

## Cálculo de la endogamia en la selección recurrente de familias en el maíz\*

### Inbreeding calculation in the recurrent family selection in maize

Fidel Márquez-Sánchez<sup>1§</sup>

<sup>1</sup>Universidad Autónoma Chapingo. Centro Regional Universitario Occidente, Guadalajara, Jalisco. <sup>§</sup>Autor para correspondencia: fidelmqz@hotmail.com.

#### Resumen

La selección recurrente garantiza el mejoramiento del rendimiento a largo plazo, además de que el fitomejorador puede obtener resultados parciales en algún ciclo de selección. La selección recurrente ha sido practicada por un sin número de investigadores en el maíz, obteniéndose respuestas por ciclo que han garantizado continuar con dicho método a través de veinte o más ciclos de selección. Sin embargo, el autor desconoce que excepto en una investigación, se haya calculado la endogamia que se va generando a través de la selección. Por lo tanto, el presente estudio tiene por objeto revisar dicha investigación y proponer un mejor enfoque.

**Palabras clave:** *Zea mays* L., selección recurrente, tipos de familias en el maíz, teoría de la endogamia.

#### Introducción

En Hallauer *et al.* (2010) se reportan varias investigaciones en la selección recurrente. En la selección modificada mazorca por surco (esencialmente selección entre y dentro de familias de medios hermanos) la ganancia por ciclo fue 5.1% en 9 investigaciones con 3 a 10 ciclos de selección; para

#### Abstract

Recurrent selection ensures an improved long-term yielding; also the breeder may obtain partial results in any selection cycle. Recurrent selection has been practiced by a number of researchers in maize, obtaining responses per cycle that have guaranteed to continue this method through twenty or more cycles of selection. However, the author is unaware that except in one research, inbreeding generated through the selection has been calculated. Therefore, the present study aims to review that research and propose a better approach.

**Key words:** *Zea mays* L., recurrent selection, family types in maize, theory of inbreeding.

#### Introduction

Hallauer *et al.* (2010) reported several studies in recurrent selection. In the modified selection ear/row (essentially a selection among and within half-sibling families), the gain per cycle was 5.1% in 9 researches with 3 to 10 cycles of selection; for the selection of full-sibling families, the gain per cycle was 5.8% in 10 investigations with 4 to 10 cycles of selection; with the lines S<sub>1</sub>, the gain per cycle was 4.5% in 18 investigations with 3 to 8 rounds of selection.

\* Recibido: julio de 2012  
Aceptado: enero de 2013

la selección de familias de hermanos completos la ganancia por ciclo fue 5.8% en 10 investigaciones con 4 a 10 ciclos de selección; con líneas  $S_1$  la ganancia por ciclo fue 4.5% en 18 investigaciones con 3 a 8 ciclos de selección.

Aunque los ciclos de selección recurrente duran 1 año en la selección mazorca por surco, 2 años en la selección de hermanos completos y 3 años en la selección de líneas  $S_1$ , puede advertirse que aun considerando la ganancia por ciclo, la selección recurrente de familias es superior a la selección masal; en ésta la ganancia por año es 3.1% en 24 investigaciones con 3 a 17 ciclos de selección. Si se calculan la ganancia por año en las investigaciones de selección de familias, ésta es 3.3%, un poco superior a la selección masal; sin embargo, la selección de familias ya es más frecuente debido a que se obtienen datos de experimentos foráneos de las unidades de selección lo cual es muy difícil de hacer en la selección masal.

En el aspecto teórico, Márquez-Sánchez (2009) calculó la endogamia en la selección de líneas  $S_1$ , en las familias de hermanos completos (HC) y en las familias de medios hermanos (MH), usando el método de Falconer (1964) para 180 familias y una intensidad de selección de 0.0555 resultante de seleccionar las 10 mejores familias. Las fórmulas de la endogamia de Márquez-Sánchez (2009) en el ciclo de selección  $t$  fueron:

$$F(S_1)_t = [1/2(nm-1)][(m-1) + 2m(n-1)F_{t-1} + (m-1)F_{t-2}] \quad [1]$$

$$F(HC)_t = [1/4(nm-1)][(m-1) + 4m(n-1)F_{t-1} + 2(m-1)F_{t-2} + (m-1)F_{t-3}] \quad [2]$$

$$F(MH)_t = [1/8(nm-1)][(m-1) + 8m(n-1)F_{t-1} + 6(m-1)F_{t-2} + (m-1)F_{t-3}] \quad [3]$$

Donde:  $n$  = a el número de familias; y  $m$  = el número de plantas por familia.

En dicha investigación la recombinación se hizo a través del diseño dialélico. Para ajustar a  $n$  y  $m$  por el número efectivo de varianza (Crossa y Venkovsky, 1997), no fue posible usar a  $n=10$  ya que se obtenían resultados negativos de  $m$ , de manera que se usaron los siguientes números de familias y de plantas por familia:  $S_1$ ,  $n=9$  y  $m=5$ ; HC,  $n=6$  y  $m=2$ ; y MH,  $m=7$  y  $m=3$ . Al obtener el número de familias seleccionadas para el ciclo de selección siguiente, la recombinación de  $n(n-1)/2$  tuvo que replicarse en cada tipo de familia resultando:  $S_1$ ,  $36 \times 1 \times 5$ ; HC,  $15 \times 3 \times 4$ ; y MH,  $21 \times 2.857 \times 3$ , en donde los números centrales corresponden a las replicaciones.

Although recurrent selection cycles last 1 year in the selection ear/row, two years in full-sib selection and three years in the selection of lines  $S_1$ , it can be seen that even considering the gain per cycle, recurrent selection of families is superior to mass selection; in this, the gain is 3.1% per year in 24 investigations with 3 to 17 cycles of selection. If the gain is calculated annually in research of family selection, it is 3.3%, slightly higher than mass selection; however, the selection of families is more common because experimental data are obtained from foreign units of selection which is quite difficult to do in mass selection.

On the theoretical side, Márquez-Sánchez (2009) estimated inbreeding in the selection of lines  $S_1$ , full-sib families (HC) and half-sib families (MH), using Falconer's (1964) method, for 180 families and a selecting intensity of 0.0555 resulting from selecting the 10 best families. The formulas of inbreeding by Márquez-Sánchez (2009) in the  $t$  selection cycle were:

$$F(S_1)_t = [1/2(nm-1)][(m-1) + 2m(n-1)F_{t-1} + (m-1)F_{t-2}] \quad [1]$$

$$F(HC)_t = [1/4(nm-1)][(m-1) + 4m(n-1)F_{t-1} + 2(m-1)F_{t-2} + (m-1)F_{t-3}] \quad [2]$$

$$F(MH)_t = [1/8(nm-1)][(m-1) + 8m(n-1)F_{t-1} + 6(m-1)F_{t-2} + (m-1)F_{t-3}] \quad [3]$$

Where:  $n$  = the number of families and  $m$  = the number of plants per family.

In this investigation, the recombination was done through a diallel design. To adjust  $n$  and  $m$  by the effective number of variance (Crossa and Venkovský, 1997), it was not possible to use  $n=10$  because negative results were obtained for  $m$ , so that the following numbers of families and plants per family were used:  $S_1$ ,  $n=9$  and  $m=5$ ; HC,  $n=6$  and  $m=2$ ; and MH,  $m=7$  and  $m=3$ . When gathering the number of families selected for the next round of selection, the recombination of  $n(n-1)/2$  had to be replicated in each family type, resulting as:  $S_1$ ,  $36 \times 1 \times 5$ ; HC,  $15 \times 3 \times 4$ ; and MH,  $21 \times 2.857 \times 3$ , where the central numbers correspond to the replications.

For the first selection cycle, inbreeding values were:  $S_1$ , 0.00298; HC, 0.01895; and MH, 0.00018, maintaining in long-term superiority of inbreeding HC on  $S_1$  and MH, a portion of which is shown in what follows for the cycle 10:  $S_1$ , 0.02932; HC, 0.16394; and MH, 0.00181.

En el primer ciclo de selección los valores de la endogamia fueron:  $S_1$ , 0.00298; HC, 0.01895; y MH, 0.00018, manteniéndose a largo plazo la superioridad de la endogamia de HC sobre  $S_1$  y MH, una parte de lo cual se muestra en lo que sigue para el ciclo 10:  $S_1$ , 0.02932; HC; 0.16394; y MH, 0.00181.

La presente investigación parte de 300 plantas de cuyas mazorca se han obtenido 300 familias, independientemente del tipo de polinización manual que se haya usado para producir las familias. En ciclos posteriores las plantas provienen de la recombinación mediante el diseño dialélico entre las  $n=20$  familias seleccionadas, obteniéndose así a las nuevas 300 familias del producto  $[n(n-1)/2]R=190R$ , en donde R es un factor que indica el número de dialélicos que se deben hacer.

Cada ciclo de selección se inicia con la experimentación de las familias y termina con la recombinación de las familias seleccionadas. Por lo tanto, ya desde el primer ciclo de selección tiene lugar la recombinación, la cual para ser óptima se hace mediante el diseño dialélico. La recombinación se hace cruzando una planta de cada familia con una planta del resto de las familias; con  $n=20$ , cada familia deberá contener 19 plantas útiles cada una de las cuales se cruzará con cada una de las 19 familias restantes, habrán por lo tanto 190 cruzamientos; sin embargo, es necesario hacer más cruzamientos para restaurar el número de 300 familias para el siguiente ciclo de selección, se necesita entonces hacer R veces el dialélico tal y como se ha dicho atrás.

Con el fin de evitar números pequeños de  $n$  y números efectivos ajustados por varianzas negativos de  $m$ , es necesario incrementar el tamaño de la población hasta  $N=300$  familias, o bien en ciclos posteriores al primero  $N=20$  familias  $\times$  15 plantas por familia = 300, además de calcular los números efectivos de varianzas  $[N_{e(v)}]$ , la intensidad de selección ( $s$ ) y los números ajustados de las recombinaciones ( $r$ ) y los números  $n$  y  $m$ .

$$N = nm = 300; n = 20 \text{ y } m = 15; s = 20/300 = 0.0666; r = 20(20-1)/2 = 190 \text{ Q} = r/m = 190/15 = 12.6666.$$

$$N_{e(v)}(AH) = N[s/(1 + 1/m)] = 300[0.0666/(1 + 1/15)] = 18.7324$$

$$N_{e(v)}(HC) = N[(2s)/(2-s)] = 300[0.1333/1.9333] = 20.6848$$

$$N_{e(v)}(MH) = N[(4s)/(4-s)] = 300[0.2666/3.9333] = 20.3183$$

This research started with 300 plants, whose ears crated 300 families, regardless of the type of manual pollination that has been used to produce families. In subsequent cycles, the plants come through recombination in a diallelic design between the  $n=20$  families selected, thus obtaining the new product 300 families  $[n(n-1)/2]R=190R$ , where R is a factor that indicates the number of diallels to be made.

Each selecting cycle begins with the experimentation of the families and ends with the recombination of the selected families. Therefore, since the first selection cycle of recombination takes place, which is optimal for diallelic designing. Recombination is crossing a plant of each plant family with another plant from the other family; with  $n=20$ , each family must contain 19 useful plants each of which will intersect with each of the remaining 19 families, having 190 crosses; however, it's necessary to make even more crosses to restore the 300 families for the next round of selection, then it's necessary to make R times the diallel as stated earlier.

In order to avoid small numbers of  $n$  and effective adjusted numbers by negative variance of  $m$ , it's necessary to increase the size of the population to  $N=300$  families, or in cycles after the first  $N=20$  families  $\times$  15 plants per family = 300, besides calculating the effective number of variance  $[N_{e(v)}]$ , the intensity of selection ( $s$ ) and the adjusted numbers of recombination ( $r$ ) and the numbers  $n$  and  $m$ .

$$N = nm = 300; n = 20 \text{ y } m = 15; s = 20/300 = 0.0666; r = 20(20-1)/2 = 190 \text{ Q} = r/m = 190/15 = 12.6666.$$

$$N_{e(v)}(AH) = N[s/(1 + 1/m)] = 300[0.0666/(1 + 1/15)] = 18.7324$$

$$N_{e(v)}(HC) = N[(2s)/(2-s)] = 300[0.1333/1.9333] = 20.6848$$

$$N_{e(v)}(MH) = N[(4s)/(4-s)] = 300[0.2666/3.9333] = 20.3183$$

Hereinafter, the recombination numbers ( $r$ ), families ( $n$ ) and family of plants ( $m$ ) is adjusted by the effective number of variance (respectively:  $r'$ ,  $n'$  and  $m'$ ) for the three types of families according to Crossa and Venkovský (1997):

For lines  $S_1$ :

$$r' = [Ne(v)(S_1)Q]^{1/2} = [18.7324 \times 12.6666]^{1/2} = 15.4037$$

$$m' = [Ne(v)(S_1)/Q]^{1/2} = [18.7324/12.6666]^{1/2} = 1.2161$$

Therefore:

$$r' = [(n'^2 - n')/2] \times 1.2161 = 15.4037$$

En lo que sigue los números de recombinaciones ( $r$ ), de familias ( $n$ ) y de plantas por familia ( $m$ ) se ajustarán por el número efectivo de varianzas (respectivamente:  $r'$ ,  $n'$  y  $m'$ ) para los tres tipos de familias, de acuerdo a Crossa y Venkovsky (1997):

Para líneas  $S_1$ :

$$r' = [N_{e(v)}(S_1)Q]^{1/2} = [18.7324 \times 12.6666]^{1/2} = 15.4037$$

$$m' = [N_{e(v)}(S_1)/Q]^{1/2} = [18.7342/12.6666]^{1/2} = 1.2161$$

Por lo tanto:

$$r' = [(n'^2 - n')/2] \times 1.2161 = 15.4037$$

Donde:

$$n' = 5.5579.$$

Para HC:

$$r' = [N_{e(v)}(HC)Q]^{1/2} = [20.6848 \times 12.6666]^{1/2} = 16.1884$$

$$m' = [N_{e(v)}(HC)/Q]^{1/2} = [20.6848/12.6666]^{1/2} = 1.2779$$

Por lo tanto,

$$r' = [(n'^2 - n')/2] \times 1.2780 \approx 16.1884$$

Donde:

$$n' = 5.5582$$

Para MH:

$$r' = [N_{e(v)}(MH)Q]^{1/2} = [20.3188 \times 12.6666]^{1/2} = 16.0427$$

$$m' = [N_{e(v)}(MH)/Q]^{1/2} = [20.3389/12.6666]^{1/2} = 1.2665$$

Por lo tanto:

$$r' = [(n'^2 - n')/2] \times 1.2665 \approx 16.0427$$

Donde:

$$n' = 5.5580$$

$$r' = [(n'^2 - n')/2] \times 1.2665 \approx 16.0427$$

Usando las Ecuaciones [1], [2] y [3], tendremos:

$$F(S_1)_t = [1/2(nm - 1)][(m - 1) + 2m(n - 1)F_{t-1} + (m - 1)F_{t-2}]$$

$$F(S_1)_t = [1/11.5179][0.2161 + 11.0857 F_{t-1} + 0.2161 F_{t-2}]$$

$$F(HC)_t = [1/4(nm - 1)][(m - 1) + 4m(n - 1)F_{t-1} + 2(m - 1)F_{t-2} + (m - 1)F_{t-3}]$$

$$F(HC)_t = [1/24.4113][0.2779 + 23.2997 F_{t-1} + 0.5558 F_{t-2} + 0.2779 F_{t-3}]$$

$$F(MH)_t = [1/8(nm - 1)][(m - 1) + 8m(n - 1)F_{t-1} + 6(m - 1)F_{t-2} + (m - 1)F_{t-3}]$$

$$F(MH)_t = [1/48.3136][0.2665 + 46.1816 F_{t-1} + 1.5990 F_{t-2} + 0.2665 F_{t-3}]$$

Where:

$$n' = 5.5579.$$

For HC

$$r' = [N_{e(v)}(HC)Q]^{1/2} = [20.6848 \times 12.6666]^{1/2} = 16.1884$$

$$m' = [N_{e(v)}(HC)/Q]^{1/2} = [20.6848/12.6666]^{1/2} = 1.2779$$

Therefore:

$$r' = [(n'^2 - n')/2] \times 1.2780 \approx 16.1884$$

Where:

$$n' = 5.5582$$

To MH

$$r' = [N_{e(v)}(MH)Q]^{1/2} = [20.3188 \times 12.6666]^{1/2} = 16.0427$$

$$m' = [N_{e(v)}(MH)/Q]^{1/2} = [20.3389/12.6666]^{1/2} = 1.2665$$

Therefore:

$$r' = [(n'^2 - n')/2] \times 1.2665 \approx 16.0427$$

Where:

$$n' = 5.5580;$$

$$r' = [(n'^2 - n')/2] \times 1.2665 \approx 16.0427$$

Using the equations [1], [2] and [3], we have:

$$F(S_1)_t = [1/2(nm - 1)][(m - 1) + 2m(n - 1)F_{t-1} + (m - 1)F_{t-2}]$$

$$F(S_1)_t = [1/11.5179][0.2161 + 11.0857 F_{t-1} + 0.2161 F_{t-2}]$$

$$F(HC)_t = [1/4(nm - 1)][(m - 1) + 4m(n - 1)F_{t-1} + 2(m - 1)F_{t-2} + (m - 1)F_{t-3}]$$

$$F(HC)_t = [1/24.4113][0.2779 + 23.2997 F_{t-1} + 0.5558 F_{t-2} + 0.2779 F_{t-3}]$$

$$F(MH)_t = [1/8(nm - 1)][(m - 1) + 8m(n - 1)F_{t-1} + 6(m - 1)F_{t-2} + (m - 1)F_{t-3}]$$

$$F(MH)_t = [1/48.3136][0.2665 + 46.1816 F_{t-1} + 1.5990 F_{t-2} + 0.2665 F_{t-3}]$$

Now, if we calculate the inbreeding for the first round of selection we have:  $S_1$ , 0.0187; HC, 0.0114; and MH, 0.0055, results expected according to the family type, and not those enunciated in Márquez- Sánchez (2009):  $S_1$ , 0.02932; HC; 0.16394; y MH, 0.00181, where the largest size of inbreeding in the first round of selection is the family HC.

In the selection of lines  $S_1$ , we have used 300 families in the first cycle to select 20; however, in order to restore the number of families for the next generation, it is necessary to consider each type of family.

Ahora, si calculamos la endogamia para el primer ciclo de selección tendremos:  $S_1$ , 0.0187; HC, 0.0114; y MH, 0.0055, resultados que son los esperados conforme el tipo de familia, y no los que enunciamos en Márquez-Sánchez (2009) que son:  $S_1$ , 0.02932; HC; 0.16394; y MH, 0.00181, en donde el mayor tamaño de la endogamia en el primer ciclo de selección es el de las familias de HC.

En la selección de líneas  $S_1$  hemos usado 300 familias en el primer ciclo de selección para seleccionar 20; sin embargo, con el objeto de restaurar el número de familias para la siguiente generación, se hace necesario considerar a cada tipo de familia.

En el caso de las líneas  $S_1$  hay 20 líneas seleccionadas para la recombinación y de éstas habrá que hacer las nuevas 300 líneas; sin embargo, como resultado de la recombinación se obtienen 190 líneas, cuyo número hay que multiplicar por un factor (R) con el fin de obtener las 300 líneas; es decir, habrá que hacer tantas recombinaciones como lo indique R; aritméticamente el factor será  $R = 300/190 = 1.58$  pero en vez de éste se usará  $R = 2$ , con el que se obtendrían 380 líneas recombinadas, de las cuales habría que descartar 80 a fin de tener las 300 nuevas líneas  $S_1$ . Estará claro que con este proceso, con las 300 nuevas líneas la recombinación será menos intensa con las 380 líneas; es decir, en éstas habrá menos variación genética que si se usaran sólo las 300 líneas.

Si se usan familias de HC al hacer la recombinación se requiere de 2 plantas para obtener una familia de HC a fin de obtener las nuevas 300 familias, y el factor sería  $R = 2$ , ya que entonces habría  $2 \times 190 = 380$  nuevas familias de las que se escogerán 300 al azar para continuar con la selección.

Para el caso de familias de MH, si cada familia se produce cruzando una planta macho por dos plantas hembras, entonces el factor es  $R = 3$ , ya que entonces  $3 \times 190 = 570$ ; pero como fue el caso de las líneas  $S_1$ , sobrarían 270 familias las que habría que descartar para quedarse con sólo 300 nuevas familias; similarmente que para líneas  $S_1$  y familias de HC, con las 570 familias habría menos variación genética que si se usaran las 300 familias.

En los tres tipos de familia se tienen que hacer extracciones de las nuevas familias para llegar al número de 300. Esto implica un gran trabajo de campo, por lo que se recomendaría que el polinizador sólo hiciera un número de familias de alrededor de las 300.

In the case of lines  $S_1$ , there are 20 selected lines for the recombination and from these, it's necessary to make 300 new lines; however, as a result of the recombination 190 lines are obtained, whose number is multiplied by a factor (R) to get 300 lines, that is, it will take as many recombinations as directed by R; arithmetically the factor is  $R = 300/190 = 1.58$ , but instead of this,  $R = 2$  will be used: obtaining 380 recombinant lines, of which 80 would be discarded in order to have 300 new lines of  $S_1$ . It's clear that in this process, with the 300 new lines the recombination will be less intense with 380 lines, *i.e.* there will be less genetic variation than using only 300 lines.

If HC families are used to make the recombination, it's required 2 plants for a family HC in order to obtain the new 300 families, and the factor is  $R = 2$ , then there would be  $2 \times 190 = 380$  new families, 300 of which will be chosen at random for further selection. In the case of MH families, if each family occurs by crossing a male plant with two females, then the factor is  $R = 3$ , then  $3 \times 190 = 570$ : as it happened in the case with the lines  $S_1$ , 270 families would be spare, which would be ruled out to stay with only 300 new families; similarly as for the lines  $S_1$  and HC families, with 570 families there would be less genetic variation than if using 300 families.

In the three types of families, we have to make withdrawals of new families to reach 300. This requires a hard work in the field, so we recommend that the pollinator should make a number of families around 300.

## Conclusions

In order to minimize the effect of inbreeding in then recurrent selection in maize, the most important factor is the total size of the population that includes both, the number of families and the number of plants per family. In order to prevent that the families number are small and obtaining negative adjusted numbers by the effective number of variance of the plants per family, the researcher should test populations larger than 200. This research used 20 families and 15 plants per family for a total size of 300; with these numbers the expected inbreeding values were obtained, in the first cycle of selection of lines  $S_1$ , of full-sibling families and half-sibling families.

*End of the English version*



**Cuadro 1. Resultados de la hibridación acumulativa de las variedades progenitoras A, B, C, D y E y frecuencias de sus progenies.**  
**Table 1. Cumulative results of hybridization parent varieties A, B, C, D and E and their frequencies progenies.**

Tipo fam.	Recombinadas	Factor	Producto	Extraer	Número de nuevas familias
S <sub>1</sub>	190	2	380	80	300
HC	190	2	380	80	300
MH	190	3	570	270	300

## Conclusiones

Para evitar al máximo el efecto de la endogamia en la selección recurrente en el maíz, el factor más importante es el tamaño total de la población que incluye tanto al número de familias como al número de plantas por familia. Con el fin de evitar que los números de familias sean pequeños y que se obtengan números negativos ajustados por el número efectivo de varianza de las plantas por familia, el investigador debe probar tamaños de población mayores que 200. La presente investigación usó 20 familias y 15 plantas por familia para un tamaño total de 300; con estos números se obtuvieron los valores esperados de la endogamia, en el primer ciclo de selección de líneas S<sub>1</sub>, de familias de hermanos completos y de familias de medios hermanos.

## Literatura citada

- Crossa, J. and Venkovsky, R. 1997. Variance effective population size for two-stage sampling of monoecious species. *Crop Sci.* 37:14-26.
- Falconer, D. S. 1964. *Introduction to quantitative genetics.* Oliver and Boyd, Edinburgh and London. 365 pp.
- Hallauer, A. R.; Carena, M. J. and Miranda F. J. B. 2010. *Quantitative genetics in maize breeding.* Springer, New York. 663 pp.
- Márquez-Sánchez, F. 2009. Expected inbreeding in recurrent selection in maize. III: selection in S<sub>1</sub> lines and full-sib and half-sib families. *Maydica* 54:109-111.