1970)** Estimation of genetic variances using inbred relatives. *Crop Sci.* 10:129-135.
- Weir B S, C C Cockerham (1977)** Two-locus theory in quantitative genetics. *In: Proc. Intnatl. Conf. Quantitative Genetics*. E Pollak, O Kempthorne, TB Bailey Jr (eds). Aug. 16-21 Ames, Iowa. pp: 247-270.