

COEFICIENTE DE ENDOGAMIA GENERAL Y EXACTO DE SINTÉTICOS DE MAÍZ FORMADOS CON CRUZAS DOBLES

GENERAL AND EXACT INBREEDING COEFFICIENT OF MAIZE SYNTHETICS DERIVED FROM DOUBLE CROSSES

Juan E. Rodríguez-Pérez, Aureliano Peña-Lomelí, Clemente Villanueva-Verduzco, Jaime Sahagún-Castellanos*

Universidad Autónoma Chapingo. Km. 38.5, carretera México-Texcoco. 56230. Chapingo, Estado de México. México. (jsahagunc@yahoo.com.mx).

RESUMEN

El desarrollo tradicional de variedades sintéticas (VS's) es difícil, porque aún con pocos progenitores potenciales las VS's posibles serían demasiadas para formarse y evaluarse. Para la solución de este problema los progenitores pueden ser híbridos. En esta investigación se evaluó el uso de d cruzas dobles (CD's) como progenitores de una VS (Sin_{CD}) en vez de que el sintético se generara con $4d$ líneas (Sin_L). Ya se han derivado coeficientes de endogamia (CE's) del Sin_{CD} ($FSin_{CD}$'s), pero falta establecer su exactitud y grado de generalidad. La hipótesis de este estudio fue que los $FSin_{CD}$'s actuales son restringidos e inexactos. Como el desempeño de los sintéticos se relaciona con su CE, el objetivo de este estudio fue derivar un $FSin_{CD}$ general y exacto. Se asumió que las $4d$ líneas iniciales no eran parientes y que su CE fue $0 \leq F \leq 1$. La muestra de cada craza doble (m plantas) se visualizó inicialmente como una serie "grande" de x grupos de 16 plantas cada uno (los 16 genotipos del arreglo genotípico de la CD correspondiente). Por el apareamiento aleatorio que genera al Sin_{CD} , cada CD produce: 1) $16x = m$ autofecundaciones; 2) $(15)(16)x = 15m$ cruzas directas y recíprocas; y 3) $256x(x-1) = m(m-16)$ cruzas entre los x grupos. Los CE's de estos tres conjuntos son $1/2$, $(3+4F)/30$ y $(1+F)/8$, respectivamente, se concluyó que el CE exacto para m "grande" es: $FSin_{CD} = (1+F)/8d$. Con base en estos resultados se derivó el $FSin_{CD}$ general (para cualquier valor de m) y exacto.

Palabras clave: *Zea mays* L., arreglo genotípico, apareamiento aleatorio.

ABSTRACT

Developing synthetic varieties (VS's) using traditional methods is a difficult task. Even using few potential parents, the possible VS's would be too many to form and evaluate. Hybrid parents can be used to find a solution for this problem. In this research, the use of d double crosses (CD's) as parents of a VS (Sin_{CD}) was evaluated, instead of using $4d$ lines (Sin_L) to generate synthetics. Inbreeding coefficients (CE's) have been derived from Sin_{CD} ($FSin_{CD}$'s); however, their accuracy and generality degree have not been established. The hypothesis of this study was that the current $FSin_{CD}$'s are restricted and inaccurate. Because the performance of synthetics is related to their CE, the objective of this study was to derive a general and exact $FSin_{CD}$. It was assumed that the initial $4d$ lines were not relatives and that their CE (F) was $0 \leq F \leq 1$. The sample of each double cross (m plants) was initially visualized as a "large" series of x groups of 16 plants each (16 genotypes of the genotypic array of the respective CD). As a result of the random mating that generates a Sin_{CD} , each CD produces: 1) $16x = m$ selfpollinations; 2) $(15)(16)x = 15m$ direct and reciprocal crosses; and 3) $256x(x-1) = m(m-16)$ crosses between the x groups. The CE's of these three sets are $1/2$, $(3+4F)/30$, and $(1+F)/8$, respectively, and it was concluded that the exact CE for "large" m is $FSin_{CD} = (1+F)/8d$. Based on these results, the general and exact $FSin_{CD}$ was derived for any value of m .

Key words: *Zea mays* L., genotypic array, random mating.

INTRODUCTION

A synthetic variety (VS) of a crop species such as maize (*Zea mays* L.) is the population resulting from the random mating of its parents (Busbice, 1970). They should be homozygous lines with a high combining ability that have survived a selection process. Due to their origin and regeneration type, resulting from their random

*Autor responsable ❖ Author for correspondence.

Recibido: diciembre, 2017. Aprobado: mayo, 2018.

Publicado como ARTÍCULO en *Agrociencia* 53: 235-244. 2019.

INTRODUCCIÓN

Una variedad sintética (VS) de un cultivo como el maíz (*Zea mays* L.) es la población resultante del apareamiento aleatorio de sus progenitores (Busbice, 1970). Ellos deberían ser líneas homocigóticas con alta aptitud combinatoria sobrevivientes a un proceso de selección. Por su origen y tipo de regeneración, ambos por apareamiento aleatorio, Wricke y Weber (1986) sugieren que las VSs parecen una alternativa para los cultivos que se reproducen, en forma natural, en esta manera. También se ha enfatizado que estas variedades contienen variabilidad genética y heterocigosidad altas, que se relacionan directamente con rendimiento y estabilidad (Wricke y Weber, 1986; Sahagún-Castellanos *et al.*, 2013). Esta estabilidad ocurre en su comportamiento, a través de los ambientes, y en su estructura genotípica de una generación a la siguiente. La estabilidad genética significa que el agricultor no tiene que comprar la semilla de una variedad sintética cada ciclo (Marquez-Sánchez, 2010).

El rendimiento de una variedad sintética depende de su coeficiente de endogamia, aptitud combinatoria y rendimiento *per se* de sus progenitores, entre otros factores. Sin embargo, las líneas progenitoras de buena calidad escasean en México, lo que ha orientado a los fitomejoradores a desarrollar cruza dobles y trilineales en lugar de híbridos de cruza simple (Espinosa *et al.*, 2002). En cualquier caso, el uso de híbridos por los agricultores de escasos recursos en México no es usual. En su lugar, frecuentemente usan semilla de variedades criollas o generaciones segregantes de híbridos cosechados en el ciclo previo. Sahagún-Castellanos y Villanueva-Verduzco (2012) mencionan que el uso de generaciones avanzadas de híbridos no necesariamente genera las variedades sintéticas que se desarrollarían con el apareamiento aleatorio de las líneas progenitoras de tales híbridos. Villanueva *et al.* (1994) evaluaron la utilidad de los cruzamientos entre híbridos comerciales de maíz, aunque no utilizaron el concepto de variedad sintética. Sahagún-Castellanos y Villanueva-Verduzco (1997) abordaron explícitamente aspectos teóricos de las variedades sintéticas formadas con cruza simple. Un estudio sobre la predicción del rendimiento de VSs formadas con cruza dobles fue realizado por Sahagún-Castellanos *et al.* (2005). La predicción del rendimiento y la determinación del coeficiente de endogamia de estas VSs fueron estudiadas por Márquez-Sánchez (2008). Márquez-Sánchez (2010) derivó fórmulas

mating, Wricke and Weber (1986) suggest that the VS's seem to be an alternative for crop species which naturally regenerate this way. It is also worthwhile to emphasize that these varieties have high genetic variability and heterozygosity, which directly relate to their performance and stability (Wricke and Weber, 1986; Sahagún-Castellanos *et al.*, 2013). This stability is reflected in their performance through the environments, and in their genotypic structure from one generation to the next. Genetic stability means that the farmer does not need to buy seed of the synthetic variety each cycle (Márquez-Sánchez, 2010).

Among other factors, the performance of a synthetic variety depends on the inbreeding coefficient, combining ability, and performance of its parents. However, good quality parental lines are scarce in Mexico; this situation has led plant breeders to develop double and three-way line crosses instead of single-cross hybrids (Espinosa *et al.*, 2002). In any case, low-income Mexican farmers do not usually use hybrids. Instead, they frequently use land races or segregant generations of hybrids, harvested in the previous cycle. Sahagún-Castellanos and Villanueva-Verduzco (2012) state that the use of advanced generations of hybrids does not necessarily generate the synthetic varieties that would develop if the parental lines of such hybrids were subject to random mating. Villanueva *et al.* (1994) evaluated the usefulness of the crosses between commercial maize hybrids, although they did not use the synthetic variety concept. Sahagún-Castellanos and Villanueva-Verduzco (1997) explicitly discussed the theoretical aspects of the synthetic varieties formed using single crosses. Sahagún-Castellanos *et al.* (2005) studied the performance prediction of VSs resulting from double crosses. Márquez-Sánchez (2008) studied the performance prediction and the inbreeding coefficient determination of these VSs. Márquez-Sánchez (2010) derived formulas for the inbreeding coefficient (CE) and for a VS performance prediction, when the VS was developed from three-way line hybrid parents, formed using pure lines.

The derivation of the inbreeding coefficient of a VS whose parents are double crosses ($F\text{Sin}_{CD}$) has been restricted to the case in which the synthetic varieties (Sin_{CD}) are developed using hybrids formed with pure lines or when their CE (F) is $0 < F < 1$. However, their accuracy and validity have been questioned when each CD is represented by a small sample of m plants. The hypothesis of this study was that the CE of Sin_{CD} that

para el coeficiente de endogamia (CE) y para la predicción del rendimiento de una VS cuyos progenitores son híbridos trilineales, formados con líneas puras.

La derivación del coeficiente de endogamia de la VS, cuyos progenitores son cruza dobles ($FSin_{CD}$), se ha restringido al caso en que las variedades sintéticas (Sin_{CD}) se construyen con híbridos formados con líneas puras, o cuando su CE $0 \leq F \leq 1$, pero existen interrogantes de su exactitud y validez cuando cada CD se representa con una muestra pequeña de m plantas. La hipótesis de este estudio fue que aún se desconoce el CE del Sin_{CD} aplicable sin error independientemente de los valores de m y F . Por esto y porque la media genotípica del Sin_{CD} se relaciona con $FSin_{CD}$, en este estudio se derivó una fórmula general (para número de representantes de cada progenitor y CE de las líneas) y exacta para el CE de las VSs formadas por el apareamiento aleatorio de d cruza dobles (Sin_{CD}).

MATERIALES Y MÉTODOS

Este estudio se basó en el modelo de un *locus* de especies diploides que se reproducen por apareamiento aleatorio (Sahagún-Castellanos *et al.*, 2005; Márquez-Sánchez, 2010). El CE del Sin_{CD} ($FSin_{CD}$) se interpretó como la probabilidad de identidad por descendencia de dos genes tomados al azar con reemplazo de los genes que forman el arreglo gamético de sus progenitores (Sahagún-Castellanos *et al.*, 2013). Además, se supuso que el CE de las líneas con que se forman las cruza dobles (CD) es $0 \leq F \leq 1$ y que son líneas no emparentadas.

Para cada cruza doble se supuso que sus cuatro líneas progenitoras son las poblaciones virtuales, por ejemplo, representadas como A_1A_2 , B_1B_2 , C_1C_2 y D_1D_2 , en las que la probabilidad (P) de que los dos genes de cada genotipo sean idénticos por descendencia (\equiv) es:

$$P(A_1 \equiv A_2) = P(B_1 \equiv B_2) = P(C_1 \equiv C_2) = P(D_1 \equiv D_2) = F$$

Si se forma la cruza doble $(A_1A_2 \times B_1B_2) \times (C_1C_2 \times D_1D_2)$, el arreglo genotípico de la progenie resultante (AG_{CD}) es:

$$AG_{CD} = (1/16)[A_1C_1 + A_1C_2 + A_1D_1 + A_1D_2 + A_2C_1 + A_2C_2 + A_2D_1 + A_2D_2 + B_1C_1 + B_1C_2 + B_1D_1 + B_1D_2 + B_2C_1 + B_2C_2 + B_2D_1 + B_2D_2] \quad (1)$$

Para formar el Sin_{CD} , cada cruza doble se visualizó como una muestra aleatoria grande de m plantas tomada con reemplazo de los genotipos de su progenie (ecuación 1), y las m plantas de cada

can be applied without error —regardless of m and F values— remains unknown. Because of this and because the genotypic mean of a Sin_{CD} is related to its $FSin_{CD}$, in this study, an exact and general (for the number of representatives of each parent and CE of the parental lines) formula was derived for the CE of VSs formed by the random mating of d double crosses (Sin_{CD}).

MATERIALS AND METHODS

This study was based on a model of a *locus* of diploid species that are reproduced by random mating (Sahagún-Castellanos *et al.*, 2005; Márquez-Sánchez, 2010). The CE of a Sin_{CD} ($FSin_{CD}$) was interpreted as the probability that two randomly selected genes taken with replacement from the gametic array of their parents are identical by descent (Sahagún-Castellanos *et al.*, 2013). Additionally, the CE of the lines that form the double crosses (CD's) was assumed to be $0 \leq F \leq 1$ and the lines were assumed to be unrelated.

For each double cross, the four parental lines were assumed to be virtual populations represented, for example, as A_1A_2 , B_1B_2 , C_1C_2 , and D_1D_2 , in which the probability (P) that the two genes of each genotype were identical by descent (\equiv) was:

$$P(A_1 \equiv A_2) = P(B_1 \equiv B_2) = P(C_1 \equiv C_2) = P(D_1 \equiv D_2) = F$$

If the double cross $(A_1A_2 \times B_1B_2) \times (C_1C_2 \times D_1D_2)$ is formed, the genotypic array of the resulting progeny (AG_{CD}) is:

$$AG_{CD} = (1/16)[A_1C_1 + A_1C_2 + A_1D_1 + A_1D_2 + A_2C_1 + A_2C_2 + A_2D_1 + A_2D_2 + B_1C_1 + B_1C_2 + B_1D_1 + B_1D_2 + B_2C_1 + B_2C_2 + B_2D_1 + B_2D_2] \quad (1)$$

In order to form the Sin_{CD} , each double cross was visualized as a large random sample of m plants taken with replacement from the genotypes of their progeny (equation 1). Additionally, the m plants of each sample were visualized as a series of x groups, each made up of 16 plants whose genotypes are those of the 16 plants shown in equation 1. It was considered that the random mating of the m plants that form the sample includes the random mating among the 16 plants of each group. The progenies produced by each group are generated due to: 1) 16 self-pollinations and 2) 16×15 direct and reciprocal crosses (C_{DR}). Additionally, the random mating of a double cross includes the $x(x-1)$ crosses between the 16 plants of each group with the 16 plants of each one of the remaining $x-1$ groups. In short, the random mating

muestra se visualizaron como una serie de x grupos, cada uno formado por 16 plantas cuyos genotipos son como los 16 que de la ecuación 1. Se consideró que el apareamiento aleatorio de las m plantas que forman la muestra incluye el que ocurre entre las 16 plantas de cada grupo. Las progenies que produce cada grupo se generan por: 1) 16 autofecundaciones y 2) 16×15 cruza directa y recíprocas (C_{DR}). Además de estos, el apareamiento aleatorio de una cruce doble incluye las $x(x-1)$ cruza entre las 16 plantas de cada grupo con las 16 de cada uno de los $x-1$ grupos restantes. En resumen, el apareamiento aleatorio de las m plantas de una CD produce: m autofecundaciones, $15m$ cruza C_{DR} y las $m(m-16)$ cruza intergrupales.

Para muestras pequeñas, a los resultados de sólo grupos completos se agregó lo correspondiente a las $m-16x$ plantas que no alcanzaron a formar un grupo.

RESULTADOS Y DISCUSIÓN

Muestras grandes

La VS resultante del apareamiento aleatorio de las d cruza dobles (Sin_{CD}), está formada por genotipos que no contribuyen y genotipos que sí contribuyen al coeficiente de endogamia ($FSin_{CD}$). Las aportaciones de cada cruce doble al $FSin_{CD}$ las hacen las autofecundaciones y los apareamientos entre plantas diferentes que en su genotipo tienen en común al menos un gen que proviene de una misma línea. Por ejemplo, $A_1C_1 \times A_1C_1$ produce una progenie cuyo coeficiente de endogamia es $1/2$; y la cruce de A_1C_1 con cada una de las 15 plantas cuyos genotipos son los restantes del AG_{CD} (ecuación 1) produce la progenie cuyo arreglo genotípico es:

$$(1/16)[A_1A_1 + A_1C_2 + C_1A_1 + C_1C_2 + A_1A_1 + A_1D_1 + C_1A_1 + C_1D_1 + A_1A_1 + A_1D_2 + C_1A_1 + C_1D_2 + \dots + A_1B_2 + A_1C_1 + C_1D_2 + C_1C_2 + A_1B_1 + A_1D_1 + C_1B_2 + C_1D_1 + A_1B_2 + A_1D_2 + C_1B_1 + C_1D_2] \quad (2)$$

El coeficiente de endogamia promedio de los genotipos que forman el arreglo genotípico de la Expresión 2 ($FCD_{A_1C_1}$) es:

$$FCD_{A_1C_1} = (1/60)[P(A_1 \equiv A_1) + P(A_1 \equiv C_2) + P(C_1 \equiv A_1) + P(C_1 \equiv C_2) + P(A_1 \equiv A_1) + P(A_1 \equiv D_1) + P(C_1 \equiv A_1) + P(C_1 \equiv D_1) + P(A_1 \equiv A_1) + \dots + P(A_1 \equiv C_2) + P(A_1 \equiv C_1) + P(C_1 \equiv B_2) + P(C_1 \equiv C_1) + P(A_1 \equiv B_1) + P(A_1 \equiv C_2) + P(A_1 \equiv B_1) + P(C_1 \equiv D_2)]$$

of the m plants of a CD produce: m selfpollinations, $15m$ C_{DR} crosses, and $m(m-16)$ intergroup crosses.

For small samples, the results of the $m-16x$ plants that were not enough to form a group were added to the results of complete groups.

RESULTS AND DISCUSSION

Large samples

The VS resulting from the random mating of the d double crosses (Sin_{CD}) is made up of genotypes that do not contribute and genotypes that do contribute to the inbreeding coefficient ($FSin_{CD}$). The selfpollination and the pairing between different plants which have at least one gene in common in their genotype that comes from the same line are the sources from each double cross to the $FSin_{CD}$. For example, the progeny of $A_1C_1 \times A_1C_1$ has an inbreeding coefficient of $1/2$, and the progeny of the cross of A_1C_1 with each one of the 15 plants whose genotypes are the remaining of the AG_{CD} (equation 1) has the following genotypic array:

$$(1/16)[A_1A_1 + A_1C_2 + C_1A_1 + C_1C_2 + A_1A_1 + A_1D_1 + C_1A_1 + C_1D_1 + A_1A_1 + A_1D_2 + C_1A_1 + C_1D_2 + \dots + A_1B_2 + A_1C_1 + C_1D_2 + C_1C_2 + A_1B_1 + A_1D_1 + C_1B_2 + C_1D_1 + A_1B_2 + A_1D_2 + C_1B_1 + C_1D_2] \quad (2)$$

The average inbreeding coefficient of the genotypes that form the genotypic array of the expression 2 ($FCD_{A_1D_1}$) is:

$$FCD_{A_1C_1} = (1/60)[P(A_1 \equiv A_1) + P(A_1 \equiv C_2) + P(C_1 \equiv A_1) + P(C_1 \equiv C_2) + P(A_1 \equiv A_1) + P(A_1 \equiv D_1) + P(C_1 \equiv A_1) + P(C_1 \equiv D_1) + P(A_1 \equiv A_1) + \dots + P(A_1 \equiv C_2) + P(A_1 \equiv C_1) + P(C_1 \equiv B_2) + P(C_1 \equiv C_1) + P(A_1 \equiv B_1) + P(A_1 \equiv C_2) + P(A_1 \equiv B_1) + P(C_1 \equiv D_2)] = (1/60)[1 + 0 + 0 + F + 0 + 0 + 0 + 1 + \dots + 0 + 0 + 0 + 0 + 0 + 0 + 0] = (3 + 4F)/30$$

As for A_1C_1 , the crosses of each of the other 15 genotypes of the AG_{CD} with the 15 remaining genotypes generate a progeny whose average inbreeding coefficient is $(3 + 4F)/30$. If the inbreeding coefficient of the $16 \times 15 = 240$ direct and reciprocal crosses are represented by F_{CDR} , then:

$$\begin{aligned}
 &= (1/60)[1+0+0+F+0+0+0+1+\dots+0 \\
 &+0+0+0+0+0+0+0] \\
 &= (3+4F)/30
 \end{aligned}$$

Así como con A_1C_1 , las cruzas de cada uno de los otros 15 genotipos del AG_{CD} con los 15 restantes generan progenies cuyo coeficiente de endogamia promedio también es $(3+4F)/30$. Si el coeficiente de endogamia de las $16 \times 15 = 240$ cruzas directas y recíprocas se representa por F_{CDR} , también:

$$F_{CDR} = \frac{3+4F}{30} \quad (3)$$

Con base en la ecuación 3 y al considerar de que los 16 eventos reproductivos del tipo $A_1C_1 \times A_1C_1$ generan progenies cuyo CE promedio es $1/2$ se debe concluir que el apareamiento aleatorio de las 16 plantas de un grupo (ecuación 1) produce una población cuyo CE (F_x) es:

$$F_x = \frac{16(1/2) + 16 \times 15(3+4F)/30}{256}$$

Evidentemente, si el número de representantes de la craza doble en cuestión (m) es $16x$ el CE de la población que resulta del apareamiento al azar de las m plantas de cualquier craza doble (F_{CD}) es:

$$F_{CD} = \frac{1+F}{8} \quad (4)$$

La ecuación 4 se puede explicar en los términos siguientes: la muestra de tamaño m de una CD se visualizó como una población formada por x grupos de 16 plantas cuyos genotipos son como los del AG_{CD} (ecuación 1). Así, el apareamiento aleatorio de la muestra y sus resultados son equivalentes al ya descrito (ecuaciones 1 a 4), pero x^2 veces y el coeficiente de endogamia promedio de la población resultante debe ser $(1+F)/8$, igual al de la ecuación 4.

Si hay d cruzas dobles progenitoras su Sim_{CD} es una población formada por d subpoblaciones, cada una como la que se acaba de describir, y $d(d-1)$ subpoblaciones formadas por cruzas entre los representantes

$$F_{CDR} = \frac{3+4F}{30} \quad (3)$$

Based on equation 3 and considering that the 16 type $A_1C_1 \times A_1C_1$ reproductive events generate progenies which have an average CE of $1/2$ the conclusion should be that random mating of the 16 plants of a group (equation 1) produces a population with the following CE (F_x):

$$F_x = \frac{16(1/2) + 16 \times 15(3+4F)/30}{256}$$

Clearly, if the number of representatives of the double cross under consideration (m) is $16x$ the CE of the population resulting from the random mating of the m plants of any double cross (F_{CD}) is, then:

$$F_{CD} = \frac{1+F}{8} \quad (4)$$

Equation 4 can be explained in the following terms: the m -size sample of a CD was visualized as a population formed by x groups of 16 plants whose genotypes are similar to those of AG_{CD} (equation 1). Therefore, random mating of the sample and its results are equivalent to those described before (equations 1 to 4), but x^2 times and the average inbreeding coefficient of the resulting population must be $(1+F)/8$, equal to the one in equation 4.

If there are d parental double crosses, their Sim_{CD} is a population formed by d subpopulations, each one the same as the one that has just been described, and $d(d-1)$ subpopulations formed by crosses between the representatives of each double cross and of each one of the remaining $d-1$ double crosses. Since their lines are not related, pairings between double crosses do not contribute to the $F_{Sin_{CD}}$. Therefore:

$$\begin{aligned}
 F_{Sin_{CD}} &= \frac{d(1+F)/8}{d^2} \\
 &= \frac{1+F}{8d}
 \end{aligned} \quad (5)$$

de cada cruce doble y la de cada una de las $d - 1$ cruces dobles restantes. Como los apareamientos entre cruces dobles no contribuyen a $FSin_{CD}$, porque las líneas no están emparentadas, entonces:

$$FSin_{CD} = \frac{d(1+F)/8}{d^2} = \frac{1+F}{8d} \quad (5)$$

Evidentemente, si $F=1$, la ecuación 5 se reduce a la forma:

$$FSin_{CD} = \frac{1}{4d} \quad (6)$$

Márquez-Sánchez (2008) señaló que si las líneas son puras ($F=1$) el coeficiente de endogamia del Sin_{CD} ($F'Sin_{CD}$) es:

$$F'Sin_{CD} = \frac{m+1}{4dm} \quad (7)$$

$F'Sin_{CD}$ tiene sesgo igual a $1/(4dm)$ de acuerdo con las ecuaciones 6 y 7. Este resultado se debe a que en $F'Sin_{CD}$ se determinó por separado la contribución de las autofecundaciones ($1/2$) y, con un error, la de las cruces intrapaternales. En efecto, el coeficiente de endogamia promedio de la progenie producida por el apareamiento aleatorio de los m representantes de una cruce doble (F_{CD}) puede expresarse en términos de lo que en promedio le aportan las m autofecundaciones ($1/2$) y las $m(m-1)$ cruces intrapaternales ($r_{0,w}$); es decir:

$$F_{CD} = \frac{m(1/2) + m(m-1)r_{0,w}}{m^2}$$

Como $F_{CD}=(1+F)/8$ (ecuación 4), resulta que:

$$r_{0,w} = \frac{m(1+F) - 4}{8(m-1)}$$

Además, si $F=1$ y $m > 1$, la coancestría intrapateral del Sin_{CD} , se reduce a:

Clearly, if $F=1$, equation 5 is reduced to the form:

$$FSin_{CD} = \frac{1}{4d} \quad (6)$$

Márquez-Sánchez (2008) indicated that if the lines are pure ($F=1$), the inbreeding coefficient of the Sin_{CD} ($F'Sin_{CD}$) is:

$$F'Sin_{CD} = \frac{m+1}{4dm} \quad (7)$$

According to equations 6 and 7, the bias of $F'Sin_{CD}$ is $1/(4dm)$. This is the result of separately determining the selfpollinations ($1/2$) and the intraparental crosses (with an error), the contributions to $F'Sin_{CD}$. In fact, the average inbreeding coefficient of the progeny produced by the random mating of the m representatives of a double cross (F_{CD}) can be expressed in terms of the average contribution of the m selfpollinations ($1/2$) and the $m(m-1)$ intraparental crosses ($r_{0,w}$):

$$F_{CD} = \frac{m(1/2) + m(m-1)r_{0,w}}{m^2}$$

Since $F_{CD}=(1+F)/8$ (equation 4), then:

$$r_{0,w} = \frac{m(1+F) - 4}{8(m-1)}$$

Additionally, if $F=1$ and $m > 1$, the intraparental coancestry of Sin_{CD} reduces to:

$$r_{0,w} = \frac{m-2}{4(m-1)} \quad (8)$$

Márquez-Sánchez (2008) indicated that the intraparental coancestry is $1/4$; based on this result, this author found that the inbreeding coefficient of Sin_{CD} was $(m+1)/[4dm]$ (equation 7). However, according to equation 8, the intraparental coancestry ($1/4$) that Márquez-Sánchez derived has a positive bias $\{1/[4(m-1)]\}$. According to equation 6 for $F=1$, this bias causes a $1/(4md)$ bias in the CE of the Sin_{CD} , that tends to zero as m increases.

$$r_{0,W} = \frac{m-2}{4(m-1)} \quad (8)$$

Márquez-Sánchez (2008) observó que la coancestría intrapaternal es $1/4$ y a partir de este resultado obtuvo que el coeficiente de endogamia del Sin_{CD} es $(m+1)/[4dm]$ (ecuación 7). Sin embargo, de acuerdo con la ecuación 8, la coancestría intrapaternal ($1/4$) que él derivó tiene un sesgo positivo $\{1/[4(m-1)]\}$. Este a su vez, según la ecuación 6 para $F=1$ causa un sesgo al CE del Sin_{CD} igual a $1/(4md)$, que tiende a cero a medida que m es más grande.

Respecto al caso general para $0 \leq F \leq 1$, Sahagún-Castellanos (2011) obtuvo que el coeficiente de endogamia del Sin_{CD} ($F \cdot Sin_{CD}$) es $[3+m(1+F)-F]/(8md)$. Este con líneas puras también tiene un sesgo, igual a $(3-F)/(8md)$ según la ecuación 5 y también se reduce a $1/(4md)$ si $F=1$.

Considérese ahora que con las $4d$ líneas que forman un Sin_{CD} se derivaron $2d$ cruzas simples y que éstas fueron las progenitoras de las d cruzas dobles, y que con cada uno de estos tres conjuntos de progenitores se forma una VS y que estos sintéticos se representan como: Sin_L , Sin_{CS} y Sin_{CD} , respectivamente. Es claro que el número de genes que de cada línea inicial llegan a la muestra de m plantas de cada progenitor de una VS es una constante para Sin_L y Sin_{CS} . En cambio, el número de genes que cada línea aporta a la muestra de una CD es una variable aleatoria que puede tomar los valores $0, 1, 2, \dots, m$ que tiene una varianza igual a $m/4$. Por esta razón, el Sin_{CD} debe ser de las tres, la VS menos estable genéticamente.

Muestras pequeñas

Cuando m no es grande, el coeficiente de endogamia del Sin_{CD} debe reflejar lo que esto causa. En este caso, m no necesariamente es múltiplo de 16. Para derivar una fórmula para el CE de un Sin_{CD} para estas circunstancias es útil el modelo siguiente:

$$m=16x+e \quad \begin{array}{l} x=0,1,2,\dots, (m-e)/16 \\ e=0,1,2,3,\dots,15 \end{array} \quad (9)$$

donde: m es el tamaño de muestra de la progenie de cada craza doble, x es el número (entero) de veces que m contiene 16 unidades y $e=m-16x$ es el número de plantas que no alcanzaron a formar un grupo.

Regarding the general case for $0 \leq F \leq 1$, Sahagún-Castellanos (2011) concluded that the inbreeding coefficient of Sin_{CD} ($F \cdot Sin_{CD}$) is $[3+m(1+F)-F]/(8md)$. According to equation 5, it also has a $(3-F)/(8md)$ bias that also reduces to $1/(4md)$ if $F=1$.

Consider now that $2d$ single crosses were derived from the $4d$ lines that form a Sin_{CD} , and that these crosses were the parents of the d double crosses; additionally, suppose that each one of these three sets of parents formed a VS and these synthetics are represented as: Sin_L , Sin_{CS} and Sin_{CD} , respectively. Clearly, the number of genes from each initial line that reaches the sample of m plants for each parent of a VS is a constant for Sin_L and Sin_{CS} . Instead, the number of genes that each line contributes to the sample of one CD is a random variable that can take the values $0, 1, 2, \dots, m$ whose initial variance is equal to $m/4$. Therefore, the Sin_{CD} must be the less genetically stable VS out of the three.

Small samples

When m is not large, the inbreeding coefficient of Sin_{CD} must show its consequences. In this case, m is not necessarily a multiple of 16. In order to derive a formula for the CE of a Sin_{CD} for this situation, the following model is useful:

$$m=16x+e \quad \begin{array}{l} x=0,1,2,\dots, (m-e)/16 \\ e=0,1,2,3,\dots,15 \end{array} \quad (9)$$

where: m is the sample size of the progeny of each double cross; x is the (whole) number of times that m contains 16 units; and $e=m-16x$ is the number of plants that did not manage to form a group.

The sources of inbreeding associated to the random mating of each group of 16 plants whose genotypes are similar to the ones of AG_{CD} (equation 1) are: 16 selfpollinations and 16×15 direct and reciprocal crosses. Other sources of inbreeding for the progeny of a double cross are the $256x(x-1)$ intergroup crosses. The average inbreeding coefficients of these three sources are: $1/2$, $(3+4F)/30$ (equation 3) and $(1+F)/8$ (equation 4). According to equation 9, during the random mating of the m plants of each double cross, these three types of reproduction events happen with

Las fuentes de endogamia asociadas al apareamiento aleatorio de cada grupo de 16 plantas cuyos genotipos son como los del AG_{CD} (ecuación 1) son: 16 autofecundaciones y 16×15 cruza directas y recíprocas. Otra fuente de endogamia de la progenie de una crusa doble son las $256x(x-1)$ cruza intergrupales. Los coeficientes de endogamia promedio de estas tres fuentes son: $1/2$, $(3+4F)/30$ (ecuación 3) y $(1+F)/8$ (ecuación 4). De acuerdo con la ecuación 9, en el apareamiento aleatorio de las m plantas de cada crusa doble estos tres tipos de eventos de reproducción ocurren con las frecuencias absolutas siguientes: m , $(15)(16)x+e(e-1)$ y $256x(x-1)+2(16)xe$, respectivamente. Según esta información, el coeficiente de endogamia de una variedad sintética formada con d cruza dobles, cada una representada por un número finito de m plantas ($F^f Sin_{CD}$) es:

$$F^f Sin_{CD} = \left\{ m(1/2) + [(15)(16)x + e(e-1)] \right. \\ \left. (3+4F)/30 + [256(x-1) + 32xe] \right. \\ \left. (1+F)/8 \right\} / m^2 d \\ x = 0, 1, 2, 3, \dots, (m-e)/16 \\ e = 0, 1, 2, \dots, 15 \tag{10}$$

Esta fórmula, que es la contribución más importante de este estudio, es general y puede generar el coeficiente de endogamia exacto para cualquier combinación posible de valores de m , x , e , F y d . Las fórmulas para el coeficiente de endogamia para un Sin_{CD} previamente derivadas (ecuaciones 5 y 6) corresponden a tamaños de muestra “grandes”. La importancia de contar con un nivel alto de precisión para cuantificar la magnitud de la endogamia de un sintético no es sólo teórica. El coeficiente de endogamia se relaciona inversamente con la media de una variable aleatoria continua (Busbice, 1970) como el rendimiento de grano.

Cuando $x=0$ la fórmula para el coeficiente de endogamia del Sin_{CD} , basada en la ecuación 10, se expresa como:

$$F^f Sin_{CD} = \frac{m(1/2) + m(m-1)(3+4F)/30}{m^2 d} \\ m = 1, 2, \dots, 15$$

the following absolute frequencies: m , $(15)(16)x+e(e-1)$ and $256x(x-1)+2(16)xe$, respectively. According to this information, the inbreeding coefficient of a synthetic variety formed with d double crosses, each one represented by a finite number of m plants ($F^f Sin_{CD}$), is the following:

$$F^f Sin_{CD} = \left\{ m(1/2) + [(15)(16)x + e(e-1)] \right. \\ \left. (3+4F)/30 + [256(x-1) + 32xe] \right. \\ \left. (1+F)/8 \right\} / m^2 d \\ x = 0, 1, 2, 3, \dots, (m-e)/16 \\ e = 0, 1, 2, \dots, 15 \tag{10}$$

This general formula is the most important contribution of this study and can generate an exact inbreeding coefficient for any possible combination of values of m , x , e , F , and d . Previously derived formulas for the inbreeding coefficient of Sin_{CD} (equations 5 and 6) match the sizes of the “large” samples. Having a high accuracy level to quantify the dimension of a synthetic inbreeding coefficient does not only have theoretical importance. The inbreeding coefficient is inversely related to the mean of a continuous random variable (Busbice, 1970) such as grain yield.

Based on equation 10, when $x=0$, the inbreeding coefficient formula for Sin_{CD} is expressed as follows:

$$F^f Sin_{CD} = \frac{m(1/2) + m(m-1)(3+4F)/30}{m^2 d} \\ m = 1, 2, \dots, 15$$

It is reduced to:

$$F^f Sin_{CD} = \frac{15 + (m-1)(3+4F)}{30md} \\ m = 1, 2, \dots, 15 \tag{11}$$

If $F=1$, equation 11 is further simplified:

$$F^f Sin_{CD} = \frac{15 + 7(m-1)}{30md} \\ m = 1, 2, \dots, 15 \tag{12}$$

Esta se reduce a:

$$F^f \text{Sin}_{CD} = \frac{15 + (m-1)(3+4F)}{30md} \quad (11)$$

$m = 1, 2, \dots, 15$

Si $F=1$, la ecuación 11 se simplifica más:

$$F^f \text{Sin}_{CD} = \frac{15 + 7(m-1)}{30md} \quad (12)$$

$m = 1, 2, \dots, 15$

Cuando m es un múltiplo de 16 ($m=16x, x=m/16$ y $e=0$), de la ecuación 10 se obtiene que:

$$F^f \text{Sin}_{CD} = \frac{m(1/2) + 15m(3+4F)/30 + 16m(m/16-1)(1+F)/8}{m^2 d}$$

Simplificada, esta queda como:

$$F^f \text{Sin}_{CD} = \frac{1+F}{8d} \quad (13)$$

La ecuación 13 coincide con la ecuación 5 generada en el contexto de los procedimientos aplicables a las muestras grandes. Esto se debe a que en sus derivaciones los supuestos en forma explícita (ecuación 13) o implícita (ecuación 5) coinciden.

CONCLUSIONES

Una fórmula exacta y general se derivó para el coeficiente de endogamia de una variedad sintética cuyos progenitores son cruza dobles. La fórmula puede aplicarse a líneas progenitoras de cualquier nivel de endogamia y tamaño de muestra de plantas que representan a cada cruza doble progenitora. La fórmula es exacta porque su versatilidad incluye las particularidades requeridas para el cálculo sin sesgo del coeficiente de endogamia del sintético cuando el tamaño de muestra es pequeño sin que se tenga que recurrir a fórmulas para muestras grandes, propias del contexto en que se enuncia la ley de Hardy-Weinberg.

When m is a multiple of 16 ($m=16x, x=m/16$, and $e=0$), equation 10 becomes:

$$F^f \text{Sin}_{CD} = \frac{m(1/2) + 15m(3+4F)/30 + 16m(m/16-1)(1+F)/8}{m^2 d}$$

The simplified, final version, is:

$$F^f \text{Sin}_{CD} = \frac{1+F}{8d} \quad (13)$$

Equation 13 matches equation 5, which was generated in the context of the procedures that can be applied to large samples. This happens because in the derivations of the suppositions of both equations match each other, whether explicitly (equation 13) or implicitly (equation 5).

CONCLUSIONS

An exact and general formula was derived for the inbreeding coefficient of a synthetic variety whose parents are double crosses. The formula can be applied to parental lines of any inbreeding level and size of plant samples that represent each parental double cross. The formula is exact, because its versatility includes the required particularities for the non-biased calculation of the synthetic's inbreeding coefficient when the sample is small, and without resorting to formulas for large samples, suitable for the context in which the Hardy-Weinberg law is enunciated.

—End of the English version—



LITERATURA CITADA

- Busbice, T. 1970. Predicting yield of synthetic varieties. *Crop Sci.* 10: 260-269.
- Espinosa, C. A., M. Sierra M., y N. Gómez M. 2002. Producción y tecnología de semillas mejoradas de maíz por el INIFAP en el escenario sin la PRONASE. *Agron. Mesoam.* 14: 117-121.
- Márquez-Sánchez, F. 2008. Endogamia y predicción de sintéticos de maíz de cruza dobles. *Rev. Fitotec. Mex.* 31: 1-4.

- Márquez-Sánchez, F. 2010. Inbreeding coefficient and mean prediction of maize synthetics of three-way lines hybrids. *Maydica* 55: 227-229.
- Sahagún-Castellanos, J. 2011. Inbreeding and yield of synthetic varieties derived from single and double cross hybrids. *Maydica* 50: 265-271.
- Sahagún-Castellanos, J., J. E. Rodríguez-Pérez, and J. L. Escalante-González. 2013. Yield prediction and inbreeding of maize synthetics generated with lines and single crosses. Classic probability. *Rev. FCA. UNCUIYO* 45: 75-84.
- Sahagún-Castellanos, J., J. E. Rodríguez-Pérez, and A. Peña-Lomelí. 2005. Predicting yield of synthetics derived from double crosses. *Maydica* 50: 129-136.
- Sahagún-Castellanos, J., y C. Villanueva-Verduzco. 1997. Teoría de las variedades sintéticas formados con híbridos de cruce simple. *Rev. Fitotec. Mex.* 20: 97-110.
- Sahagún-Castellanos, J., C. Villanueva-Verduzco. 2012. ¿Variedades sintéticas derivadas de cruces simples o de cruces dobles?. *Rev. Chapingo Ser. Hortic.* 18: 279-282.
- Villanueva V., C., F. Castillo G., y J. D. Molina G. 1994. Aprovechamiento de cruzamientos dialélicos entre híbridos comerciales de maíz: análisis de progenitores y cruces. *Rev. Fitotec. Mex.* 17: 175-185.
- Wricke, G., and W. E. Weber. 1986. *Quantitative Genetics and Selection in Plant Breeding*. Walter de Gruyter, Berlin. New York. USA. 406 p.
- Wright, A. J. 1981. The quantitative genetics of diploid synthetic varieties. *In: Gallais A. Proceedings of the fourth meeting of the section Biometrics in Plant Breeding*. Poitiers, France, Sept. 2-4.1981.