

# Genetic stability of synthetics derived from double-cross or three-way line hybrids

Estabilidad genética de sintéticos formados con cruzas dobles o trilineales

Jaime Sahagún-Castellanos

Instituto de Horticultura, Departamento de Fitotecnia, Universidad Autónoma Chapingo.  
Carretera México-Texcoco km 38.5, Chapingo, Estado de México. C.P. 56230. MÉXICO  
Correo- e: jsahagunc@yahoo.com.mx

## Abstract

In the development of the theory of synthetic varieties (SVs) of crop species such as maize (*Zea mays* L.) and onion (*Allium cepa* L.), formed with single-cross hybrids, it has been shown that genetic erosion may occur during their formation. This erosion increases inbreeding and reduces yield. This gene loss is caused by the mating of heterozygous genotypes, the finite number of their progeny, and the randomness of the genetic mechanism. The main objective of this study was to determine the number of non-identical by descent (NIBD) genes lost during the development of the individuals that represent the double-cross (DC) or three-way (TW) line hybrids that in turn will be parents of a SV. The initial lines were assumed to be inbred and unrelated and each hybrid derived from them was represented by  $m$  individuals, and formulae for the mean, variance and number of lost NIBD genes per hybrid were derived. The number of lost NIBD genes of a DC (TW) was expressed as the number of NIBD genes in the 4 (3) initial lines minus the mean of NIBD genes forming the genotypes of the  $m$  individuals representing each DC (TW). It was found that each DC and TW loses on average  $4/2^m$  and  $2/2^m$  NIBD genes, respectively. The magnitudes of these losses reflect the effect of the sample size ( $m$ ) and the number of loss sources (the single crosses) of the parents of the DCs (2) and TWs (1). The largest NIBD gene losses per parent were 2 (DC) and 1 (TW), which occur when  $m = 1$ . However, when  $m$  is large ( $m \geq 12$ ), as should occur in reality, the losses and the variance of the number of NIBD genes are nearly zero.

**Keywords:** *Zea mays* L., *Allium cepa* L., non-identical by descent genes, inbreeding coefficient, genotypic mean.

## Resumen

En el desarrollo de teoría de las variedades sintéticas (VSs), de especies como el maíz (*Zea mays* L.) y la cebolla (*Allium cepa* L.), formadas con híbridos de cruce simple se ha evidenciado que durante su formación puede ocurrir erosión genética. Ésta incrementa la endogamia y reduce el rendimiento. Los causantes de esta pérdida de genes son el apareamiento de genotipos heterocigotos, el número finito de sus progenies y el azar del mecanismo genético. El principal objetivo de este trabajo fue determinar la pérdida de genes no idénticos por descendencia (NIPD), que ocurre en el desarrollo de los progenitores de VSs formadas con híbridos de cruce doble (CD) o triples (CT). Se consideró el uso de líneas puras no emparentadas para formar los híbridos; cada uno representado por  $m$  individuos, y se derivaron fórmulas para la media y la varianza del número de genes NIPD que se pierden. El número de genes NIPD perdidos en una CD (CT) se expresó como el número de genes NIPD en las 4 (3) líneas iniciales menos la media de genes NIPD de los  $m$  individuos que representan cada CD (CT). Se encontró que en cada CD y CT se pierden en promedio  $4/2^m$  y  $2/2^m$  genes NIPD, respectivamente. Las magnitudes de estas pérdidas reflejan el efecto del tamaño de muestra ( $m$ ) y del número de fuentes de pérdida (las cruces simples) de los progenitores de las CDs (2) y de las CTs (1). Las mayores pérdidas de genes NIPD por progenitor fueron 2 (CD) y 1 (CT), que ocurren cuando  $m = 1$ . Sin embargo, cuando  $m$  es grande ( $m \geq 12$ ), como debe suceder en la práctica, las pérdidas y la varianza del número de genes NIPD se reducen prácticamente a cero en ambos casos.

**Palabras clave:** *Zea mays* L., *Allium cepa* L., genes no idénticos por descendencia, coeficiente de endogamia, media genotípica.



## Introduction

Sahagún-Castellanos and Villanueva-Verduzco (1997) conducted a theoretical study of synthetic varieties (SVs) derived from a set of single-cross (SC) hybrids. This way of forming synthetics, which can be applied to species such as maize (*Zea mays* L.) and onion (*Allium cepa* L.), was later extended to double-cross (DC) hybrids (Sahagún-Castellanos, Rodríguez-Pérez, & Peña-Lomeli, 2005; Márquez-Sánchez, 2008), three-way (TW) line hybrids (Márquez-Sánchez, 2010) and mixtures of different types of hybrids (Sahagún-Castellanos & Rodríguez-Pérez, 2011; Sahagún-Castellanos & Villanueva-Verduzco, 2012). The idea of forming synthetic varieties in this way was inspired by practices carried out by some farmers to avoid the high annual seed cost of hybrid maize varieties. These farmers grow advanced generation hybrids or mixtures thereof, which have been viewed as SVs that would be formed with their parental lines. It is expected that these lines have already successfully passed through a selection process due to their combining ability, among other desirable characteristics. In the case of double crosses already released, it is expected that their parental lines will have good general and specific combining ability, which should be expressed in the population generated by randomly mating them in the form of a synthetic variety with good yield in the farmer's field.

However, there is evidence that a SV formed with double crosses cannot be the same as a SV that would be formed with the single crosses that are parents of the double crosses, and from that derived from the parental lines of such single crosses (Sahagún-Castellanos & Villanueva-Verduzco, 2012). These differences include changes in the inbreeding coefficient and the genotypic array. However, when the parents are double crosses, it is not precisely known whether these changes are reflected in the frequencies or even in the loss of some parental genes. The objectives of this study were to determine the mean and variance of the number of non-identical by descent (NIBD) genes passing from initial lines to double-cross or three-way line hybrids that will form a SV, and derive the number of NIBD genes that every parent loses in this process.

## Materials and methods

This study was based on the model of a locus of diploid species reproduced by random mating. The starting point to form DC and TW hybrids was a set of  $L$  unrelated inbred lines (completely homozygous).

It should be noted that in the formation of  $L/2$  SCs, NIBD genes cannot be lost. This is because each individual of the progeny of the cross between two inbred lines always carries two NIBD genes, each from a parental line. However, in the formation of  $L/4$  DCs this may be

## Introducción

Sahagún-Castellanos y Villanueva-Verduzco (1997) hicieron un estudio teórico de las variedades sintéticas (VSs) derivadas de un conjunto de híbridos de cruce simple (CS). Esta vía para formar sintéticos, que se puede aplicar a especies como el maíz (*Zea mays* L.) y la cebolla (*Allium cepa* L.), se extendió posteriormente a las cruces dobles (CD) (Sahagún-Castellanos, Rodríguez-Pérez, & Peña-Lomeli, 2005; Márquez-Sánchez, 2008), trilineales (CT) (Márquez-Sánchez, 2010) y mezclas de diferentes tipos de híbridos (Sahagún-Castellanos & Rodríguez-Pérez, 2011; Sahagún-Castellanos & Villanueva-Verduzco, 2012). La idea de formar variedades sintéticas de esta manera fue inspirada en prácticas que realizan algunos agricultores para evitar el alto costo anual de la semilla de variedades híbridas de maíz, los cuales cultivan generaciones avanzadas de híbridos o de mezclas de éstos, que han sido visualizadas como VSs que se formarían con sus líneas progenitoras. Es de esperarse que estas líneas ya pasaran exitosamente por un proceso de selección por su aptitud combinatoria; entre otras características deseables. En el caso de cruces dobles ya liberadas, de sus líneas progenitoras se espera que tengan buena aptitud combinatoria general y específica, que deben expresarse en la población que genera su apareamiento aleatorio en la forma de variedad sintética con rendimiento satisfactorio en el campo del agricultor.

Sin embargo, hay evidencias de que una VS formada con cruces dobles puede no ser la misma a la que formarían las cruces simples progenitoras de las cruces dobles, ni la que formarían las líneas progenitoras de tales cruces simples (Sahagún-Castellanos & Villanueva-Verduzco, 2012). Entre las diferencias encontradas se mencionan cambios en el coeficiente de endogamia y en el arreglo genotípico. Sin embargo, en las cruces dobles no se sabe con precisión si estos cambios son el reflejo en las frecuencias o aún en la pérdida de algunos genes de los progenitores. El objetivo de este estudio fue determinar la media y la varianza del número de genes no idénticos por descendencia (NIPD) que pasan de las líneas iniciales a los híbridos de cruce doble o trilineales que formarán una VS, y derivar el número de genes NIPD que pierde cada progenitor en este proceso.

## Materiales y métodos

Este estudio se basó en el modelo de un locus de especie diploide que se reproduce por apareamiento aleatorio. El punto de partida para formar los híbridos de CD y CT fue un conjunto de  $L$  líneas puras (completamente homocigóticas) no emparentadas.

Debe considerarse que en la formación de  $L/2$  CSs no puede haber pérdida de genes NIPD. Esto se debe a

different since each gamete from a *SC* can carry the gene of either line. For this reason, the number of *NIBD* genes reaching the genotypes that will represent each *DC* is a random variable. Likewise, genotypes formed with *TWs* contain a number of *NIBD* genes, which is a variable that takes random values because one parent of each three-way line hybrid is a single cross.

As in a previous study (Sahagún-Castellanos, Rodríguez-Pérez, & Escalante-González, 2013), this work was based on the concepts of gametic and genotypic array. If in a population the frequencies of the gene  $A_i$  ( $i = 1, 2, \dots, a$ ) and genotype  $A_i A_j$  are  $p_i$  and  $p_{ij}$ , the gametic (*GAA*) and genotypic (*GEA*) arrays, respectively, are defined as follows:

$$GAA = \sum_{i=1}^a p_i A_i \quad \text{and} \quad GEA = \sum_{i=1}^a \sum_{j=1}^a p_{ij} A_i A_j$$

If the gametic arrays of the two populations *P* and *Q* are  $\sum_{i=1}^a p_i A_i$  and  $\sum_{j=1}^a q_j A_j$ , then the genotypic array of the progeny produced by the cross between these two populations [*GEA(PXQ)*] is:

$$GEA(PXQ) = \sum_{i=1}^a \sum_{j=1}^a p_i q_j A_i A_j$$

The initial number of candidate *NIBD* genes to be a part of the genotype of a hybrid parent of the *SV* is the number of lines which form it because they are inbred and unrelated. The number of *NIBD* genes lost by each parent of the *SV* was considered as the difference between the initial number and the average or expected number of *NIBD* genes carried by the genotypes of the  $m$  individuals representing this parent. If the number of *NIBD* genes which on average reach the  $m$  representatives of a *DC* and a *TW* are symbolized as  $E(D_m)$  and  $E(T_m)$ , respectively, and the respective numbers of lost genes are expressed as  $(NIBD)_{pD}$  and  $(NIBD)_{pT}$ , then:

$$(NIBD)_{pD} = 4 - E(D_m) \quad (1)$$

and

$$(NIBD)_{pT} = 3 - E(T_m) \quad (2)$$

The  $m$  individuals representing a parent of the *SV* were considered as a random sample taken with replacement from the population of the genotypes forming the genotypic array of this parent (Sahagún et al., 2013).

## Deriving results

### Double-cross hybrids

The formation of single crosses (*SCs*) such as  $(AA) \times (BB)$  and  $(CC) \times (DD)$  was considered for each of the *L/4* four-line subsets. In addition, the genotypic array of the *DC*

que cada individuo de la progenie, de la cruza entre dos líneas puras, siempre lleva dos genes *NIPD* que provienen de sendas líneas progenitoras. Sin embargo, en la formación de las *L/4* *CDs* esto puede ser diferente puesto que cada gameto que emite una *CS* puede portar el gen de una u otra línea. Por esta razón, el número de genes *NIPD* que lleguen a los genotipos que representarán a cada *CD* es una variable aleatoria. Asimismo, los genotipos que se forman mediante *CTs* contienen un número de genes *NIPD*, que es una variable que toma valores al azar debido a que un progenitor de cada cruza trilineal es una cruza simple.

Como en un trabajo previo (Sahagún-Castellanos, Rodríguez-Pérez, & Escalante-González, 2013), este estudio se apoyó en los conceptos de arreglo gamético y genotípico. Si en una población las frecuencias del gen  $A_i$  ( $i = 1, 2, \dots, a$ ) y del genotipo  $A_i A_j$  son  $p_i$  y  $p_{ij}$ , sus arreglos gamético (*AGA*) y genotípico (*AGE*), respectivamente, se definen en la forma:

$$AGA = \sum_{i=1}^a p_i A_i \quad \text{y} \quad AGE = \sum_{i=1}^a \sum_{j=1}^a p_{ij} A_i A_j$$

También se consideró que si los arreglos gaméticos de dos poblaciones *P* y *Q* son  $\sum_{i=1}^a p_i A_i$  y  $\sum_{j=1}^a q_j A_j$ , entonces el arreglo genotípico de la progenie que produce la cruza entre estas dos poblaciones [*AGE(PXQ)*] es:

$$AGE(PXQ) = \sum_{i=1}^a \sum_{j=1}^a p_i q_j A_i A_j$$

El número inicial de genes *NIPD* candidatos a llegar a un híbrido progenitor de la *VS* es el número de líneas que lo forman ya que éstas son puras y no están emparentadas. Mientras que, el número de genes *NIPD* que se pierden por cada progenitor de la *VS* fue considerado como la diferencia entre el número inicial y el promedio o número esperado de genes *NIPD* que portan los genotipos de los  $m$  individuos que representan tal progenitor. Si el número de genes *NIPD*, que en promedio llegan a los  $m$  representantes de una *CD* y *CT*, se simbolizan como  $E(D_m)$  y  $E(T_m)$ , correspondientemente, y el respectivo número de genes que se pierden se expresan como  $(NIPD)_{pD}$  y  $(NIPD)_{pT}$ , entonces:

$$(NIPD)_{pD} = 4 - E(D_m) \quad (1)$$

y

$$(NIPD)_{pT} = 3 - E(T_m) \quad (2)$$

Los  $m$  individuos que representan un progenitor de la *VS* fueron considerados como el resultado de un muestreo aleatorio con reemplazo de tamaño  $m$  de la población, formada con los genotipos del progenitor (Sahagún et al., 2013).

produced by crossing these two SCs  $[(GEA)_{DC}]$  is as follows:

$$(GEA)_{DC} = \frac{1}{4}AC + \frac{1}{4}AD + \frac{1}{4}BC + \frac{1}{4}BD \quad (3)$$

Each genotype comprises two *NIBD* genes in Equation 3. Therefore, if  $m = 1$  the number of *NIBD* genes in the genotype sampled must be two. However, if  $m = 2$  the number of *NIBD* genes in both genotypes of the sample is a variable because when the sample captures the same genotype twice the number of *NIBD* genes is two, but when the genotypes captured are different, the number of *NIBD* genes carried can be three (with *AC* and *AD*, for example) or four (with *AC* and *BD*, or with *AD* and *BC*).

According to Equation 3, the sampling results, necessary to derive the mean of the number of *NIBD* genes in the  $m$  genotypes comprising the sample of a double cross, are:

- 1) The same genotype  $m$  times. This result can occur in four different ways, but always with the contribution of two *NIBD* genes.
- 2) Two different genotypes that contain three *NIBD* genes. The pairs of possible genotypes are: (*AC* and *AD*), (*AC* and *BC*), (*AD* and *BD*) and (*BC* and *BD*). To find the number of ways in which any of these pairs can occur, for example *AC* and *AD*, consider that if one of them (e.g. *AC*) occurs  $n$  times, the other (*AD*) must occur the remaining  $m-n$  times ( $n = 1, 2, 3, \dots, m-1$ ). The number of forms of occurrence of *AC*  $n$  times is the number of combinations  $\binom{m}{n}$ . Moreover, as *AC* can appear in the sample 1, 2, 3, ..., and up to  $m-1$  times, the number of times in which *AC* and *AD* occurs in the sample ( $N_{AC,AD}$ ) is as follows:

$$N_{AC,AD} = \sum_{n=1}^{m-1} \binom{m}{n}$$

Since:

$$\sum_{n=0}^m \binom{m}{n} = 2^m,$$

$$N_{AC,AD} = 2^m - 2$$

Thus, since each of the four pairs of genotypes carries three *NIBD* genes, the total number of ways in which two genotypes carrying three *NIBD* genes can occur in the sample is  $4(2^m - 2)$ .

- 3) Of the four genotypes of the double cross (Equation 3), six different pairs can be formed: the four mentioned in the previous paragraph and the remaining two: (*AC* and *BD*) and (*AD* and *BC*). Each one of the last two carries four *NIBD* genes. In addition, in the case of four pairs of two genotypes with three *NIBD* genes, the number of ways in which the remaining two pairs can occur is  $2(2^m - 2)$ .

## Derivación de resultados

### Cruzas dobles

De cada uno de  $L/4$ , subconjuntos de cuatro líneas, se consideró la formación de las (*CSs*) de la forma (*AA*) X (*BB*) y (*CC*) X (*DD*); además, por el cruzamiento entre estas dos *CSs* se generó la *CD*, cuyo arreglo genotípico  $[(AGE)_{CD}]$  es:

$$(AGE)_{CD} = \frac{1}{4}AC + \frac{1}{4}AD + \frac{1}{4}BC + \frac{1}{4}BD \quad (3)$$

En la Ecuación 3 cada genotipo está formado por dos genes *NIPD*. Por esta razón, si  $m = 1$  el número de genes *NIPD*, en el genotipo muestreado, debe ser dos. En cambio, si  $m = 2$  el número de genes *NIPD*, en ambos genotipos de la muestra, es una variable: porque cuando la muestra captura el mismo genotipo dos veces el número de genes *NIPD* es dos. En tanto que si los genotipos capturados son diferentes, el número de genes *NIPD* que portan puede ser tres (con *AC* y *AD*, por ejemplo) o cuatro (con *AC* y *BD*, o con *AD* y *BC*).

De acuerdo con la Ecuación 3, los resultados del muestreo, necesarios para derivar la media del número de genes *NIPD* en los  $m$  genotipos que constituyen la muestra de una crusa doble, son:

- 1) El mismo genotipo las  $m$  veces. Este resultado puede ocurrir en cuatro formas diferentes, pero siempre con aportación de dos genes *NIPD*.
- 2) Dos genotipos diferentes que aportan en cada caso tres genes *NIPD*. Los pares de genotipos posibles son: (*AC* y *AD*), (*AC* y *BC*), (*AD* y *BD*) y (*BC* y *BD*). Para encontrar el número de formas en que puede ocurrir cualquiera de estos pares, por ejemplo *AC* y *AD*, considérese que mientras que uno de ellos (por ejemplo *AC*), puede ocurrir  $n$  veces, el otro (*AD*) debe ocurrir las  $m-n$  veces restantes ( $n = 1, 2, 3, \dots, m-1$ ). También debe considerarse que el número de formas de aparición de *AC*  $n$  veces es el número de combinaciones  $\binom{m}{n}$ . Además, como *AC* puede aparecer en la muestra 1, 2, 3, ..., y hasta  $m-1$  veces, el número de formas en que *AC* y *AD* pueden ocurrir en la muestra ( $N_{AC,AD}$ ) es:

$$N_{AC,AD} = \sum_{n=1}^{m-1} \binom{m}{n}$$

Como:

$$\sum_{n=0}^m \binom{m}{n} = 2^m,$$

$$N_{AC,AD} = 2^m - 2$$

De esta manera, como también hay cuatro parejas de genotipos que portan tres genes *NIPD*, el número total de formas que en la muestra pueden ocurrir dos genotipos, que portan tres genes *NIPD*, es  $4(2^m - 2)$ .



4) To complete the derivation it should be considered that the total number of equally possible and mutually exclusive outcomes produced by the random sampling of size  $m$  with replacement is  $4^m$ . Of these, those not yet discussed are  $4^m - 4(2^m - 2) - 2(2^m - 2) - 4$ , each of which corresponds to a sample of genotypes carrying four *NIBD* genes.

Considering the four possible results of the sampling, the mean of *NIBD* genes in samples of size  $m$  [ $E(D_m)$ ] is:

$$E(D_m) = \frac{4(2) + 4[2^m - 2](3) + 2[2^m - 2](4)}{4^m} + \frac{[4^m - 6(2^m - 2) - 4](4)}{4^m} = 4 - 2^{2-m} \quad (4)$$

Since each double cross is based on four lines which together carry four *NIBD* genes, the number of *NIBD* genes lost by every double cross [ $(NIBD)_{pD}$ ], according to Equations 1 and 4, is:

$$(NIBD)_{pD} = 4 - \{4 - [2^{2-m}]\} = 2[2 / 2^m] \quad (5)$$

Equation 5 means that in the extreme case where  $m = 1$  two *NIBD* genes are always lost. This is because a genotype of a single cross is formed by two *NIBD* genes, and the formation of the genotype of an individual generated by the double cross also requires two *NIBD* genes; of the four *NIBD* genes from this *DC*, the remaining two will always be lost. When  $m = 2$ , randomness becomes a factor regarding the number of *NIBD* genes. Even though one gene on average is lost (Equation 5), two genes are lost when the sample "captures" the same genotype both times.

Regarding the genetic erosion of each of the  $L/2$  single crosses that are parents of a synthetic variety, Sahagún et al. (2013) found no loss of *NIBD* genes. Instead, according to Equation 5 in the case of  $(L/4)$  *DCs* a total loss of  $(L/4)(4/2^m) = L(1/2)^m$  *NIBD* genes is expected. This means that in terms of the cost of forming a synthetic with  $L/4$  double crosses instead of one with  $L$  lines the loss of  $L(1/2)^m$  *NIBD* genes should be included.

Given the information used to derive the mean of the number of *NIBD* genes of the  $m$  representatives of the double-cross parents of a *SV* (Equation 4), the variance [ $Var(D_m)$ ] is:

$$Var(D_m) = \{4(2)^2 + 4(2^m - 2)(3)^2 + [4^m - 4 - 4(2^m - 2)](4)^2\} / 4^m - [4 - 2^{2-m}]^2 = 4[2^m - 2]/4^m \quad (6)$$

The most evident is: with  $m = 1$  there is no variability in the number of *NIBD* genes in the sample; consistent with the previous observation, according to which if  $m = 1$ , the number of *NIBD* genes is always two.

3) De los cuatro genotipos de la cruce doble (Ecuación 3), seis parejas diferentes pueden ser formadas: las cuatro mencionadas en el párrafo previo y las dos restantes: (*AC* y *BD*) y (*AD* y *BC*). Cada una de estas últimas es portadora de cuatro genes *NIPD*. Por argumentos similares a los utilizados en el caso de las cuatro parejas de dos genotipos con tres genes *NIPD*, el número de formas en que pueden ocurrir las dos parejas restantes es  $2(2^m - 2)$ .

4) Para concluir la derivación, debe considerarse que el número total de resultados igualmente posibles y mutuamente excluyentes que puede producir el muestreo aleatorio de tamaño  $m$  con reemplazo es  $4^m$ . De éstos, los no discutidos aún son  $4^m - 4(2^m - 2) - 2(2^m - 2) - 4$ , y corresponden a sendas muestras de genotipos portadoras de cuatro genes *NIPD*.

Por los cuatro resultados posibles del muestreo, la media de genes *NIPD* en las muestras de tamaño  $m$  [ $E(D_m)$ ] es:

$$E(D_m) = \frac{4(2) + 4[2^m - 2](3) + 2[2^m - 2](4)}{4^m} + \frac{[4^m - 6(2^m - 2) - 4](4)}{4^m} = 4 - 2^{2-m} \quad (4)$$

Como cada cruce doble tiene su base en cuatro líneas que, juntas, portan cuatro genes *NIPD*, el número de genes *NIPD* que se pierden por cada cruce doble [ $(NIBD)_{pD}$ ], según las Ecuaciones 1 y 4, es:

$$(NIBD)_{pD} = 4 - \{4 - [2^{2-m}]\} = 2[2 / 2^m] \quad (5)$$

La Ecuación 5 implica que en el caso extremo en que  $m = 1$  siempre se pierden dos genes *NIPD*. Esto es así porque un genotipo de una cruce simple se forma con dos genes *NIPD*, y la formación del genotipo de un individuo generado por la cruce doble también requiere dos genes *NIPD*; de los cuatro genes *NIPD* de esa *CD*, los dos restantes siempre se perderán. Cuando  $m = 2$  el azar empieza a ser un factor respecto a número de genes *NIPD*. Si bien se pierde un gen en promedio (Ecuación 5), se pierden dos cuando la muestra "captura" al mismo genotipo las dos veces.

Respecto a la erosión genética de cada una de las  $L/2$  cruces simples progenitoras de una variedad sintética, Sahagún et al. (2013) encontraron que no hay pérdida de genes *NIPD*. En cambio, de acuerdo con la Ecuación 5 en el caso de las  $(L/4)$  *CD* se espera una pérdida total de  $(L/4)(4/2^m) = L(1/2)^m$  genes *NIPD*. Esto significa que en el costo de la facilidad de formar un sintético con  $L/4$  cruces dobles en lugar de uno con  $L$  líneas debe incluirse la pérdida de  $L(1/2)^m$  genes *NIPD*.

Considerando la información utilizada para derivar la media del número de genes *NIPD*, de los  $m$

Moreover, although when  $m$  is great the variance is zero, it is noteworthy that the greatest variance occurs when  $m = 2$ .

### Three-way line hybrids

With three inbred lines with genotypes  $AA$ ,  $BB$  and  $CC$ , a three-way line hybrid can be, without loss of generality,  $(AA \times BB) \times CC$ . This cross generates the progeny whose genotypic array  $[(GEA)_{TL}]$  is:

$$(GEA)_{TL} = (1/2)AC + (1/2)BC \quad (7)$$

The sampling size  $m$  with replacement of this genotypic array has only two possible outcomes in terms of the number of *NIBD* genes: 1) this number is two when the  $m$  genotypes of the sample are  $AC$  or  $BC$ ; and 2) the other possible number of *NIBD* genes in the sample is three (with  $m \geq 2$ ). It occurs in all possible samples except in the last two. Accordingly, the number of equally possible and mutually exclusive ways in which three *NIBD* genes can occur is  $2^m - 2$ . Therefore, the mean of the number of *NIBD* genes in the sample  $[E(T_m)]$  is:

$$\begin{aligned} E(T_m) &= [2(2) + (2^m - 2)(3)]/2^m \\ &= [3(2^m) - 2]/2^m \\ &= 3 - 2^{1-m} \end{aligned} \quad (8)$$

Furthermore, in terms of Equation 8, the variance of the number of *NIBD* genes in the sample  $[Var(T_m)]$  is:

$$\begin{aligned} Var(T_m) &= E[T_m^2] - [E(T_m)]^2 \\ &= 2[2^m - 2] / 4^m \end{aligned} \quad (9)$$

Evidently then,  $E(T_m)$ , the mean of the number of *NIBD* genes in the sample (Equation 8), is always equal to two when  $m = 1$ . In this case there is no variability, which is consistent with the value equal to zero of the variance (Equation 9). If  $m = 2, 3, 4, \dots$ , the respective average numbers of *NIBD* genes captured by the samples are: 2.5, 2.75, 2.875, ..., respectively (Equation 8). For their part, the variances of the number of *NIBD* genes for these sample sizes are, respectively: 1/4, 3/16, 7/64, ... These series, and Equations 8 and 9, mean that when the sample size is large the mean and variance of the number of *NIBD* genes in the sample representing a three-way line hybrid are 3 and 0, respectively.

Regarding the number of *NIBD* genes lost per each three-way line hybrid  $[(NIBD)_{PT}]$ , according to the previous result and Equation 8, we have the following:

$$\begin{aligned} (NIBD)_{PT} &= 3 - [3 - 2 / 2^m] \\ &= 2 / 2^m \end{aligned} \quad (10)$$

representantes de las cruzas dobles progenitoras de una *VS* (Ecuación 4), la varianza  $[Var(D_m)]$  es:

$$\begin{aligned} Var(D_m) &= \{4(2)^2 + 4(2^m - 2)(3)^2 + [4^m - 4 - 4(2^m - 2)](4)^2\} / \\ &\quad 4^m - [4 - 2^{2-m}]^2 \\ &= 4[2^m - 2] / 4^m \end{aligned} \quad (6)$$

De lo más evidente: con  $m = 1$  no hay variabilidad en el número de genes *NIBD* en la muestra; en consistencia con la observación previa según la cual con  $m = 1$  el número de genes *NIBD* es siempre dos. Además, si bien cuando  $m$  es grande la varianza es igual a cero, es consiguiente que la mayor varianza ocurre cuando apenas  $m = 2$ .

### Cruzas trilineales

Con tres líneas puras con genotipos  $AA$ ,  $BB$  y  $CC$  la cruce trilineal puede ser, sin pérdida de generalidad:  $(AA \times BB) \times CC$ . Esta cruce genera la progenie cuyo arreglo genotípico  $[(AGE)_{TL}]$  es:

$$(AGE)_{TL} = (1/2)AC + (1/2)BC \quad (7)$$

El muestreo de tamaño  $m$  con reemplazo de este arreglo genotípico sólo tiene dos resultados posibles en términos del número de genes *NIBD*: 1) este número es dos cuando los  $m$  genotipos de la muestra o son  $AC$ , o son  $BC$ ; y 2) el otro número posible de genes *NIBD* en la muestra es tres (con  $m \geq 2$ ). Éste ocurre en todas las muestras posibles excepto en las dos anteriores. Según esto, el número de formas igualmente posibles y mutuamente excluyentes en que pueden ocurrir tres genes *NIBD* es  $2^m - 2$ . Por lo tanto, la media del número de genes *NIBD* en la muestra  $[E(T_m)]$  es:

$$\begin{aligned} E(T_m) &= [2(2) + (2^m - 2)(3)]/2^m \\ &= [3(2^m) - 2]/2^m \\ &= 3 - 2^{1-m} \end{aligned} \quad (8)$$

Además, en términos de la Ecuación 8, la varianza del número de genes *NIBD* en la muestra  $[Var(T_m)]$  es:

$$\begin{aligned} Var(T_m) &= E[T_m^2] - [E(T_m)]^2 \\ &= 2[2^m - 2] / 4^m \end{aligned} \quad (9)$$

Evidentemente,  $E(T_m)$  la media del número de genes *NIBD* en la muestra (Ecuación 8), siempre es igual a dos cuando  $m = 1$ . En este caso no hay variabilidad, lo que es consistente con el valor igual a cero de la varianza (Ecuación 9). Si  $m = 2, 3, 4, \dots$ , los números promedio respectivos de genes *NIBD* que capturan las muestras son: 2.5, 2.75, 2.875, ..., respectivamente (Ecuación 8). Por su parte, las varianzas del número de genes *NIBD* para estos tamaños de muestra son, en forma respectiva: 1/4, 3/16, 7/64, ... Estas series, y las

In Equation 10, losses of *NIBD* genes of each three-way line hybrid are 1, 1/2, 1/4, ..., when  $m = 1, 2, 3, \dots$ , respectively. These losses are half of those occurring in a double cross (Equation 5).

## Discussion

To compare gene losses of double-cross and three-way line hybrids, consider that when  $m = 1, 2, 3, \dots$ , in a double-cross hybrid, 2,1,1/2, ..., *NIBD* genes (Equation 5), respectively, are lost, while from a three-way line hybrid, *NIBD* gene losses (Equation 10) are, in the same order, 1, 1/2, 1/4, .... According to these results and Equations 5 and 10, the number of *NIBD* genes lost with a double cross is twice that lost with a three-way line hybrid. This is because, whereas in double crosses each of the two parental single crosses (due to their heterozygous condition) is a source of *NIBD* gene loss in the formation of  $m$  representatives, in the case of a three-way line hybrid only one parent has those characteristics; the other, an inbred line, has no possibility of contributing to the loss of *NIBD* genes since all the gametes it produces have the same gene.

The behavior of the variances of the number of *NIBD* genes in the samples of double-cross and three-way line hybrids (Equations 6 and 9) is, in general, very similar. When  $m = 1$  both a double-cross and a three-way line hybrid always produce a genotype formed by two *NIBD* genes; that is, in both cases the variance is zero. With  $m = 2$ , however, both variances reach their peak as evidenced very clearly by Equations 6 and 9. From  $m = 3$  the variances begin their descent to zero, faster in the case of three-way line hybrids. From  $m = 12$ , stability is almost reached and *NIBD* gene loss is about zero (Equations 5 and 10). These results are consistent with those of Butrón, Tarrío, Revilla, Malvar, and Ordás (2003) who, in order to save time and effort in developing a maize synthetic, proposed the formation of *DCs* and, based on a molecular evaluation, concluded that the method is appropriate, as it assures that the lines have equal contribution. They used  $m = 40$ .

According to Equations 6 and 9, the variance of the *NIBD* genes of a double cross is twice that of a *TW*. This is because the sources of variation of the number of *NIBD* genes are the two parental single crosses of each double cross and that of each *TW*. Every single cross produces the same variability and these sources of variability are independent.

Regarding inbreeding, which is inversely related to the genotypic mean, Sahagún-Castellanos et al. (2013) found that synthetics made with mixtures of lines and single crosses have inbreeding coefficients greater than those of synthetics whose parents are either lines or single crosses; this is due to the fact that while in these two cases the gene frequencies are balanced,

Ecuaciones 8 y 9, implican que cuando la muestra es grande la media y la varianza del número de genes *NIPD*, en la muestra que representa una cruce trilineal, son 3 y 0, respectivamente.

Respecto al número de genes *NIPD* perdidos por cada cruce trilineal  $[(NIPD)_{PT}]$ , según el resultado previo y la Ecuación 8, se tiene lo siguiente:

$$\begin{aligned} (NIPD)_{PT} &= 3 - [3 - 2 / 2^m] \\ &= 2 / 2^m \end{aligned} \quad (10)$$

En la Ecuación 10, las pérdidas de genes *NIPD* de cada cruce trilineal son 1, 1/2, 1/4, ..., cuando  $m = 1, 2, 3, \dots$ , respectivamente. Estas pérdidas son la mitad de las que se producen en una cruce doble (Ecuación 5).

## Discusión

Para comparar las pérdidas de genes de una cruce doble con las de una cruce trilineal, considérese que cuando  $m = 1, 2, 3, \dots$ , en un híbrido de cruce doble se pierden 2,1,1/2, ..., genes *NIPD* (Ecuación 5), respectivamente; en tanto que de una cruce trilineal las pérdidas de genes *NIPD* (Ecuación 10) son, en el mismo orden, 1, 1/2, 1/4, .... De acuerdo con estos resultados y con las Ecuaciones 5 y 10, el número de genes *NIPD* que se pierden con una cruce doble es el doble del que se pierde con una cruce triple. Esto se debe a que, mientras que en una cruce doble cada una de las dos cruces simples progenitoras (por su condición heterocigótica) es una fuente de pérdida de genes *NIPD* en la formación de los  $m$  representantes, en una cruce trilineal sólo un progenitor tiene esas características; el otro, una línea pura, no tiene posibilidad de contribuir a la pérdida de genes *NIPD*, puesto que todos los gametos que emite llevan el mismo gen.

El comportamiento de las varianzas del número de genes *NIPD* en las muestras de híbridos de cuatro y tres líneas (Ecuaciones 6 y 9) es, en términos generales, muy similar. Cuando  $m = 1$  tanto una cruce doble como una trilineal producen siempre un genotipo formado por dos genes *NIPD*; es decir, en ambos casos las varianzas son iguales a cero. Con  $m = 2$ , en cambio, ambas varianzas alcanzan su máximo como lo evidencian muy claramente las Ecuaciones 6 y 9. A partir de  $m = 3$  inician su descenso hasta cero, más rápidamente en las cruces trilineales. A partir de  $m = 12$ , prácticamente se alcanza la estabilidad y la pérdida de genes *NIPD* se reduce a cero (Ecuaciones 5 y 10). Estos resultados son consistentes con los de Butrón, Tarrío, Revilla, Malvar, y Ordás (2003) quienes para ahorrar tiempo y esfuerzo en el desarrollo de un sintético de maíz propusieron la formación de *CDs* y, con base en una evaluación molecular, concluyeron que el método es apropiado, pues asegura que las líneas tengan igual contribución. Ellos usaron  $m = 40$ .

this does not happen in the mixtures. In a *TW* there is also an imbalance of gene frequencies, and although there is no imbalance in the case of double crosses, they suffer the greatest loss of *NIBD* genes, which is directly related to inbreeding (Sahagún-Castellanos & Villanueva-Verduzco, 2012), and this with a reduction in the genotypic mean (Busbice, 1970). The inbreeding coefficients of the *SVs* derived from three double-cross hybrids reported by Sahagún-Castellanos and Villanueva-Verduzco (2012) are slightly lower than those of the *SVs* of the four three-way line hybrids calculated with the formula reported by Márquez-Sánchez (2010). These numbers of double and single crosses, due to the line numbers involved, are consistent with those expected to be optimal for the derivation of a synthetic (Kutka & Smith, 2007).

It should be considered that the derivation of a synthetic variety is not over until the formation of the representatives of each parent, and that the rest of the process may involve an additional loss of *NIBD* genes. In individual representatives of parents, when these are hybrids, a necessary condition always occurs for gene loss: presence of heterozygous genotypes. However, when synthetic parents are inbred lines, there can be no loss of genes because each of these lines always transmits the same gene to their descendants.

## Conclusions

The loss of non-identical by descent (*NIBD*) genes, which occurs during the derivation of the *m* individuals which will represent each parental *DC* of a synthetic variety, is twice that of one *TW*. The peak loss occurs in both cases when  $m = 1$ ; in this case each *DC* always loses two *NIBD* genes and a *TW* loses one. Also, the loss of a *DC* is always equal to twice the loss of a *TW*, because the only sources of erosion are the *SCs*; two in one *DC* and one in a *TW*. However, when  $m = 12$ , the losses in both cases are approximately equal to zero. This means that in practice the risk of the genetic erosion discussed in this study should not be a problem with large numbers of *m*.

*End of English version*

## Literature citation / Literatura citada

- Busbice, T. H. (1970). Predicting yield of synthetic varieties. *Crop science*, 10(3), 265-269. doi:10.2135/cropsci1970.0011183X001000030017x
- Butrón, A., Tarrío, R., Revilla, P., Malvar, R., & Ordás, A. (2003). Molecular evaluation of two methods for developing maize synthetic varieties. *Molecular Breeding*, 12(4), 329-333. doi: 10.1023/B:MOLB.0000006718.11324.4f

Según las Ecuaciones 6 y 9, la varianza de los genes *NIPD* de una craza doble es dos veces la de una *CT*. Esto se debe a que las fuentes de variación del número de genes *NIPD* son las dos cruza simples progenitoras de cada craza doble y la de cada *CT*. Cada craza simple produce la misma variabilidad y estas fuentes de variabilidad son independientes.

Respecto a endogamia, que se relaciona inversamente con la media genotípica, Sahagún-Castellanos et al. (2013) encontraron que los sintéticos hechos con mezclas de líneas y *CSs* tienen coeficientes endogámicos mayores que los de sintéticos cuyos progenitores son sólo líneas o cruza simples; lo anterior debido a que mientras en estos dos casos las frecuencias génicas están balanceadas en las mezclas no sucede esto. En una *CT* también hay desbalance de frecuencias génicas, y aunque en las cruza dobles no, éstas sufren la mayor pérdida de genes *NIPD* que se relaciona directamente con la endogamia (Sahagún-Castellanos, & Villanueva-Verduzco, 2012), y ésta con una reducción en la media genotípica (Busbice, 1970). Los coeficientes de endogamia de las *VSs*, formadas con tres cruza dobles que reportan Sahagún-Castellanos y Villanueva-Verduzco (2012) son ligeramente menores que los de las *VSs* de cuatro cruza trilineales calculadas con la fórmula que reporta Márquez-Sánchez (2010). Estos números de cruza dobles y simples, por los números de líneas involucradas, son compatibles con los que se espera que sean los óptimos para la derivación de un sintético (Kutka & Smith, 2007).

Debe considerarse que hasta la formación de los representantes de cada progenitor no ha concluido la creación de la variedad sintética, y que en la parte faltante del proceso puede haber una pérdida adicional de genes *NIPD*. En los individuos representantes de los progenitores, cuando éstos son híbridos, siempre ocurre una condición necesaria para la pérdida de genes: presencia de genotipos heterocigóticos. En cambio, cuando los progenitores del sintético son líneas puras, no puede haber pérdida de genes debido a que cada una de estas líneas siempre transmite el mismo gen a sus descendientes.

## Conclusiones

La pérdida de genes no idénticos por descendencia (*NIPD*), que ocurre durante la formación de los *m* individuos que van a representar a cada *CD* progenitora de una variedad sintética, es el doble en relación con la de una *CT*. La pérdida máxima ocurre en ambos casos cuando  $m = 1$ ; en este caso cada *CD* y *CT* siempre pierden dos y un gen *NIPD*, respectivamente. Además, la pérdida de una *CD* siempre es igual al doble de la pérdida de una *CT*, debido a que las únicas fuentes de erosión son las *CSs*; dos en una *CD* y una en una *CT*. Sin embargo, a partir de  $m = 12$ , las pérdidas en ambos casos son



- Kutka, F. J., & Smith, M. E. (2007). How many parents give the highest yield in predicted synthetic and composite populations of maize? *Crop science*, 47(5), 1905-1913. doi: 10.2135/cropsci2006.12.0802sc
- Márquez-Sánchez, F. (2008). Endogamia y predicción de sintéticos de maíz de cruas dobles. *Revista Fitotecnia Mexicana*, 31(3), 1-4. Recuperado de <http://www.redalyc.org/articulo.oa?id=61009701>
- Márquez-Sánchez, F. (2010). Inbreeding coefficient and mean prediction of maize synthetics of three-way lines hybrids. *Maydica*, 55(3), 227-229. Recuperado de [http://www.maydica.org/articles/55\\_227.pdf](http://www.maydica.org/articles/55_227.pdf)
- Sahagún-Castellanos, J., Rodríguez-Pérez, J. E., & Escalante-González, J. L. (2013). Yield prediction and inbreeding of maize synthetics generated with lines and single crosses. Classic probability. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo*, 45(2), 75-84. Recuperado de [http://revista.fca.uncu.edu.ar/images/stories/pdfs/2013-02/T45\\_2\\_06\\_Sahagun-Castellanos.pdf](http://revista.fca.uncu.edu.ar/images/stories/pdfs/2013-02/T45_2_06_Sahagun-Castellanos.pdf)
- Sahagún-Castellanos, J., Rodríguez-Pérez, J. E., & Peña-Lomeli, A. (2005). Predicting yield of synthetics derived from double crosses. *Maydica*, 50(2), 129-136. Recuperado de [http://www.maydica.org/articles/50\\_129.pdf](http://www.maydica.org/articles/50_129.pdf)
- aproximadamente iguales a cero. Esto implica que en la práctica el peligro de la erosión genética estudiada en este trabajo no debe ser un problema con números grandes de *m*.

*Fin de la versión en español*

- Sahagún-Castellanos, J., & Rodríguez-Pérez, J. E. (2011). Inbreeding of synthetic varieties derived from lines and single crosses. *Revista Chapingo Serie Horticultura*, 17(3), 107-115. doi:10.5154/r.rchsh.2011.17.022
- Sahagún-Castellanos, J., & Villanueva-Verduzco, C. (1997). Teoría de las variedades sintéticas formadas con híbridos de crusa simple. *Revista Fitotecnia Mexicana*, 20(1).
- Sahagún-Castellanos, J., & Villanueva-Verduzco, C. (2012). ¿Variedades sintéticas derivadas de cruas simples o de cruas dobles? *Revista Chapingo Serie Horticultura*, 18(3), 279-289. doi: 10.5154/r.rchsh.2012.06.031